

From Coupling to Spectral Independence and Blackbox Comparison with the Down-Up Walk

Kuikui Liu ✉

University of Washington, Seattle, WA, USA

Abstract

We show that the existence of a “good” coupling w.r.t. Hamming distance for any local Markov chain on a discrete product space implies rapid mixing of the Glauber dynamics in a blackbox fashion. More specifically, we only require the expected distance between successive iterates under the coupling to be summable, as opposed to being one-step contractive in the worst case. Combined with recent local-to-global arguments [16], we establish asymptotically optimal lower bounds on the standard and modified log-Sobolev constants for the Glauber dynamics for sampling from spin systems on bounded-degree graphs when a curvature condition [44] is satisfied. To achieve this, we use Stein’s method for Markov chains [10, 46] to show that a “good” coupling for a local Markov chain yields strong bounds on the spectral independence of the distribution in the sense of [6].

Our primary application is to sampling proper list-colorings on bounded-degree graphs. In particular, combining the coupling for the flip dynamics given by [49, 13] with our techniques, we show optimal $O(n \log n)$ mixing for the Glauber dynamics for sampling proper list-colorings on any bounded-degree graph with maximum degree Δ whenever the size of the color lists are at least $(\frac{11}{6} - \epsilon) \Delta$, where $\epsilon \approx 10^{-5}$ is small constant. While $O(n^2)$ mixing was already known before, our approach additionally yields Chernoff-type concentration bounds for Hamming Lipschitz functions in this regime, which was not known before. Our approach is markedly different from prior works establishing spectral independence for spin systems using spatial mixing [6, 14, 15, 30], which crucially is still open in this regime for proper list-colorings.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains

Keywords and phrases Markov chains, Approximate counting, Spectral independence

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2021.32

Category RANDOM

Funding *Kuikui Liu*: The author is supported by NSF grants CCF-1552097, CCF-1907845.

Acknowledgements The author would like to thank their advisor Shayan Oveis Gharan for comments on a preliminary draft of this paper. We also thank Nima Anari and Pierre Youssef for informing us that a conjecture posed in a preliminary draft of the paper was already known to be false. We finally thank the anonymous reviewers for delivering valuable feedback on this paper.

1 Introduction

Given a probability distribution μ on a collection of subsets of a finite universe U with a fixed size n , one would like to generate (approximate) samples from μ . This problem is widely encountered in machine learning, statistical physics, and theoretical computer science, and encompasses many problems as special cases, including distributions over bases of matroids, discrete probabilistic graphical models, etc. A popular approach used in practice is to run a Markov chain on $\text{supp}(\mu)$ whose stationary distribution is μ . The main question then becomes how quickly does the distribution of the Markov chain converge to stationarity, i.e. does it mix rapidly?

A particularly natural Markov chain known as the “down-up walk” (or “high-order walk”) originally studied in the high-dimensional expander community [39, 21, 40, 45, 1] has recently received a lot of attention due to applications to sampling from discrete log-concave



© Kuikui Liu;

licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021).

Editors: Mary Wootters and Laura Sanità; Article No. 32; pp. 32:1–32:21



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

distributions [4, 18, 3, 5] and spin systems in statistical physics [1, 6, 14, 15, 30, 2, 16]. For sampling bases of matroids, the down-up walk recovers exactly the bases exchange walk first studied in [29, 43], and for sampling from discrete graphical models, the down-up walk recovers exactly the classical Glauber dynamics. One of the main insights in this area is that to prove rapid mixing of the down-up walk, it suffices to look only at pairwise correlations between elements, albeit for all conditional distributions of μ (see Definition 13). One additional advantage behind this approach is that one can prove local-to-global results not just for the spectral gap [40, 1], but also for the rate of entropy decay [18, 34, 16] and even for the rate of decay for arbitrary f -divergences [2]. This has led to asymptotically optimal mixing times for many problems [18, 5, 16] as well as Chernoff-type concentration bounds for Lipschitz functions.

However, establishing sufficiently strong bounds on pairwise correlations (or, more precisely, pairwise influences; see Definition 13), remains a challenging problem. Prior works typically rely on one of three techniques: Oppenheim’s trickle-down theorem [45, 4], spatial mixing (or correlation decay) [6, 14, 15, 30, 16], or the absence of roots for the multivariate generating polynomial of μ in a sufficiently large region of the complex plane [2, 17]. However, there are settings, such as proper list-colorings when the number of colors is less than twice the maximum degree of the graph, where the trickle-down theorem fails, and where spatial mixing and the existence of a nice root-free region are not known.

In this work, we show that the classical technique of (path) coupling can be used to bound these pairwise correlations. In fact, we will show that the existence of a “good” coupling for **any** sufficiently “local” dynamics with stationary distribution μ implies spectral independence for μ in the sense of [6], and hence, rapid mixing of the down-up walk. Hence, one can view our main result as a blackbox comparison result between any local Markov chain and the down-up walk.

As our main concrete application, we use the variable-length path coupling devised by [13] for the flip dynamics, building off work of Vigoda [49], to show $O(n \log n)$ mixing of the Glauber dynamics for sampling proper list-colorings on graphs of maximum degree $\Delta \leq O(1)$ whenever the number of available colors is at least $(\frac{11}{6} - \epsilon) \Delta$, where $\epsilon \approx 10^{-5}$ is a small constant. This mixing time is asymptotically optimal [36]. While $O(n^2)$ mixing was known earlier [31] (see also [49, 13]) by using a spectral gap comparison argument [19], our approach yields optimal bounds on the rate of entropy decay as well as Chernoff-type concentration inequalities. As mentioned earlier, strong spatial mixing and the existence of sufficiently large root-free regions are not known in this regime for proper list-colorings. Along the way, we make an additional conceptual contribution by answering the natural question of if Dobrushin-type mixing conditions imply spectral independence.

1.1 Our Contributions

To state our blackbox comparison result, let us first define the down-up walk for sampling from distributions over homogeneous set systems. We eschew the use of much of the terminology of high-dimensional expanders so as to simplify the exposition. Let μ be a distribution over $\binom{U}{n} = \{S \subseteq U : |S| = n\}$ for a finite set U and a positive integer $n \geq 1$ ¹. The down-up walk

¹ One can view μ as being a distribution over the facets of a “pure simplicial complex” weighted by μ . These are a generalization of usual graphs which are studied in geometry, topology, and combinatorics. The notion of spectral independence (Definition 13) was derived from a high-dimensional notion of “expansion” for simplicial complexes known as “local spectral expansion” first discovered by [21, 40, 45].

is described by the following two-step process. If the current state of the chain is $S^{(t)}$, then we select the next state $S^{(t+1)}$ as follows:

1. Select a uniformly random element $i \in S$.
2. Sample a set $S \in \text{supp}(\mu)$ satisfying $S \supseteq S^{(t)} \setminus \{i\}$ with probability proportional to $\mu(S)$ and transition to $S^{(t+1)} = S$.

As special cases, this class of Markov chains includes the bases exchange walk for matroids [43] and the Glauber dynamics for distributions over discrete product spaces.

We also define our notion of a “good” coupling and locality precisely here.

► **Definition 1 (Amortized Convergent Coupling).** Fix an irreducible transition probability matrix P which is reversible w.r.t. a distribution π on a finite state space Ω . Further endow Ω with a metric $d(\cdot, \cdot)$. We say a coupling of two faithful copies of the chain $(X^{(t)})_{t \geq 0}, (Y^{(t)})_{t \geq 0}$ is **C -amortized convergent w.r.t. $d(\cdot, \cdot)$** if the following holds for all $x, y \in \Omega$:

$$\sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)}, Y^{(t)}} \left[d(X^{(t)}, Y^{(t)}) \mid \begin{matrix} X^{(0)}=x \\ Y^{(0)}=y \end{matrix} \right] \leq C \cdot d(x, y).$$

► **Definition 2 (Locality of Dynamics).** Fix an irreducible transition probability matrix P which is reversible w.r.t. a distribution π on a finite state space Ω . Further endow Ω with a metric $d(\cdot, \cdot)$. For a positive real number $\ell > 0$, we say the dynamics P is **ℓ -local w.r.t. $d(\cdot, \cdot)$** if

$$\max_{x, y \in \Omega: P(x, y) > 0} d(x, y) \leq \ell.$$

Throughout the paper, unless stated otherwise, we work with Hamming distance. With these notions in hand, we now state our blackbox comparison result.

► **Theorem 3 (Blackbox Comparison with Down-Up Walk).** Let μ be a distribution on $\binom{U}{n}$, where U is a finite universe and $n \geq 1$ is a positive integer. For each $A \subseteq U$ with $|A| \leq n - 2$ and $A \subseteq S$ for some $S \in \text{supp}(\mu)$, let $P_{\mu|A}$ be some Markov chain on $\text{supp}(\mu \mid A)$ with stationary distribution $\mu \mid A$. Assume the family of Markov chains $\{P_{\mu|A}\}$ satisfy the following:

1. *Locality:* For some $\ell \geq 0$, $P_{\mu|A}$ is ℓ -local w.r.t. Hamming distance for all A .
2. *Good Coupling:* For some $C_{n-k} > 0$, $P_{\mu|A}$ admits a C_{n-k} -amortized convergent coupling w.r.t. Hamming distance for all k and A with $|A| = k$.
3. *Bounded Differences Between Chains:* For some $C'_{n-k} > 0$, we have the bound

$$\max_{S \in \text{supp}(\mu|(A \cup i))} \left\{ \sum_{T \neq S} |P_{\mu|A}(S \rightarrow T) - P_{\mu|(A \cup i)}(S \rightarrow T)| \right\} \leq C'_{n-k},$$

for all k, i and A with $|A| = k$.

