OPTIMAL MIXING OF MARKOV CHAINS FOR SPIN SYSTEMS VIA SPECTRAL INDEPENDENCE

A Dissertation
Presented to
The Academic Faculty

By

Zongchen Chen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Algorithms, Combinatorics, and Optimization (ACO) Program

School of Computer Science
Georgia Institute of Technology

August 2021

OPTIMAL MIXING OF MARKOV CHAINS FOR SPIN SYSTEMS VIA SPECTRAL INDEPENDENCE

Thesis committee:

Dr. Eric Vigoda (Advisor)
School of Computer Science
Georgia Institute of Technology

Dr. Prasad Tetali
School of Mathematics
Georgia Institute of Technology

Dr. Dana Randall
School of Computer Science
Georgia Institute of Technology

Dr. Santosh Vempala
School of Computer Science
Georgia Institute of Technology

Dr. Daniel Štefankovič (Reader)
Department of Computer Science
University of Rochester

Date approved: July 16, 2021

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my advisor Eric Vigoda for his supports and encouragements during the five years of my Ph.D. study. I am grateful to his patient and valuable guidance in research as well as in academic life. It is my fortune to have Eric as my advisor. I also want to thank Daniel Štefankovič, who I have been working with since my first year and have taught me a lot of valuable things. I want to thank my committee members for their time and service: Dana Randall, Daniel Štefankovič, Prasad Tetali, Santosh Vempala, and Eric Vigoda.

I would like to thank my collaborators for their discussions, inspirations, and sharing: Ivona Bezáková, Antonio Blanca, Pietro Caputo, Amin Coja-Oghlan, Charilaos Efthymiou, Andreas Galanis, Leslie Ann Goldberg, Kuikui Liu, Daniel Parisi, Will Perkins, Daniel Štefankovič, James Stewart, Prasad Tetali, Santosh Vempala, Eric Vigoda.

I would like to thank all my friends for their company and helps. I would like to thank everyone at Georgia Tech who has provided suggestions or helps to me in the five years. I also want to thank the ACO program for providing great resources and academic environment. Especially, I would like to thank Robin Thomas and Prasad Tetali for leading the program and helping me in various occasions during the five years.

Last but not least, I would like to thank my parents for their selfless and unconditional love, and also all my family members for their love and supports.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Spin systems, or *undirected graphical models*, are important tools for modeling joint distributions of discrete random variables, and are broadly used in statistical physics, machine learning, and theoretical computer science. One of the most important computational tasks for spin systems is to generate random samples approximately from the equilibrium distribution of the model called the *Gibbs distribution*. The approximate sampling problem could be challenging in the high-dimensional setting, where the dimension is the number of variables and the state space is exponentially large in dimension. We study the single-site update Markov chain known as the *Glauber dynamics* or *Gibbs sampling* for sampling from the Gibbs distribution. In each step, the dynamics picks a single variable uniformly at random and updates it conditional on all other variables. The Glauber dynamics is widely used in many practical situations due to its simplicity and efficiency. Despite its popularity, theoretical guarantees of the convergence rate of the Glauber dynamics are only known in a few special cases.

We prove optimal mixing time of the Glauber dynamics in a variety of settings. As an application of our results, for the hardcore model (weighted independent sets) on any n-vertex graph of constant maximum degree, we establish $O(n \log n)$ mixing time of the Glauber dynamics when the fugacity (vertex weight) lies in the tree uniqueness region. For the Ising model, and more generally any antiferromagnetic 2-spin system, we prove $O(n \log n)$ mixing time of the Glauber dynamics on any bounded degree graph in the corresponding tree uniqueness region. Our results apply more broadly; for example, we also obtain $O(n \log n)$ mixing for q-colorings on graphs of maximum degree Δ when the number of colors satisfies $q > (\frac{11}{6} - \varepsilon_0)\Delta$ where $\varepsilon_0 \approx 10^{-5}$, and $O(n \log n)$ mixing for the monomer-dimer model (weighted matchings) and weighted partial even subgraphs (corresponding to the ferromagnetic Ising model with nonzero external fields) on any graph with bounded degrees.

Our work presents an improved version of the spectral independence approach of Anari, Liu, and Oveis Gharan (2020). Roughly speaking, a distribution is said to be *spectrally independent* if the maximum eigenvalues of the associated influence matrices are upper bounded. We show $O(n \log n)$ mixing time of the Glauber dynamics for spectrally independent spin systems on any n-vertex graph of bounded degrees. Furthermore, we prove optimal mixing results for arbitrary heat-bath block dynamics and (for ferromagnetic Ising and Potts models) the Swendsen-Wang dynamics. We also show that spectral independence can be established using current algorithmic tools of approximate sampling and counting, including coupling of Markov chain Monte Carlo (MCMC) methods, correlation decay approach, and polynomial interpolation approach. This in particular demonstrates the importance and power of spectral independence for connecting different aspects in the study of spin systems.

CHAPTER 1
INTRODUCTION

In many scientific settings we often use a statistical model to describe the dependencies among a set of variables. This can be represented by a graph which contains a set of nodes corresponding to the variables of the system and a set of edges connecting pairs of nodes which represent the interrelationship between these variables; these models are known as *spin systems* or *undirected graphical models*. Spin systems arise in a wide variety of scientific fields; for example they correspond to phylogenetic trees in evolutionary biology and are used in network science for community detection. More recently, spin systems are extensively applied in machine learning and data science, such as the restricted Boltzmann machine and Bayesian network.

Despite the wide usage of spin systems, many technical questions remain open even for the basic ones. One fundamental task in the study of spin systems is to simulate the equilibrium state of the system; more specifically, to sample from the corresponding Gibbs distribution of the model. The *Markov chain Monte Carlo (MCMC)* method is a standard and popular approach for sampling from the equilibrium distribution. In every step, the algorithm will update the current configuration randomly under appropriate rules, such that the distribution will eventually converge to the desired Gibbs distribution. The mixing time of a MCMC dynamics is defined to be the number of steps it takes to be sufficiently close to the stationary distribution in total variation distance.

MCMC methods are usually simple, elegant, and very efficient in practice for sampling Gibbs distributions of spin systems. However, few rigorous guarantees are known in the literature on the convergence rate of MCMC, and typically heuristic arguments are used to measure convergence with no assurance on the statistical accuracy. The main contribution of this thesis is to establish a well-known conjecture that the Glauber dynamics converges

very quickly to its stationary distribution in the tree uniqueness region, i.e., decay of correlations region. The Glauber dynamics is the quintessential example of a local Markov chain, and its convergence rate is of great interest due to its simplicity and wide applicability.

Our setting is the general framework of spin systems. Spin systems capture many combinatorial models of interest, including the hardcore model on weighted independent sets, the Ising model, and colorings, and are equivalent to undirected graphical models. For integer $q \geq 2$, a q-state spin system is defined by a $q \times q$ interaction matrix A. For a given graph $G = (V, E)$ with $n = |V|$ vertices, the configurations of the model are the collection Ω of assignments $\sigma : V \to [q]$ of spins to the vertices of the graph. Each configuration $\sigma \in \Omega$ has an associated weight $w(\sigma)$ defined by the pairwise interactions weighted by the interaction matrix A, see Section 2.1 for a detailed definition.

The Gibbs distribution μ is the probability distribution over the collection Ω of configurations and is defined as $\mu(\sigma) = w(\sigma)/Z$ where $Z = \sum_\sigma w(\sigma)$ is the normalizing factor known as the partition function. Approximately sampling from the Gibbs distribution is polynomial-time equivalent to approximating the partition function [79, 122]. Given an $\varepsilon > 0$ and $\delta > 0$, a fully polynomial-time randomized approximation scheme (FPRAS) for the partition function outputs a $(1 \pm \varepsilon)$-relative approximation of the partition function with probability $\geq 1 - \delta$, whereas a fully polynomial-time approximation scheme (FPTAS) is the deterministic analog (i.e., it achieves $\delta = 0$).

The canonical example of a spin system in statistical physics is the Ising model. The Ising model is a 2-spin system (i.e., $q = 2$); the spin space is denoted as $\{+, -\}$ and the configurations of the model are the 2^n assignments of spins $\{+, -\}$ to the vertices of the underlying graph. In the simpler case without an external field the Ising model has a single parameter $\beta > 0$ corresponding to the inverse temperature. A configuration $\sigma \in \Omega$ has weight $w(\sigma) = \beta^{m(\sigma)}$ where $m(\sigma) = |\{(u, v) \in E : \sigma(u) = \sigma(v)\}|$ is the number of monochromatic edges in σ. When $\beta > 1$ then the model is *ferromagnetic* as the two

fully monochromatic configurations have maximum weight, whereas when $\beta < 1$ then the model is *antiferromagnetic*.

The hardcore model is a natural combinatorial example of an antiferromagnetic 2-spin system. The model is parameterized by a fugacity $\lambda > 0$. For a graph $G = (V, E)$, configurations of the model are the collection Ω of independent sets of G, and the weight of an independent set σ is $w(\sigma) = \lambda^{|\sigma|}$.

In general, a 2-spin system is defined by three parameters $\beta, \gamma \geq 0$ and $\lambda > 0$. A spin configuration $\sigma \in \{0,1\}^V$ is assigned weight: $w(\sigma) = \beta^{m_1(\sigma)}\gamma^{m_0(\sigma)}\lambda^{n_1(\sigma)}$, where, for $s \in \{0,1\}$, $m_s(\sigma)$ is the number of edges where both endpoints receive spin s and $n_s(\sigma)$ is the number of vertices assigned spin s. Note the Ising model corresponds to the case $\beta = \gamma$ where λ is the external field, and the hardcore model corresponds to $\beta = 0, \gamma = 1$. The model is ferromagnetic when $\beta\gamma > 1$ and antiferromagnetic when $\beta\gamma < 1$ (the model is trivial when $\beta\gamma = 1$).

The Glauber dynamics is a simple Markov chain (X_t) designed for sampling from the Gibbs distribution μ. The transitions $X_t \to X_{t+1}$ update a randomly chosen vertex as follows: (i) select a vertex v uniformly at random; (ii) for all $u \neq v$, set $X_{t+1}(u) = X_t(u)$; and (iii) choose $X_{t+1}(v)$ from the marginal distribution for the spin at v conditional on the configuration $X_{t+1}(N(v))$ on the neighbors $N(v)$ of v. It is straightforward to verify that the chain is ergodic (in the cases considered here, see the definition of totally-connected in Section 2.1) and the unique stationary distribution is the Gibbs distribution.

The *mixing time* is the number of transitions, for the worst initial state X_0, to guarantee that X_t is within total variation distance $\leq 1/4$ of the Gibbs distribution; for a formal statement, see Eq. (2.1). We say the chain is *rapidly mixing* when the mixing time is polynomial in $n = |V|$. Hayes and Sinclair [71] established that the mixing time of the Glauber dynamics is $\Omega(n \log n)$ for a family of bounded-degree graphs, and hence we say that the Glauber dynamics has *optimal* mixing time when the mixing time is $O(n \log n)$.

The computational complexity of approximating the partition function is closely con-

nected to statistical physics phase transitions. For $\Delta \geq 3$, consider the tree T_ℓ of height ℓ where all of the internal vertices have degree Δ, and let r denote its root. The uniqueness vs non-uniqueness phase transition captures whether the leaves influence the root, in the limit as the height grows.

The uniqueness/non-uniqueness phase transition is nicely illustrated for the Ising model which has two extremal boundaries: the all $+$ boundary and all $-$ boundary. For $s \in \{+,-\}$, let p_ℓ^s denote the marginal probability that the root has spin $+$ in the Gibbs distribution on T_ℓ conditional on all leaves having spin s. The model is in the uniqueness phase iff $\lim_{\ell\to\infty} p_\ell^+ = \lim_{\ell\to\infty} p_\ell^-$. For the Ising model (without an external field) the uniqueness/non-uniqueness phase transition occurs at $\beta_c(\Delta) = (\Delta-2)/\Delta$ for the antiferromagnetic case and $\overline{\beta}_c(\Delta) = \Delta/(\Delta-2)$ for the ferromagnetic case. For the hardcore model, the critical fugacity is $\lambda_c(\Delta) := (\Delta-1)^{\Delta-1}/(\Delta-2)^\Delta$. This phase transition on the Δ-regular tree is connected to the complexity of approximating the partition function on graphs of maximum degree Δ.

For the hardcore model, for constant Δ, for any $\delta > 0$, Weitz [127] presented an FPTAS for the partition function on graphs of maximum degree Δ when $\lambda < (1-\delta)\lambda_c(\Delta)$. In contrast, when $\lambda > \lambda_c(\Delta)$, Sly [120] (see also [121, 58]) showed that, unless NP $=$ RP, there is no FPRAS for approximating the partition function on graphs of maximum degree Δ. Sinclair, Srivastava, and Thurley [119] extended Weitz's correlation decay algorithmic approach to the antiferromagnetic Ising model in the tree uniqueness region, and Li, Lu, and Yin [89] further generalized it to all antiferromagnetic 2-spin systems when the system is up-to-Δ unique. One important caveat to these correlation decay approaches is that the running time depends exponentially on $\log\Delta$ and $1/\delta$.

Despite the algorithmic successes of the correlation decay approach, establishing rapid mixing of the Glauber dynamics in the same tree uniqueness region was a vexing open problem. Anari, Liu, and Oveis Gharan [4] introduced the spectral independence approach based on the theory of high-dimensional expanders [80, 49, 82, 108, 1], and established

rapid mixing of the Glauber dynamics for the hardcore model on any graph of maximum degree Δ when $\lambda < (1-\delta)\lambda_c(\Delta)$ for $\delta > 0$. However, while the mixing time had polynomial dependence on Δ, it also had doubly exponential dependence on $1/\delta$. In [44] the authors established rapid mixing for all antiferromagnetic 2-spin systems when the system is up-to-Δ-unique with gap δ and improved the mixing time to an exponential dependence on $1/\delta$. Here, roughly speaking, up-to-Δ uniqueness with gap δ means (multiplicative) gap δ from the uniqueness threshold on the Δ-regular tree for all $d \leq \Delta$; see Section 5.3 for a precise statement, and [89] for more discussion.

In this thesis, we not only establish a fixed polynomial upper bound on the mixing time, but we also prove optimal mixing of the Glauber dynamics. Our approach holds for general spin systems. The spectral independence approach, first introduced for 2-spins in [4] and subsequently extended to q-spins in [43, 55], considers the $qn \times qn$ influence matrix. For spins $i, j \in [q]$ and vertices $u, v \in V$, the entry $((u,i),(v,j))$ of the influence matrix measures the effect of vertex u having spin i on the marginal probability that vertex v has spin j, see Definition 2.1.3 for a precise statement. Here we prove that if the maximum eigenvalue of the influence matrix is upper bounded and the marginal probabilities are lower bounded then the mixing time is $O(n \log n)$ where the only dependence on $1/\delta$ and Δ is in the constant factor captured by the big-O notation.

We establish optimal mixing time of $O(n \log n)$ by proving that the Glauber dynamics contracts relative entropy (with respect to the Gibbs distribution) at a constant rate. This is analogous to establishing a modified log-Sobolev constant for the Glauber dynamics; there are several recent results in other contexts also proving entropy decay for various Markov chains [47, 37, 24]. In contrast, previous works utilizing the spectral independence approach [4, 44] and related works on high-dimensional expanders [80, 49, 108, 82, 1] consider the spectral gap (or analogously, decay of variance); such an approach is unable to establish optimal mixing time.

We give our main results in Section 1.1 regarding optimal mixing of Glauber dynamics

for many classes of spin systems. In Section 1.2, we introduce the notion of spectral independence, discuss its implications for MCMC, and show various methods of establishing spectral independence. The structure of this thesis can be found in Section 1.3.

1.1 Optimal Mixing of Glauber Dynamics

The application of our results is nicely illustrated for the particular case of antiferromagnetic 2-spin systems. We prove $O(n \log n)$ mixing time of the Glauber dynamics when the system is up-to-Δ-unique. This is the same region where the correlation decay results of [89] and the rapid mixing results of [44] hold, which matches the hardness results of [121]. Note, a mixing time of $O(n \log n)$ implies an $\widetilde{O}(n^2)$ time FPRAS for approximating the partition function [122, 83]. For the case of the hardcore model we have the following result.

Theorem 1.1.1 (Hard-core Model)**.** *Let $\Delta \geq 3$ be an integer and let $\delta \in (0,1)$ be a real. For every n-vertex graph G of maximum degree Δ and every $0 < \lambda \leq (1-\delta)\lambda_c(\Delta)$, the mixing time of the Glauber dynamics for the hardcore model on G with fugacity λ is at most $Cn \log n$ where $C = C(\Delta, \delta)$ is a constant independent of n.*

For the case of the Ising model in both the antiferromagnetic and ferromagnetic case, we obtain optimal mixing whenever β is between the two thresholds $\beta_c(\Delta) = \frac{\Delta-2}{\Delta}$ and $\overline{\beta}_c(\Delta) = \frac{\Delta}{\Delta-2}$.

Theorem 1.1.2 (Ising model)**.** *Let $\Delta \geq 3$ be an integer and let $\delta \in (0,1)$ be a real. For every n-vertex graph G of maximum degree Δ, every $\beta \in [\frac{\Delta-2+\delta}{\Delta-\delta}, \frac{\Delta-\delta}{\Delta-2+\delta}]$, and every $\lambda > 0$, the mixing time of the Glauber dynamics for the Ising model on G with edge activity β and external field λ is at most $Cn \log n$ where $C = C(\Delta, \delta)$ is a constant independent of n.*

Recall that the above results are tight as there is no efficient approximation algorithm in the tree non-uniqueness region which corresponds to $\lambda > \lambda_c(\Delta)$ for the hardcore model and $\beta < \beta_c(\Delta)$ for the antiferromagnetic Ising model. The only analog of the above results

establishing optimal mixing time in the entire tree uniqueness region was the work of Mossel and Sly [106] for the ferromagnetic Ising model. Their proof utilizes the monotonicity properties of the ferromagnetic Ising model which allows the use of the censoring inequality of Peres and Winkler [111]. The algorithm of Jerrum and Sinclair [77] gives an FPRAS for the ferromagnetic Ising model for any β and any G, but the polynomial exponent is a large constant.

Both Theorems 1.1.1 and 1.1.2 are special cases of the following optimal mixing result for general antiferromagnetic 2-spin systems in the entire tree uniqueness region. For general 2-spin systems the appropriate tree phase transition is more complicated as there are models where the tree uniqueness threshold is not monotone in Δ. Hence the appropriate notion is "up-to-Δ uniqueness" as considered by [89]. Roughly speaking, we say uniqueness with gap $\delta \in (0,1)$ holds on the d-regular tree if for every integer $\ell \geq 1$, all vertices at distance ℓ from the root have total "influence" $\lesssim (1-\delta)^\ell$ on the marginal of the root. We say up-to-Δ uniqueness with gap δ holds if uniqueness with gap δ holds on the d-regular tree for all $1 \leq d \leq \Delta$; see Section 5.3 for the precise definition.

Theorem 1.1.3 (Antiferromagnetic 2-Spin Systems)**.** *Let $\Delta \geq 3$ be an integer and let $\delta \in (0,1)$ be a real. Let (β, γ, λ) with $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$ be parameters specifying an antiferromagnetic 2-spin system which is up-to-Δ unique with gap δ. For every n-vertex graph G of maximum degree Δ, the mixing time of the Glauber dynamics for the antiferromagnetic 2-spin system on G with parameters (β, γ, λ) is at most $Cn \log n$ where $C = C(\Delta, \delta, \beta, \gamma, \lambda)$ is a constant independent of n.*

For general ferromagnetic 2-spin systems the existing picture is not as clear as for antiferromagnetic systems. Our work extends to ferromagnetic 2-spin systems, proving $O(n \log n)$ mixing time for the same range of parameters as the previously best known bounds [67, 118, 44]. In particular, we recover Theorems 26 and 27 in [44] with $O(n \log n)$ mixing time.

Our results hold for multi-spin systems as well. The most notable example of a multi-

spin system is the q-colorings problem, namely, proper vertex q-colorings. Given a graph $G = (V, E)$ of maximum degree Δ, can we approximate the number of q-colorings of G, or generate a q-coloring of G approximately uniformly at random? Jerrum [74] proved $O(n \log n)$ mixing time of the Glauber dynamics whenever $q > 2\Delta$. This was further improved in [42, 125] to $O(n^2)$ mixing time when $q > (11/6 - \varepsilon_0)\Delta$ for some small $\varepsilon_0 > 0$. There are several further improvements with various assumptions on the girth or maximum degree, c.f. [51]. On the hardness side, Galanis et al. [57] proved that unless $\mathsf{NP} = \mathsf{RP}$ there is no FPRAS for approximating the number of q-colorings when q is even and $q < \Delta$. For triangle-free graphs, a recent pair of works [55, 43] extended the spectral independence approach to establish rapid mixing of the Glauber dynamics when $q > (\alpha^* + \delta)\Delta$ for any $\delta > 0$ where $\alpha^* \approx 1.763$; however the polynomial exponent in the mixing time depends on $1/\delta$ in these results.

We prove $O(n \log n)$ mixing time of the Glauber dynamics under the same conditions as previous results mentioned above.

Theorem 1.1.4 (Colorings)**.** *Let $\Delta \geq 3$ be an integer and let $\delta > 0$ be a real. For every n-vertex graph G of maximum degree Δ and every integer $q \geq 3$, if one of the following holds:*

(1) $q \geq (\frac{11}{6} - \varepsilon_0 + \delta)\Delta$ where $\varepsilon_0 \approx 10^{-5}$ is a universal constant given in [42];

(2) G is triangle-free and $q \geq (\alpha^* + \delta)\Delta$ where $\alpha^* \approx 1.763$ is the unique solution to $x = \exp(1/x)$;

then the mixing time of the Glauber dynamics for sampling uniformly random q-colorings of G is at most $Cn \log n$ where $C = C(\Delta, \delta)$ is a constant independent of n.

We prove optimal mixing time bounds for the monomer-dimer model on all matchings of a graph with constant maximum degree. Given a graph $G = (V, E)$ and a fugacity $\lambda > 0$, the Gibbs distribution μ for the monomer-dimer model is defined on the collection $\mathcal{M}$ of all matchings of G where $\mu(M) = w(M)/Z$ for $w(M) = \lambda^{|M|}$. The Glauber dynamics for

the monomer-dimer model adds or deletes a random edge in each step. In particular, from $X_t \in \mathcal{M}$, choose an edge e uniformly at random from E and let $X' = X_t \oplus e$. If $X' \in \mathcal{M}$ then let $X_{t+1} = X'$ with probability $w(X')/(w(X')+w(X_t))$ and otherwise let $X_{t+1} = X_t$. We prove $O(n \log n)$ mixing time for the Glauber dynamics for sampling matchings on bounded-degree graphs with n vertices. A classical result of Jerrum and Sinclair [76] yields rapid mixing of the Glauber dynamics for any graph even with unbounded degrees, but the best mixing time bound was $O(n^2 m \log n)$ [75] where m is the number of edges.