If $\ell \cdot C_{n-k} \cdot C'_{n-k} \leq O(1)$ for all k , then the down-up walk has spectral gap at least $n^{-O(1)}$. If, in addition, μ is the Gibbs distribution of a spin system (see Section 2.1) on a bounded-degree graph, then the spectral gap, standard and modified log-Sobolev constants Equation (4) for the down-up walk are all $\Omega(1/n)$.

We refer the reader to Section A and references therein for the importance of lower bounding the spectral gap, standard and modified log-Sobolev constants, and in particular, their relation to mixing and concentration.

► **Remark 4.** While initially it may seem inconvenient to first build an entire family of Markov chains, one for each conditional distribution, this is very natural for many classes of distributions, in particular those which are closed under conditioning. As we will see, in practice, it is easy to obtain bounded differences between chains with $C'_{n-k} \lesssim \frac{1}{n-k}$ simply via brute force calculation. While $C_{n-k} \gtrsim n-k$ is often unavoidable, particularly for ℓ -local chains with $\ell \leq O(1)$, we will see that in many settings, we have $C_{n-k} \lesssim n-k$ as well. If additionally our dynamics are ℓ -local with $\ell \leq O(1)$, then the above yields a $n^{-O(1)}$ spectral gap for the down-up walk. It will turn out that our notion of ℓ -locality can also be relaxed; see Remark 10.

Our primary concrete application is to sampling proper list-colorings on graphs via the Glauber dynamics, which may be realized as a down-up walk. In this setting, we compare with another useful Markov chain known as the flip dynamics. The flip dynamics is ℓ -local w.r.t. unweighted Hamming distance with $\ell \leq 12$, and was analyzed in [49], who gave a greedy coupling which is one-step contractive whenever the number of available colors is at least $\frac{11}{6}\Delta$, implying it is C -amortized convergent with $C \leq O(n)$. [13] tweaked the parameters of the flip dynamics slightly while preserving locality, and further constructed a variable-length coupling which contracts by a constant factor every expected $O(n)$ steps whenever the number of available colors is at least $(\frac{11}{6} - \epsilon)\Delta$ for a small constant $\epsilon \approx 10^{-5}$. We will show this variable-length coupling is also C -amortized convergent with $C \leq O(n)$, and deduce optimal mixing for list-colorings in this regime.

► **Theorem 5.** *Let (G, \mathcal{L}) be a list-coloring instance where $G = (V, E)$ is a graph of maximum degree $\Delta \leq O(1)$ and $\mathcal{L} = (L(v))_{v \in V}$ is a collection of color lists. Then for some absolute constant $\epsilon \approx 10^{-5}$, if $|L(v)| \geq (\frac{11}{6} - \epsilon)\Delta$ for all $v \in V$, then the uniform distribution over proper list-colorings for (G, \mathcal{L}) is $(\eta_0, \dots, \eta_{n-2})$ -spectrally independent where $\eta_k \leq O(1)$ for all k . Furthermore, the spectral gap, standard and modified log-Sobolev constants Equation (4) for the Glauber dynamics are all $\Omega(1/n)$, and the mixing time is $O(n \log n)$.*

► **Remark 6.** Our running time dependence on Δ is roughly $\Delta^{O(\Delta^c)}$ for a mild constant c , which is rather poor. The main bottleneck in improving this dependence lies in the local-to-global result of [16], although our spectral independence bound, which depends polynomially on Δ , can also be significantly improved.

To prove Theorem 3, we leverage recent local-to-global results [1, 16] (see Theorems 14 and 15 for formal statements), which show that if one has sufficiently strong upper bounds on the total pairwise correlation $\sum_{j \in U} |\Pr_{S \sim \mu}[j \in S] - \Pr_{S \sim \mu|i}[j \in S]|$, then one can deduce rapid mixing for the down-up walk [39, 21, 40, 45]. To upper bound these correlations, we considerably generalize a result simultaneously due to [10, 46], which was discovered in the context of bounding the Wasserstein 1-distance between Ising models, or more generally, two measures on the discrete hypercube $\{-1, +1\}^n$. More specifically, we extend their results in several different directions:

1. We replace the Glauber dynamics by any local dynamics.
2. We allow the dynamics to admit a coupling which in a sense “contracts on average”, as opposed to a step-wise contraction in the worst-case.

► **Theorem 7.** *Let μ be a distribution on $\binom{U}{n}$, where U is some finite universe and $n \geq 1$ is a positive integer. Fix an arbitrary $i \in U$. Let P_μ (resp. $P_{\mu|i}$) be the transition kernel of any irreducible Markov chain on $\text{supp}(\mu)$ (resp. $\text{supp}(\mu | i)$) which is reversible w.r.t. μ (resp. $\text{supp}(\mu | i)$). Suppose that P_μ is ℓ -local and admits a C -amortized convergent coupling, both*

w.r.t. the Hamming metric $d_H(\cdot, \cdot)$. Then we have the bound

$$\sum_{j \in U} \left| \Pr_{S \sim \mu} [j \in S] - \Pr_{S \sim \mu|i} [j \in S] \right| \leq C \cdot \ell \cdot \max_{S \in \text{supp}(\mu|i)} \left\{ \sum_{T \neq S} |P_\mu(S \rightarrow T) - P_{\mu|i}(S \rightarrow T)| \right\}.$$

1.2 Main Technical Result

We now state our main technical result, which provides the most general bound on the difference between marginals of two distributions μ, ν . We immediately use it to deduce Theorem 7.

► **Theorem 8 (Main Technical).** *Let μ, ν be any two distributions on 2^U for a finite set U with $\text{supp}(\nu) \subseteq \text{supp}(\mu)$, where U is a finite universe and $n \geq 1$ is a positive integer. Further, let P_μ (resp. P_ν) be the transition kernel of any Markov chain on $\text{supp}(\mu)$ (resp. $\text{supp}(\nu)$) with stationary distribution μ (resp. ν). Assume P_μ is irreducible and reversible w.r.t. μ . Then we may bound both $\sum_{j \in U} |\Pr_{S \sim \mu}[j \in S] - \Pr_{S \sim \nu}[j \in S]|$ and the 1-Wasserstein distance $W_1(\mu, \nu)$ (see Definition 11) by the following quantity:*

$$\mathbb{E}_{S \sim \nu} \left[\sum_{T \neq S} |P_\mu(S \rightarrow T) - P_\nu(S \rightarrow T)| \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)}, Y^{(t)}} \left[d_H(X^{(t)}, Y^{(t)}) \mid \begin{matrix} X^{(0)}=S \\ Y^{(0)}=T \end{matrix} \right] \right],$$

where $(X^{(t)}, Y^{(t)})_{t=0}^{\infty}$ is a coupling of the Markov chain P_μ .

► **Remark 9.** The technical condition $\text{supp}(\nu) \subseteq \text{supp}(\mu)$ is just for convenience, as it ensures the transition probability $P_\mu(S \rightarrow T)$ also makes sense when $S \sim \nu$. This assumption is certainly satisfied in our application where ν is a conditional distribution of μ .

Proof of Theorem 7. We use Theorem 8 with $\nu = \mu | i$ to obtain the upper bound

$$\begin{aligned} & \mathbb{E}_{S \sim \mu|i} \left[\sum_{T \neq S} |P_\mu(S \rightarrow T) - P_{\mu|i}(S \rightarrow T)| \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)}, Y^{(t)}} \left[d_H(X^{(t)}, Y^{(t)}) \mid \begin{matrix} X^{(0)}=S \\ Y^{(0)}=T \end{matrix} \right] \right] \\ & \leq \max_{S \in \text{supp}(\mu|i)} \left\{ \sum_{T \neq S} |P_\mu(S \rightarrow T) - P_{\mu|i}(S \rightarrow T)| \right\} \\ & \quad \cdot \underbrace{\mathbb{E}_{S \sim \mu|i} \left[\max_{T: P_\mu(S \rightarrow T) > 0} \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)}, Y^{(t)}} \left[d_H(X^{(t)}, Y^{(t)}) \mid \begin{matrix} X^{(0)}=S \\ Y^{(0)}=T \end{matrix} \right] \right]}_{(*)}. \end{aligned}$$

It suffices to bound $(*)$ by $C \cdot \ell$. Since P_μ admits a C -amortized convergent coupling, we have that

$$\sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)}, Y^{(t)}} \left[d_H(X^{(t)}, Y^{(t)}) \mid \begin{matrix} X^{(0)}=S \\ Y^{(0)}=T \end{matrix} \right] \leq C \cdot d_H(S, T).$$

Hence,

$$(*) \leq C \cdot \mathbb{E}_{S \sim \mu|i} \left[\max_{T: P_\mu(S \rightarrow T) > 0} d_H(S, T) \right] \leq C \cdot \ell. \quad \blacktriangleleft$$

► Remark 10. One can see from the proof that we only needed that

$$\mathbb{E}_{S \sim \mu | i} \left[\max_{T: P_\mu(S \rightarrow T)} d_H(S, T) \right] \leq \ell,$$

as opposed to the stronger notion of ℓ -locality, where we have $\max_{S, T: P_\mu(S \rightarrow T)} d_H(S, T) \leq \ell$. Thus, in some sense, we only need the dynamics to make local moves “on average”. We leave it to future work to exploit this additional flexibility.

1.3 Independent Work

The results we obtain here were also independently discovered in [7].

1.4 Organization of the Paper

We state preliminaries on spin systems, spectral independence, etc. in Section 2. We then move to the proof of our main technical result (Theorem 8) in Section 3. In Section 4, we apply our techniques to distributions on discrete product spaces. In Section 5, we combine our techniques with couplings constructed in prior works to obtain spectral independence for proper list-colorings.

2 Preliminaries

For a positive integer $n \geq 1$, we write $[n] = \{1, \dots, n\}$. For a distribution μ on some finite state space Ω , we write $\text{supp}(\mu) = \{x \in \Omega : \mu(x) > 0\}$ for the support of μ . For a matrix A , we write $\|A\|_\infty = \max_i \sum_j |A(i, j)|$ for the maximum absolute row sum, and if A has real eigenvalues, we write $\lambda_{\max}(A)$ for the largest eigenvalue of A .

Throughout, we write $G = (V, E)$ for an undirected graph, and we will write Δ for the maximum degree of G . For a finite universe U and $S \subseteq U$, we write \mathbb{I}_S for the $\{0, 1\}$ -indicator function of S ; for an element $j \in U$, we write \mathbb{I}_j as opposed to $\mathbb{I}_{\{j\}}$. If μ is a distribution over $\binom{U}{n}$ and $S \subseteq U$, then we write $\mu | S$ for the conditional distribution of μ on $\binom{U \setminus S}{n - |S|}$, where $(\mu | S)(T) \propto \mu(S \cup T)$ whenever $S \cup T \in \text{supp}(\mu)$, $S \cap T = \emptyset$, and $(\mu | S)(T) = 0$ otherwise.

We will measure convergence of our Markov chains using total variation distance, defined as

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sum_x |\mu(x) - \nu(x)| = \sup_{S \subseteq \Omega} |\mu(S) - \nu(S)|$$

for two distributions μ, ν on a common state space Ω . We define the ϵ -mixing time of a Markov chain P on a state space Ω with stationary distribution π as

$$t_{\text{mix}}(\epsilon) \stackrel{\text{def}}{=} \max_{x \in \Omega} \min\{t \geq 0 : d_{\text{TV}}(\mathbb{I}_x P^t, \pi) \leq \epsilon\}.$$

The mixing time of the chain is defined as $t_{\text{mix}}(1/4)$. For the reader’s convenience, in Section A, we record the relation between mixing, spectral gap, modified and standard log-Sobolev constants. Finally, we also define the 1-Wasserstein distance.

► **Definition 11** (1-Wasserstein Distance). *Given two probability measures μ, ν on a common state space Ω endowed with a metric $d(\cdot, \cdot)$, we define the 1-Wasserstein distance $W_1(\mu, \nu)$ w.r.t. $d(\cdot, \cdot)$ by*

$$W_1(\mu, \nu) = \sup_f |\mathbb{E}_\mu f - \mathbb{E}_\nu f|,$$

where the supremum is over functions $f : \Omega \rightarrow \mathbb{R}$ which are 1-Lipschitz w.r.t. $d(\cdot, \cdot)$ (i.e. $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in \Omega$).

► **Remark 12.** By Kantorovich duality, one may equivalently define the 1-Wasserstein distance as

$$W_1(\mu, \nu) = \inf_{\gamma} \mathbb{E}_{(x,y) \sim \gamma} [d(x, y)],$$

where the infimum is over all couplings γ of μ, ν on $\Omega \times \Omega$.

2.1 Spin Systems

Fix an undirected graph $G = (V, E)$, and a positive integer $q \geq 2$. We view $[q]$ as a collection of possible “spin assignments” for the vertices of G . We also fix a symmetric nonnegative matrix $A \in \mathbb{R}_{\geq 0}^{q \times q}$ of “edge interaction activities” and a positive vector $h \in \mathbb{R}_{> 0}^q$ of “external fields”. The Gibbs distribution of the spin system on $G = (V, E)$ with parameters A, h is the distribution $\mu = \mu_{G,A,h}$ over configurations $\sigma : V \rightarrow [q]$ given by

$$\mu(\sigma) \propto \prod_{\{u,v\} \in E} A(\sigma(u), \sigma(v)) \prod_{v \in V} h(\sigma(v)),$$

where the constant of proportionality is the partition function of the system, given by

$$Z_G(A, h) = \sum_{\sigma: V \rightarrow [q]} \prod_{\{u,v\} \in E} A(\sigma(u), \sigma(v)) \prod_{v \in V} h(\sigma(v)).$$

Many classical models in statistical physics as well as distributions over often-studied combinatorial objects on graphs may be found as special cases. Notable examples include the Ising model on cuts, the hardcore gas model on independent sets, the monomer-dimer model on matchings, and the zero-temperature antiferromagnetic Potts model on proper colorings.

We call a configuration $\sigma : V \rightarrow [q]$ feasible if $\mu(\sigma) > 0$. For instance, if A has all positive entries, then all configurations $\sigma : V \rightarrow [q]$ are feasible. We call a partial configuration $\xi : S \rightarrow [q]$, where $S \subseteq V$ is a subset of vertices, a boundary condition. For such a boundary condition, we write $\mu \mid \xi$ for the conditional Gibbs distribution on $V \setminus S$ given by taking μ and conditioning on the event that the sampled $\sigma \sim \mu$ satisfies $\sigma(v) = \xi(v)$ for all $v \in S$.

2.2 Discrete Product Spaces and Homogeneous Set Systems

Fix a collection of finite sets $(\Omega(v))_{v \in V}$, where V is some finite index set with $|V| = n$, and consider a measure μ on the product space $\prod_{v \in V} \Omega(v)$. For instance, if $\Omega(v) = \{-1, +1\}$ for each $v \in V$, then μ is just a measure on the discrete hypercube $\{-1, +1\}^V$. An important subclass of examples which we will discuss at length include discrete probabilistic graphical models, where the index set V is the set of vertices of a (hyper)graph, and the measure μ designed in such a way that the (hyper)edges represent local interactions between vertices of the model; see Section 2.1 for more details.

As done in [6, 14, 15, 30], we view μ as a measure on $\binom{U}{n}$ where

$$U = \{(v, \omega(v)) : v \in V, \omega(v) \in \Omega(v)\}.$$

Note the usual Hamming distance $d_H(\cdot, \cdot)$ on $\binom{U}{n}$ is twice the usual Hamming distance typically associated with a discrete product space.

We will often write a single vertex-assignment pair (v, c) , where $v \in V$ and $c \in \Omega(v)$, as simply vc . Here, each configuration $\sigma \in \prod_{v \in V} \Omega(v)$ corresponds to the set $\{(v, \sigma(v)) : v \in V\}$. In this setting, the down-up walk is precisely the **Glauber dynamics** (or **Gibbs sampler**) for sampling from μ . For each configuration $\sigma \in \prod_{v \in V} \Omega(v)$, we transition to the next configuration by the following process:

1. Select a uniformly random coordinate $v \in V$.
2. Resample $\sigma(v)$ according to μ conditioned on σ_{-v} .

Let us make this more concrete. For each $\sigma \in \prod_{v \in V} \Omega(v)$ and $v \in V$, we write σ_{-v} for the partial subconfiguration of σ which only excludes $\sigma(v)$. For $c \in \Omega(v)$, we also write σ_{vc} for the configuration obtained by flipping the coordinate of v from $\sigma(v)$ to c . We may then write $\mu^v(\cdot | \sigma_{-v})$ for the marginal distribution of $\sigma(v)$ under μ conditioned on σ_{-v} . The transition kernel of the Glauber dynamics may then be written as

$$P_\mu(\sigma \rightarrow \sigma_{vc}) = \frac{1}{n} \cdot \mu^v(c | \sigma_{-v}).$$

2.3 Spectral Independence and The Down-Up Walk

Here, we formalize spectral independence and its connection with rapid mixing of the down-up walk. Throughout the paper, we will assume the following connectivity/nondegeneracy condition. Assuming Theorem 7 holds, we will also give a proof of Theorem 3.

Assumption: The down-up walk for μ and all of its conditional distributions is connected. In the context of spin systems, this condition is guaranteed by “total connectivity” of the system parameters (A, h) [16]. For instance, this is satisfied by all “soft-constraint” models (i.e. those with $A > 0$), and many “hard-constraint” models such as the hardcore model and the uniform distribution over proper colorings when $q \geq \Delta + 2$.

Let us now formalize spectral independence.

► **Definition 13** (Pairwise Influence and Spectral Independence [6]). *Fix a finite universe U and a positive integer $n \geq 1$. Fix a distribution μ on $\binom{U}{n} = \{S \subseteq U : |S| = n\}$. We define the **pairwise influence** of an element i on another element j by*

$$\mathcal{I}_\mu(i \rightarrow j) \stackrel{\text{def}}{=} \Pr_{S \sim \mu} [j \in S | i \in S] - \Pr_{S \sim \mu} [j \in S].$$

We write $\mathcal{I}_\mu \in \mathbb{R}^{U \times U}$ defined by $\mathcal{I}_\mu(i, j) = \mathcal{I}_\mu(i \rightarrow j)$ for the **pairwise influence matrix** of μ . We say the distribution μ is **η -spectrally independent** if $\lambda_{\max}(\mathcal{I}_\mu) \leq \eta + 1$. We say the distribution μ is **$(\eta_0, \dots, \eta_{n-2})$ -spectrally independent** if μ is η_0 -spectrally independent, $\mu | i$ is η_1 -spectrally independent for all $i \in U$, and so on.