Theorem 1.1.5 (Monomer-Dimer Model)**.** *Let $\Delta \geq 3$ be an integer and let $\lambda > 0$ be a real. For every n-vertex graph G of maximum degree Δ, the mixing time of the Glauber dynamics for the monomer-dimer model on G with fugacity λ is at most $Cn \log n$ where $C = C(\Delta, \lambda)$ is a constant independent of n.*

We also establish optimal mixing time of the Glauber dynamics for sampling random weighted (partial) even subgraphs, which corresponds to the Gibbs distribution for the high-temperature expansion of the ferromagnetic Ising model. Let $G = (V, E)$ be a graph. The Gibbs distribution μ for weighted even subgraphs is defined on all subsets of edges. For $\lambda > 0$ and $\rho \in [0, 1]$, every subset $S \subseteq E$ has probability density $\mu(S) \propto \rho^{|\mathsf{odd}(S)|}\lambda^{|S|}$ where $\mathsf{odd}(S)$ is the set of odd-degree vertices in the subgraph (V, S). The weighted even subgraphs model corresponds to the ferromagnetic Ising model by the relations $\beta_{\text{Ising}} = \frac{1+\lambda}{1-\lambda}$ and $\lambda_{\text{Ising}} = \frac{1+\rho}{1-\rho}$, for which one can easily transform a subset of edges from μ to a sample from μ_{Ising} [64]. Notice that for $\rho = 0$ one obtains the ferromagnetic Ising model without external fields (i.e., $\lambda_{\text{Ising}} = 1$).

In [77], an MCMC algorithm is presented to sample weighted even subgraphs of an arbitrary (unbounded-degree) graph in time $O(m^3 \text{poly}(1/\rho))$ where m is the number of edges of the graph. Here, we prove that the Glauber dynamics converges in $O(n \log n)$ steps for bounded-degree graphs when $\rho > 0$; unfortunately the constant hidden in the big-O notation depends exponentially on $1/\rho$.

Theorem 1.1.6 (Weighted Even Subgraphs; Ferromagnetic Ising Model)**.** *Let $\Delta \geq 3$ be an integer and let $\lambda > 0$, $\rho \in (0, 1]$ be reals. For every n-vertex graph G of maximum degree Δ, the mixing time of the Glauber dynamics for sampling random weighted even subgraphs of G with parameters (λ, ρ) is at most $Cn \log n$ where $C = C(\Delta, \lambda, \rho)$ is a constant independent of n. In particular, we get an approximate sampling algorithm with running time $O(n \log n)$ for the ferromagnetic Ising model with edge activity $\beta_{\text{Ising}} = \frac{1+\lambda}{1-\lambda}$ and external field $\lambda_{\text{Ising}} = \frac{1+\rho}{1-\rho}$.*

Finally, we mention that our techniques imply asymptotically optimal bounds (up to constant factors) on both the standard and modified log-Sobolev constants of the Glauber dynamics for spin systems on bounded degree graphs in all of the regimes mentioned above. This also applies for certain problems where prior works have obtained rapid mixing via other techniques such as path coupling and canonical paths.

1.2 Spectral Independence: A Powerful Tool for Analyzing MCMC

The central tool for establishing optimal mixing time bounds for the Glauber dynamics is the notion of spectral independence, which was first proposed by Anari, Liu, and Oveis Gharan [4] building upon techniques for high-dimensional expanders [1].

1.2.1 Implications of Spectral Independence

To introduce spectral independence, we consider the hardcore model as a representative example and refer to Section 2.1 for general settings and definitions. Recall that, for a graph $G = (V, E)$ and a real $\lambda > 0$, the Gibbs distribution μ_G of the hardcore model on G with fugacity λ is a distribution over $\mathcal{I}(G)$, the collection of all independent sets of G. Every independent set $I \in \mathcal{I}(G)$ is assigned probability density $\mu_G(I) = \lambda^{|I|}/Z_G$ where the normalizing constant $Z_G = \sum_{I \in \mathcal{I}(G)} \lambda^{|I|}$ is called the partition function.

Spectral independence describes how much vertices interact among each other under the Gibbs distribution μ_G in a global and spectral manner. Let I be a random (weighted)

independent set generated from μ_G. For two distinct vertices $u, v \in V$, we define the (pairwise) influence of u on v to be

$$\Psi_G(u, v) = \mu_G(v \in I \mid u \in I) - \mu_G(v \in I \mid u \notin I);$$

in other words, $\Psi_G(u, v)$ is the difference of the probabilities of $v \in I$ given that u is contained in I or not. Intuitively, the influence becomes small when the graph distance between u and v is large, and it is maximized when u, v are adjacent. We further let $\Psi_G(v, v) = 0$ for each $v \in V$. The influence matrix Ψ_G is a $|V| \times |V|$ square matrix with entries given as above. It is not hard to check that all eigenvalues of Ψ_G are real; see [4].

Roughly speaking, η-spectral independence means that the maximum eigenvalue of the influence matrix Ψ_G, denoted by $\lambda_1(\Psi_G)$, is upper bounded by η. To understand this, one may consider the infinity norm $\|\Psi_G\|_\infty = \max_{u \in V} \sum_{v \in V} |\Psi_G(u, v)|$ which is no less than the maximum eigenvalue. If $\|\Psi_G\|_\infty = O(1)$, then one has $O(1)$-spectral independence. Therefore, spectral independence holds with an $O(1)$ constant if, for every vertex u, the sum of absolute influences of u on all other vertices is small, which happens when μ_G is close to a product distribution or satisfies the decay of correlations. In fact, in many applications we establish spectral independence by upper bounding $\|\Psi_G\|_\infty$.

To be more precise, for spectral independence we also need to consider the distribution under pinnings. For the hardcore model, this means that we will fix some vertices to be included in I and some to be excluded, and consider the distribution conditioned on this event. For hardcore, this is convenient for us, since fixing $v \notin I$ is equivalent to removing v from G and fixing $v \in I$ corresponds to removing both v and all its neighbors. Hence, we only need to consider all induced subgraphs of G. For $U \subseteq V$, let $H = G[U]$ be the subgraph induced by U. We consider the hardcore model on H with fugacity λ. The Gibbs distribution is denoted by μ_H and the influence matrix is denoted by Ψ_H. We say μ_G is *η-spectrally independent* if for every $U \subseteq V$ and $H = G[U]$, one has $\lambda_1(\Psi_H) \leq \eta$.

For our main result we also require a bound on marginals at every vertex under arbitrary pinnings. We say μ_G is *b-marginally bounded* if for every $U \subseteq V$ and $H = G[U]$, one has $\min\{\mu_H(v \in I), \mu_H(v \notin I)\} \geq b$ for all $v \in U$. This marginal boundedness condition is easily satisfied for spin systems defined on bounded-degree graphs.

Our main result is that if the Gibbs distribution on a bounded-degree graph is both marginally bounded and spectrally independent, then the Glauber dynamics satisfies the modified log-Sobolev inequality with constant $\Omega(1/n)$ (see Definition 2.2.3) and mixes in $O(n \log n)$ steps, where n is the number of vertices of the graph.

Theorem 1.2.1. *Let $\Delta \geq 3$ be an integer and $b, \eta > 0$ be reals. Suppose that $G = (V, E)$ is an n-vertex graph of maximum degree at most Δ and μ is a totally-connected Gibbs distribution of some spin system on G. If μ is both b-marginally bounded and η-spectrally independent, then the Glauber dynamics for sampling from μ satisfies the modified log-Sobolev inequality with constant $\frac{1}{C_1 n}$ where*

$$C_1 = \left(\frac{\Delta}{b}\right)^{O\left(\frac{\eta}{b}+1\right)}.$$

Furthermore, the mixing time of the Glauber dynamics satisfies

$$T_{\text{mix}}(P_{\text{GL}}, \varepsilon) = \left(\frac{\Delta}{b}\right)^{O\left(\frac{\eta}{b}+1\right)} \times O\left(n \log\left(\frac{n}{\varepsilon}\right)\right).$$

Previous results [4, 1] could only obtain $\text{poly}(\Delta) \times n^{\eta+2}$ mixing but without the assumption of marginal boundedness. In the setting of spin systems, we always have b-marginal boundedness with b depending only on the parameters of the spin system and the maximum degree Δ of the graph, and so our results supersede those of [4, 1] in the bounded degree regime.

We further show that spectral independence implies optimal mixing of arbitrary heat-bath block dynamics and for ferromagnetic Ising/Potts models the Swendsen-Wang dy-

namics; see Chapter 4 for definitions and results.

1.2.2 Establishing Spectral Independence

For approximate sampling and counting in the setting of spin systems, there are three disparate algorithmic methods: MCMC, correlation decay, and polynomial interpolation. In this thesis, we show that tools and properties establishing these methods can also be used to establish spectral independence and thus optimal mixing time results of the Glauber dynamics.

The correlation decay method was presented in [127] to obtain an FPTAS for approximating the partition function of the hardcore model up to the tree uniqueness threshold. It was extended to the antiferromagnetic Ising model [119] and to all antiferromagnetic 2-spin systems [89] in the corresponding tree uniqueness regions. These results build upon a key property of the Gibbs distribution known as *strong spatial mixing (SSM)* or *correlation decay*. We do not formally define SSM in this thesis since it is not explicitly used. Roughly speaking, for SSM we consider the effect of a pair of boundary conditions on the marginal distribution of a specified vertex, and at a high level this is analogous to spectral independence. SSM can be proved by showing that the tree recursion is contracting with a carefully chosen potential function. In Chapter 5, we establish spectral independence for 2-spin systems in the uniqueness region by combining the potential function method [89] and the self-avoiding walk tree identity [127]. For the monomer-dimer model (matchings) and colorings on triangle-free graphs, we also prove spectral independence by modifying proofs of SSM, see Chapter 6.

Theorem 1.2.2 (Informal Version of Theorem 5.2.2; Spectral Independence by Strong Spatial Mixing Approach)**.** *If there exists a "good" potential function such that the tree recursion, equipped with the potential function, is contractive (from which one can deduce strong spatial mixing property), then μ is $O(1)$-spectrally independent.*

Rapid mixing of MCMC algorithms can also be proved through coupling arguments,

especially by the path coupling method [30]. A distribution is said to be contractive for a Markov chain if, roughly speaking, the distance of two configurations contracts in expectation after one step of the chain; see Chapter 7 for details. We show spectral independence for contractive distributions with respect to some local Markov chain, and thus establish optimal mixing time for the Glauber dynamics. This result allows us to transfer coupling arguments for some local dynamics to optimal mixing results of the Glauber dynamics. One example is sampling random q-colorings on graphs of maximum degree Δ where $q > (11/6 - \varepsilon_0)\Delta$: the uniform distribution over all colorings was known to be contractive with respect to the flip dynamics [125, 42]; however, previously it was not known how to deduce optimal mixing results for the Glauber dynamics in this parameter regime.

Theorem 1.2.3 (Informal Version of Theorem 7.2.3; Spectral Independence by Coupling Methods)**.** *If μ is contractive with respect to some "local" Markov chain $\mathcal{M}$ by (path) coupling arguments (from which one can get optimal mixing of $\mathcal{M}$ but not Glauber dynamics), then μ is $O(1)$-spectrally independent.*

The study of zero-free regions of the partition function has a long and rich history in the analysis of phase transitions in statistical physics models. The well-known work of Lee and Yang [86] utilizes zeros of the partition function to study phase transitions for the ferromagnetic Ising model. The polynomial interpolation method was introduced by Barvinok [10] and refined by Patel and Regts [110]. This approach utilizes the absence of zeros of the partition function in the complex plane to efficiently approximate a suitable transformation of the logarithm of the partition function using Taylor approximation. For a spin system on a graph with n vertices and constant maximum degree Δ, the polynomial interpolation method yields a running time of $O(n^C)$ where the constant C depends on Δ and parameters of the model and is usually pretty large. In Chapter 8, we prove that a zero-free region of the partition function implies spectral independence. This immediately yields several new optimal mixing results for MCMC methods with significantly improved running times.

Theorem 1.2.4 (Informal Version of Theorem 8.2.1; Spectral Independence by Zero-Free Results)**.** *If the multivariate partition function of the external fields is nonzero in a "large" region on the complex plane (from which one can apply the polynomial interpolation method for approximate counting), then μ is $O(1)$-spectrally independent.*

1.3 Thesis Organization

The rest of this thesis is organized as follows. In Chapter 2 we give necessary definitions and well-known facts that are used in this thesis. In Chapter 3 we show optimal mixing of the Glauber dynamics for spin systems that are spectrally independent. We further show optimal mixing results for arbitrary block dynamics and the Swendsen-Wang dynamics in Chapter 4. We establish spectral independence for antiferromagnetic 2-spin systems in Chapter 5, and for the monomer-dimer model and colorings on triangle-free graphs in Chapter 6, by modifying the proofs of correlation decay. In Chapter 7, we show how to deduce spectral independence from a coupling result for any local Markov chain and in particular establish spectral independence for colorings on any bounded degree graph. In Chapter 8 we relate spectral independence and zero-free region of the partition function; as applications we prove spectral independence for weighted edge covers and weighted even subgraphs.

CHAPTER 2

PRELIMINARIES

2.1 Spin Systems

Let $q \geq 2$ be an integer and $[q] = \{1, \ldots, q\}$. Given a graph $G = (V, E)$, we consider the q-spin system on G parameterized by a symmetric interaction matrix $A \in \mathbb{R}^{q \times q}_{\geq 0}$ representing "interaction strengths" and a field vector $B \in \mathbb{R}^{q}_{>0}$ representing "external fields". A configuration $\sigma \in [q]^V$ is an assignment of spins to vertices. The *Gibbs distribution* $\mu = \mu_{G,A,B}$ over all configurations is given by

$$\mu(\sigma) = \frac{1}{Z_G(A,B)} \prod_{\{u,v\} \in E} A(\sigma_u, \sigma_v) \prod_{v \in V} B(\sigma_v), \qquad \forall \sigma \in [q]^V$$

where

$$Z_G(A,B) = \sum_{\sigma \in [q]^V} \prod_{\{u,v\} \in E} A(\sigma_u, \sigma_v) \prod_{v \in V} B(\sigma_v)$$

is called the *partition function*.

We recall some classical examples of spin system.

1. The Ising/Potts model at inverse temperature $\beta \in \mathbb{R}$ corresponds to the interaction $A(a, a') = \exp(\beta \mathbf{1}(a = a'))$ and $B(a) = \exp(h(a))$ where $h \in \mathbb{R}^q$ is a vector of external fields, with $q = 2$ for the Ising model and $q \geq 3$ for the Potts model.

2. The hardcore (or independent sets) model with parameter $\lambda > 0$ is obtained with $q = 2$, $A(a, a') = 0$ if $a = a' = 1$ and $A(a, a') = 1$ otherwise, and $B(a) = \lambda$ if $a = 1$ and $B(a) = 1$ if $a = 2$.

3. The q-colorings model corresponds to $A(a, a') = \mathbf{1}(a \neq a')$ and $B(a) = 1$.

2.1.1 Pinnings and Conditional Distributions

Let μ be an arbitrary distribution over $[q]^V$. A configuration $\sigma \in [q]^V$ is said to be *feasible* with respect to μ if $\mu(\sigma) > 0$. Let $\Omega = \Omega(\mu)$ denote the collection of all feasible configurations (we omit μ when it is clear from the context); namely, Ω is the support of μ. For $U \subseteq V$, we use the notation $\sigma_U = (\sigma_u)_{u \in U}$ and let $\Omega_U = \{\tau \in [q]^U : \exists \sigma \in \Omega, \sigma_U = \tau\}$ be the set of all possible pinnings on U. Observe that $\Omega_V = \Omega$. Note, for $u \in V$, Ω_u is the set of feasible spin assignments for vertex u. Denote the collection of all pinnings by $\mathcal{T} = \cup_{U \subseteq V} \Omega_U$ and denote the set of all feasible vertex-spin pairs by $\mathcal{P} = \{(u, j) : u \in V, j \in \Omega_u\}$. For $\tau \in \mathcal{T}$, let μ^τ denote the conditional Gibbs distribution $\mu(\cdot \mid \sigma_U = \tau)$.

For a pinning $\tau \in \Omega_U$ for $U \subseteq V$, let $\Omega^\tau = \{\sigma \in \Omega : \sigma_U = \tau\}$ denote the corresponding state space; i.e., Ω^τ is the support of μ^τ. We also define $\Omega^\tau_W = \{\phi \in [q]^W : \exists \sigma \in \Omega^\tau, \sigma_W = \phi\}$ for $W \subseteq V \setminus U$ and $\mathcal{P}^\tau = \{(u, j) : u \in V \setminus U, j \in \Omega^\tau_u\}$. We say Ω^τ is connected if the graph on Ω^τ with edges connecting pairs at Hamming distance 1 is connected. The distribution μ over Ω is said to be *totally-connected* if for every $\tau \in \mathcal{T}$, the set Ω^τ is connected. Throughout this thesis, we will assume the distribution μ is totally-connected as this is necessary for the Glauber dynamics to be ergodic for all conditional measures μ^τ.

We remark that all soft-constraint models (i.e., $A(i, j) > 0$ for all $i, j \in [q]$) are totally-connected spin systems and common hard-constraint models, including the hard-core model, q-colorings when $q \geq \Delta + 2$, matchings, and other models studied in this thesis, all satisfy this assumption as well.

The marginal bound for a distribution μ over $[q]^V$ is the minimum nonzero marginal probability of a vertex receiving a spin under any pinning.

Definition 2.1.1 (Bounded Marginals)**.** We say a distribution μ over $[q]^V$ is *b-marginally bounded* if for every $\Lambda \subsetneq V$ and $\tau \in \Omega_\Lambda$, it holds for every $v \in V \setminus \Lambda$ and $i \in \Omega^\tau_v$ that,

$$\mu(\sigma_v = i \mid \sigma_\Lambda = \tau) \geq b.$$

2.1.2 Spectral Independence

The notion of spectral independence was first proposed in [4] for binary product space and later generalized to multi-state product space in [43, 55].

Definition 2.1.2 (Influence Matrix for $q = 2$)**.** Let $q = 2$. Given $\Lambda \subsetneq V$ and $\tau \in \Omega_\Lambda$, let

$$\widetilde{V}^\tau = \{u \in V \setminus \Lambda : \mu(\sigma_u = 1 \mid \sigma_\Lambda = \tau) > 0, \mu(\sigma_u = 0 \mid \sigma_\Lambda = \tau) > 0\}.$$

For every $u, v \in \widetilde{V}^\tau$ with $u \neq v$, we define the *(pairwise) influence* of u on v conditioned on τ by

$$\Psi_\mu^\tau\big(u, v\big) = \mu(\sigma_v = 1 \mid \sigma_u = 1, \sigma_\Lambda = \tau) - \mu(\sigma_v = 1 \mid \sigma_u = 0, \sigma_\Lambda = \tau).$$

Furthermore, let $\Psi_\mu^\tau\big(v, v\big) = 0$ for $v \in \widetilde{V}^\tau$. We call Ψ_μ^τ the *(pairwise) influence matrix* conditioned on τ.

Definition 2.1.3 (Influence Matrix for $q \geq 3$)**.** Let $q \geq 3$. Given $\Lambda \subsetneq V$ and $\tau \in \Omega_\Lambda$, let

$$\widetilde{V}^\tau = \{(u, i) : u \in V \setminus \Lambda, \mu(\sigma_u = i \mid \sigma_\Lambda = \tau) > 0\}.$$

For every $(u, i), (v, j) \in \widetilde{V}^\tau$ with $u \neq v$, we define the *(pairwise) influence* of (u, i) on (v, j) conditioned on τ by

$$\Psi_\mu^\tau\big((u, i), (v, j)\big) = \mu(\sigma_v = j \mid \sigma_u = i, \sigma_\Lambda = \tau) - \mu(\sigma_v = j \mid \sigma_\Lambda = \tau).$$

Furthermore, let $\Psi_\mu^\tau\big((v, i), (v, j)\big) = 0$ for all $(v, i), (v, j) \in \widetilde{V}^\tau$. We call Ψ_μ^τ the *(pairwise) influence matrix* conditioned on τ.

Note that all eigenvalues of the influence matrix Ψ_μ^τ are real; see [4, 43, 23].

Definition 2.1.4 (Spectral Independence)**.** We say a distribution μ over $[q]^V$ is η-*spectrally independent* if for every $\Lambda \subsetneq V$ and $\tau \in \Omega_\Lambda$, the largest eigenvalue $\lambda_1(\Psi_\mu^\tau)$ of the influence matrix Ψ_μ^τ satisfies

$$\lambda_1(\Psi_\mu^\tau) \leq \eta.$$

The work of [55] defined another version of influence matrix by

$$\Psi_\mu^\tau(u,v) = \max_{i,j \in \Omega_u^\tau} \left\| \mu(\sigma_v = \cdot \mid \sigma_u = i, \sigma_\Lambda = \tau) - \mu(\sigma_v = \cdot \mid \sigma_u = j, \sigma_\Lambda = \tau) \right\|_{\mathrm{TV}},$$

and the spectral independence correspondingly. We remark that Definition 2.1.4 is weaker than the notion of spectral independence given in [55], and for all current applications here in this thesis, both definitions work.

2.2 Markov Chains

2.2.1 Glauber Dynamics

The *Glauber dynamics*, also known as the *Gibbs sampling*, is a simple, natural, and popular Markov chain for sampling from a distribution μ over $[q]^V$. The dynamics starts with some (possibly random) configuration X_0. For every $t \geq 1$, a new random configuration X_{t+1} is generated from X_t as follows: pick a coordinate $v \in V$ uniformly at random, set $X_{t+1}(u) = X_t(u)$ for all $u \in V \setminus \{v\}$, and sample $X_{t+1}(v)$ from the conditional distribution $\mu(\sigma_v = \cdot \mid \sigma_{V \setminus \{v\}} = X_t(V \setminus \{v\}))$. Denote the transition matrix of the Glauber dynamics by P_{GL}. If μ is totally-connected, then the Glauber dynamics is ergodic (i.e., irreducible and aperiodic) and has stationary distribution μ.

Let P be the transition matrix of an ergodic Markov chain (X_t) on a finite state space Ω with stationary distribution μ. For $t \geq 0$ and $\sigma \in \Omega$, let $P^t(\sigma, \cdot)$ denote the distribution of X_t when starting the chain with $X_0 = \sigma$. For $\varepsilon \in (0,1)$, the *mixing time* of P is defined

as

$$T_{\text{mix}}(P, \varepsilon) = \max_{\sigma \in \Omega} \min \left\{ t \geq 0 : \left\| P^t(\sigma, \cdot) - \mu \right\|_{\text{TV}} \leq \varepsilon \right\}. \tag{2.1}$$

In particular, we write $T_{\text{mix}}(P) = T_{\text{mix}}(P, 1/4)$.