Often in practice, and in this paper, instead of bounding $\lambda_{\max}(\mathcal{I}_\mu)$, we will bound

$$\|\mathcal{I}_\mu\|_\infty = \max_{i \in U} \sum_{j \in U} |\mathcal{I}_\mu(i \rightarrow j)|,$$

which is sufficient since it is well-known that $\lambda_{\max}(A) \leq \|A\|_\infty$ for any matrix A with real eigenvalues. The main usefulness of spectral independence is that it implies rapid mixing of the down-up walk, while only requiring bounds on pairwise correlations. We state the main local-to-global results most relevant to us here. In the most general setting, we may deduce an inverse polynomial spectral gap from sufficiently strong spectral independence [21, 40, 1].

► **Theorem 14** ([1], [6]). *Let U be a finite universe, and $n \geq 1$ a positive integer. Let μ be a distribution on $\binom{U}{n}$ which is $(\eta_0, \dots, \eta_{n-2})$ -spectrally independence. Then the down-up walk on $\binom{U}{n}$ for sampling from μ has spectral gap at least*

$$\frac{1}{n} \prod_{k=0}^{n-2} \left(1 - \frac{\eta_k}{n - k - 1}\right).$$

In particular, if $\eta_k \leq O(1)$ for all $k = 0, \dots, n-2$, then we have $n^{O(1)}$ -mixing of the down-up walk.

Subject to a certain mild technical condition on the marginals of the distribution μ , one can transfer spectral independence bounds to “local entropy decay” bounds, and then employ versions of the local-to-global result for entropy decay [34, 16, 2]. In the setting of spin systems, one can further take advantage of the bounded-degree assumption to obtain $O(n \log n)$ mixing time upper bounds [16], which are asymptotically optimal [36]. We state the current state-of-the-art for spin systems on bounded-degree graphs here, as we will need it in our application to proper list-colorings.

► **Theorem 15** ([16]). *Let (A, h) be the parameters of a spin system, and let $G = (V, E)$ be a graph with maximum degree at most $\Delta \leq O(1)$. If the Gibbs distribution $\mu = \mu_{G,A,h}$ is $(\eta_0, \dots, \eta_{n-2})$ -spectrally independent where $\eta_k \leq O(1)$ for all k , then both the standard and modified log-Sobolev constants of the Glauber dynamics (i.e. the down-up walk) for sampling from μ are at least $\Omega(1/n)$.*

We conclude this section with a proof of Theorem 3.

Proof of Theorem 3. By Theorems 14 and 15, it suffices to establish $(\eta_0, \dots, \eta_{n-2})$ -spectral independence for $\eta_k \leq O(1)$ for all k . Fix an arbitrary $A \subseteq U$ with $|A| = k \leq n - 2$ and $A \subseteq S$ for some $S \in \text{supp}(\mu)$. With the locality and coupling assumptions, Theorem 7 shows that for each $i \in U$, the absolute row sum of $\mathcal{I}_{\mu|A}$ for row i is upper bounded by

$$\ell \cdot C_{n-k} \cdot \max_{S \in \text{supp}(\mu|_i)} \left\{ \sum_{T \neq S} |P_\mu(S \rightarrow T) - P_{\mu|_i}(S \rightarrow T)| \right\}.$$

Bounded differences between chains then yields the upper bound $\lambda_{\max}(\mathcal{I}_{\mu|A}) \leq \|\mathcal{I}_{\mu|A}\|_\infty \leq \ell \cdot C_{n-k} \cdot C'_{n-k}$, which is $O(1)$ by assumption. As this holds for all such A , it follows that $\eta_k \leq O(1)$. ◀

3 Stein’s Method for Markov Chains

Our goal in this section is to prove Theorem 8. We follow [10, 46], using what is known as **Stein’s method for Markov chains**. Historically, Stein’s method [48] was developed as a method to bound distances between probability measures, with the primary motivation being to prove quantitative central limit theorems. [10, 46] adapted this method to bound the distance between two probability measures μ, ν on the discrete hypercube $\{-1, +1\}^n$ assuming the Glauber dynamics of either measure admits a contractive coupling. Our main intuition lies in viewing spectral independence (see Definition 13) as a measure of distance between different conditionings of the same distribution. Thus, one can try to apply this method to bound the spectral independence of a distribution. Let us now elucidate this method.

For a fixed function $f : \Omega \rightarrow \mathbb{R}$, we will construct an auxiliary function $h : \Omega \rightarrow \mathbb{R}$ which satisfies the **Poisson equation**

$$h - P_\mu h = f - \mathbb{E}_\mu f.$$

Questions concerning $\mathbb{E}_\mu f$ may then be studied by looking at $P_\mu h$. The following lemma constructs h more explicitly.

► **Lemma 16** (see Lemma 2.1 [10], Lemma 2.3 [46]). *Fix an irreducible transition probability matrix P which is reversible w.r.t. a distribution π on a finite state space Ω . Let $(X^{(t)})_{t=0}^\infty$ be the Markov chain generated by P , and for a fixed function $f : \Omega \rightarrow \mathbb{R}$, define $h : \Omega \rightarrow \mathbb{R}$ by*

$$h(x) = \sum_{t=0}^\infty \mathbb{E} \left[f(X^{(t)}) - \mathbb{E}_\pi f \mid X^{(0)} = x \right].$$

32:10 Coupling to Spectral Independence

Then h is well-defined as a function, and further satisfies the Poisson equation

$$h - Ph = f - \mathbb{E}_\pi f.$$

With this lemma in hand, we can immediately prove Theorem 8.

Proof of Theorem 8. Fix a function $f : 2^U \rightarrow \mathbb{R}$, and let h be the solution to the Poisson equation $h - P_\mu h = f - \mathbb{E}_\mu f$ given in Lemma 16. Then since ν is stationary w.r.t. P_ν , we have $\mathbb{E}_\nu P_\nu h = \mathbb{E}_\nu h$, so that using the Poisson equation yields

$$\mathbb{E}_\nu(P_\nu - P_\mu)h = \mathbb{E}_\nu h - \mathbb{E}_\nu[h - f + \mathbb{E}_\mu f] = \mathbb{E}_\nu f - \mathbb{E}_\mu f.$$

Hence, by the Triangle Inequality, we have that $|\mathbb{E}_\mu f - \mathbb{E}_\nu f| \leq \mathbb{E}_\nu |(P_\nu - P_\mu)h|$.

Now, let us bound $|(P_\nu - P_\mu)h|$ entrywise. For each $S \in \text{supp}(\nu)$, using that $P_\mu(S \rightarrow S) = 1 - \sum_{T \neq S} P_\mu(S \rightarrow T)$ (and analogously for P_ν),

$$\begin{aligned} & (P_\nu - P_\mu)h(S) \\ &= \sum_T (P_\nu(S \rightarrow T) - P_\mu(S \rightarrow T)) \cdot h(T) \\ &= \sum_{T \neq S} (P_\nu(S \rightarrow T) - P_\mu(S \rightarrow T)) \cdot (h(T) - h(S)) \\ &= \sum_{T \neq S} (P_\nu(S \rightarrow T) - P_\mu(S \rightarrow T)) \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)}, Y^{(t)}} \left[f(Y^{(t)}) - f(X^{(t)}) \mid \begin{matrix} X^{(0)}=S \\ Y^{(0)}=T \end{matrix} \right]. \end{aligned} \tag{Lemma 16}$$

It follows by the Triangle Inequality that

$$\begin{aligned} |(P_\nu - P_\mu)h(S)| &\leq \sum_{T \neq S} |P_\mu(S \rightarrow T) - P_\nu(S \rightarrow T)| \\ &\quad \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)}, Y^{(t)}} \left[\left| f(X^{(t)}) - f(Y^{(t)}) \right| \mid \begin{matrix} X^{(0)}=S \\ Y^{(0)}=T \end{matrix} \right]. \end{aligned} \tag{1}$$

Taking expectations w.r.t. ν finally yields a bound on $|\mathbb{E}_\mu f - \mathbb{E}_\nu f|$. The bound on the 1-Wasserstein distance follows immediately by taking f to be an arbitrary function which is 1-Lipschitz the metric $d_H(\cdot, \cdot)$. To obtain the bound on the total difference between marginals $\sum_{j \in U} |\Pr_{S \sim \mu}[j \in S] - \Pr_{S \sim \nu}[j \in S]|$, we apply the above inequality to $f = \mathbb{I}_j$ for each $j \in U$ and sum over all $j \in U$, noting that $d_H(S, T) = \sum_{j \in U} |\mathbb{I}_j(S) - \mathbb{I}_j(T)|$ and $\mathbb{E}_\mu f = \mathbb{E}_\mu \mathbb{I}_j = \Pr_{S \sim \mu}[j \in S]$ (and analogously for ν). ◀

4 Discrete Ricci Curvature on Product Spaces

In this section, we discuss applications of our results to general distributions on discrete product spaces. We show that the existence of a contractive coupling w.r.t. Hamming distance for the Glauber dynamics implies $O(1)$ -spectral independence. Such a condition is known as a discrete Ricci curvature condition for the dynamics in the sense of [44]. This also shows that the Dobrushin uniqueness condition implies $O(1)$ -spectral independence. When combined with the local-to-global result of [16], we resolve an unpublished conjecture due Peres and Tetali for spin systems on bounded-degree graphs; see [28] and references therein for recent progress on this conjecture on general graphs. We also give an alternative proof of the $\Omega(1/n)$ lower bound on the standard and modified log-Sobolev constants of the Glauber dynamics in this setting when a Dobrushin-type condition is satisfied, recovering a result of [42].