2.2.2 Expectation, Variance, and Entropy

In the following definition, we assume the underlying distribution μ is fixed and omit it from the subscript.

Definition 2.2.1. Let Ω be a finite set and μ be a distribution over Ω. For all functions $f, g : \Omega \to \mathbb{R}$:

(a) The *expectation* of f is defined as $\mu(f) = \sum_{x \in \Omega} \mu(x) f(x)$;

(b) The *variance* of f is defined as $\text{Var}(f) = \mu[(f - \mu(f))^2] = \mu(f^2) - \mu(f)^2$;

(c) The *covariance* of f and g is defined as $\text{Cov}(f, g) = \mu[(f - \mu(f))(g - \mu(g))] = \mu(fg) - \mu(f)\mu(g)$;

(d) If $f \geq 0$, the *entropy* of f is defined as $\text{Ent}(f) = \mu\left[f \log\left(\frac{f}{\mu(f)}\right)\right] = \mu(f \log f) - \mu(f) \log \mu(f)$ with the convention that $0 \log 0 = 0$.

For two distributions μ, ν over a finite set Ω, the Kullback–Leibler divergence (KL divergence), also called relative entropy, is defined as

$$D_{\text{KL}}(\nu \parallel \mu) = \sum_{x \in \Omega} \nu(x) \log\left(\frac{\nu(x)}{\mu(x)}\right).$$

Let $f = \nu/\mu$ be the relative density of ν with respect to μ; i.e., $f(x) = \nu(x)/\mu(x)$ for all $x \in \Omega$. Then $\text{Ent}(f) = D_{\text{KL}}(\nu \parallel \mu)$. The following is a well-known fact; see, e.g., [50].

Fact 2.2.2 (Donsker-Varadhan's Variational Representation)**.** *For two distributions μ, ν over a finite set Ω, the KL divergence admits the following variational formula:*

$$D_{\mathrm{KL}}(\nu \parallel \mu) = \sup_{f:\Omega\to\mathbb{R}} \left\{\nu(f) - \log \mu(e^f)\right\}.$$

2.2.3 Functional Inequalities for Markov Chains

We now review some standard functional inequalities, and refer to [25, 105] for further background.

Definition 2.2.3. Let Ω be a finite set and μ be a distribution over Ω. Let P denote the transition matrix of an ergodic, reversible Markov chain on Ω with stationary distribution μ.

(a) The Dirichlet form of P is defined as for every $f, g : \Omega \to \mathbb{R}$,

$$\mathcal{E}_P(f, g) = \frac{1}{2} \sum_{x,y\in\Omega} \mu(x)P(x, y)(f(x) - f(y))(g(x) - g(y)).$$

In particular, if $\Omega \subseteq [q]^V$ and $P = P_{\mathrm{GL}}$ is the Glauber dynamics for μ, then we can write

$$\mathcal{E}_{P_{\mathrm{GL}}}(f, g) = \frac{1}{n} \sum_{v\in V} \mu[\mathrm{Cov}_v(f, g)].$$

(b) We say the *Poincaré inequality* holds with constant λ if for every $f : \Omega \to \mathbb{R}$,

$$\lambda \operatorname{Var}(f) \le \mathcal{E}_P(f, f).$$

The spectral gap of P is $\lambda(P) = \inf\left\{ \frac{\mathcal{E}_P(f,f)}{\operatorname{Var}(f)} \;\middle|\; f : \Omega \to \mathbb{R}, \operatorname{Var}(f) \neq 0 \right\}$.

(c) We say the *standard log-Sobolev inequality* holds with constant ρ if for every $f : \Omega \to \mathbb{R}_{\ge 0}$,

$$\rho \operatorname{Ent}(f) \le \mathcal{E}_P(\sqrt{f}, \sqrt{f}).$$

The standard log-Sobolev constant of P is

$$\rho(P) = \inf\Big\{\frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\mathrm{Ent}(f)}\Big| f:\Omega\to\mathbb{R}_{\geq 0}, \mathrm{Ent}(f)\neq 0\Big\}.$$

(d) We say the *modified log-Sobolev inequality* holds with constant ρ_0 if for every $f:\Omega\to\mathbb{R}_{\geq 0}$,

$$\rho_0\,\mathrm{Ent}(f)\leq \mathcal{E}_P(f,\log f).$$

The modified log-Sobolev constant of P is $\rho_0(P) = \inf\Big\{\frac{\mathcal{E}_P(f,\log f)}{\mathrm{Ent}(f)} \;\Big|\; f:\Omega\to\mathbb{R}_{\geq 0}, \mathrm{Ent}(f)\neq 0\Big\}$.

(e) We say *the relative entropy decays* with rate α if for every distribution ν over Ω,

$$D_{\mathrm{KL}}(\nu P \parallel \mu)\leq (1-\alpha)\,D_{\mathrm{KL}}(\nu\parallel\mu).$$

We recall some well known facts about its relation to the other two inequalities and its implications for mixing times.

Lemma 2.2.4. *If (P,μ) satisfies the standard LSI with constant s then it satisfies the MLSI with constant $\rho = 2s$. If it satisfies the discrete time relative entropy decay with rate $\delta > 0$, then it satisfies the MLSI with constant $\rho = \delta$. Finally, if it satisfies the discrete time relative entropy decay with rate $\delta > 0$, then*

$$T_{\mathrm{mix}}(P)\leq 1+\delta^{-1}[\log(8)+\log\log(1/\mu_*)]\,, \tag{2.2}$$

where $\mu_* = \min_{\sigma\in\Omega}\mu(\sigma)$.

We refer to e.g. [24, Section 2] for a proof. Note that we have not assumed reversibility of P in the above lemma. If (P,μ) is reversible, then one can additionally show that the standard LSI with constant s implies the discrete time relative relative entropy decay with rate $\delta = s$.

Next, we consider the case that $\Omega \subseteq [q]^V$ for a finite set V. Let $S \subseteq V$ and $\tau \in \Omega_{V\setminus S}$. Recall that for every function $f : \Omega \to \mathbb{R}_{\geq 0}$, we write $\mu_S^\tau(f)$ and $\mathrm{Ent}_S^\tau(f) = \mathrm{Ent}_{\mu_S^\tau}(f)$ for the expectation and entropy of f under the conditional distribution $\mu_S^\tau(\cdot) = \mu(\sigma_S = \cdot \mid \sigma_{V\setminus S} = \tau)$, where $f = f_\tau$ is understood as a function of the configuration on S with τ fixed outside S. We think of $\mu_S^\tau(f)$ and $\mathrm{Ent}_S^\tau(f)$ as functions of τ, and we will use, for example, $\mathrm{Ent}[\mu_S(f)]$ to represent the entropy of $\mu_S^\tau(f)$ under μ, and $\mu[\mathrm{Ent}_S(f)]$ for the expectation of $\mathrm{Ent}_S^\tau(f)$. We give below a useful property of the expectation and entropy; see, e.g., [100] for proofs.

Fact 2.2.5. *Let $S \subseteq V$ and $\tau \in \Omega_{V\setminus S}$. For every function $f : \Omega \to \mathbb{R}_{\geq 0}$, we have*

$$\mu(f) = \mu[\mu_S(f)] \qquad \textit{and} \qquad \mathrm{Ent}(f) = \mu[\mathrm{Ent}_S(f)] + \mathrm{Ent}[\mu_S(f)].$$

2.3 Entropy Tensorization and Factorization

2.3.1 Approximate Tensorization of Entropy

The notion of approximate tensorization of entropy is formally defined as follows.

Definition 2.3.1 (Approximate Tensorization)**.** We say that a distribution μ over $[q]^V$ satisfies the *approximate tensorization of entropy* (with constant C_1) if for all $f : \Omega \to \mathbb{R}_{\geq 0}$ we have

$$\mathrm{Ent}(f) \leq C_1 \sum_{v \in V} \mu[\mathrm{Ent}_v(f)]. \tag{2.3}$$

Approximate tensorization can be understood as closeness of μ to a product distribution, or weak dependency of variables. In fact, if μ is exactly a product distribution (e.g., the Gibbs distribution on an empty graph), then approximate tensorization holds with constant $C_1 = 1$; e.g., see [41, 36]. If μ satisfies approximate tensorization with a constant C_1 independent of n, then the Glauber dynamics for sampling from μ mixes in $O(n \log n)$ steps. In fact, given approximate tensorization, one can deduce tight bounds on all of the following quantities: the spectral gap, both standard and modified log-Sobolev constants,

relative entropy decay rate, mixing time, and concentration bounds. We summarize here a few corollaries of approximate tensorization of entropy for arbitrary distributions over discrete product spaces.

Fact 2.3.2. *Let V be a set of size n and μ be a distribution over $[q]^V$. If μ satisfies the approximate tensorization of entropy with constant C_1, then the Glauber dynamics for μ satisfies all of the following:*

(1) The Poincaré inequality holds with constant $\lambda = \frac{1}{C_1 n}$;

(2) The modified log-Sobolev inequality holds with constant $\rho_0 = \frac{1}{C_1 n}$;

(3) The relative entropy decays with rate $\alpha = \frac{1}{C_1 n}$;

(4) The mixing time of the Glauber dynamics satisfies

$$T_{\mathrm{mix}}(P_{\mathrm{GL}}, \varepsilon) \leq \left\lceil C_1 n \left(\log\log \frac{1}{\mu_{\min}} + \log \frac{1}{2\varepsilon^2} \right) \right\rceil$$

where $\mu_{\min} = \min_{\sigma \in \Omega} \mu(\sigma)$; If furthermore μ is b-marginally bounded, then we have $\mu_{\min} \geq b^n$ and thus

$$T_{\mathrm{mix}}(P_{\mathrm{GL}}, \varepsilon) \leq \left\lceil C_1 n \left(\log n + \log\log \frac{1}{b} + \log \frac{1}{2\varepsilon^2} \right) \right\rceil;$$

(5) For every $f : \Omega \to \mathbb{R}$ which is c-Lipschitz with respect to the Hamming distance on $[q]^V$ and every $a \geq 0$, we have the concentration inequality

$$\Pr_{\sigma \sim \mu} \left[|f(\sigma) - \mu(f)| \geq a \right] \leq 2 \exp\left(-\frac{a^2}{2c^2 C_1 n} \right);$$

(6) If furthermore μ is b-marginally bounded, then the standard log-Sobolev inequality holds with constant $\rho = \frac{1-2b}{\log(1/b-1)} \cdot \frac{1}{C_1 n}$ when $b < \frac{1}{2}$, or $\rho = \frac{1}{2C_1 n}$ when $b = \frac{1}{2}$.

The implications in Fact 2.3.2 are all known and have been widely used, often implicitly. In the proof below, we give references where explicit statements or direct proofs are available.

Proof of Fact 2.3.2. Item 1 and Item 2 are proved in [36, Proposition 1.1]. To show Item 3, let P_v be the transition matrix corresponding to updating the spin at v conditioned on all other vertices. Thus, we have the decomposition

$$P_{\text{GL}} = \frac{1}{n} \sum_{v \in V} P_v.$$

Let $f = \nu/\mu$ be the relative density of ν with respect to μ. Then we get

$$\begin{aligned} &D_{\text{KL}}(\nu P_{\text{GL}} \parallel \mu) = D_{\text{KL}} \left(\frac{1}{n} \sum_{v \in V} \nu P_v \,\middle\|\, \mu \right) \leq \frac{1}{n} \sum_{v \in V} D_{\text{KL}}(\nu P_v \parallel \mu) = \frac{1}{n} \sum_{v \in V} \text{Ent}(P_v f) \\ &= \frac{1}{n} \sum_{v \in V} \text{Ent}[\mu_v(f)] = \frac{1}{n} \sum_{v \in V} \text{Ent}(f) - \mu[\text{Ent}_v(f)] = \text{Ent}(f) - \frac{1}{n} \sum_{v \in V} \mu[\text{Ent}_v(f)] \\ &\leq \left(1 - \frac{1}{C_1 n}\right) \text{Ent}(f) = \left(1 - \frac{1}{C_1 n}\right) D_{\text{KL}}(\nu \parallel \mu). \end{aligned}$$

Item 4 can be deduced from Item 3 as shown by [24, Lemma 2.4]; see also [25, Corollary 2.8] for the continuous time setting. Item 5 follows from Item 2 and [47, Lemma 15]. Finally, Item 6 follows by an application of [48, Theorem A.1]. □

2.3.2 General Block Factorization of Entropy

Caputo and Parisi [37] introduced the notion of *general block factorization* of entropy which generalizes approximate tensorization, and is useful for analyzing more general classes of Markov chains. Let $\alpha = (\alpha_B)_{B \subseteq V}$ be an arbitrary probability distribution over subsets of V, and define the minimum "coverage probability" of a vertex by

$$\delta = \delta(\alpha) = \min_{u \in V} \sum_{B : B \ni u} \alpha_B. \tag{2.4}$$

Definition 2.3.3 (General Block Factorization)**.** We say that a distribution μ over $[q]^V$ satisfies the *general block factorization of entropy* (with constant C) if for all weights α, for all $f : \Omega \to \mathbb{R}_{\geq 0}$ we have

$$\delta(\alpha)\mathrm{Ent} f \leq C \sum_{B \subseteq V} \alpha_B \, \mu[\mathrm{Ent}_B f]. \tag{2.5}$$

Recall that $\mu[\mathrm{Ent}_B f] = \mu[f \log(f/\mu_B f)]$ is the expected value of the conditional entropy $\tau \mapsto \mathrm{Ent}(f|\tau)$ for τ a spin configuration on $V \setminus B$. Entropy tensorization in Definition 2.3.1 is the special case when $\alpha_B = 1/n$ for every block of size 1 and $\alpha_B = 0$ for larger blocks. The choice of the constant $\delta(\alpha)$ in this inequality is motivated by the fact that when μ is a product measure then (Equation 4.5) holds with $C = 1$, in which case it is known as the Shearer inequality; see [36]. The block factorization of entropy is a statement concerning the equilibrium distribution μ which has deep consequences for several natural sampling algorithms. In particular, it implies optimal mixing and optimal entropy decay for arbitrary block dynamics and constitutes a key concept in the proof of Theorems 4.1.1 and 4.1.2.

Fix a probability distribution α over subsets of V and observe that the α-weighted heat bath block dynamics defined in the introduction is the Markov chain with transition matrix P_α on Ω such that for any real function f

$$P_\alpha f = \sum_{B \subseteq V} \alpha_B \, \mu_B(f) \, . \tag{2.6}$$

To clarify the above notation, if we evaluate the left hand side at a spin configuration $\sigma \in \Omega$ then each for each B the term $\mu_B f$ in the right hand side is given by $\mu_B^\tau f$ where $\tau = \sigma_{V \setminus B}$. If $\alpha_B = n^{-1} \, \mathbf{1}(|B| = 1)$, then Eq. (2.6) is the Glauber dynamics for μ.

The α-weighted heat bath block dynamics Eq. (2.6) defines a reversible pair (P_α, μ).

Moreover, its Dirichlet form satisfies

$$\mathcal{D}_\alpha(f,g) = \sum_{B\subseteq V} \alpha_B\, \mu[f(1-\mu_B)g] = \sum_{B\subseteq V} \alpha_B\, \mu\left[\mathrm{Cov}_B(f,g)\right]\,, \tag{2.7}$$

where $\mathrm{Cov}_B(f,g) = \mu_B\left[(f-\mu_B f)(g-\mu_B g)\right]$ denotes the covariance functional.

Lemma 2.3.4. *If the spin system satisfies the general block factorization with constant C then for all α the Markov chain (P_α, μ) satisfies*

1. *the modified log-Sobolev inequality with constant $\rho = \frac{\delta(\alpha)}{C}$;*
2. *the discrete time relative entropy decay with rate $\delta = \frac{\delta(\alpha)}{C}$;*
3. *$T_{\mathrm{mix}}(P_\alpha) \le 1 + \frac{C}{\delta(\alpha)}[\log(8) + \log\log(1/\mu_*)]$, where $\mu_* = \min_{\sigma\in\Omega}\mu(\sigma)$.*

Proof. In view of Fact 2.3.2 it is sufficient to prove item 2. We note that the relative entropy decay with rate δ is equivalent to the entropy contraction

$$\mathrm{Ent}(P_\alpha f) \le (1-\delta)\mathrm{Ent}(f), \tag{2.8}$$

for all $f \ge 0$. By convexity of $x \mapsto x\log x$ one has

$$\begin{aligned}\mathrm{Ent}(P_\alpha f) &= \mu[P_\alpha f \log(P_\alpha f)] - \mu[f]\log\mu[f] \\ &\le \sum_B \alpha_B\, \mu[\mu_B(f)\log(\mu_B(f))] - \mu[f]\log\mu[f] = \sum_B \alpha_B \mathrm{Ent}(\mu_B(f))].\end{aligned} \tag{2.9}$$

From the decomposition in Lemma 4.3.1 it follows that

$$\mathrm{Ent}(P_\alpha f) \le \mathrm{Ent}(f) - \sum_B \alpha_B \mu[\mathrm{Ent}_B(f)]. \tag{2.10}$$

By definition of block factorization we conclude

$$\mathrm{Ent}(P_\alpha f) \le (1-\delta(\alpha)/C)\mathrm{Ent}(f). \qquad \square$$

2.4 Simplicial Complexes

A *simplicial complex* $\mathfrak{X}$ is a collection of subsets (called faces) of a ground set $\mathcal{U}$ which is downwards closed; that is, if $\sigma \in \mathfrak{X}$ and $\tau \subseteq \sigma$ then $\tau \in \mathfrak{X}$. The dimension of a face is its size, and the dimension of $\mathfrak{X}$ is defined to be the maximum dimension of its faces. We say an n-dimensional simplicial complex $\mathfrak{X}$ is *pure* if every face is contained in a maximal face of size n. We write $\mathfrak{X}(k)$ for the collection of faces of size k. For a k-dimensional face $\tau \in \mathfrak{X}(k)$, we can define a pure $(n-k)$-dimensional simplicial subcomplex $\mathfrak{X}_\tau$ by taking $\mathfrak{X}_\tau = \{\xi \subseteq \mathcal{U} \setminus \tau : \tau \cup \xi \in \mathfrak{X}\}$.

For a pure n-dimensional simplicial complex $\mathfrak{X}$, consider a positive weight function $w : \mathfrak{X}(n) \to \mathbb{R}_{>0}$, which induces a distribution π_n on $\mathfrak{X}(n)$ with $\pi_n(\sigma) \propto w(\sigma)$. Furthermore, we can also define a distribution π_k over $\mathfrak{X}(k)$ for each nonnegative integer $k < n$ via the following process: sample σ from π_n, and select a uniformly random subset of size k. For $\tau \in \mathfrak{X}(k)$, the weight function w induces the weights for the simplicial subcomplex $\mathfrak{X}_\tau$ by $w_\tau(\xi) = w(\tau \cup \xi)$ for each $\xi \in \mathfrak{X}_\tau(n-k)$. The distribution $\pi_{\tau,j}$ is also defined accordingly for each nonnegative integer $j \leq n-k$.

As noticed in [4], there is a natural way to represent every distribution μ over $[q]^V$ with $|V| = n$ as a pure n-dimensional weighted simplicial complex $(\mathfrak{X} = \mathfrak{X}^\Omega, \mu)$, which is defined as follows. The ground set of $\mathfrak{X}$ consists of pairs

$$\widetilde{V} = \{(v, i) : v \in V, i \in \Omega_v\}.$$

The maximal faces of $\mathfrak{X}$ consist of collections of n pairs forming a valid configuration $\sigma \in \Omega$; i.e., every configuration $\sigma \in \Omega$ corresponds to a maximal face $\{(v, \sigma_v) : v \in V\}$. The rest of $\mathfrak{X}$ is generated by taking downwards closure so that $\mathfrak{X}$ is pure by construction. Namely, every $U \subseteq V$ and $\tau \in \Omega_U$ corresponds to a face $\{(v, \tau_v) : v \in U\}$; we shall denote it by (U, τ) for simplicity. Note that the faces of intermediate dimension can be thought of as partial configurations. Now, if there is a weight function $w : \Omega \to \mathbb{R}_{>0}$

associated with μ such that $\mu(\sigma) \propto w(\sigma)$ for each $\sigma \in \Omega$, then it also gives a weight function $w : \mathfrak{X}(n) \to \mathbb{R}_{>0}$ by the one-to-one correspondence between Ω and $\mathfrak{X}(n)$, and thus induces the associated distribution π_n on $\mathfrak{X}(n)$. Observe that π_n is exactly the distribution μ. Moreover, for each $k < n$, the distribution π_k on $\mathfrak{X}(k)$ is given by

$$\pi_k(U, \tau) = \frac{1}{\binom{n}{k}} \mu(\sigma_U = \tau)$$

for every $U \subseteq V$ and $\tau \in \Omega_U$.

For simplicial complexes, the global down-up and up-down walks between faces of distinct dimensions have attracted a lot of attention in recent years [80, 49, 82, 108, 5, 47, 1]. For integers $0 \leq r < s \leq n$, define the *order-(s, r) (global) down-up* walk with transition matrix denoted by $P^{\vee}_{s,r}$ to be the following random walk over $\mathfrak{X}(s)$: in each step we remove $s - r$ elements, chosen uniformly at random, from the current face $\sigma_t \in \mathfrak{X}(s)$ to obtain a face $\tau_t \in \mathfrak{X}(r)$, and then pick $\xi_{t+1} \in \mathfrak{X}_{\tau_t}(s - r)$ from the distribution $\pi_{\tau_t, s-r}$ and set $\sigma_{t+1} = \tau_t \cup \xi_{t+1}$. The stationary distribution of $P^{\vee}_{s,r}$ is π_s. In particular, observe that the Glauber dynamics for a distribution μ over $[q]^V$ is the same as the order-$(n, n-1)$ down-up walk for the weighted simplicial complex $(\mathfrak{X}, \mu)$. Similarly, the *order-(r, s) (global) up-down* walk with transition matrix $P^{\wedge}_{r,s}$ is a random walk over $\mathfrak{X}(r)$ with stationary distribution π_r: given the current face $\tau_t \in \mathfrak{X}(r)$, sample $\xi_{t+1} \in \mathfrak{X}_{\tau_t}(s - r)$ from $\pi_{\tau_t, s-r}$, set $\sigma_{t+1} = \tau_t \cup \xi_{t+1}$, and finally remove $s - r$ elements from σ_{t+1} uniformly at random to obtain $\tau_{t+1} \in \mathfrak{X}(r)$.