Classical work on Dobrushin-type conditions [25, 22, 23, 24, 35, 26] yield relatively simple and direct criteria for rapid mixing of the Glauber dynamics [11, 12]. The main idea here is intuitively similar to that of spectral independence (although the notion of Dobrushin influence here historically precedes spectral independence): so long as some measure of “total influence” is small, then μ is close in some sense to a product distribution, for which rapid mixing holds. However, prior to our work, the precise relationship between Dobrushin influence and the notion of pairwise influence used in spectral independence was unclear. This is an additional conceptual contribution of our work.

► **Definition 17** (Discrete Ricci Curvature [44]). *Fix an irreducible transition probability matrix P which is reversible w.r.t. a distribution π on a finite state space Ω . Further, endow Ω with a metric $d(\cdot, \cdot)$. We define the **discrete Ricci curvature** of the Markov chain P w.r.t. the metric space (Ω, d) by*

$$\alpha = \inf_{x, y \in \Omega: x \neq y} \left\{ 1 - \frac{W_1(P(x \rightarrow \cdot), P(y \rightarrow \cdot))}{d(x, y)} \right\},$$

where $W_1(\cdot, \cdot)$ is again the 1-Wasserstein distance w.r.t. $d(\cdot, \cdot)$. In other words, for every pair $x, y \in \Omega$, there is a coupling of the transitions $P(x \rightarrow \cdot), P(y \rightarrow \cdot)$ such that the expected distance $d(\cdot, \cdot)$ under the coupling contracts by a $(1 - \alpha)$ -multiplicative factor. In this case, we will say P admits a **$(1 - \alpha)$ -contractive coupling w.r.t. $d(\cdot, \cdot)$** .

► **Fact 18.** *Suppose P admits a $(1 - \alpha)$ -contractive coupling w.r.t. $d(\cdot, \cdot)$. Then this coupling is C -amortized convergent with $C = \frac{1}{\alpha}$.*

The following is an immediate application of Theorem 7, and yields a positive resolution to the Peres-Tetali conjecture for spin systems on bounded-degree graphs.

► **Theorem 19** (Curvature Implies Spectral Independence on Product Spaces). *Let μ be a measure on a discrete product space $\Omega = \prod_{v \in V} \Omega(v)$, where V is a finite index set and $\Omega(v)$ is finite for all $v \in V$. Endow Ω with the Hamming metric $d_H(\cdot, \cdot)$, and let α be the discrete Ricci curvature of the Glauber dynamics w.r.t. (Ω, d_H) . Then, the distribution is $(\eta_0, \dots, \eta_{n-2})$ -spectrally independent where $\eta_k \leq \frac{4}{\alpha n} - 1$ for all k . In particular, if $\alpha \geq \Omega(1/n)$, then the Glauber dynamics has spectral gap $n^{-O(1)}$. If additionally the measure μ is the Gibbs distribution of a spin system on a bounded-degree graph, then the spectral gap, standard and modified log-Sobolev constants for the Glauber dynamics are all $\Omega(1/n)$.*

Note that since the Glauber dynamics only updates the assignment to a single $v \in V$ in each step, it must be that $\alpha \leq O(1/n)$. Theorem 19 follows almost immediately from Fact 18 and a straightforward calculation involving the entries of the transition matrix of the Glauber dynamics. In the interest of space, we provide the proof in Section B.

4.1 Dobrushin Uniqueness and Spectral Independence

We now use Theorem 19 to show that Dobrushin’s uniqueness condition implies spectral independence.

► **Definition 20** (Dobrushin Influence). *Fix a probability measure on a finite product space $\prod_{v \in V} \Omega(v)$, where V is a finite indexing set. For each $u \in V$, let D_u be the collection of pairs $\tau, \sigma \in \prod_{v \in V} \Omega(v)$ such that $\tau_{-u} = \sigma_{-u}$ while $\tau(u) \neq \sigma(u)$. For distinct $u, v \in V$, we may then define the **Dobrushin influence** of u on v by*

$$\rho_\mu(u \rightarrow v) = \max_{(\tau, \sigma) \in D_u} d_{TV}(\mu^v(\cdot \mid \tau_{-v}), \mu^v(\cdot \mid \sigma_{-v})).$$

32:12 Coupling to Spectral Independence

We write $\rho_\mu = (\rho_\mu(u \rightarrow v))_{u,v} \in \mathbb{R}^{V \times V}$ for the **Dobrushin influence matrix**. We say the distribution μ satisfies the **Dobrushin uniqueness condition** if

$$\|\rho_\mu\|_1 \stackrel{\text{def}}{=} \max_{u \in V} \sum_{v \in V} \rho_\mu(u \rightarrow v) < 1.$$

A straightforward application of the path coupling technique of [11, 12] shows that if $\|\rho_\mu\|_1 < 1$, then there is a coupling for the Glauber dynamics which is one-step contractive w.r.t. Hamming distance. We state this well-known implication formally here, and refer to [26] for the proof.

► **Fact 21.** *Let μ be a distribution on some finite product space $\prod_{v \in V} \Omega(v)$, where V is a finite index set. If $\|\rho_\mu\|_1 \leq \gamma < 1$, then the Glauber dynamics is $(1 - \alpha)$ -contractive w.r.t. Hamming distance with $\alpha = \frac{1}{n}(1 - \gamma)$.*

In particular, combining Theorem 19 and Fact 21 immediately yields spectral independence under the Dobrushin uniqueness condition. Combined with Theorem 15, this additionally recovers a version of a result due to [42], which says that a weaker ℓ_2 -version of the Dobrushin uniqueness condition (see also [35, 26]) implies a $\Omega(1/n)$ log-Sobolev constant for the Glauber dynamics.

► **Corollary 22 (Dobrushin Uniqueness Implies Spectral Independence).** *Let μ be a distribution on some finite product space $\prod_{v \in V} \Omega(v)$, where V is a finite index set. If $\|\rho_\mu\|_1 \leq \gamma < 1$, then μ is $(\eta_0, \dots, \eta_{n-2})$ -spectrally independent with $\eta_k \leq \frac{4}{1-\gamma} - 1$ for all k . If additionally the measure μ is the Gibbs distribution of a spin system on a bounded-degree graph, then the spectral gap, standard and modified log-Sobolev constants for the Glauber dynamics are all $\Omega(1/n)$.*

5 Spectral Independence for Proper List-Colorings

We now specialize to the setting of proper list-colorings of a graph. Formally, we fix a graph $G = (V, E)$, a collection of color lists $(L(v))_{v \in V}$. We call a configuration $\sigma \in \prod_{v \in V} L(v)$ a list-coloring of G . We say a list-coloring σ is proper if $\sigma(u) \neq \sigma(v)$ whenever $u \neq v$ are neighbors. Throughout, we will let Δ denote the maximum degree of G , and we assume $\Delta \leq O(1)$. We also assume there is a positive integer $q \geq \Delta + 2$ such that $L(v) \subseteq [q]$ for all $v \in V$.

A well-known result due to [38] using path coupling shows that if $|L(v)| > 2\Delta$ for all $v \in V$, then there is a contractive one-step coupling for the Glauber dynamics which yields $O(n \log n)$ mixing. As noted in [16], one can adapt the argument of [32] to obtain strong spatial mixing when $|L(v)| > 2\Delta$, and use the arguments of [15, 30] to deduce spectral independence in this regime. However, it is still open whether one can obtain strong spatial mixing below the 2Δ threshold; see [32, 27] for results going below 2Δ on special classes of graphs.

In the seminal work of Vigoda [49], it was shown that there is a contractive one-step coupling for a different local Markov chain known as the flip dynamics whenever $|L(v)| \geq \frac{11}{6}\Delta$. This threshold was further improved to $|L(v)| \geq (\frac{11}{6} - \epsilon)\Delta$ in a recent breakthrough by [13], this time using a more sophisticated variable-length coupling. Both works further showed that Glauber dynamics mixes in $O(n^2)$ time in this regime using a spectral gap comparison argument [19].

Our goal is to use these coupling results along with Theorem 7 to obtain spectral independence for the uniform distribution over proper list-colorings in the regime $|L(v)| \geq$

$(\frac{11}{6} - \epsilon) \Delta$. Combined with Theorem 15, we improve the previous $O(n^2)$ mixing time bound to the optimal $O(n \log n)$, as well as show Chernoff-type concentration bounds for Lipschitz functions, which were not known before.

5.1 The Flip Dynamics

We follow the presentation in [13], which generalizes the flip dynamics analyzed in [49] to list-colorings. Fix a list-coloring σ . We say a path $u = w_1, \dots, w_\ell = v$ in G is an alternating path from u to v using colors $\sigma(u), c$ if for all i , we have $\sigma_{w_i} \in \{\sigma(u), c\}$ and $\sigma_{w_i} \neq \sigma_{w_{i+1}}$. For a fixed list-coloring σ , $v \in V$ and color c , we define the **Kempe component for σ, v, c** by the following subset of vertices.

$$S_\sigma(u, c) = \left\{ v \in V : \begin{array}{c} \exists \text{ alternating path from } u \\ \text{to } v \text{ using } \sigma(u), c \end{array} \right\}.$$

Given σ and a Kempe component $S = S_\sigma(u, c)$, we define σ_S to be the coloring obtained by “flipping” the color assigned to vertices in $\{v \in S : \sigma(v) = \sigma(u)\}$ to c , and the color assigned to vertices in $\{v \in S : \sigma(v) = c\}$ to $\sigma(u)$. Note that σ_S need not be a proper list-coloring; we say a Kempe component $S = S_\sigma(u, c)$ is **flippable** if the coloring σ_S is a proper list-coloring.

For each $j \in \mathbb{N}$, let $0 \leq p_j \leq 1$ be a tunable parameter to be determined later. We define the flip dynamics with flip parameters $\{p_j\}_{j \in \mathbb{N}}$ for sampling proper list-colorings as follows: Given the current list-coloring $\sigma^{(t-1)}$, we generate the next list-coloring $\sigma^{(t)}$ by the following two-step process:

1. Select a uniformly random vertex $v^{(t)} \in V$, and a uniformly random color $c^{(t)} \in L(v^{(t)})$.
2. If the Kempe component $S = S_{\sigma^{(t-1)}}(v^{(t)}, c^{(t)})$ is flippable, set $\sigma^{(t)} = \sigma_S^{(t-1)}$ with probability $\frac{p_j}{j}$ and $\sigma^{(t)} = \sigma^{(t-1)}$ otherwise, where $j = |S|$.

We write $P_{\mu, \text{flip}}$ for the transition probability matrix of the flip dynamics. It is straightforward to verify that the stationary distribution of the flip dynamics is uniform over proper list-colorings, regardless of the choice of the flip parameters. One can recover the Wang-Swendsen-Kotecký Markov chain by setting $p_j = j$ for all $j \in \mathbb{N}$ [50].

[49] showed that with flip parameters

$$p_1 = 1 \quad p_2 = \frac{13}{42} \quad p_3 = \frac{1}{6} \quad p_4 = \frac{2}{21} \quad p_5 = \frac{1}{21} \quad p_6 = \frac{1}{84} \quad p_j = 0, \forall j \geq 7, \tag{2}$$

there is a one-step coupling which is contractive w.r.t. Hamming distance whenever $|L(v)| \geq \frac{11}{6} \Delta$. [13] showed using linear programming arguments that this is optimal in the sense that when $|L(v)| < \frac{11}{6}$, there is no choice of the flip parameters which has a one-step contractive coupling w.r.t. Hamming distance. They additionally construct an explicit family of hard instances witnessing optimality.

One of the key insights of [13] is that the optimal choice of flip parameters comes out of the solution to a linear program, with the objective value of the program governing the contraction properties of the coupling. By solving this linear program, they show that for the following choice of flip parameters

$$\begin{aligned} \hat{p}_1 = 1 \quad \hat{p}_2 \approx 0.296706 \quad \hat{p}_3 \approx 0.166762 \quad \hat{p}_4 \approx 0.101790 \\ \hat{p}_5 \approx 0.058475 \quad \hat{p}_6 = 0.025989 \quad p_j = 0, \forall j \geq 7, \end{aligned} \tag{3}$$

there is a variable-length coupling such that the Hamming distance contracts by a constant factor every $O(n)$ steps in expectation. One can thus expect that the coupling is C -amortized convergent with $C \leq O(n)$.

We formalize their main coupling result in the following subsection. For the moment, we state two intermediate lemmas, and show how they imply Theorem 5.

► **Lemma 23.** *Assume the input graph $G = (V, E)$ has maximum degree $\Delta \leq O(1)$. Then, the flip dynamics with parameters given in Equation (3) satisfy the following:*

$$\max_{\tau \in \text{supp}(\mu|_{uc})} \left\{ \sum_{\sigma \neq \tau} |P_{\mu, \text{flip}}(\tau \rightarrow \sigma) - P_{\mu|_{uc}, \text{flip}}(\tau \rightarrow \sigma)| \right\} \leq O(1/n).$$

► **Lemma 24.** *Let (G, \mathcal{L}) be a list-coloring instance, where $\Delta \leq O(1)$ and $|L(v)| \geq \lambda^* \Delta$ for all $v \in V$, where $\lambda^* = \frac{11}{6} - \epsilon$ and $\epsilon \approx 10^{-5}$ is a small constant. Then the flip dynamics with parameters given in Equation (3) admits a C -amortized convergent coupling w.r.t. Hamming distance where $C \leq O(n)$.*

Proof of Theorem 5. The flip dynamics is clearly $O(1)$ -local w.r.t. Hamming distance since only Kempe components of size at most 6 can be flipped. $(\eta_0, \dots, \eta_{n-2})$ -spectral independence where $\eta_k \leq O(1)$ for all k then follows immediately by combining Lemma 23 and Lemma 24 with Theorem 3. The lower bounds on the spectral gap, standard and modified log-Sobolev constants then follow from Theorem 15. ◀

We provide the proof of Lemma 23 in Section B. At this point, all that remains is to prove Lemma 24, which we do using the variable-length path coupling constructed in [13].

5.2 Variable-Length Path Coupling: Proof of Lemma 24

To begin, we first define the notion of variable-length coupling following [37, 13].

► **Definition 25** (Path-Generating Set). *For a finite state space Ω , a **path generating set** is a subset $S \subseteq \binom{\Omega}{2}$ such that the undirected graph (Ω, S) is connected. We let $d_S(\cdot, \cdot)$ denote the induced shortest-path metric on Ω , and write $d(\cdot, \cdot)$ when the path generating set S is clear from context. We also write $x \sim y$ whenever $\{x, y\} \in S$.*

► **Definition 26** (Variable-Length Path Coupling [37]). *Fix an irreducible transition probability matrix P which is reversible w.r.t. a distribution π on a finite state space Ω , and let $d(\cdot, \cdot)$ be a metric on Ω induced by a path generating set $S \subseteq \binom{\Omega}{2}$. For every pair of starting states $x^{(0)}, y^{(0)} \in \Omega$ with $x^{(0)} \sim y^{(0)}$, we let $(\bar{x}, \bar{y}, T) = (\bar{x}(x^{(0)}, y^{(0)}), \bar{y}(x^{(0)}, y^{(0)}), T(x^{(0)}, y^{(0)}))$ denote a random variable where T is a (potentially random) nonnegative integer and $\bar{x} = (x^{(0)}, x^{(1)}, \dots, x^{(T)})$, $\bar{y} = (y^{(0)}, y^{(1)}, \dots, y^{(T)})$ are length- T sequences of states in Ω .*

*For every integer $t \geq 0$ and every pair of neighboring states $x^{(0)} \sim y^{(0)}$, define random variables x_t, y_t by the following experiment. Sample (\bar{x}, \bar{y}, T) , and set $x_t = x^{(t)}, y_t = y^{(t)}$ if $t \leq T$, and sample $x_t \sim P^{t-T}(x^{(T)}, \cdot), y_t \sim P^{t-T}(y^{(T)}, \cdot)$ if $t > T$. We say the random variable (\bar{x}, \bar{y}, T) is a **variable-length path coupling for P** if $x_t \sim P^t(x^{(0)}, \cdot), y_t \sim P^t(y^{(0)}, \cdot)$ for every integer $t \geq 0$ and every pair of neighboring states $x^{(0)} \sim y^{(0)}$. In this case, we say that \bar{x}, \bar{y} are individually **faithful copies**. If $T = t$ with probability 1 for some nonnegative integer $t \geq 0$, we say that (\bar{x}, \bar{y}, T) is a **t -step path coupling**.*

► **Remark 27.** In our application to colorings, the random time T will be a stopping time in the sense that its value only depends on the past, i.e. $x^{(0)}, y^{(0)}, \dots, x^{(t)}, y^{(t)}$ for $t \leq T$.

Given a variable-length path coupling, [37] showed one can construct a full coupling, generalizing the original path coupling theorem of [11, 12]. Furthermore, the contraction properties of the full coupling are inherited from the path coupling. While the original statement in [37] merely states rapid mixing given a variable-length path coupling, its proof implies the following.

► **Theorem 28** (Proof of Corollary 4 from [37]). *Let (\bar{x}, \bar{y}, T) be a variable-length path coupling w.r.t. a path generating set S for a reversible Markov chain P on a state space Ω with stationary distribution π . Let*

$$\begin{aligned}\alpha &\stackrel{\text{def}}{=} 1 - \max_{\{x^{(0)}, y^{(0)}\} \in S} \mathbb{E}[d_H(x^{(T)}, y^{(T)})] \\ W &\stackrel{\text{def}}{=} \max_{\{x^{(0)}, y^{(0)}\} \in S, t \leq T} d_H(x^{(t)}, y^{(t)}) \\ \beta &\stackrel{\text{def}}{=} \max_{\{x^{(0)}, y^{(0)}\} \in S} \mathbb{E}[T].\end{aligned}$$

Assume $0 < \alpha < 1$. Then there is a full M -step coupling with $M = \lceil \frac{2\beta W}{\alpha} \rceil$ such that for all pairs $x^{(0)}, y^{(0)}$, which need not be neighbors in S , we have the inequality

$$\mathbb{E}[d_H(x^{(M)}, y^{(M)}) \mid x^{(0)}, y^{(0)}] \leq \left(1 - \frac{\alpha}{2}\right) \cdot d_H(x^{(0)}, y^{(0)}).$$

Given this, all we need now is a good variable-length path coupling. This is given by the following result due to [13].

► **Theorem 29** ([13]). *Let (G, \mathcal{L}) be a list-coloring instance, where $G = (V, E)$ is a graph with maximum degree $\Delta \leq O(1)$, and $\mathcal{L} = (L(v))_{v \in V}$ is a collection of color lists. Let the path generating set S be given by the set of pairs $\{\tau, \sigma\}$ such that τ, σ differ on the coloring of exactly one vertex. Assume $|L(v)| \geq \lambda^* \Delta$ for all $v \in V$ where $\lambda^* = \frac{11}{6} - \epsilon$ for an absolute constant $\epsilon \approx 10^{-5}$. Then there exists a variable-length path coupling $(\bar{\tau}, \bar{\sigma}, T)$ for the flip dynamics w.r.t. S with flip parameters given in Equation (3), where T is the first time such that the Hamming distance changes, such that $\alpha = \frac{q - \lambda^* \Delta}{q - \Delta - 2} = \Theta(1)$, $W = 13$ and $\beta \leq \frac{qn}{q - \Delta - 2} \leq O(n)$*

With these tools in hand, we may now finally prove Lemma 24 and complete the proof of Theorem 5.