Definition 2.4.1 (Bounded Marginals)**.** We say a pure n-dimensional weighted simplicial complex $(\mathfrak{X}, w)$ is $(b_0, \ldots, b_{n-1})$-*marginally bounded* if for all $0 \leq k \leq n - 1$, every $\tau \in \mathfrak{X}(k)$, and every $i \in \mathfrak{X}_\tau(1)$, we have

$$\pi_{\tau,1}(i) \geq b_k.$$

Claim 2.4.2. *If a distribution μ over $[q]^V$ is b-marginally bounded, then the weighted sim-*

plicial complex $(\mathfrak{X}, \mu)$ *is* $(b_0, \ldots, b_{n-1})$*-marginally bounded with* $b_k = \frac{b}{n-k}$ *for each* k*.*

The proof of Claim 2.4.2 can be found in Section 3.4.1.

The global walks in simplicial complexes can be studied by decomposition into local walks which we define now. For every $0 \leq k \leq n-2$ and every face $\tau \in \mathfrak{X}(k)$, the *local walk* at τ with transition matrix P_τ is the following random walk over $\mathfrak{X}_\tau(1)$: given the current element $i \in \mathfrak{X}_\tau(1)$, the next element is generated from the distribution $\pi_{\tau\cup\{i\},1}$. One can relate mixing properties of the local walks to the mixing properties of the global walks; see [82, 5, 47, 1]. In nearly all prior works, such a relation was quantified using the spectral gap of the walks. Like in [47], while our ultimate goal is to show the modified log-Sobolev inequality of the global walks, we will still need the notion of local spectral expansion for local walks. Let us now capture this idea using the following definition, taking after [80, 49, 108, 82, 1, 81].

Definition 2.4.3 (Local Spectral Expansion [1])**.** We say a pure n-dimensional weighted simplicial complex $(\mathfrak{X}, w)$ is a $(\zeta_0, \ldots, \zeta_{n-2})$*-local spectral expander* if for every $0 \leq k \leq n-2$ and every $\tau \in \mathfrak{X}(k)$, we have

$$\lambda_2(P_\tau) \leq \zeta_k.$$

Claim 2.4.4. *If a distribution* μ *over* $[q]^V$ *is* η*-spectrally independent, then the weighted simplicial complex* $(\mathfrak{X}, \mu)$ *is a* $(\zeta_0, \ldots, \zeta_{n-2})$*-local spectral expander with* $\zeta_k = \frac{\eta}{n-k-1}$ *for each* k*.*

Proof. This is Theorem 8 from [43]. □

2.5 Complex Analysis

2.5.1 Stability

For n sets $\Gamma_1, \ldots, \Gamma_n$, let $\prod_{\ell=1}^n \Gamma_\ell = \Gamma_1 \times \cdots \times \Gamma_n$ denote the Cartesian product of them.

Definition 2.5.1 (Stability)**.** For an integer $n \geq 1$ and $\mathcal{K} \subseteq \mathbb{C}^n$, we say a multivariate polynomial $P \in \mathbb{C}[z_1, \ldots, z_n]$ is $\mathcal{K}$-stable if $P(z_1, \ldots, z_n) \neq 0$ whenever $(z_1, \ldots, z_n) \in \mathcal{K}$. In particular, if $\mathcal{K} = \prod_{\ell=1}^{n} \Gamma$ for some $\Gamma \subseteq \mathbb{C}$, then we simply say P is Γ-stable.

Theorem 2.5.2 (Hurwitz' Theorem)**.** *Let $n \geq 1$ be an integer and $\mathcal{K} \subseteq \mathbb{C}^n$ be an open connected set. Suppose that $\{f_m\}_{m=1}^{\infty}$ is a sequence of non-vanishing analytic functions on $\mathcal{K}$ that converges to f uniformly on compact subsets of $\mathcal{K}$. Then f is either non-vanishing on $\mathcal{K}$ or else identically zero.*

2.5.2 Complex Plane

We refer to subsets of the complex plane as regions. Let $\Gamma \subseteq \mathbb{C}$ be a region. Denote the complement of Γ by $\Gamma^{\mathrm{c}} = \mathbb{C} \setminus \Gamma$, its interior by Γ^{o}, its closure by $\overline{\Gamma}$, and its boundary by $\partial\Gamma$. We say Γ is *unbounded* if for any $M \in \mathbb{R}_+$ there exists $z \in \Gamma$ such that $|z| > M$; otherwise it is called bounded. For $z \in \mathbb{C}$ let $\operatorname{dist}(z, \Gamma) = \inf_{w \in \Gamma} |w - z|$ be the distance from z to Γ on the complex plane.

For a region $\Gamma \subseteq \mathbb{C}$ and $z \in \mathbb{C}$, we define $\Gamma + z = \{w + z : w \in \Gamma\}$, $z\Gamma = \{zw : w \in \Gamma\}$, and $\Gamma^{-1} = (\Gamma \setminus \{0\})^{-1} = \{w^{-1} : w \in \Gamma \setminus \{0\}\}$. For $\Gamma_1, \Gamma_2 \subseteq \mathbb{C}$, let $\Gamma_1 \cdot \Gamma_2 = \{zw : z \in \Gamma_1, w \in \Gamma_2\}$ denote their Minkowski product; in particular, for $\Gamma \subseteq \mathbb{C}$ let $\Gamma^2 = \Gamma \cdot \Gamma = \{zw : z, w \in \Gamma\}$ (note that we write $\prod_{\ell=1}^{2} \Gamma = \Gamma \times \Gamma = \{(z, w) : z, w \in \Gamma\}$ for the Cartesian product).

For $z \in \mathbb{C}$ and $r \in \mathbb{R}_+$, let $\mathbb{D}(z, r) = \{w \in \mathbb{C} : |w - z| < r\}$ denote the open disk centered at z with radius r, and let $\overline{\mathbb{D}}(z, r) = \{w \in \mathbb{C} : |w - z| \leq r\}$ denote the closed disk. For $\varepsilon \in \mathbb{R}_+$, let $\mathbb{H}_\varepsilon = \{x + iy : x < -\varepsilon\}$ and $\overline{\mathbb{H}}_\varepsilon = \{x + iy : x \leq -\varepsilon\}$ be open and closed half-planes.

Let $\Gamma \subseteq \mathbb{C}$ be a non-empty open region on the complex plane. We say $w, z \in \Gamma$ are (path-)connected in Γ if there exists a continuous map $\gamma : [0, 1] \to \Gamma$ such that $\gamma(0) = w$ and $\gamma(1) = z$. Observe that connectivity in Γ is an equivalence relation, and we call each equivalence class a (path-)connected component of Γ. The region Γ is said to be

(path-)connected if every two points from Γ are connected in Γ; namely, Γ has a unique connected component which is itself. If Γ is open then every connected component of Γ is also open.

A non-empty open connected region $\Gamma \subseteq \mathbb{C}$ is called *simply connected* if its complement in the Riemann sphere ($\mathbb{C} \cup \infty$) is also connected. A Jordan curve (simple closed curve) is a continuous map $\gamma : [0,1] \to \mathbb{C}$ such that $\gamma(0) = \gamma(1)$ and the restriction of γ to $[0,1)$ is injective. The Jordan curve theorem states that for a Jordan curve γ, the complement of its image on the complex plane consists of exactly two open connected components; one of these components is bounded called the interior and the other is unbounded called the exterior. A non-empty open connected region $\Gamma \subseteq \mathbb{C}$ is simply connected if and only if for every Jordan curve γ whose image is contained in Γ, the interior of γ is also contained in Γ.

2.5.3 Useful Theorems

Throughout this thesis, we select the principal branch for the complex functions $z \mapsto \log z$ and $z \mapsto z^{1/d}$.

Theorem 2.5.3 (Schwarz-Pick Theorem)**.** *Let $f : \mathbb{D}(0,1) \to \mathbb{D}(0,1)$ be a holomorphic function. Then*

$$|f'(0)| \leq 1 - |f(0)|^2 \leq 1.$$

For open regions $\Gamma_1, \Gamma_2 \subseteq \mathbb{C}$, a function $f : \Gamma_1 \to \Gamma_2$ is said to be biholomorphic if f is a bijective holomorphic function whose inverse is also holomorphic.

Theorem 2.5.4 (Riemann Mapping Theorem)**.** *Let $\Gamma \subseteq \mathbb{C}$ be a non-empty open simply connected region that is not $\mathbb{C}$. Then for any $z \in \Gamma$ there exists a unique biholomorphic mapping $f : \Gamma \to \mathbb{D}(0,1)$ such that*

$$f(z) = 0 \quad \textit{and} \quad f'(z) \in \mathbb{R}_+.$$

Theorem 2.5.5 (Koebe's One-Quarter Theorem)**.** *Let* $\Gamma \subseteq \mathbb{C}$ *and let* $f : \mathbb{D}(0,1) \to \Gamma$ *be an injective holomorphic function. Then*

$$\mathbb{D}\left(f(0), \frac{1}{4}|f'(0)|\right) \subseteq \Gamma.$$

Theorem 2.5.6 (Multivariate Open Mapping Theorem, [84, Theorem 1.8.1])**.** *Let* $n \geq 1$ *be an integer and let* $\mathcal{K} \subseteq \mathbb{C}^n$ *be a non-empty open connected subset of* $\mathbb{C}^n$*. Let* $f : \mathcal{K} \to \mathbb{C}$ *be a non-constant holomorphic function. Then the image of* f *is an open connected region.*

CHAPTER 3
OPTIMAL MIXING OF GLAUBER DYNAMICS VIA SPECTRAL INDEPENDENCE

In this chapter, we show that for a spin system defined on a bounded-degree graph with constant marginal bounds, spectral independence implies asymptotically optimal mixing time of the Glauber dynamics. This chapter is based on [45].

3.1 Main Result: Optimal Mixing of Glauber Dynamics

Our main result in this chapter is a general statement regarding the Glauber dynamics for an arbitrary spin system satisfying marginal bounds and spectral independence. We show that if the Gibbs distribution on a bounded-degree graph is both marginally bounded and spectrally independent, then the Glauber dynamics satisfies the modified log-Sobolev inequality with constant $\Omega(1/n)$ and mixes in $O(n \log n)$ steps, where n is the number of vertices of the graph. Note that the mixing time of the Glauber dynamics is $\Omega(n \log n)$ for a family of bounded-degree graphs [71].

Theorem 1.2.1. *Let $\Delta \geq 3$ be an integer and $b, \eta > 0$ be reals. Suppose that $G = (V, E)$ is an n-vertex graph of maximum degree at most Δ and μ is a totally-connected Gibbs distribution of some spin system on G. If μ is both b-marginally bounded and η-spectrally independent, then the Glauber dynamics for sampling from μ satisfies the modified log-Sobolev inequality with constant $\frac{1}{C_1 n}$ where*

$$C_1 = \left(\frac{\Delta}{b}\right)^{O\left(\frac{\eta}{b}+1\right)}.$$

Furthermore, the mixing time of the Glauber dynamics satisfies

$$T_{\mathrm{mix}}(P_{\mathrm{GL}}, \varepsilon) = \left(\frac{\Delta}{b}\right)^{O\left(\frac{\eta}{b}+1\right)} \times O\left(n \log\left(\frac{n}{\varepsilon}\right)\right).$$

Previous results [4, 44, 43, 55] could obtain $\mathrm{poly}(\Delta) \times n^{O(\eta)}$ mixing but without the assumption of marginal boundedness. In the setting of spin systems, we always have b-marginal boundedness with b depending only on the parameters A, h of the spin system and the maximum degree Δ of the graph, and so our results supersede those of [4, 44, 43, 55] in the bounded degree regime.

This chapter is organized as follows. In Section 3.2 we outline our proof approach, emphasizing the relations between approximate tensorization and uniform block factorization of entropy, and how spectral independence implies these notions. In Section 3.3, we show how to reduce approximate tensorization to uniform block factorization with linear-sized blocks. Finally, Section 3.4 includes the proof of uniform block factorization of entropy assuming spectral independence in the more general framework of pure weighted simplicial complexes and deduce global entropy contraction from local spectral expansion.

3.2 Proof Outline

In this section, we outline our proofs of Theorem 1.2.1.

3.2.1 Approximate Tensorization and Factorization of Entropy

One way of establishing rapid mixing of the Glauber dynamics is to show that the Gibbs distribution satisfies the *approximate tensorization of entropy*. This approach has been (implicitly) used in many literature to establish the log-Sobolev inequalities, from which one can deduce an optimal bound on the mixing time. Before giving the formal definition, we first review some standard definitions.

In many cases, especially on the integer lattice $\mathbb{Z}^d$, log-Sobolev inequalities for the

Glauber dynamics are established through the approximate tensorization of entropy, which is more intuitive and easier to handle; e.g., see [98, 65, 41, 37]. Despite the success on $\mathbb{Z}^d$, there is not much study for spin systems on bounded-degree graphs. The works of [36, 101] considered approximate tensorization for general discrete product spaces, and gave sufficient conditions to derive it; however, for spin systems these results do not cover the whole uniqueness region.

One can regard approximate tensorization of entropy as factorizing entropy into all single vertices. Motivated by tools from high dimensional simplicial complexes [1, 4] and study on general block factorization of entropy [37] (see Definition 2.3.3), we consider in this thesis a more general notion of entropy factorization, where the entropy is factorized into subsets of vertices of a fixed size. The formal definition is given as follows.

Definition 3.2.1 (Uniform Block Factorization)**.** We say that a distribution μ over $[q]^V$ satisfies the *ℓ-uniform block factorization of entropy* (with constant C) if for all $f : \Omega \to \mathbb{R}_{\geq 0}$ we have

$$\frac{\ell}{n} \operatorname{Ent}(f) \leq C \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu[\operatorname{Ent}_S(f)]. \tag{3.1}$$

We remark that uniform block factorization of entropy is a special case of block factorization given by equation (1.3) in [37]; there, the entropy factorizes into arbitrary blocks with arbitrary weights. Also observe that 1-uniform block factorization is the same as approximate tensorization of entropy. Just as the approximate tensorization corresponds to the single-site Glauber dynamics, the ℓ-uniform block factorization corresponds to the heat-bath block dynamics where in each step a subset of vertices of size ℓ is chosen uniformly at random and gets updated. Moreover, similar results as in Fact 2.3.2 can be deduced for this block dynamics.

Our first key result is a reduction from approximate tensorization to uniform block factorization. For b-marginally bounded Gibbs distributions on graphs with maximum degree $\leq \Delta$, we show that approximate tensorization is implied by ℓ-uniform block factorization

for $\ell = \lceil \theta n \rceil$ and an appropriate constant θ depending on b and Δ. This is given by the following lemma.

Theorem 3.2.2. *Let $\Delta \geq 3$ be an integer and $b > 0$ be a real. Consider the Gibbs distribution μ on an n-vertex graph G of maximum degree at most Δ and assume that μ is b-marginally bounded. Suppose there exist positive reals $\theta \leq \frac{b^2}{12\Delta}$ and C such that μ satisfies the $\lceil \theta n \rceil$-uniform block factorization of entropy with constant C. Then μ satisfies the approximate tensorization of entropy with constant*

$$C_1 = \frac{18 \log(1/b)}{b^4} C.$$

To see the intuition behind Theorem 3.2.2, we consider the $\lceil \theta n \rceil$-uniform block dynamics which in each step simultaneously updates $\lceil \theta n \rceil$ distinct vertices chosen uniformly at random conditioned on the configuration of unpicked vertices. Since the graph is sparse (maximum degree Δ), when $\theta \ll 1/\Delta$ the chosen vertices will, with high probability, form many connected components, each of which has constant size in expectation. A key property of spin systems is that given the configuration outside the chosen vertices, the conditional Gibbs distribution is a product over all components. Hence, it suffices to consider approximate tensorization for each component which is of constant size in expectation. The proof of Theorem 3.2.2 is given in Section 3.3.

Remark 3.2.3. The notion of approximate tensorization and uniform block factorization with respect to variance is also meaningful. In fact, for variance these definitions are equivalent to bounding the spectral gap of the corresponding chains. Moreover, Theorem 3.2.2 holds for variance as well, which can already provide a tight bound on the spectral gap of the Glauber dynamics combining results from [4, 44, 43, 55]. See [45] for details.

3.2.2 Simplicial Complexes and Entropy Contraction

Our next goal is to establish ℓ-uniform block factorization of entropy for $\ell = \Theta(n)$, which relies on the spectral independence property. The following lemma holds for all distributions over $[q]^V$, not only Gibbs distributions.

Theorem 3.2.4. *Let $b, \eta > 0$ be reals. Then for every real $\theta \in (0,1)$ and every integer $n \geq \frac{2}{\theta}(\frac{4\eta}{b^2} + 1)$ the following holds.*

Let V be a set of size n and μ be a distribution over $[q]^V$. If μ is both b-marginally bounded and η-spectrally independent, then μ satisfies $\lceil \theta n \rceil$-uniform block factorization of entropy with constant

$$C = \left(\frac{2}{\theta}\right)^{\frac{4\eta}{b^2}+1}.$$

The recent work [4] studied spin systems, and more generally any distribution over $[q]^V$, in a novel way by viewing full and partial configurations as a high dimensional simplicial complex and utilizing tools such as high-dimensional expansion. Subsequent works [44, 43, 55] follow the same path as well. In this chapter we also study spin systems in the framework of simplicial complexes.

Recall that there is a natural correspondence between a distribution μ over $[q]^V$ and the weighted simplicial complex $(\mathfrak{X}, \mu)$. For general weighted simplicial complexes, one property studied in [47] is how the entropy of a function defined on faces contracts when it projects down from higher dimensions to lower. This can be captured by the definition below. For a pure n-dimensional weighted simplicial complex $(\mathfrak{X}, w)$ and a nonnegative integer $k < n$, let $P_k^{\uparrow}$ denote the $|\mathfrak{X}(k)| \times |\mathfrak{X}(k+1)|$ dimensional transition matrix corresponding to adding a random element $i \notin \tau$ to some $\tau \in \mathfrak{X}(k)$ where i is distributed as $\pi_{\tau,1}$. Also for any $0 \leq r < s \leq n$ and any function $f^{(s)} : \mathfrak{X}(s) \to \mathbb{R}_{\geq 0}$, define $f^{(r)} : \mathfrak{X}(r) \to \mathbb{R}_{\geq 0}$ by $f^{(r)} = P_r^{\uparrow} \cdots P_{s-1}^{\uparrow} f^{(s)}$.

Definition 3.2.5 (Global Entropy Contraction)**.** We say a pure n-dimensional weighted simplicial complex $(\mathfrak{X}, w)$ satisfies the *order-(r,s) global entropy contraction* with rate

$\kappa = \kappa(r, s)$ if for all $f^{(s)} : \mathfrak{X}(s) \to \mathbb{R}_{\geq 0}$ we have

$$\mathrm{Ent}_{\pi_r}(f^{(r)}) \leq (1 - \kappa)\,\mathrm{Ent}_{\pi_s}(f^{(s)}).$$

It turns out, as a remarkable fact, that uniform block factorization of entropy for a distribution μ over $[q]^V$ is *equivalent* to global entropy contraction for the weighted simplicial complex $(\mathfrak{X}, \mu)$.

Lemma 3.2.6. *A distribution μ over $[q]^V$ satisfies the ℓ-uniform block factorization of entropy with some constant C if and only if the corresponding weighted simplicial complex $(\mathfrak{X}, \mu)$ satisfies order-$(n - \ell, n)$ global entropy contraction with rate κ, where $C\kappa = \ell/n$.*

The proof of Lemma 3.2.6 can be found in Section 3.4.1. As a consequence, to prove Theorem 3.2.4, it suffices to establish global entropy contraction for the weighted simplicial complex $(\mathfrak{X}, \mu)$.

Just like approximate tensorization and uniform block factorization having many implications for the corresponding single-site and block dynamics (e.g., see Fact 2.3.2), the notion of global entropy contraction can provide for weighted simplicial complexes meaningful bounds on the spectral gap, modified log-Sobolev constant, relative entropy decay rate, mixing time, and concentration bounds; see Fact 3.4.2 for details. In Lemma 11 of [47], the authors established order-(r, s) global entropy contraction with rate $\kappa = \frac{s-r}{s}$ for simplicial complexes with respect to homogeneous strongly log-concave distributions. From this, they deduced the modified log-Sobolev inequality for the down-up and up-down walks and showed rapid mixing of it.

We then show that for an arbitrary weighted simplicial complex $(\mathfrak{X}, w)$, one can deduce global entropy contraction from local spectral expansion whenever the marginals of the induced distributions are nicely bounded. For this, we prove a local-to-global result for entropy contraction in the spirit of [1]. If we additionally know that the marginals are nicely bounded, we can further reduce the local entropy contraction to local spectral expansion.

Lemma 3.2.7. *Let $(\mathfrak{X}, w)$ be a pure n-dimensional weighted simplicial complex. Suppose that $(\mathfrak{X}, w)$ is $(b_0, \ldots, b_{n-1})$-marginally bounded and has $(\zeta_0, \ldots, \zeta_{n-2})$-local spectral expansion. Then for all $0 \leq r < s \leq n$, $(\mathfrak{X}, w)$ satisfies order-(r, s) global entropy contraction with rate $\kappa = \kappa(r, s)$ given as in Theorem 3.2.9 below.*

We remark that Lemma 3.2.7 recovers Lemma 11 of [47] for simplicial complexes corresponding to discrete log-concave distributions, since there one has $\zeta_k = 0$ for all k as shown in [5].

We present next the proof of Theorem 3.2.4, which follows directly from Lemmas 3.2.6 and 3.2.7.