Proof of Lemma 24. First, note that the path generating set S generates the Hamming metric $d_H(\cdot, \cdot)$ on proper list-colorings. Now, given the variable-length path coupling furnished by Theorem 29, we use Theorem 28 to construct an M -step coupling with $M = \lceil \frac{2\beta W}{\alpha} \rceil \leq O(n)$ which contracts with rate $1 - \alpha$ every M steps, where α is a constant independent of n . Under this coupling, for every $k = 0, \dots, M - 1$ and every positive integer j , we have that

$$\begin{aligned}&\mathbb{E}_{\tau^{(jM+k)}, \sigma^{(jM+k)}} \left[d_H(\tau^{(jM+k)}, \sigma^{(jM+k)}) \mid \tau^{(k)}, \sigma^{(k)} \right] \\ &\leq \left(1 - \frac{\alpha}{2}\right) \mathbb{E}_{\tau^{((j-1)M+k)}, \sigma^{((j-1)M+k)}} \left[d_H(\tau^{((j-1)M+k)}, \sigma^{((j-1)M+k)}) \mid \tau^{(k)}, \sigma^{(k)} \right] \\ &\leq \dots \\ &\leq \left(1 - \frac{\alpha}{2}\right)^j \cdot d_H(\tau^{(k)}, \sigma^{(k)}),\end{aligned}$$

where $\tau^{(0)} = \tau, \sigma^{(0)} = \sigma$ are arbitrary starting states, which need not be neighbors under S .

It follows that

$$\begin{aligned}
& \sum_{t=0}^{\infty} \mathbb{E}_{\tau^{(t)}, \sigma^{(t)}} \left[d_H(\tau^{(t)}, \sigma^{(t)}) \mid \tau^{(0)} = \tau, \sigma^{(0)} = \sigma \right] \\
& \leq \sum_{k=0}^{M-1} \sum_{j=0}^{\infty} \mathbb{E}_{\tau^{(jM+k)}, \sigma^{(jM+k)}} \left[d_H(\tau^{(jM+k)}, \sigma^{(jM+k)}) \mid \tau^{(k)}, \sigma^{(k)} \right] \\
& \leq \sum_{k=0}^{M-1} \mathbb{E} \left[d_H(\tau^{(k)}, \sigma^{(k)}) \mid \tau^{(0)}, \sigma^{(0)} \right] \sum_{j=0}^{\infty} \left(1 - \frac{\alpha}{2}\right)^j \\
& = \frac{2}{\alpha} \sum_{k=0}^{M-1} \mathbb{E} \left[d_H(\tau^{(k)}, \sigma^{(k)}) \mid \tau^{(0)}, \sigma^{(0)} \right] \\
& \leq \frac{2M}{\alpha} d_H(\tau^{(0)}, \sigma^{(0)}) \tag{*} \\
& \leq O(n) \cdot d_H(\tau^{(0)}, \sigma^{(0)}).
\end{aligned}$$

To justify (*), note that T is the first time the Hamming distance changes, and that each time the Hamming distance changes, the expected Hamming distance contracts by a factor of $1 - \alpha$. ◀

6 Future Directions

Two concrete open problems are to bring down the required number of colors from $(\frac{11}{6} - \epsilon) \Delta$ to $\Delta + 2$, and to remove the bounded-degree assumption, both in this work and in [16]. Another interesting question is if spectral independence implies any useful notion of correlation decay, such as strong spatial mixing, or the absence of zeros for partition function in a large region. This is relevant particularly for proper list-colorings, where we showed spectral independence when $q \geq (\frac{11}{6} - \epsilon) \Delta$, but correlation decay and absence of zeros are both open in general when number of colors is below 2Δ .

We also reiterate that one feature of our approach which we haven't exploited is that in order to obtain $O(1)$ -spectral independence, it suffices for the Markov chain admitting the nice coupling to merely update $O(1)$ -coordinates in a single move “on average”, as opposed to the worst-case starting state; see Remark 10. We leave it to future work to see if this can be exploited.

References

- 1 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*, STOC 2020, page 1198–1211, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3357713.3384317.
- 2 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, page 433–446, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451123.
- 3 Nima Anari and Michal Dereziński. Isotropy and log-concave polynomials: Accelerated sampling and high-precision counting of matroid bases. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1331–1344, Los Alamitos, CA, USA, November 2020. IEEE Computer Society.

- 4 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 the Theory Computing*, STOC 2019, pages 1–12, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3313276.3316385.
- 5 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 the Theory Computing*, STOC 2021, New York, NY, USA, 2021. Association for Computing Machinery.
- 6 Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020. doi:10.1109/FOCS46700.2020.00125.
- 7 Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of markov chains: Coupling, spectral independence, and entropy factorization. *arXiv preprint arXiv:2103.07459*, 2021.
- 8 Sergey Bobkov and Prasad Tetali. Modified log-sobolev inequalities, mixing and hypercontractivity. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 287–296, New York, NY, USA, 2003. Association for Computing Machinery.
- 9 Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 1 edition, 2016.
- 10 Guy Bresler and Dheeraj Nagaraj. Stein’s method for stationary distributions of markov chains and application to ising models. *Annals of Applied Probability*, 29(5), 2019.
- 11 R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, FOCS '97, pages 223–, Washington, DC, USA, 1997. IEEE Computer Society. URL: <http://dl.acm.org/citation.cfm?id=795663.796353>.
- 12 R. Bubley and M. E. Dyer. Path coupling, dobrushin uniqueness, and approximate counting. Technical report, University of Leeds, 1997.
- 13 Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '19, page 2216–2234, USA, 2019. Society for Industrial and Applied Mathematics.
- 14 Z. Chen, K. Liu, and E. Vigoda. Rapid mixing of glauber dynamics up to uniqueness via contraction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1307–1318, 2020.
- 15 Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. *Rapid Mixing for Colorings via Spectral Independence*, page 1548–1557. Society for Industrial and Applied Mathematics, USA, 2021.
- 16 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, page 1537–1550, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451035.
- 17 Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to holant-type problems. *arXiv preprint arXiv:2106.03366*, 2021.
- 18 M. Cryan, H. Guo, and G. Mousa. Modified log-sobolev inequalities for strongly log-concave distributions. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1358–1370, 2019.
- 19 Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible markov chains. *Annals of Applied Probability*, 3(3):696–730, 1993.
- 20 Persi Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *Annals of Applied Probability*, 6(3), August 1996.

- 21 I. Dinur and T. Kaufman. High dimensional expanders imply agreement expanders. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 974–985, 2017.
- 22 R. L. Dobrushin and S. B. Shlosman. *Completely Analytical Gibbs Fields*, pages 371–403. Birkhäuser Boston, Boston, MA, 1985.
- 23 R. L. Dobrushin and S. B. Shlosman. *Constructive Criterion for the Uniqueness of Gibbs Field*, pages 347–370. Birkhäuser Boston, Boston, MA, 1985.
- 24 R. L. Dobrushin and S. B. Shlosman. Completely analytical interactions: Constructive description. *Journal of Statistical Physics*, 46(5):983–1014, 1987.
- 25 Roland Lvovich Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability and its Applications*, 15(3):458–486, 1970.
- 26 Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Matrix norms and rapid mixing for spin systems. *The Annals of Applied Probability*, 19(1):71–107, 2009.
- 27 Charilaos Efthymiou, Andreas Galanis, Thomas P. Hayes, Daniel Stefankovic, and Eric Vigoda. Improved Strong Spatial Mixing for Colorings on Trees. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, volume 145 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 48:1–48:16, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.48.
- 28 Ronen Eldan, James R. Lee, and Joseph Lehec. Transport-entropy inequalities and curvature in discrete-space markov chains. *A Journey Through Discrete Mathematics*, pages 391–406, 2017.
- 29 Tomás Feder and Milena Mihail. Balanced matroids. In *STOC*, pages 26–38, 1992.
- 30 Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. *Rapid Mixing from Spectral Independence beyond the Boolean Domain*, page 1558–1577. Society for Industrial and Applied Mathematics, USA, 2021.
- 31 Alan Frieze and Eric Vigoda. A survey on the use of markov chains to randomly sample colourings. *Combinatorica*, January 2007. doi:10.1093/acprof:oso/9780198571278.003.0004.
- 32 David Gamarnik, Dmitriy Katz, and Sidhant Misra. Strong spatial mixing of list coloring of graphs. *Random Struct. Algorithms*, 46(4):599–613, 2015. doi:10.1002/rsa.20518.
- 33 Sharad Goel. Modified logarithmic sobolev inequalities for some models of random walk. *Stochastic Processes and their Applications*, 114:51–79, 2004.
- 34 Heng Guo and Giorgos Mousa. Local-to-global contraction in simplicial complexes. *arXiv preprint arXiv:2012.14317*, 2020.
- 35 Thomas P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. *FOCS*, pages 39–46, 2006.
- 36 Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *Annals of Applied Probability*, 17(3):931–952, 2007.
- 37 Thomas P. Hayes and Eric Vigoda. Variable length path coupling. *Random Structures and Algorithms*, 31:251–272, 2007.
- 38 Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Struct. Algorithms*, 7(2):157–165, 1995.
- 39 Tali Kaufman and David Mass. High dimensional random walks and colorful expansion. In *ITCS*, pages 4:1–4:27, 2017.
- 40 Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. In *APPROX/RANDOM*, pages 47:1–47:17, 2018.
- 41 David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2 edition, 2017.
- 42 Katalin Marton. Logarithmic sobolev inequalities in discrete product spaces. *Combinatorics, Probability and Computing*, 28(6):919–935, 2019. doi:10.1017/S0963548319000099.

- 43 M. Mihail and U. Vazirani. On the expansion of 0/1 polytopes. *Journal of Combinatorial Theory*, 1989.
- 44 Yann Ollivier. Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis*, 256:810–864, February 2009.
- 45 Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders part i: Descent of spectral gaps. *Discrete and Computational Geometry*, 59(2):293–330, 2018.
- 46 Gesine Reinert and Nathan Ross. Approximating stationary distributions of fast mixing glauber dynamics, with applications to exponential random graphs. *Annals of Applied Probability*, 29, 2019.
- 47 Marcus D. Sammer. *Aspects of Mass Transportation in Discrete Concentration Inequalities*. PhD thesis, Georgia Institute of Technology, April 2005.
- 48 Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, pages 583–602, 1972.
- 49 Eric Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41, 2000.
- 50 Jian-Sheng Wang, Robert H. Swendsen, and Roman Kotecký. Antiferromagnetic potts models. *Phys. Rev. Lett.*, 63:109–112, July 1989. doi:10.1103/PhysRevLett.63.109.

A Variance and Entropy Decay

While we primarily use prior results on the spectral gap, and standard and modified log-Sobolev constants as blackboxes, to keep this paper self-contained, we define these constants here, and state their relevance to mixing and concentration. Fix a Markov kernel P on a finite state space Ω which reversible w.r.t. a distribution π . We may define an inner product using π by $\langle f, g \rangle_\pi = \mathbb{E}_\pi[fg]$. This inner product together with the kernel P induces a positive semidefinite quadratic form known as the Dirichlet form, defined as $\mathcal{E}_P(f, g) = \langle f, (I - P)g \rangle_\pi$. The variance of a function $f : \Omega \rightarrow \mathbb{R}$ is given by $\text{Var}_\pi(f) = \mathbb{E}_\pi(f^2) - \mathbb{E}_\pi(f)^2$, while the entropy of a function $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$ is given by $\text{Ent}_\pi(f) = \mathbb{E}_\pi(f \log f) - \mathbb{E}_\pi(f) \log \mathbb{E}_\pi(f)$.

With these notions in hand, we may now define the following constants:

$$\begin{aligned}
 \text{(Spectral Gap)} \quad \lambda(P) &\stackrel{\text{def}}{=} \inf_{f \neq 0} \frac{\mathcal{E}(f, f)}{\text{Var}_\pi(f)} \\
 \text{(Modified Log-Sobolev Constant)} \quad \rho(P) &\stackrel{\text{def}}{=} \inf_{f \geq 0} \frac{\mathcal{E}(f, \log f)}{\text{Ent}_\pi(f)} \\
 \text{(Standard Log-Sobolev Constant)} \quad \kappa(P) &\stackrel{\text{def}}{=} \inf_{f \geq 0} \frac{\mathcal{E}(\sqrt{f}, \sqrt{f})}{\text{Ent}_\pi(f)}. \tag{4}
 \end{aligned}$$

It is known that $4\kappa(P) \leq \rho(P) \leq 2\lambda(P)$ [8], with lower bounds on $\kappa(P)$ being the most difficult to establish. For the reader's convenience, we collect some well-known relations between these constants, mixing, and concentration.

► **Proposition 30 (Mixing and Concentration).** *We have the following bounds on the mixing time of a Markov chain with transition probability matrix P and stationary distribution π .*

$$t_{\text{mix}}(\epsilon) \leq \frac{1}{\lambda(P)} \left(\frac{1}{2} \log \frac{1}{\pi_{\min}} + \log \frac{1}{2\epsilon} \right) \tag{41}$$

$$t_{\text{mix}}(\epsilon) \leq \frac{1}{\rho(P)} \left(\log \log \frac{1}{\pi_{\min}} + \log \frac{1}{2\epsilon^2} \right) \tag{8}$$

$$t_{\text{mix}}(\epsilon) \leq \frac{1}{4\kappa(P)} \left(\log \log \frac{1}{\pi_{\min}} + \log \frac{1}{2\epsilon^2} \right). \tag{20}$$

32:20 Coupling to Spectral Independence

Furthermore, for every function $f : \Omega \rightarrow \mathbb{R}$ which is 1-Lipschitz w.r.t. graph distance under P , we have the following Chernoff-type concentration inequalities [33, 47, 9].

$$\Pr[f \geq \mathbb{E}_\pi(f) + \epsilon] \leq \exp\left(-\frac{\rho(P)\epsilon^2}{2v(f)}\right)$$

$$\Pr[f \leq \mathbb{E}_\pi(f) - \epsilon] \leq \exp\left(-\frac{\kappa(P)\epsilon^2}{2v(f)}\right),$$

where

$$v(f) \stackrel{\text{def}}{=} \max_{x \in \Omega} \left\{ \sum_{y \in \Omega} P(x \rightarrow y) \cdot (f(x) - f(y))^2 \right\}.$$

B Missing Proofs

Proof of Theorem 19. We show that $\eta_0 \leq \frac{4}{\alpha n} - 1$. The bound $\eta_k \leq \frac{4}{\alpha n} - 1$ follows by the same argument by instead considering the Glauber dynamics for the conditional distributions $\mu \mid A$ of μ . Because the Glauber dynamics only updates at most one coordinate in each step, it is 2-local w.r.t. $d_H(\cdot, \cdot)$. By Fact 18, we also have there is a C -amortized convergent coupling with $C = \frac{1}{\alpha}$. It follows from Theorem 7 that

$$\sum_{v \in V} \sum_{c' \in \Omega(v)} \left| \Pr_{\sigma \sim \mu} [\sigma(v) = c' \mid \sigma(u) = c] - \Pr_{\sigma \sim \mu} [\sigma(v) = c'] \right|$$

$$\leq \frac{2}{\alpha} \max_{\sigma \in \text{supp}(\mu \mid uc)} \sum_{\tau \neq \sigma} |P_\mu(\sigma \rightarrow \tau) - P_{\mu \mid uc}(\sigma \rightarrow \tau)|.$$

Now, by the definition of the Glauber dynamics, for each $\sigma \in \text{supp}(\mu \mid uc)$, we have

$$\sum_{\tau \neq \sigma} |P_\mu(\sigma \rightarrow \tau) - P_{\mu \mid uc}(\sigma \rightarrow \tau)|$$

$$= \sum_{v \in V} \sum_{c' \in L(v): c' \neq \sigma(v)} \left| \frac{1}{n} \mu^v(c' \mid \sigma_{-v}) - \frac{1}{n-1} \mu_{uc}^v(c' \mid \sigma_{-v}) \right|$$

$$= \sum_{v \in V: v \neq u} \sum_{c' \in L(v): c' \neq \sigma(v)} \left(\frac{1}{n-1} - \frac{1}{n} \right) \mu^v(c' \mid \sigma_{-v}) + \sum_{c' \in L(u): c' \neq c} \frac{1}{n} \mu^v(c' \mid \sigma_{-v})$$

$$\leq \frac{2}{n}.$$

The claim for the spectral gap in the case $\alpha \geq \Omega(1/n)$ follows by combining with Theorem 14. The final claim for spin systems on bounded-degree graphs follows by combining with Theorem 15. \blacktriangleleft

Proof of Lemma 23. The main detail one must be careful of is that the flip dynamics for sampling from $\mu \mid uc$ always leaves the color for u fixed to c . Hence, flipping any Kempe component containing u leads to potentially different list-colorings under $P_{\mu, \text{flip}}$ versus $P_{\mu \mid uc, \text{flip}}$. However, since we only flip components of $O(1)$ -size, this isn't an issue for us.

Fix a τ with $\tau(u) = c$, and let $B(u, 6)$ denote the set of vertices of shortest path distance at most 6 away from u in G . Since we only flip Kempe components of size at most 6, we have that for any $v \in V \setminus B(u, 6)$ and $c \in L(u)$, the flippable Kempe component $S_\tau(v, c')$

does not contain u , and hence, flipping it leads to the same list-coloring under $P_{\mu, \text{flip}}$ and $P_{\mu|vc, \text{flip}}$. Hence, we have

$$\begin{aligned}
& \sum_{\sigma \neq \tau} |P_{\mu, \text{flip}}(\tau \rightarrow \sigma) - P_{\mu|uc, \text{flip}}(\tau \rightarrow \sigma)| \\
&= \sum_{v \in V: v \notin B(u, 6)} \sum_{c' \in L(v)} \frac{1}{|L(v)|} \cdot \left(\frac{1}{n} - \frac{1}{n-1} \right) \cdot p_{|S_\tau(v, c')|} \\
&\quad + \sum_{v \in B(u, 6)} \sum_{c' \in [q]} |P_{\mu, \text{flip}}(\tau \rightarrow \sigma) - P_{\mu|uc, \text{flip}}(\tau \rightarrow \sigma)| \\
&\leq \frac{n - |B(u, 6)|}{n(n-1)} + \frac{|B(u, 6)|}{n} \\
&\leq \frac{|B(u, 6)| + 1}{n} \\
&\lesssim \frac{\Delta^6}{n} \\
&\leq O(1/n). \qquad \qquad \qquad \text{(Bounded-degree assumption)}
\end{aligned}$$

◀

► Remark 31. As one can see in the proof from the factor of Δ^6 , we have made no attempt to optimize constants.