Proof of Theorem 3.2.4. From Claims 2.4.2 and 2.4.4 we know that the weighted simplicial complex $(\mathfrak{X}, \mu)$ corresponding to μ is $(b_0, \ldots, b_{n-1})$-marginally bounded with $b_k = \frac{b}{n-k}$ and has $(\zeta_0, \ldots, \zeta_{n-2})$-local spectral expansion with $\zeta_k = \frac{\eta}{n-k-1}$. Then, Lemma 3.2.7 implies that $(\mathfrak{X}, \mu)$ satisfies order-$(n-\ell, n)$ global entropy contraction for $\ell = \lceil \theta n \rceil$ with rate

$$\kappa = \frac{\sum_{k=n-\ell}^{n-1} \Gamma_k}{\sum_{k=0}^{n-1} \Gamma_k}$$

where $\Gamma_0 = 1$, $\Gamma_k = \prod_{j=0}^{k-1} \alpha_j$, and $\alpha_k = \max\left\{1 - \frac{4\eta}{b^2(n-k-1)}, \frac{1-\eta/(n-k-1)}{4+2\log((n-k)(n-k-1)/(2b^2))}\right\}$. Define an integer $R = \left\lceil \frac{4\eta}{b^2} \right\rceil$ and observe that $n \geq \ell \geq \theta n \geq 2R$ by our assumption. Thus, we have

$$\alpha_k \geq \hat{\alpha}_k := \max\left\{1 - \frac{R}{n-k-1}, 0\right\}.$$

Notice that κ, when viewed as a function of α_k's, is monotone increasing with each α_k. Thus, to lower bound κ, we can plug in the lower bounds $\hat{\alpha}_k$'s and get

$$\kappa \geq \frac{\sum_{k=n-\ell}^{n-1} \hat{\Gamma}_k}{\sum_{k=0}^{n-1} \hat{\Gamma}_k}$$

where $\hat{\Gamma}_0 = 1$ and $\hat{\Gamma}_k = \prod_{j=0}^{k-1} \hat{\alpha}_j$ for each $k \geq 1$. We will show that for every $0 \leq k \leq n-1$

one actually has

$$\hat{\Gamma}_k = \frac{(n-k-1)(n-k-2)\cdots(n-k-R)}{(n-1)(n-2)\cdots(n-R)}. \tag{3.2}$$

For $k=0$ we have $\hat{\Gamma}_0 = 1$ and Eq. (3.2) holds. For $1 \le j \le n-R-2$ we have

$$\hat{\alpha}_j = \max\left\{\frac{n-j-1-R}{n-j-1}, 0\right\} = \frac{n-j-1-R}{n-j-1}$$

and thus for $1 \le k \le n-R-1$

$$\hat{\Gamma}_k = \prod_{j=0}^{k-1} \frac{n-j-1-R}{n-j-1} = \frac{(n-k-1)(n-k-2)\cdots(n-k-R)}{(n-1)(n-2)\cdots(n-R)}.$$

Finally, since $\hat{\alpha}_j = 0$ when $n-R-1 \le j \le n-2$, we have $\hat{\Gamma}_k = 0$ for $n-R \le k \le n-1$. Therefore, Eq. (3.2) is true for all k. It then follows that

$$\begin{aligned}\kappa &\ge \frac{\sum_{k=n-\ell}^{n-1}(n-k-1)(n-k-2)\cdots(n-k-R)}{\sum_{k=0}^{n-1}(n-k-1)(n-k-2)\cdots(n-k-R)} \\ &= \frac{\sum_{j=0}^{\ell-1} j(j-1)\cdots(j-R+1)}{\sum_{j=0}^{n-1} j(j-1)\cdots(j-R+1)}.\end{aligned}$$

The following is a standard equality which can be proved by induction:

$$\sum_{j=0}^{N-1} j(j-1)\cdots(j-R+1) = \frac{1}{R+1} N(N-1)\cdots(N-R).$$

Hence, we obtain

$$\kappa \ge \frac{\ell(\ell-1)\cdots(\ell-R)}{n(n-1)\cdots(n-R)}.$$

Finally, we deduce from Lemma 3.2.6 that

$$C \le \frac{\ell}{n}\cdot\frac{1}{\kappa} \le \frac{(n-1)\cdots(n-R)}{(\ell-1)\cdots(\ell-R)} \le \left(\frac{n-R}{\ell-R}\right)^R \le \left(\frac{2n}{\ell}\right)^R \le \left(\frac{2}{\theta}\right)^{\frac{4\eta}{b^2}+1}$$

where we use our assumption $\ell \geq \theta n \geq 2R$. □

3.2.3 Wrapping up: Proof of Main Theorem 1.2.1

Combining Theorems 3.2.2 and 3.2.4, we establish approximate tensorization of entropy with a constant independent of n, when the Gibbs distribution is marginally bounded and spectrally independent. This is stated in the following theorem.

Theorem 3.2.8. *Let $\Delta \geq 3$ be an integer and $b, \eta > 0$ be reals. Suppose that $G = (V, E)$ is an n-vertex graph of maximum degree at most Δ and μ is a totally-connected Gibbs distribution of some spin system on G. If μ is both b-marginally bounded and η-spectrally independent and $n \geq \frac{24\Delta}{b^2}(\frac{4\eta}{b^2}+1)$, then μ satisfies the approximate tensorization of entropy with constant*

$$C_1 = \frac{18\log(1/b)}{b^4}\left(\frac{24\Delta}{b^2}\right)^{\frac{4\eta}{b^2}+1}.$$

Theorem 1.2.1 then follows immediately from Theorem 3.2.8 and Fact 2.3.2.

Proof of Theorem 1.2.1. Follows from Theorem 3.2.8 and Fact 2.3.2. □

3.2.4 Mixing Results for Simplicial Complexes

By Lemma 3.2.7, we obtain new bounds on the mixing time and modified log-Sobolev constant of the global down-up and up-down walks for arbitrary pure weighted simplicial complexes.

We establish the modified log-Sobolev inequality and give meaningful bounds on the mixing time for the down-up and up-down walks for arbitrary weighted simplicial complexes. Our proof utilizes the local-to-global scheme as in [1] and establishes contraction of entropy extending the result of [47]. Recall that, as we can see from Claims 2.4.2 and 2.4.4, our requirements of marginal boundedness and spectral independence in Theorem 1.2.1 is translated from the corresponding conditions needed for simplicial complexes.

We then show that for any pure weighted simplicial complexes, the modified log-Sobolev inequality holds for down-up and up-down walks if the marginal probabilities of the simplicial complex are bounded away from zero and all local walks have good expansion properties. This also bounds the mixing times of these random walks.

Theorem 3.2.9. *Let $(\mathfrak{X}, w)$ be a pure n-dimensional weighted simplicial complex. If $(\mathfrak{X}, w)$ is $(b_0, \ldots, b_{n-1})$-marginally bounded and has $(\zeta_0, \ldots, \zeta_{n-2})$-local spectral expansion, then for every $0 \leq r < s \leq n$, both the order-(s, r) down-up walk and the order-(r, s) up-down walk satisfy the modified log-Sobolev inequality with constant $\kappa = \kappa(r, s)$ defined as*

$$\kappa = \frac{\sum_{k=r}^{s-1} \Gamma_k}{\sum_{k=0}^{s-1} \Gamma_k}$$

where: $\Gamma_0 = 1$; for $1 \leq k \leq s-1$, $\Gamma_k = \prod_{j=0}^{k-1} \alpha_j$; and for $0 \leq k \leq s-2$,

$$\alpha_k = \max\left\{1 - \frac{4\zeta_k}{b_k^2(s-k)^2}, \frac{1-\zeta_k}{4 + 2\log(\frac{1}{2b_k b_{k+1}})}\right\}.$$

Furthermore, the mixing time of the order-(s, r) down-up walk is bounded by

$$T_{\mathrm{mix}}(P^{\vee}_{s,r}, \varepsilon) \leq \left\lceil \frac{1}{\kappa}\left(\log\log\frac{1}{\pi_s^*} + \log\frac{1}{2\varepsilon^2}\right)\right\rceil \tag{3.3}$$

where $\pi_s^* = \min_{\sigma \in \mathfrak{X}(s)} \pi_s(\sigma)$. The mixing time of the order-(r, s) up-down walk is also bounded by Eq. (3.3) *with π_s^* replaced by π_r^*.*

Theorem 3.2.9 follows immediately from Lemma 3.2.7 and Fact 3.4.2.

Theorem 3.2.9 generalizes both the result of [47] for simplicial complexes with respect to strongly log-concave distributions and the result of [1] for the Poincaré inequality (i.e., bounding the spectral gap). It in some sense answers a question of [47] on local-to-global modified log-Sobolev inequalities in high-dimensional expanders, at least in the bounded marginals setting.

Even though Theorem 3.2.9 can give a bound on the mixing time of the Glauber dynamics, which is the order-$(n, n-1)$ down-up walk in the corresponding weighted simplicial complex, our main result Theorem 1.2.1 does *not* follow directly from Theorem 3.2.9. In fact, we will only consider the order-$(n, n-\ell)$ down-up walk for $\ell = \Theta(n)$, which corresponds to the heat-bath block dynamics that updates a uniformly random subset of ℓ vertices in every step. One of our main technical contributions is to compare this block dynamics with the single-site Glauber dynamics; this is showed in Section 3.2.1. Nevertheless, we find Theorem 3.2.9 interesting of its own and possible for future applications in other problems.

3.3 Approximate Tensorization via Uniform Block Factorization

In this section we prove Theorem 3.2.2 by showing approximate tensorization given uniform block factorization of entropy.

Fix a graph G on n vertices of maximum degree at most Δ, and assume that μ is a b-marginally bounded Gibbs distribution defined on G satisfying the $\lceil \theta n \rceil$-uniform block factorization of entropy with constant C where $\theta \leq b^2/(4e\Delta)$; i.e., for $\ell = \lceil \theta n \rceil$ and all $f : \Omega \to \mathbb{R}_{\geq 0}$ it holds that

$$\frac{\ell}{n} \operatorname{Ent}(f) \leq C \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu[\operatorname{Ent}_S(f)].$$

We will show that μ also satisfies the approximate tensorization of entropy with constant $\Theta(C)$, which establishes Theorem 3.2.2.

The intuition behind our approach is that for ℓ as large as θn, if one picks a uniformly random subset $S \subseteq V$ satisfying $|S| = \ell$, then the induced subgraph $G[S]$ of G on vertex set S is disconnected into many small connected components, each of which has constant size in expectation and at most $O(\log n)$ with high probability. Since the conditional Gibbs distribution μ_S^τ is a product distribution of each connected component, we can use entropy

factorization for product distributions to reduce approximate tensorization on G to that on small connected subgraphs of G. This allows us to upper bound the optimal approximate tensorization constant with a converging series.

Towards fulfilling this intuition, for any $S \subseteq V$, let $\mathcal{C}(S)$ denote the set of connected components of $G[S]$, with each connected component being viewed as a subset of vertices of S. Note that $\mathcal{C}(S)$ is a partition of S. For any $v \in S$, let S_v denote the (unique) connected component in $\mathcal{C}(S)$ containing v; for $v \notin S$, take $S_v = \emptyset$. The following is a well-known fact regarding the factorization of entropy for product measures; see, e.g., [41, 36].

Lemma 3.3.1. *For every subset $S \subseteq V$, every boundary condition $\tau \in \Omega_{V \setminus S}$, and every function $f : \Omega_S^\tau \to \mathbb{R}_{\geq 0}$, we have*

$$\mathrm{Ent}_S^\tau(f) \leq \sum_{U \in \mathcal{C}(S)} \mu_S^\tau[\mathrm{Ent}_U(f)].$$

Recall that $\mathrm{Ent}_U(f) = \mathrm{Ent}_U^\phi(f)$ is regarded as a function of the boundary condition $\phi \in \Omega_{S \setminus U}^\tau$ on $S \setminus U$, and $\mu_S^\tau[\mathrm{Ent}_U(f)]$ is the expectation of it under the conditional Gibbs measure μ_S^τ.

We also need the following crude exponential upper bound on the approximate tensorization constant for a Gibbs distribution with bounded marginals.

Lemma 3.3.2. *If μ is b-marginally bounded, then for every subset $U \subseteq V$, every boundary condition $\xi \in \Omega_{V \setminus U}$, and every function $f : \Omega_U^\xi \to \mathbb{R}_{\geq 0}$, we have*

$$\mathrm{Ent}_U^\xi(f) \leq \frac{3|U|^2 \log(1/b)}{2b^{2|U|+2}} \sum_{v \in U} \mu_U^\xi[\mathrm{Ent}_v(f)].$$

Finally, the lemma below shows that when a uniformly random and sufficiently small subset of vertices is selected, the size of the connected component containing a given vertex is small with high probability.

Lemma 3.3.3. *Let $G = (V, E)$ be an n-vertex graph of maximum degree at most Δ. Then for every $k \in \mathbb{N}^+$ we have*

$$\mathbb{P}_S(|S_v| = k) \leq \frac{\ell}{n} \cdot (2e\Delta\theta)^{k-1},$$

where the probability $\mathbb{P}$ is taken over a uniformly random subset $S \subseteq V$ of size $\ell = \lceil \theta n \rceil$.

Before presenting the proofs of Lemmas 3.3.2 and 3.3.3, we first give the proof of Theorem 3.2.2.

Proof of Theorem 3.2.2. Combining everything in this section, we deduce that

$$\begin{aligned}
\mathrm{Ent}(f) &\leq C \cdot \frac{n}{\ell} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu[\mathrm{Ent}_S(f)] && (\ell\text{-uniform block factorization}) \\
&\leq C \cdot \frac{n}{\ell} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \sum_{U \in \mathcal{C}(S)} \mu[\mathrm{Ent}_U(f)] && (\text{Lemma 3.3.1}) \\
&\leq C \cdot \frac{n}{\ell} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \sum_{U \in \mathcal{C}(S)} \frac{3|U|^2 \log(1/b)}{2b^{2|U|+2}} \sum_{v \in U} \mu[\mathrm{Ent}_v(f)] && (\text{Lemma 3.3.2}) \\
&= \frac{3C \log(1/b)}{2b^4} \cdot \frac{n}{\ell} \sum_{v \in V} \mu[\mathrm{Ent}_v(f)] \sum_{k=1}^{\ell} \mathbb{P}_S(|S_v| = k) \cdot \frac{k^2}{b^{2(k-1)}} && (\text{rearranging}) \\
&\leq \frac{3C \log(1/b)}{2b^4} \sum_{v \in V} \mu[\mathrm{Ent}_v(f)] \sum_{k=1}^{\ell} k^2 \left(\frac{2e\Delta\theta}{b^2}\right)^{k-1} && (\text{Lemma 3.3.3}) \\
&\leq \frac{3C \log(1/b)}{2b^4} \sum_{k=1}^{\ell} \frac{k^2}{2^{k-1}} \sum_{v \in V} \mu[\mathrm{Ent}_v(f)] && (\theta \leq \tfrac{b^2}{12\Delta}) \\
&\leq \frac{18C \log(1/b)}{b^4} \sum_{v \in V} \mu[\mathrm{Ent}_v(f)]. && (\textstyle\sum_{k=1}^{\infty} \frac{k^2}{2^{k-1}} = 12)
\end{aligned}$$

This establishes the lemma. □

We next prove Lemma 3.3.2 which gives a crude bound on the approximate tensorization constant for any subset and boundary condition.

Proof of Lemma 3.3.2. Fix a subset $U \subseteq V$ of size $k \geq 1$ and some boundary condition

$\xi \in \Omega_{V\setminus U}$. Let $C_1 = C_1(U,\xi)$ be the optimal constant of approximate tensorization for μ_U^ξ; hence, for every function $f : \Omega_U^\xi \to \mathbb{R}_{\geq 0}$ one has

$$\mathrm{Ent}_U^\xi(f) \leq C_1 \sum_{v\in U} \mu_U^\xi[\mathrm{Ent}_v(f)].$$

Let $\lambda = \lambda(U,\xi)$ be the spectral gap of the Glauber dynamics for μ_U^ξ, and let $\rho = \rho(U,\xi)$ be the standard log-Sobolev constant. Thus, for every function $f : \Omega_U^\xi \to \mathbb{R}_{\geq 0}$ it holds that

$$\lambda \,\mathrm{Var}_U^\xi(f) \leq \frac{1}{k}\sum_{v\in U} \mu_U^\xi[\mathrm{Var}_v(f)];$$
$$\rho \,\mathrm{Ent}_U^\xi(f) \leq \frac{1}{k}\sum_{v\in U} \mu_U^\xi[\mathrm{Var}_v(\sqrt{f})].$$

Since $\mathrm{Var}_v(\sqrt{f}) \leq \mathrm{Ent}_v(f)$, we have

$$C_1 \leq \frac{1}{\rho k}; \tag{3.4}$$

see also [36, Proposition 1.1]. Next, [48, Corollary A.4] gives a comparison between the standard log-Sobolev constant and the spectral gap:

$$\rho \geq \frac{(1-2\mu^*)}{\log(1/\mu^* - 1)}\lambda$$

where $\mu^* = \min_{\sigma\in\Omega_U^\xi} \mu_U^\xi(\sigma)$. Since μ is b-marginally bounded, we have $\mu^* \geq b^k$. Also, notice that $|\Omega_U^\xi| = 1$ and $|\Omega_U^\xi| = 2$ corresponds to trivial cases where we have $C_1 \leq 1$, so we may assume that $|\Omega_U^\xi| \geq 3$ which makes $\mu^* \leq 1/3$. It follows that

$$\rho \geq \frac{\lambda}{3k\log(1/b)}. \tag{3.5}$$

Finally, Cheeger's inequality yields

$$\lambda \geq \frac{\Phi^2}{2} \tag{3.6}$$

where Φ is the conductance of the Glauber dynamics defined by

$$\Phi = \min_{\substack{\Omega_0 \subseteq \Omega_U^\xi \\ \mu_U^\xi(\Omega_0) \leq \frac{1}{2}}} \Phi_{\Omega_0}, \quad \Phi_{\Omega_0} = \frac{P_{\text{GL}}(\Omega_0, \Omega_U^\xi \setminus \Omega_0)}{\mu_U^\xi(\Omega_0)} = \frac{1}{\mu_U^\xi(\Omega_0)} \sum_{\sigma \in \Omega_0} \sum_{\tau \in \Omega_U^\xi \setminus \Omega_0} \mu_U^\xi(\sigma) P_{\text{GL}}(\sigma, \tau).$$

Our assumption that μ is totally-connected guarantees $\Phi_{\Omega_0} > 0$ for every $\Omega_0 \subseteq \Omega_U^\xi$ with $\mu_U^\xi(\Omega_0) \leq \frac{1}{2}$. Furthermore, since μ is b-marginally bounded, for every $\sigma \in \Omega_0$ and $\tau \in \Omega_U^\xi \setminus \Omega_0$ such that $P_{\text{GL}}(\sigma, \tau) > 0$ we have

$$\mu_U^\xi(\sigma) P_{\text{GL}}(\sigma, \tau) \geq b^k \cdot \frac{b}{k} = \frac{b^{k+1}}{k}.$$

This gives

$$\Phi \geq \frac{2b^{k+1}}{k}. \tag{3.7}$$

Combining Eqs. (3.4) to (3.7), we finally conclude that

$$C_1 \leq \frac{3k^2 \log(1/b)}{2b^{2k+2}},$$

as claimed. □

Finally, we establish Lemma 3.3.3. We use the following lemma concerning the number of connected induced subgraphs in a bounded degree graph.

Lemma 3.3.4 ([28, Lemma 2.1])**.** *Let $G = (V, E)$ be a graph with maximum degree at most Δ, and $v \in V$. Then for every $k \in \mathbb{N}^+$, the number of connected induced subgraphs of G containing v with k vertices is at most $(e\Delta)^{k-1}$.*

We then prove Lemma 3.3.3.

Proof of Lemma 3.3.3. If $\mathcal{A}_v(k)$ denotes the collection of subsets of vertices $U \subseteq V$ such

that $|U| = k$, $v \in U$, and $G[U]$ is connected, then by the union bound, we have

$$\begin{aligned}
\mathbb{P}_S(|S_v| = k) &\leq \mathbb{P}_S(\exists U \in \mathcal{A}_v(k) : U \subseteq S) \\
&\leq \sum_{U \in \mathcal{A}_v(k)} \mathbb{P}_S(U \subseteq S) \\
&= |\mathcal{A}_v(k)| \cdot \frac{\ell}{n} \cdot \frac{\ell-1}{n-1} \cdots \frac{\ell-k+1}{n-k+1} \\
&\leq |\mathcal{A}_v(k)| \cdot \frac{\ell}{n} \cdot \left(\frac{\ell-1}{n-1}\right)^{k-1}.
\end{aligned}$$

We may assume that $n \geq 2$ (when $n = 1$ the lemma holds trivially), and thus

$$\frac{\ell-1}{n-1} \leq \frac{\theta n}{n-1} \leq 2\theta.$$

The lemma then follows immediately from $|\mathcal{A}_v(k)| \leq (e\Delta)^{k-1}$ by Lemma 3.3.4. □

3.4 Uniform Block Factorization via Spectral Independence

In this section, we establish uniform block factorization of entropy in the framework of pure weighted simplicial complexes. We show that spectral independence, along with suitable marginal bounds, implies uniform block factorization and thus prove Lemma 3.2.7.

We prove Lemma 3.2.7 by establishing global entropy contraction when the simplicial complex is a local spectral expander. We give preliminaries in Section 3.4.1 for simplicial complexes. In Section 3.4.2 we present a very general local-to-global scheme for entropy contraction. Finally, in Section 3.4.3 we reduce local entropy contraction to local spectral expansion.

3.4.1 Preliminaries for Simplicial Complexes

Following the definitions and notations from Section 2.4, here we give a few more definitions, examples, and references, emphasizing on the global and local walks in weighted

simplicial complexes. The proofs of Claim 2.4.2 and Lemma 3.2.6 are included in Section 3.4.1.

Consider a pure n-dimensional weighted simplicial complex $(\mathfrak{X}, w)$. We say $\mathfrak{X}$ is n-*partite* if the ground set $\mathcal{U}$ of $\mathfrak{X}$ admits a partition $\mathcal{U} = \mathcal{U}_1 \cup \cdots \cup \mathcal{U}_n$ such that every face of $\mathfrak{X}$ has at most one element from each part $\mathcal{U}_1, \ldots, \mathcal{U}_n$. For a distribution μ over $[q]^V$ where $|V| = n$, the corresponding weighted simplicial complex $(\mathfrak{X}, \mu)$ is n-partite.

Global Walks

For $0 \leq k \leq n$, the distribution π_k on $\mathfrak{X}(k)$ is given by for every $\tau \in \mathfrak{X}(k)$,

$$\pi_k(\tau) = \frac{1}{\binom{n}{k}} \sum_{\sigma \in \mathfrak{X}(n): \sigma \supseteq \tau} \pi_n(\sigma)$$

Moreover, we have the following equality:

$$\pi_k(\tau) = \frac{1}{k+1} \sum_{\sigma \in \mathfrak{X}(k+1): \sigma \supseteq \tau} \pi_{k+1}(\sigma).$$

We may define simple random walks on $\mathfrak{X}(k)$ for sampling from the distributions π_k given by "down-up" and "up-down" motions in $\mathfrak{X}$. These walks were first introduced in [80], and further studied in [49, 82] in the context of high-dimensional expanders. Recent works [5, 47, 4, 1, 44, 55, 43] have leveraged these walks to study mixing times of Markov chains.

For each $0 \leq k \leq n-1$, define the *order-k (global) up operator* $P_k^{\uparrow} \in \mathbb{R}^{\mathfrak{X}(k) \times \mathfrak{X}(k+1)}$ and the *order-$(k+1)$ (global) down operator* $P_{k+1}^{\downarrow} \in \mathbb{R}^{\mathfrak{X}(k+1) \times \mathfrak{X}(k)}$ as the row stochastic matrices with the following entries:

$$P_k^{\uparrow}(\tau, \sigma) = \begin{cases} \frac{\pi_{k+1}(\sigma)}{(k+1)\pi_k(\tau)}, & \text{if } \tau \subseteq \sigma \\ 0, & \text{o.w.} \end{cases}$$

$$P^{\downarrow}_{k+1}(\sigma,\tau) = \begin{cases} \frac{1}{k+1}, & \text{if } \tau \subseteq \sigma \\ 0, & \text{o.w.} \end{cases}$$

In words, the action of $P^{\uparrow}_k$ can be described as starting with $\tau \in \mathfrak{X}(k)$ and adding a random element $i \notin \tau$ according to the conditional distribution $\pi_{\tau,1}(i) = \frac{\pi_{k+1}(\tau \cup \{i\})}{(k+1)\pi_k(\tau)}$. In particular, we have $\pi_k P^{\uparrow}_k = \pi_{k+1}$. Similarly, the action of $P^{\downarrow}_{k+1}$ can be described as starting with $\sigma \in \mathfrak{X}(k+1)$ and removing a uniformly random element, and we have $\pi_{k+1}P^{\downarrow}_{k+1} = \pi_k$.

From here, for $0 \le j < k \le n$ we may define the order-(j,k) (global) up-down walk on $\mathfrak{X}(j)$ as $P^{\wedge}_{j,k} = P^{\uparrow}_j \cdots P^{\uparrow}_{k-1} P^{\downarrow}_k \cdots P^{\downarrow}_{j+1}$ and the order-(k,j) (global) down-up walk on $\mathfrak{X}(k)$ as $P^{\vee}_{k,j} = P^{\downarrow}_k \cdots P^{\downarrow}_{j+1} P^{\uparrow}_j \cdots P^{\uparrow}_{k-1}$. The stationary distributions of $P^{\wedge}_{j,k}$ and $P^{\vee}_{k,j}$ are π_j and π_k respectively.

For example, for any distribution μ over $[q]^V$, the Glauber dynamics for sampling from μ is precisely the order-$(n, n-1)$ down-up walk on the corresponding weighted simplicial complex $(\mathfrak{X}, \mu)$, and the heat-bath block dynamics which updates a random subset of ℓ vertices in each step is precisely the order-$(n, n-\ell)$ down-up walk.

Let $0 < k \le n$ and let $f^{(k)} : \mathfrak{X}(k) \to \mathbb{R}$ be an arbitrary function. Then for each $0 \le j < k$ we define the function $f^{(j)} : \mathfrak{X}(j) \to \mathbb{R}$ by $f^{(j)} = P^{\uparrow}_j \cdots P^{\uparrow}_{k-1} f^{(k)}$. One can think of $f^{(j)}$ as the projection of $f^{(k)}$ onto $\mathfrak{X}(j)$.

Local Walks

One of the beautiful properties of simplicial complexes is that they facilitate a nice decomposition of the global walks into local walks. Recall that for every face $\tau \in \mathfrak{X}(k)$ there is a $(n-k)$-dimensional weighted simplicial subcomplex $(\mathfrak{X}_\tau, w_\tau)$ where $\mathfrak{X}_\tau = \{\xi \subseteq \mathcal{U} \setminus \tau : \tau \cup \xi \in \mathfrak{X}\}$ and $w_\tau(\xi) = w(\tau \cup \xi)$ for each $\xi \in \mathfrak{X}_\tau$. The subcomplex is known as the *link* of $\mathfrak{X}$ with respect to the face τ. The induced distribution $\pi_{\tau,n-k}$ over $\mathfrak{X}_\tau(n-k)$ can be thought of as the distribution of a face σ obtained from π_n conditioned on the event that $\tau \subseteq \sigma$.

Again, consider as an example the weighted simplicial complex $(\mathfrak{X}, \mu)$ with respect to a distribution μ over $[q]^V$. For every $U \subseteq V$ of size k and every $\tau \in \Omega_U$, one can check that the distribution $\pi_{(U,\tau),n-k}$ over maximal faces of the link $(\mathfrak{X}_{(U,\tau)}, w_{(U,\tau)})$ is exactly the conditional distribution $\mu^{\tau}_{V\setminus U}$ over configurations on $V \setminus U$.

With the notion of link, we may now define the local walks rigorously. For each $0 \leq k \leq n-2$ and $\tau \in \mathfrak{X}(k)$, we define the *local walk* at τ as $P_\tau = 2P^{\wedge}_{\tau,1,2} - I$ where $P^{\wedge}_{\tau,1,2} = P^{\uparrow}_{\tau,1}P^{\downarrow}_{\tau,2}$ is the order-$(1,2)$ up-down walk on the link $(\mathfrak{X}_\tau, w_\tau)$. More specifically, for $i, j \in \mathfrak{X}_\tau(1)$, we have

$$P_\tau(i,j) = \begin{cases} \pi_{\tau\cup\{i\},1}(j) = \frac{\pi_{\tau,2}(\{i,j\})}{2\pi_{\tau,1}(i)}, & \text{if } \{i,j\} \in \mathfrak{X}_\tau(2) \\ 0, & \text{o.w.} \end{cases}$$

Note that the stationary distribution of P_τ is $\pi_{\tau,1}$. One way to think of P_τ is being the random walk matrix corresponding to the following weighted graph: take the vertices of the weighted graph to be the elements of $\mathfrak{X}_\tau(1)$, and take the edges to be the vertex-pairs in $\mathfrak{X}_\tau(2)$, with weights given by $\pi_{\tau,2}$.

We can establish mixing properties of global walks by decomposing them into local walks, which are usually easier to study; we refer the readers to [82, 5, 47, 1] for more details.

Let $0 < k \leq n$ and let $f : \mathfrak{X}(k) \to \mathbb{R}$ be an arbitrary function. Then for every $0 \leq j < k$ and every $\tau \in \mathfrak{X}(j)$, we define the function $f^{(k-j)}_\tau : \mathfrak{X}_\tau(k-j) \to \mathbb{R}$ by $f^{(k-j)}_\tau(\xi) = f^{(k)}(\tau \cup \xi)$ for each $\xi \in \mathfrak{X}_\tau(k-j)$. One can think of $f^{(k-j)}_\tau$ as the restriction of $f^{(k)}$ to the link $\mathfrak{X}_\tau(k-j)$.

Finally, we mention that the notion of global and local walks has a similar flavor as the restriction-projection framework used in [97, 95, 78] for decomposition of general Markov chains. For instance, the analog of the "restriction chains" in the setting of simplicial complexes are the Glauber dynamics of the conditional distributions. However, we note

that these analogous "restriction chains" significantly overlap, in contrast to the partitioning condition in [97, 95, 78].

Proofs of Claim 2.4.2 and Lemma 3.2.6

We present here the proofs of Claim 2.4.2 and Lemma 3.2.6.

Proof of Claim 2.4.2. For all $0 \le k \le n-1$, every $\tau \in \mathfrak{X}(k)$, and every $i \in \mathfrak{X}_\tau(1)$, we have

$$\pi_{\tau,1}(i) = \frac{\pi_{k+1}(\tau \cup \{i\})}{\sum_{j \in \mathfrak{X}_\tau(1)} \pi_{k+1}(\tau \cup \{j\})} = \frac{\pi_{k+1}(\tau \cup \{i\})}{(k+1)\pi_k(\tau)}.$$

Now, consider the weighted simplicial complex $(\mathfrak{X}, \mu)$ corresponding to a distribution μ over $[q]^V$. Every face in $\mathfrak{X}(k)$ is in the form (U, τ) where $U \subseteq V$, $|U| = k$, and $\tau \in \Omega_U$, and every element $i \in \mathfrak{X}_\tau(1)$ is in the form (v, i) for some $v \in V \setminus U$ and $i \in \Omega_v^\tau$. Hence, the equality above implies that

$$\begin{aligned}
\pi_{(U,\tau),1}\big((v,i)\big) &= \frac{\pi_{k+1}\big((U \cup \{v\}, \tau \cup \{i\})\big)}{(k+1)\pi_k\big((U,\tau)\big)} \\
&= \frac{\frac{1}{\binom{n}{k+1}}\,\mu(\sigma_U = \tau, \sigma_v = i)}{(k+1)\cdot\frac{1}{\binom{n}{k}}\,\mu(\sigma_U = \tau)} \\
&= \frac{1}{n-k}\,\mu(\sigma_v = i \mid \sigma_U = \tau) \\
&\ge \frac{b}{n-k}
\end{aligned}$$

where the last inequality follows from the marginal boundedness of μ. This shows the claim. □

From the proof we see that $\pi_{(U,\tau),1}\big((v,i)\big) \le \frac{1}{n-k}$ always holds. This in fact is true for all weighted simplicial complexes.

To prove Lemma 3.2.6, we need the following entropy decomposition result in [47].

Lemma 3.4.1 (Entropy Decomposition in Simplicial Complexes, [47])**.** *Let $(\mathfrak{X}, w)$ be a pure n-dimensional weighted simplicial complex, and let $1 \le j < k \le n$. Then the*

following decomposition of entropy holds: for every $f^{(k)} : \mathfrak{X}(k) \to \mathbb{R}_{\geq 0}$*,*

$$\mathrm{Ent}_{\pi_k}(f^{(k)}) = \mathrm{Ent}_{\pi_j}(f^{(j)}) + \sum_{\tau \in \mathfrak{X}(j)} \pi_j(\tau) \cdot \mathrm{Ent}_{\pi_{\tau,k-j}}(f_\tau^{(k-j)}).$$

Proof. Since $\pi_k(f^{(k)}) = \pi_j(f^{(j)})$ and the equality is scale invariant, we may assume without loss that $\pi_k(f^{(k)}) = \pi_j(f^{(j)}) = 1$; hence $\mathrm{Ent}_{\pi_k}(f^{(k)}) = \pi_k(f^{(k)} \log f^{(k)})$ and $\mathrm{Ent}_{\pi_j}(f^{(j)}) = \pi_j(f^{(j)} \log f^{(j)})$. Using the law of conditional expectation, we have

$$\begin{aligned}
&\pi_k\left(f^{(k)} \log f^{(k)}\right) \\
&= \sum_{\tau \in \mathfrak{X}(j)} \pi_j(\tau) \cdot \pi_{\tau,k-j}\left(f_\tau^{(k-j)} \log f_\tau^{(k-j)}\right) \\
&= \sum_{\tau \in \mathfrak{X}(j)} \pi_j(\tau) \cdot \underbrace{\pi_{\tau,k-j}(f_\tau^{(k-j)})}_{=f^{(j)}(\tau)} \log \underbrace{\pi_{\tau,k-j}(f_\tau^{(k-j)})}_{=f^{(j)}(\tau)} + \sum_{\tau \in \mathfrak{X}(j)} \pi_j(\tau) \cdot \mathrm{Ent}_{\pi_{\tau,k-j}}(f_\tau^{(k-j)}) \\
&= \pi_j\left(f^{(j)} \log f^{(j)}\right) + \sum_{\tau \in \mathfrak{X}(j)} \pi_j(\tau) \cdot \mathrm{Ent}_{\pi_{\tau,k-j}}(f_\tau^{(k-j)})
\end{aligned}$$

as claimed. □

Proof of Lemma 3.2.6. Since there is a one-to-one correspondence between Ω and $\mathfrak{X}(n)$, every function $f : \Omega \to \mathbb{R}_{\geq 0}$ for the spin system is equivalent to a function $f^{(n)} : \mathfrak{X}(n) \to \mathbb{R}_{\geq 0}$ for the simplicial complex. Moreover, for every $0 \leq k \leq n-1$, every $U \subseteq V$ of size k, and every $\tau \in \Omega_U$, the function $f_\tau : \Omega_{V \setminus U} \to \mathbb{R}_{\geq 0}$ (restriction of f to configurations on $V \setminus U$ with U fixed to be τ) is the same as the function $f_{(U,\tau)}^{(n-k)} : \mathfrak{X}_{(U,\tau)}(n-k) \to \mathbb{R}_{\geq 0}$ for the link with respect to $(U, \tau) \in \mathfrak{X}(k)$. Thus, we can get

$$\begin{aligned}
\frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu[\mathrm{Ent}_S(f)] &= \sum_{S \in \binom{V}{\ell}} \sum_{\tau \in \Omega_{V \setminus S}} \frac{1}{\binom{n}{\ell}} \cdot \mu(\sigma_{V \setminus S} = \tau) \cdot \mathrm{Ent}_S^\tau(f) \\
&= \sum_{U \in \binom{V}{n-\ell}} \sum_{\tau \in \Omega_U} \frac{1}{\binom{n}{\ell}} \cdot \mu(\sigma_U = \tau) \cdot \mathrm{Ent}_{V \setminus U}^\tau(f_\tau) \\
&= \sum_{(U,\tau) \in \mathfrak{X}(n-\ell)} \pi_{n-\ell}(U, \tau) \cdot \mathrm{Ent}_{\pi_{(U,\tau),\ell}}(f_{(U,\tau)}^{(\ell)})
\end{aligned}$$

$$= \mathrm{Ent}_{\pi_n}(f^{(n)}) - \mathrm{Ent}_{\pi_{n-\ell}}(f^{(n-\ell)})$$

where the last equality follows from Lemma 3.4.1. The lemma then follows. □

Implications of Global Entropy Contraction

We summarize here a few corollaries of global entropy contraction for arbitrary weighted simplicial complexes.

Fact 3.4.2. *Let $(\mathfrak{X}, w)$ be a pure n-dimensional weighted simplicial complex. If $(\mathfrak{X}, w)$ satisfies the order-(r, s) global entropy contraction with rate κ, then the order-(s, r) down-up walk and order-(r, s) up-down walk satisfy all of the following:*

(1) The Poincaré inequality holds with constant $\lambda = \kappa$;

(2) The modified log-Sobolev inequality holds with constant $\rho_0 = \kappa$;

(3) The relative entropy decays with rate $\alpha = \kappa$;

(4) The mixing time of the order-(s, r) down-up walk satisfies

$$T_{\mathrm{mix}}(P^{\vee}_{s,r}, \varepsilon) \leq \left\lceil \frac{1}{\kappa} \left(\log\log \frac{1}{\pi^*_s} + \log \frac{1}{2\varepsilon^2} \right) \right\rceil$$

where $\pi^*_s = \min_{\sigma \in \mathfrak{X}(s)} \pi_s(\sigma)$; The same mixing time bound holds for the order-(r, s) up-down walk as well with π^*_s replaced by π^*_r.

(5) For every $f : \mathfrak{X}(s) \to \mathbb{R}$ which is c-Lipschitz w.r.t. the shortest path metric induced by the order-(s, r) down-up walk on $\mathfrak{X}(s)$, and every $a \geq 0$, we have the concentration inequality

$$\Pr_{\sigma \sim \pi_s} \left[|f(\sigma) - \mu(f)| \geq a \right] \leq 2 \exp\left(-\frac{\kappa a^2}{2c^2} \right).$$

These implications are shown in [47]. We remark that Fact 2.3.2 can also be viewed as a consequence of Fact 3.4.2 by Lemma 3.2.6. Technically we can also show a stan-

dard log-Sobolev inequality under the assumption of marginal boundedness, like Item 6 of Fact 2.3.2. However, the constant is complicated to state in this setting and so we omit it here.

Proof. Item 1 can be proved by the linearization argument in the proof of [36, Proposition 1.1]. The remaining results are all shown in [47]: Item 3 follows from [47, Corollary 13]; Item 2 follows by Item 3 and [47, Theorem 7]; Item 4 can be deduced from Item 3 and the direct proof of [47, Corollary 8]; finally, Item 5 follows from Item 2 and [47, Lemma 15]. □

3.4.2 Local-to-Global Entropy Contraction

As mentioned earlier, one of the beautiful properties of simplicial complexes is that they admit a local-to-global phenomenon, where mixing of the local walks P_τ imply mixing for the global walks in a quantitative sense. For instance, $(\frac{C}{n-1}, \frac{C}{n-2}, \dots, C)$-local spectral expansion implies $\Omega(n^{-(1+C)})$ spectral gap for the Glauber dynamics [1] (although we note weaker results of this type were previously known due to [49, 82]). It turns out this local-to-global spectral result is completely equivalent to the property that local contraction of variance implies global contraction of variance; see [45] for more discussion. The goal of this section is to prove an analogous local-to-global result for entropy. First, we formalize our notion of local entropy contraction.

Definition 3.4.3 (Local Entropy Contraction)**.** We say a pure n-dimensional weighted simplicial complex $(\mathfrak{X}, w)$ satisfies $(\alpha_0, \dots, \alpha_{n-2})$-*local entropy contraction* if for every (global) function $f^{(n)} : \mathfrak{X}(n) \to \mathbb{R}_{\geq 0}$, every $0 \leq k \leq n-2$, and every $\tau \in \mathfrak{X}(k)$, we have

$$\mathrm{Ent}_{\pi_{\tau,2}}(f_\tau^{(2)}) \geq (1+\alpha_k)\,\mathrm{Ent}_{\pi_{\tau,1}}(f_\tau^{(1)}).$$

Recall that $f_\tau^{(2)}(\xi) = f^{(k+2)}(\tau \cup \xi) = (P_{k+2}^{\uparrow} \cdots P_{n-1}^{\uparrow} f^{(n)})(\tau \cup \xi)$ for each $\xi \in \mathfrak{X}_\tau(2)$, and $f_\tau^{(1)}$ is defined similarly.

One might want to replace the conditions in Definition 3.4.3 with "for every $0 \leq k \leq n-2$, every $\tau \in \mathfrak{X}(k)$, and every (local) function $f_\tau^{(2)} : \mathfrak{X}_\tau(2) \to \mathbb{R}_{\geq 0}$", resulting in a stronger notion of local entropy contraction. Indeed, this is the case for local spectral expansion or local variance contraction. In contrast, our Definition 3.4.3 only considers local functions originating from global functions. This is sufficient since in the end we are only interested in the order-(n,k) down-up walk $P^{\vee}_{n,n-\ell}$ and only need to consider global functions $f^{(n)} : \mathfrak{X}(n) \to \mathbb{R}_{\geq 0}$. Moreover, using this weaker notion makes it easier for us to deduce local entropy contraction from local spectral expansion in Section 3.4.3.

With the notion of local entropy contraction, we are able to prove the following local-to-global entropy contraction result.

Theorem 3.4.4 (Local-to-Global Entropy Contraction)**.** *Suppose a pure n-dimensional weighted simplicial complex $(\mathfrak{X}, w)$ satisfies $(\alpha_0, \ldots, \alpha_{n-2})$-local entropy contraction. For every function $f^{(n)} : \mathfrak{X}(n) \to \mathbb{R}_{\geq 0}$ and all $1 \leq k \leq n-1$, we have*

$$\frac{\mathrm{Ent}_{\pi_{k+1}}(f^{(k+1)})}{\sum_{i=0}^{k} \Gamma_i} \geq \frac{\mathrm{Ent}_{\pi_j}(f^{(k)})}{\sum_{i=0}^{k-1} \Gamma_i} \tag{3.8}$$

where $\Gamma_i = \prod_{j=0}^{i-1} \alpha_j$ for $1 \leq i \leq n-1$ and $\Gamma_0 = 1$. In particular, $(\mathfrak{X}, w)$ satisfies the order-(k,n) global entropy contraction with rate

$$\kappa = \frac{\sum_{i=k}^{n-1} \Gamma_i}{\sum_{i=0}^{n-1} \Gamma_i}.$$

Remark 3.4.5. After we posted a preliminary version of [45], it was brought to our attention that Guo and Mousa had also independently obtained this result [68]. The analogous result for variance, which was presented in [45], was also independently obtained by Kaufman and Mass [81]. Finally, we mention the recent work [3] which established local-to-global arguments for any f-divergence in a very general setting.

In [47], the authors showed that the simplicial complexes with respect to homogeneous

strongly log-concave distributions satisfy $(1, \ldots, 1)$-local entropy contraction; see Lemma 10 of [47]. That is, $\alpha_k = 1$ for all k and thus $\Gamma_k = 1$ for all k. Then Theorem 3.4.4 implies that for every k we have

$$\mathrm{Ent}_{\pi_{k+1}}(f^{(k+1)}) \geq \frac{k+1}{k} \mathrm{Ent}_{\pi_k}(f^{(k)}),$$

which recovers Lemma 11 of [47].

Proof of Theorem 3.4.4. We prove Eq. (3.8) by induction. When $k = 1$, Eq. (3.8) is equivalent to

$$\mathrm{Ent}_{\pi_2}(f^{(2)}) \geq (1 + \alpha_0) \mathrm{Ent}_{\pi_1}(f^{(1)}),$$

which holds by assumption. Now suppose Eq. (3.8) holds for $k - 1$; i.e.,

$$\frac{\mathrm{Ent}_{\pi_k}(f^{(k)})}{\sum_{i=0}^{k-1} \Gamma_i} \geq \frac{\mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)})}{\sum_{i=0}^{k-2} \Gamma_i}.$$

By Lemma 3.4.1 we have

$$\begin{aligned}
\mathrm{Ent}_{\pi_{k+1}}(f^{(k+1)}) - \mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) &= \sum_{\tau \in \mathfrak{X}(k-1)} \pi_{k-1}(\tau) \cdot \mathrm{Ent}_{\pi_{\tau,2}}(f_\tau^{(2)}) \\
&\geq (1 + \alpha_{k-1}) \sum_{\tau \in \mathfrak{X}(k-1)} \pi_{k-1}(\tau) \cdot \mathrm{Ent}_{\pi_{\tau,1}}(f_\tau^{(1)}) \\
&= (1 + \alpha_{k-1}) \left(\mathrm{Ent}_{\pi_k}(f^{(k)}) - \mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) \right).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{Ent}_{\pi_{k+1}}(f^{(k+1)}) &\geq (1 + \alpha_{k-1}) \mathrm{Ent}_{\pi_k}(f^{(k)}) - \alpha_{k-1} \mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) \\
&\geq (1 + \alpha_{k-1}) \mathrm{Ent}_{\pi_k}(f^{(k)}) - \alpha_{k-1} \cdot \frac{\sum_{i=0}^{k-2} \Gamma_i}{\sum_{i=0}^{k-1} \Gamma_i} \cdot \mathrm{Ent}_{\pi_k}(f^{(k)}) \quad \text{(Induction)} \\
&= \left(1 + \alpha_{k-1} \cdot \frac{\Gamma_{k-1}}{\sum_{i=0}^{k-1} \Gamma_i} \right) \mathrm{Ent}_{\pi_k}(f^{(k)})
\end{aligned}$$

$$= \frac{\sum_{i=0}^{k} \Gamma_i}{\sum_{i=0}^{k-1} \Gamma_i} \cdot \mathrm{Ent}_{\pi_k}(f^{(k)}).$$

This proves the theorem. □

3.4.3 Local Entropy Contraction via Local Spectral Expansion

Thanks to Theorem 3.4.4, it remains to show local entropy contraction in order to prove Lemma 3.2.7. In general, proving local entropy contraction is a difficult task. Our goal here is to show that one can deduce local entropy contraction from local spectral expansion if we further enforce certain bounds on the marginals of the distribution.

Theorem 3.4.6. *Suppose that $(\mathfrak{X}, w)$ is a pure n-dimensional weighted simplicial complex that is $(b_0, \ldots, b_{n-1})$-marginally bounded and has $(\zeta_0, \ldots, \zeta_{n-2})$-local spectral expansion. Then $\mathfrak{X}$ satisfies $(\alpha_0, \ldots, \alpha_{n-2})$-local entropy contraction where for $0 \le k \le n-2$,*

$$\alpha_k = \max\left\{1 - \frac{4\zeta_k}{b_k^2(n-k)^2}, \frac{1-\zeta_k}{4 + 2\log(\frac{1}{2b_k b_{k+1}})}\right\}. \tag{3.9}$$

With Theorem 3.4.6, we are able to prove Lemma 3.2.7.

Proof of Lemma 3.2.7. First notice that it suffices to prove the lemma for the case $s = n$, since when $s < n$ we can consider the s-dimensional simplicial complex $\mathfrak{X}' = \{\sigma \in \mathfrak{X} : |\sigma| \le s\}$ instead. When $s = n$, the lemma follows immediately from Theorems 3.4.4 and 3.4.6. □

The rest of this section aims to prove Theorem 3.4.6. We will show separately the two bounds in Eq. (3.9) on the rate of local entropy contraction, and we will refer to them as *the first bound* and *the second bound*. The first bound is more subtle and indicates that $\alpha_k = 1 - \Theta(\zeta_k)$ for the weighted simplicial complex $(\mathfrak{X}, \mu)$ corresponding to a marginally bounded distribution μ over $[q]^V$ (see Claim 2.4.2). The second bound is crude but may still be helpful when the first bound is bad.

Proof of the First Bound

We break the proof of it into the following lemmas, which are the essence of our approach.

First we generalize Lemma 10 of [47] as follows.

Lemma 3.4.7. *Suppose the local walk $P_\emptyset$ with stationary distribution π_1 satisfies $\lambda_2(P_\emptyset) \leq \zeta$. Then for every $f^{(2)} : \mathfrak{X}(2) \to \mathbb{R}_{\geq 0}$ and taking $f^{(1)} = P_1^{\uparrow} f^{(2)}$, we have*

$$\mathrm{Ent}_{\pi_2}(f^{(2)}) - 2\,\mathrm{Ent}_{\pi_1}(f^{(1)}) \geq -\zeta \cdot \frac{\mathrm{Var}_{\pi_1}(f^{(1)})}{\pi_1(f^{(1)})}.$$

The following lemma shows that for marginally bounded simplicial complexes, for any global function $f^{(n)}$, the induced local functions $f_\tau^{(1)}$ are "balanced" in the sense that the values of $f_\tau^{(1)}$ cannot be too large compared to the expectation of it.

Lemma 3.4.8. *If $\mathfrak{X}$ is a $(b_0, \ldots, b_{n-1})$-marginally bounded simplicial complex, then for every function $f^{(n)} : \mathfrak{X}(n) \to \mathbb{R}_{\geq 0}$, all $0 \leq k \leq n-1$, every $\tau \in \mathfrak{X}(k)$ such that $f^{(k)}(\tau) > 0$, and every $i \in \mathfrak{X}_\tau(1)$, we have*

$$\frac{f_\tau^{(1)}(i)}{\pi_{\tau,1}(f_\tau^{(1)})} \leq \frac{1}{b_k(n-k)}.$$

Finally, we show that for "balanced" functions the entropy and variance differ only by a constant factor after normalization.

Lemma 3.4.9. *Let π be a distribution over a finite set Ω, and let $f : \Omega \to \mathbb{R}_{\geq 0}$ such that $\pi(f) > 0$. If $f(x) \leq c \cdot \pi(f)$ for all $x \in \Omega$, then*

$$\frac{\mathrm{Var}_\pi(f)}{\pi(f)} \leq 4c^2\,\mathrm{Ent}_\pi(f).$$

Note that $c \geq 1$ and we always have the inequality $\mathrm{Ent}_\pi(f) \leq \frac{\mathrm{Var}_\pi(f)}{\pi(f)}$.

We then show how to use these three lemmas to prove the first bound in Eq. (3.9).

Proof of the first bound in Eq. (3.9). Let $0 \leq k \leq n-2$, $\tau \in \mathfrak{X}(k)$ and $f^{(n)} : \mathfrak{X}(n) \to \mathbb{R}_{\geq 0}$ be arbitrary. Then applying, Lemma 3.4.7 to the link $(\mathfrak{X}_\tau, w_\tau)$ and using $\lambda_2(P_\tau) \leq \zeta_k$, we have the inequality

$$\mathrm{Ent}_{\pi_{\tau,2}}(f_\tau^{(2)}) - 2\,\mathrm{Ent}_{\pi_{\tau,1}}(f_\tau^{(1)}) \geq -\zeta_k \cdot \frac{\mathrm{Var}_{\pi_{\tau,1}}(f_\tau^{(1)})}{\pi_{\tau,1}(f_\tau^{(1)})} \tag{3.10}$$

where the functions $f_\tau^{(2)} : \mathfrak{X}_\tau(2) \to \mathbb{R}_{\geq 0}$ and $f_\tau^{(1)} : \mathfrak{X}_\tau(1) \to \mathbb{R}_{\geq 0}$ are derived from $f^{(n)}$. By Lemma 3.4.8, we have $f_\tau^{(1)}(i) \leq \frac{1}{b_k(n-k)} \cdot \pi_{\tau,1}(f_\tau^{(1)})$ for all $i \in \mathfrak{X}_\tau(1)$. It follows by Lemma 3.4.9 that

$$\frac{\mathrm{Var}_{\pi_{\tau,1}}(f_\tau^{(1)})}{\pi_{\tau,1}(f_\tau^{(1)})} \leq \frac{4}{b_k^2(n-k)^2} \cdot \mathrm{Ent}_{\pi_{\tau,1}}(f_\tau^{(1)}). \tag{3.11}$$

Hence, it follows by combining Eqs. (3.10) and (3.11) that

$$\mathrm{Ent}_{\pi_{\tau,2}}(f_\tau^{(2)}) - 2\,\mathrm{Ent}_{\pi_{\tau,1}}(f_\tau^{(1)}) \geq -\frac{4\zeta_k}{b_k^2(n-k)^2} \cdot \mathrm{Ent}_{\pi_{\tau,1}}(f_\tau^{(1)}).$$

Rearranging, we obtain

$$\mathrm{Ent}_{\pi_{\tau,2}}(f_\tau^{(2)}) \geq \left(1 + \left(1 - \frac{4\zeta_k}{b_k^2(n-k)^2}\right)\right) \mathrm{Ent}_{\pi_{\tau,1}}(f_\tau^{(1)}).$$

As this holds for all $\tau \in \mathfrak{X}(k)$ and all $0 \leq k \leq n-2$, we obtain the first bound. □

It remains to prove Lemmas 3.4.7 to 3.4.9. We note that these lemmas are logically independent of each other.

Proof of Lemma 3.4.7. Observe that the desired inequality is scale invariant, and hence we may assume without loss that $\pi_2(f^{(2)}) = \pi_1(f^{(1)}) = 1$. We shall write ij to represent $\{i, j\} \in \mathfrak{X}(2)$ for simplicity. Let us rewrite $2\,\mathrm{Ent}_{\pi_1}(f^{(1)})$ in a form which is more conve-

nient to compare with $\mathrm{Ent}_{\pi_2}(f^{(2)})$. Observe that

$$\begin{aligned}
\mathrm{Ent}_{\pi_1}(f^{(1)}) &= \sum_{i \in \mathfrak{X}(1)} \pi_1(i) f^{(1)}(i) \log f^{(1)}(i) \\
&= \sum_{i \in \mathfrak{X}(1)} \pi_1(i) \left(\sum_{j \in \mathfrak{X}(1) : ij \in \mathfrak{X}(2)} \frac{\pi_2(ij)}{2\pi_1(i)} \cdot f^{(2)}(ij) \right) \log f^{(1)}(i) \\
&= \sum_{i \in \mathfrak{X}(1)} \sum_{j \in \mathfrak{X}(1) : ij \in \mathfrak{X}(2)} \frac{\pi_2(ij)}{2} \cdot f^{(2)}(ij) \log f^{(1)}(i) \\
&= \frac{1}{2} \sum_{ij \in \mathfrak{X}(2)} \pi_2(ij) \cdot f^{(2)}(ij) \log \left(f^{(1)}(i) f^{(1)}(j) \right).
\end{aligned}$$

By the inequality $a \log \frac{a}{b} \geq a - b$ for any $a \geq 0$ and $b > 0$, we can get

$$\begin{aligned}
\mathrm{Ent}_{\pi_2}(f^{(2)}) - 2\,\mathrm{Ent}_{\pi_1}(f^{(1)}) &= \sum_{ij \in \mathfrak{X}(2)} \pi_2(ij) \cdot f^{(2)}(ij) \left(\log f^{(2)}(ij) - \log \left(f^{(1)}(i) f^{(1)}(j) \right) \right) \\
&\geq \sum_{ij \in \mathfrak{X}(2)} \pi_2(ij) \cdot \left(f^{(2)}(ij) - f^{(1)}(i) f^{(2)}(j) \right) \\
&= \underbrace{\sum_{ij \in \mathfrak{X}(2)} \pi_2(ij) f^{(2)}(ij)}_{= \pi_2(f^{(2)}) = 1} - \sum_{ij \in \mathfrak{X}(2)} \pi_2(ij) f^{(1)}(i) f^{(1)}(j) \\
&= 1 - (f^{(1)})^\top W f^{(1)}
\end{aligned}$$

where $W \in \mathbb{R}_{\geq 0}^{\mathfrak{X}(1) \times \mathfrak{X}(1)}$ is the symmetric matrix with entries $W(i,j) = \frac{\pi_2(ij)}{2}$ whenever $ij \in \mathfrak{X}(2)$, and $W(i,j) = 0$ otherwise. Note that $W = \mathrm{diag}(\pi_1) P_\emptyset$ since recall that $P_\emptyset$ has entries $P_\emptyset(i,j) = \frac{\pi_2(ij)}{2\pi_1(i)}$ whenever $ij \in \mathfrak{X}(2)$, and $P_\emptyset(i,j) = 0$ otherwise. Now, we analyze W spectrally assuming knowledge of the spectral gap of $P_\emptyset$. Observe that since $\pi_1(f^{(1)}) = 1$, we have that

$$\begin{aligned}
1 - (f^{(1)})^\top W f^{(1)} &= \underbrace{\pi_1 \left[f^{(1)} \right]^2 - \pi_1 \left[(f^{(1)})^2 \right]}_{-\mathrm{Var}_{\pi_1}(f^{(1)})} + \underbrace{\pi_1 \left[(f^{(1)})^2 \right] - (f^{(1)})^\top \mathrm{diag}(\pi_1) P_\emptyset f^{(1)}}_{= \mathcal{E}_{P_\emptyset}(f^{(1)}, f^{(1)})} \\
&= \mathcal{E}_{P_\emptyset}(f^{(1)}, f^{(1)}) - \mathrm{Var}_{\pi_1}(f^{(1)})
\end{aligned}$$

$$\geq (1-\zeta)\cdot \mathrm{Var}_{\pi_1}(f^{(1)}) - \mathrm{Var}_{\pi_1}(f^{(1)})$$

(Spectral Gap or Poincaré Inequality)

$$= -\zeta \cdot \frac{\mathrm{Var}_{\pi_1}(f^{(1)})}{\pi_1(f^{(1)})} \qquad (\pi_1(f^{(1)}) = 1)$$

and we are done. □

Proof of Lemma 3.4.8. Without loss of generality, we may assume $\pi_n(f^{(n)}) = 1$, which also implies that $\pi_k(f^{(k)}) = 1$ for all k. It follows that if we define $\nu_k = f^{(k)}\pi_k$ for $0 \leq k \leq n$, then ν_k is a distribution on $\mathfrak{X}(k)$. For intuition, note that $\nu_k = \nu_n P_n^{\downarrow} \cdots P_{k+1}^{\downarrow}$, and hence we can regard these distributions as from the weighted simplicial complex $(\mathfrak{X}, \nu_n)$. We then have

$$\frac{f_\tau^{(1)}(i)}{\pi_{\tau,1}(f_\tau^{(1)})} = \frac{f^{(k+1)}(\tau \cup \{i\})}{f^{(k)}(\tau)} = \frac{\frac{\nu_{k+1}(\tau\cup\{i\})}{(k+1)\,\nu_k(\tau)}}{\frac{\pi_{k+1}(\tau\cup\{i\})}{(k+1)\,\pi_k(\tau)}} = \frac{\nu_{\tau,1}(i)}{\pi_{\tau,1}(i)}.$$

Trivially we have $\nu_{\tau,1}(i) \leq \frac{1}{n-k}$; see the proof of Claim 2.4.2 in Section 3.4.1 and the remark after it. Furthermore, by assumption, we have $\pi_{\tau,1}(i) \geq b_k$. The claim follows. □

Proof of Lemma 3.4.9. Without loss of generality, we may assume that $\pi(f) = 1$. Then, $f(x) \leq c$ for all $x \in \Omega$. Let $\nu = f\pi$, so f is the relative density of ν with respect to π. We need to show that

$$\mathrm{Var}_\pi(f) \leq 4c^2\,\mathrm{Ent}_\pi(f).$$

Fact 2.2.2 implies that for every function $g : \Omega \to \mathbb{R}$, we have

$$\nu(g) \leq D_{\mathrm{KL}}(\nu \parallel \pi) + \log \pi(e^g) = \mathrm{Ent}_\pi(f) + \log \pi(e^g).$$

Let $g = t(f-1)$ for some parameter $t > 0$ to be determined. Then

$$\nu(g) = t\,(\nu(f) - 1) = t\,(\pi(f^2) - 1) = t\,\mathrm{Var}_\pi(f).$$

Hence, we obtain that

$$\mathrm{Var}_\pi(f) \le \frac{1}{t}\,\mathrm{Ent}_\pi(f) + \frac{1}{t}\,\log\pi\left(e^{t(f-1)}\right).$$

This is known as the entropy inequality [100].

Notice that $c \ge 1$ always and $\mathrm{Ent}_\pi(f) \le \mathrm{Var}_\pi(f)$ when $\pi(f) = 1$ (see, e.g., [115]). Consider first the case that $1 \le c \le 2$. We shall pick

$$t = \sqrt{\frac{\mathrm{Ent}_\pi(f)}{\mathrm{Var}_\pi(f)}} \le 1.$$

Then $t(f-1) \le c - 1 \le 1$. Since $e^x \le 1 + x + x^2$ when $x \le 1$, we get

$$\begin{aligned}\log\pi\left(e^{t(f-1)}\right) &\le \log\pi\left(1 + t(f-1) + t^2(f-1)^2\right)\\ &= \log\left(1 + t^2\,\mathrm{Var}_\pi(f)\right)\\ &\le t^2\,\mathrm{Var}_\pi(f).\end{aligned}$$

It follows that

$$\mathrm{Var}_\pi(f) \le \frac{1}{t}\,\mathrm{Ent}_\pi(f) + t\,\mathrm{Var}_\pi(f).$$

For our choice of t, we obtain

$$\mathrm{Var}_\pi(f) \le 4\,\mathrm{Ent}_\pi(f) \le 4c^2\,\mathrm{Ent}_\pi(f).$$

Next, consider the case that $c > 2$. This time we pick

$$t = \sqrt{\frac{\mathrm{Ent}_\pi(f)}{\mathrm{Var}_\pi(f)}} \cdot \frac{2\ln c}{c} \le \frac{2\ln c}{c}.$$

Then $t(f-1) \le 2\ln c$. For all $x \le 2\ln c$, it holds that $e^x \le 1 + x + (\frac{c}{2\ln c})^2 x^2$. Hence, we

get

$$\begin{aligned}\log \pi \left(e^{t(f-1)}\right) &\leq \log \pi \left(1 + t(f-1) + \left(\frac{c}{2 \ln c}\right)^2 t^2 (f-1)^2\right) \\ &= \log \left(1 + t^2 \left(\frac{c}{2 \ln c}\right)^2 \operatorname{Var}_\pi(f)\right) \\ &\leq t^2 \left(\frac{c}{2 \ln c}\right)^2 \operatorname{Var}_\pi(f).\end{aligned}$$

We then deduce that

$$\begin{aligned}& \operatorname{Var}_\pi(f) \leq \frac{1}{t} \operatorname{Ent}_\pi(f) + t \left(\frac{c}{2 \ln c}\right)^2 \operatorname{Var}_\pi(f) \\ \implies \quad & \operatorname{Var}_\pi(f) \leq \left(\frac{c}{\ln c}\right)^2 \operatorname{Ent}_\pi(f) \leq 4c^2 \operatorname{Ent}_\pi(f).\end{aligned}$$

This establishes the lemma. □

Proof of the Second Bound

Here we prove the second bound in Eq. (3.9). We do this by reducing entropy contraction to bounding the standard log-Sobolev constant. Since the marginals are bounded, a comparison inequality between the standard log-Sobolev constant and spectral gap then finishes the proof.

We will show that for every (local) function $f^{(2)} : \mathfrak{X}(2) \to \mathbb{R}_{\geq 0}$, we have

$$\operatorname{Ent}_{\pi_2}(f^{(2)}) \geq (1 + \alpha_0) \operatorname{Ent}_{\pi_1}(f^{(1)}), \tag{3.12}$$

where

$$\alpha_0 = \frac{1 - \zeta_0}{4 + 2 \log(\frac{1}{2 b_0 b_1})}.$$

This establishes the second bound for the case $k = 0$ and $\tau = \emptyset$. For arbitrary $0 \leq k \leq n-2$ and arbitrary $\tau \in \mathfrak{X}(k)$, we can just consider the link $(\mathfrak{X}_\tau, w_\tau)$ at τ instead and achieve the results.

Recall that for a reversible Markov chain P with stationary distribution π on a finite state space Ω, the standard logarithmic Sobolev constant is defined as

$$\rho(P) = \inf \left\{ \frac{\mathcal{E}_P(\sqrt{f}, \sqrt{f})}{\mathrm{Ent}_\pi(f)} \;\middle|\; f : \Omega \to \mathbb{R}_{\geq 0},\ \mathrm{Ent}_\pi(f) \neq 0 \right\}.$$

We prove Eq. (3.12) by the following two lemmas. First we relate α_0 with the standard log-Sobolev constant of the order-$(2,1)$ down-up walk $P^\vee_{2,1}$.

Lemma 3.4.10. *For every (local) function $f^{(2)} : \mathfrak{X}(2) \to \mathbb{R}_{\geq 0}$, we have*

$$\mathrm{Ent}_{\pi_1}(f^{(1)}) \leq \left(1 - \rho(P^\vee_{2,1})\right)\ \mathrm{Ent}_{\pi_2}(f^{(2)}).$$

The following lemma is equation (3.9) from [48] which compares the standard log-Sobolev constant and the spectral gap.

Lemma 3.4.11 ([48])**.** *For every reversible Markov chain P with stationary distribution π on a finite state space Ω, we have the inequality*

$$\rho(P) \geq \frac{\lambda(P)}{2 + \log(1/\pi^*)}$$

where $\pi^* = \min_{x \in \Omega} \pi(x)$.

We are now ready to prove the second bound of Eq. (3.9).

Proof of the second bound in Eq. (3.9). From Lemma 3.4.10 we deduce that for every (local) function $f^{(2)} : \mathfrak{X}(2) \to \mathbb{R}_{\geq 0}$,

$$\mathrm{Ent}_{\pi_2}(f^{(2)}) \geq \frac{1}{1 - \rho(P^\vee_{2,1})}\, \mathrm{Ent}_{\pi_1}(f^{(1)}) \geq \left(1 + \rho(P^\vee_{2,1})\right) \mathrm{Ent}_{\pi_1}(f^{(1)}). \tag{3.13}$$

Meanwhile, Lemma 3.4.11 gives

$$\rho(P^\vee_{2,1}) \geq \frac{\lambda(P^\vee_{2,1})}{2 + \log(1/\pi^*_2)} \geq \frac{1 - \zeta_0}{4 + 2\log(1/\pi^*_2)}, \tag{3.14}$$

where the last inequality follows from

$$\lambda(P_{2,1}^{\vee}) = \lambda(P_{1,2}^{\wedge}) = 1 - \lambda_2(P_{1,2}^{\wedge}) = \frac{1}{2}\left(1 - \lambda_2(P_{\emptyset})\right) \geq \frac{1-\zeta_0}{2}.$$

Also, the marginal boundedness of $(\mathfrak{X}, w)$ implies that for every $\{i,j\} \in \mathfrak{X}(2)$,

$$\pi_2(\{i,j\}) = 2 \cdot \pi_1(i) \cdot \pi_{\{i\},1}(j) \geq 2b_0b_1. \tag{3.15}$$

Combining Eqs. (3.13) to (3.15), we get Eq. (3.12). The bounds on local entropy contraction rate for any $0 \leq k \leq n-2$ and $\tau \in \mathfrak{X}(k)$ follows by considering the link $(\mathfrak{X}_\tau, w_\tau)$. □

Let us now prove Lemma 3.4.10. We follow the proof of [104, Proposition 6], using the following technical lemma.

Lemma 3.4.12 ([104, Lemma 5])**.** *For real numbers $t \geq 0$ and $s \geq -t$, we have the inequality*

$$(t+s)\log(t+s) \geq t\log t + s(1+\log t) + (\sqrt{t+s} - \sqrt{t})^2.$$

Proof of Lemma 3.4.10. For convenience, we write $P^{\uparrow} = P_1^{\uparrow}$, $P^{\downarrow} = P_2^{\downarrow}$, and $P^{\vee} = P_{2,1}^{\vee} = P_2^{\downarrow}P_1^{\uparrow}$ in the proof. Since $\pi_1(f^{(1)}) = \pi_2(f^{(2)})$ and the inequality is scale invariant, we may assume these expectations are 1. Towards proving the desired contraction inequality, we first prove the following intermediate inequality: for all $i \in \mathfrak{X}(1)$,

$$\left(P^{\uparrow} f^{(2)} \log f^{(2)}\right)(i) \geq \left(P^{\uparrow} f^{(2)}\right)(i) \cdot \log\left(P^{\uparrow} f^{(2)}\right)(i) + \left(P^{\uparrow} f^{(2)}\right)(i) - \left(P^{\uparrow}\sqrt{f^{(2)}}\right)(i)^2. \tag{3.16}$$

Let us first see how to use this inequality to prove the desired contraction inequality. Observe from Eq. (3.16) that

$$\mathrm{Ent}_{\pi_1}(f^{(1)}) = \mathrm{Ent}_{\pi_1}\left(P^{\uparrow} f^{(2)}\right)$$

$$
\begin{aligned}
&= \sum_{i\in\mathfrak{X}(1)} \pi_1(i)\cdot\left(P^{\uparrow}f^{(2)}\right)(i)\cdot\log\left(P^{\uparrow}f^{(2)}\right)(i)\\
&\leq \sum_{i\in\mathfrak{X}(1)} \pi_1(i)\left(P_1^{\uparrow}f^{(2)}\log f^{(2)}\right)(i) - \sum_{i\in\mathfrak{X}(1)}\pi_1(i)\left(\left(P^{\uparrow}f^{(2)}\right)(i) - \left(P^{\uparrow}\sqrt{f^{(2)}}\right)(i)^2\right)\\
&= \left\langle \mathbf{1}, P^{\uparrow}f^{(2)}\log f^{(2)}\right\rangle_{\pi_1} - \left\langle \mathbf{1}, P^{\uparrow}f^{(2)}\right\rangle_{\pi_1} + \left\langle P^{\uparrow}\sqrt{f^{(2)}}, P^{\uparrow}\sqrt{f^{(2)}}\right\rangle_{\pi_1}\\
&= \left\langle \mathbf{1}, f^{(2)}\log f^{(2)}\right\rangle_{\pi_2} - \left\langle \mathbf{1}, f^{(2)}\right\rangle_{\pi_2} + \left\langle \sqrt{f^{(2)}}, P^{\vee}\sqrt{f^{(2)}}\right\rangle_{\pi_2}\\
&= \mathrm{Ent}_{\pi_2}(f^{(2)}) - \mathcal{E}_{P^{\vee}}\left(\sqrt{f^{(2)}}, \sqrt{f^{(2)}}\right)\\
&\leq (1-\rho(P^{\vee}))\ \mathrm{Ent}_{\pi_2}(f^{(2)}).
\end{aligned}
$$

All that remains is to prove Eq. (3.16). For every $i\in\mathfrak{X}(1)$, taking $t = \left(P^{\uparrow}f^{(2)}\right)(i)$, we deduce from Lemma 3.4.12 that

$$
\begin{aligned}
&\left(P^{\uparrow}f^{(2)}\log f^{(2)}\right)(i)\\
&= \sum_{\sigma\in\mathfrak{X}(2)} P^{\uparrow}(i,\sigma)f^{(2)}(\sigma)\log f^{(2)}(\sigma)\\
&= \sum_{\sigma\in\mathfrak{X}(2)} P^{\uparrow}(i,\sigma)\left(t + f^{(2)}(\sigma) - t\right)\log\left(t + f^{(2)}(\sigma) - t\right)\\
&\geq \sum_{\sigma\in\mathfrak{X}(2)} P^{\uparrow}(i,\sigma)\left(t\log t + (f^{(2)}(\sigma)-t)(1+\log t) + \left(\sqrt{f^{(2)}(\sigma)} - \sqrt{t}\right)^2\right)\\
&= \left(P^{\uparrow}f^{(2)}\right)(i)\log\left(P^{\uparrow}f^{(2)}\right)(i) + \sum_{\sigma\in\mathfrak{X}(2)} P^{\uparrow}(i,\sigma)\underbrace{\left(\sqrt{f^{(2)}(\sigma)} - \sqrt{\left(P^{\uparrow}f^{(2)}\right)(i)}\right)^2}_{\text{Expand}}\\
&= \left(P^{\uparrow}f^{(2)}\right)(i)\log\left(P^{\uparrow}f^{(2)}\right)(i) + \underbrace{2\left(P^{\uparrow}f^{(2)}\right)(i) - 2\sqrt{\left(P^{\uparrow}f^{(2)}\right)(i)}\cdot\left(P^{\uparrow}\sqrt{f^{(2)}}\right)(i)}_{(*)}.
\end{aligned}
$$

Let us now lower bound $(*)$. We observe that

$$
(*) - \left(\left(P^{\uparrow}f^{(2)}\right)(i) - \left(P^{\uparrow}\sqrt{f^{(2)}}\right)(i)^2\right) = \left(\sqrt{\left(P^{\uparrow}f^{(2)}\right)(i)} - \left(P^{\uparrow}\sqrt{f^{(2)}}\right)(i)\right)^2 \geq 0.
$$

Eq. (3.16) then follows and we are done. □

CHAPTER 4

OTHER IMPLICATIONS OF SPECTRAL INDEPENDENCE

In this chapter, we show that spectral independence implies optimal mixing time bounds for, in addition to the single-site Glauber dynamics, arbitrary heat-bath block dynamics and (for ferromagnetic Ising/Potts models) the Swendsen-Wang dynamics. This chapter is based on [23].

4.1 Introduction and Our Results

Our results apply broadly to the general class of heat-bath block dynamics. Let $\mathcal{B} = \{B_1, \ldots, B_\ell\}$ be any collection of sets (or blocks) such that $V = \cup_i B_i$ and let $\alpha = (\alpha_B)_{B \in \mathcal{B}}$ be a probability distribution on $\mathcal{B}$. A step of the heat-bath block dynamics operates by choosing a block B with probability α_B and updating the configuration in B with a sample from the Gibbs distribution conditional on the configuration on $V \setminus B$. Note that the Glauber dynamics corresponds to setting the blocks to individual vertices with uniform weights, and for a bipartite graph the even-odd chain (also known as the alternating scan dynamics) corresponds to uniform weighting for two blocks corresponding to the two parts. By extending the weight to $\alpha_B = 0$ if $B \notin \mathcal{B}$ we think of α as a distribution over all subsets of V and speak of the α-weighted heat-bath block dynamics.

Given α, define the minimum "coverage probability" of a vertex by

$$\delta = \delta(\alpha) = \min_{u \in V} \sum_{B : B \ni u} \alpha_B. \tag{4.1}$$

We say that the block dynamics have *optimal mixing* when there exists a constant C such that for all weights α the mixing time of the α-weighted heat-bath block dynamics is at most $C\delta(\alpha)^{-1} \log n$. Similarly, we say that the block dynamics have *optimal entropy decay*

if the modified log-Sobolev constant of the α-weighted heat-bath block dynamics is at least $\delta(\alpha)/C$. Note that the constant C may depend on the parameters defining the spin system and on the maximum degree Δ, but it does not depend on n and it is independent of the choice of weights α. In this generality, these bounds are optimal up to the value of the constant C. Indeed, for the Glauber dynamics we have $\delta(\alpha) = 1/n$ and the mixing time matches the $\Omega(n \log n)$ lower bound established by Hayes and Sinclair [71] for bounded-degree graphs. Moreover, by restricting to test functions of a single spin it is not hard to check that the spectral gap of the α-weighted block dynamics is always at most $\delta(\alpha)$, and therefore the lower bound $\delta(\alpha)/C$ on the modified log-Sobolev constant of the block dynamics is optimal up to the multiplicative constant $1/C$; see e.g. [25] for standard relations between spectral gap and modified log-Sobolev constant.

Our first result is a substantial extension of Theorem 1.2.1 to the block dynamics with arbitrary weighted blocks.

Theorem 4.1.1. *For an arbitrary spin system on a graph of maximum degree Δ, if the system is η-spectrally independent and b-marginally bounded, then general block factorization of entropy (see Definition 2.3.3) holds with constant $C = C(b, \eta, \Delta)$. Moreover, all heat-bath block dynamics have optimal mixing and optimal entropy decay. The constant C satisfies $C = \left(\frac{2}{b}\right)^{O\left(\Delta\left(\frac{\eta}{b}+1\right)\right)}$.*

Recall, for the Glauber dynamics $\delta(\alpha) = 1/n$ and hence we recover Theorem 1.2.1 as a special case of the above result, though with a worse constant. As another example, for a bipartite graph Theorem 4.1.1 implies $O(\log n)$ mixing time of the even-odd dynamics which updates the two parts of the bipartite graph alternatively.

Recent work of Blanca et al. [24] utilizes block factorization of entropy into the even and odd sublattices of $\mathbb{Z}^d$ to obtain tight mixing time bounds for the Swendsen-Wang dynamics on boxes of $\mathbb{Z}^d$ in the high-temperature region. Following the approach presented in [24], here we prove optimal mixing time of the Swendsen-Wang dynamics when spectral independence holds on arbitrary bounded-degree graphs. This can be formalized in the

following statement.

Theorem 4.1.2. *For the ferromagnetic Ising and Potts models on a graph of maximum degree Δ, if the system is η-spectrally independent and b-marginally bounded, then there exists a constant $C = C(b, \eta, \Delta)$ such that the mixing time of the Swendsen-Wang dynamics is at most $C \log n$ and the modified log-Sobolev constant is at least C^{-1}. The constant C satisfies $C = \left(\frac{2}{b}\right)^{O\left(\Delta(1+\frac{\eta}{b})\right)}$.*

This chapter is organized as follows. In Section 4.2 we discuss our proof approach and techniques. We summarize a few key properties of entropy in Section 4.3. In Section 4.4 we deduce optimal mixing of the block dynamics with arbitrary weighted blocks from spectral independence. In Section 4.5 we consider the Swendsen-Wang dynamics for ferromagnetic Ising and Potts model. Finally, in Section 4.6 we present a direct proof of Theorem 3.2.4 without utilizing simplicial complexes.

4.2 Proof Approach and Discussions

The key step in the proof of Theorem 1.2.1 is the implication

$$\text{Spectral Independence} \implies \text{Approximate Tensorization of Entropy.} \tag{4.2}$$

Recall that approximate tensorization of entropy says that there exists a constant $C \geq 1$, such that for any function $f : \Omega \to \mathbb{R}_+$,

$$\mathrm{Ent}(f) \leq C \sum_{u \in V} \mu[\mathrm{Ent}_u(f)], \tag{4.3}$$

where $\mu[f] = \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma)$ and $\mathrm{Ent}(f) = \mu[f \log(f/\mu[f])]$ denote the mean and entropy of f with respect to the measure μ. In particular, $\mathrm{Ent}(f)$ is the relative entropy of the probability measure $f\mu/\mu[f]$ with respect to μ, while $\mu[\mathrm{Ent}_u f] = \mu[f \log(f/\mu_u[f])]$ is the expected value according to μ of the conditional entropy $\tau \mapsto \mathrm{Ent}(f|\tau)$ for τ a

spin configuration on $V \setminus \{u\}$. To make some intuitive sense of approximate tensorization, notice that if μ is a product distribution over V then Eq. (4.3) holds with $C = 1$. In general, approximate tensorization is easily seen to imply the desired bounds on the modified log-Sobolev constant and the mixing time of the Glauber dynamics; see e.g. [36]. In the setting of spin systems on the lattice $\mathbb{Z}^d$, approximate tensorization estimates are known to hold under the so-called strong spatial mixing condition; this follows from the logarithmic Sobolev inequalities established in [123, 99, 41].

We present an alternative proof of some of the key steps for the implication Eq. (4.2) in Section 4.6; see Theorem 4.6.1. The analogous result Theorem 3.2.4 in Chapter 3 is proved in the more general framework of simplicial complexes and generalizes the result of [47] for homogeneous strongly log-concave distributions; see also [72] for related results. Our proof is completely framed in the setting of spin systems and is devoid of any work on simplicial complexes. This new approach may be conceptually simpler to some readers, and it enables us to present a self-contained proof of our main results. As a byproduct we also obtain an incremental improvement in the resulting mixing time bound improving the exponent in the constant C from $O(1 + \eta/b^2)$ (see Theorem 1.9 in [45]) to $O(1 + \eta/b)$ as stated in Theorem 1.2.1.

One of our results in this chapter is the following substantial extension of Eq. (4.2):

$$\text{Spectral Independence} \implies \text{General Block Factorization of Entropy.} \tag{4.4}$$

Caputo and Parisi [37] introduced the notion of *general block factorization* of entropy which generalizes approximate tensorization, and is useful for analyzing more general classes of Markov chains. Let $\alpha = (\alpha_B)_{B \subseteq V}$ be an arbitrary probability distribution over subsets of V, and set $\delta(\alpha) = \min_{u \in V} \sum_{B : B \ni u} \alpha_B$ as in Eq. (4.1). General block factoriza-

tion of entropy holds with constant C if for all weights α, for all $f : \Omega \to \mathbb{R}_+$:

$$\delta(\alpha)\mathrm{Ent} f \leq C \sum_{B \subseteq V} \alpha_B \, \mu[\mathrm{Ent}_B f], \tag{4.5}$$

where $\mu[\mathrm{Ent}_B f] = \mu[f \log(f/\mu_B f)]$ is the expected value of the conditional entropy $\tau \mapsto \mathrm{Ent}(f|\tau)$ for τ a spin configuration on $V \setminus B$. Entropy tensorization Eq. (4.3) is the special case when $\alpha_B = 1/n$ for every block of size 1 and $\alpha_B = 0$ for larger blocks. The choice of the constant $\delta(\alpha)$ in this inequality is motivated by the fact that when μ is a product measure then Eq. (4.5) holds with $C = 1$, in which case it is known as the Shearer inequality; see [36]. The block factorization of entropy is a statement concerning the equilibrium distribution μ which has deep consequences for several natural sampling algorithms. In particular, it implies optimal mixing and optimal entropy decay for arbitrary block dynamics.

When the spin system satisfies Eq. (4.5) with α the uniform distribution over all subsets of a given size ℓ we refer to this as *ℓ-uniform block factorization of entropy* or ℓ-UBF for short. In Chapter 3, an important step in the proof of Theorem 1.2.1 is establishing ℓ-UBF with $\ell \sim \theta n$ for some $\theta \in (0, 1)$. To prove Theorem 4.1.1 for arbitrary blocks we establish that ℓ-UBF implies general block factorization of entropy, see Theorem 4.4.2 for a detailed statement.

We turn to a further interesting consequence of spectral independence:

$$\text{Spectral Independence} \implies \text{Approximate Subadditivity of Entropy}. \tag{4.6}$$

We say that the approximate subadditivity of entropy holds with constant C if

$$\sum_{x \in V} \mathrm{Ent}(f_x) \leq C\mathrm{Ent}(f), \tag{4.7}$$

where, for any nonnegative function f, the functions f_x are defined by $f_x(a) = \mu(f \,|\, \sigma_x =$

a). Notice that when $\mu(f) = 1$ then $\nu = f\mu$ is a probability measure and, if μ_x denotes the marginal of μ on x, then $f_x\mu_x$ gives the marginal of ν on x. The inequality Eq. (4.7) is known to be equivalent to a Brascamp-Lieb type inequality for the measure μ [40, 39]. In particular, it implies that for any collection of functions $\varphi_u : [q] \mapsto \mathbb{R}$, $u \in V$, one has

$$\mu\left(\prod_{u\in V}\varphi_u(\sigma_u)\right) \leq \prod_{u\in V}\mu\left(|\varphi_u(\sigma_u)|^C\right)^{1/C}, \tag{4.8}$$

where C is the same constant as above. For a general discussion of subadditivity of entropy, Brascamp-Lieb type inequalities, and their applications, see for instance [8] and the references therein. In Theorem 4.6.1 below we shall see that for an arbitrary spin system on a graph of maximum degree Δ, if the system is η-spectrally independent and b-marginally bounded, then Eq. (4.7) holds with $C = O(1 + \eta/b)$. The question of the validity of such inequalities in the context of high temperature spin systems was raised in [36] but as far as we know there are no prior results in this direction.

4.3 Basic Properties of Entropy

To compute the relative entropy with respect to a pinned measure μ^τ_Λ it is convenient to use the notation

$$\mathrm{Ent}_\Lambda(f) = \mu_\Lambda\left[f\log\left(f/\mu_\Lambda[f]\right)\right], \tag{4.9}$$

with the understanding that if we evaluate the left hand side at a given pinning τ on $\Lambda^c = V \setminus \Lambda$ we then evaluate the expectations in the right hand side with respect to μ^τ_Λ. To emphasize the dependence on the pinning we sometimes write $\mathrm{Ent}^\tau_\Lambda(f)$. The expectation $\mu[\mathrm{Ent}_\Lambda f]$ is obtained by averaging with respect to μ over the pinning τ on Λ^c, and satisfies

$$\mu[\mathrm{Ent}_\Lambda(f)] = \sum_{\tau\in\Omega_{\Lambda^c}}\mu(\sigma_{\Lambda^c} = \tau)\,\mathrm{Ent}^\tau_\Lambda(f) = \mu\left[f\log\left(f/\mu_\Lambda[f]\right)\right]. \tag{4.10}$$

The following lemma summarizes a standard decomposition of the relative entropy; see e.g. [37, Lemma 3.1] for a proof.

Lemma 4.3.1. *For any $\Lambda \subseteq V$, for any $f : \Omega \to \mathbb{R}_+$:*

$$\mathrm{Ent}(f) = \mu\left[\mathrm{Ent}_\Lambda(f)\right] + \mathrm{Ent}\left(\mu_\Lambda[f]\right). \tag{4.11}$$

More generally, for any $\Lambda_0 \subseteq \Lambda_1 \subseteq \cdots \subseteq \Lambda_w \subseteq V$, for any $f : \Omega \to \mathbb{R}_+$:

$$\sum_{i=1}^{w} \mu\left[\mathrm{Ent}_{\Lambda_i}(\mu_{\Lambda_{i-1}}[f])\right] = \mu\left[\mathrm{Ent}_{\Lambda_w}(\mu_{\Lambda_0}[f])\right]. \tag{4.12}$$

The following monotonicity property of the entropy functional is an immediate consequence of the previous lemma.

Lemma 4.3.2. *For all $A \subseteq B \subseteq V$,*

$$\mu[\mathrm{Ent}_A(f)] \leq \mu[\mathrm{Ent}_B(f)]. \tag{4.13}$$

Next, we recall the definition of general block factorization of entropy from Definition 2.3.3. The spin system is said to satisfy the *general block factorization of entropy* with constant C if for all $f \geq 0$, for all probability distribution α over subsets of V,

$$\delta(\alpha)\mathrm{Ent} f \leq C \sum_{B \subseteq V} \alpha_B \, \mu[\mathrm{Ent}_B f], \tag{4.14}$$

where $\delta(\alpha) = \min_{x \in V} \sum_{B:\, B \ni x} \alpha_B$.

We will often consider independent sets Λ of V, that is sets of vertices whose induced subgraph in G has no edge; in those cases, μ_Λ is a product measure $\mu_\Lambda = \otimes_{x \in \Lambda} \mu_x$ and the following lemma will be useful.

Lemma 4.3.3. *Fix $\Lambda \subseteq V$ and suppose that μ_Λ is a product measure on $\mu_\Lambda = \otimes_{x \in \Lambda} \mu_x$.*

Then, for any distribution α over the subsets of Λ, and any $f : \Omega \to \mathbb{R}_+$:

$$\delta(\alpha)\,\mathrm{Ent}_\Lambda(f) \leq \sum_{B\subseteq\Lambda} \alpha_B\,\mu_\Lambda[\mathrm{Ent}_B(f)]\,, \tag{4.15}$$

that is μ_Λ satisfies the general block factorization of entropy with constant $C = 1$.

The above statement is a consequence of the weighted Shearer inequality for the Shannon entropy; see Lemma 4.2 in [37]. The following properties will also be used.

Lemma 4.3.4. *Let $\Lambda = A \cup B$ and assume that μ_Λ is a product $\mu_\Lambda = \mu_A \otimes \mu_B$. Then, for all $f \geq 0$:*

$$\mathrm{Ent}_\Lambda(\mu_B(f)) = \mu_\Lambda[\mathrm{Ent}_A(\mu_B(f))], \tag{4.16}$$

and for all $U \subseteq B$,

$$\mu_\Lambda[\mathrm{Ent}_A(\mu_B(f))] \leq \mu_\Lambda[\mathrm{Ent}_A(\mu_U(f))]. \tag{4.17}$$

Proof. From the decomposition in Lemma 4.3.1 it follows that

$$\mathrm{Ent}_\Lambda(\mu_B(f)) - \mu_\Lambda[\mathrm{Ent}_A(\mu_B(f))] = \mathrm{Ent}_\Lambda(\mu_A\mu_B(f)) = \mathrm{Ent}_\Lambda(\mu_\Lambda(f)) = 0.$$

This proves (Equation 4.16). To prove (Equation 4.17) notice that by definition

$$\mu_\Lambda\left[\mathrm{Ent}_A(\mu_B(f))\right] = \mu_\Lambda\left[\mu_B(f)\log\left(\frac{\mu_B(f)}{\mu_A\mu_B(f))}\right)\right].$$

For any $U \subseteq B$, $\mu_B(f) = \mu_B\mu_U(f)$ and the product structure $\mu_\Lambda = \mu_A \otimes \mu_B$ implies the commutation relation $\mu_A\mu_B\mu_U = \mu_B\mu_A\mu_U$. Therefore,

$$\mu_\Lambda\left[\mathrm{Ent}_A(\mu_B(f))\right] = \mu_\Lambda\left[\mu_B\mu_U(f)\log\left(\frac{\mu_B\mu_U(f)}{\mu_B\mu_A\mu_U(f)}\right)\right]$$

$$= \mu_\Lambda \left[\mu_U(f) \log \left(\frac{\mu_B \mu_U(f)}{\mu_B \mu_A \mu_U(f))} \right) \right]$$
$$= \mu_\Lambda \left[\mu_A \left[\mu_U(f) \log \left(\frac{\mu_B \mu_U(f)}{\mu_A \mu_B \mu_U(f)} \right) \right] \right].$$

It remains to observe that

$$\mu_A \left[\mu_U(f) \log \left(\frac{\mu_B \mu_U(f)}{\mu_A \mu_B \mu_U(f)} \right) \right] \leq \mathrm{Ent}_A(\mu_U(f)).$$

The latter estimate follows from the well known variational principle

$$\mathrm{Ent}_A(g) = \sup \left\{ \mu_A(gh)\,,\ \mu_A(e^h) \leq 1 \right\} \tag{4.18}$$

valid for any A and any function $g \geq 0$; see, e.g. [85, Proposition 2.2]. □

4.4 Optimal Mixing of Arbitrary Block Dynamics

A key step in the proof of Theorem 4.1.1 is the proof that uniform block factorization (UBF) implies general block factorization (GBF).

We begin with the formal definition of UBF. For a positive integer $\ell \leq n$, let $\binom{V}{\ell}$ denote the collection of all subsets of V of size ℓ.

Definition 4.4.1 (Uniform Block Factorization (UBF))**.** We say that the spins system μ satisfies the *ℓ-uniform block factorization (ℓ-UBF) of entropy* with constant C_{UBF} if for all $f : \Omega \to \mathbb{R}_+$

$$\frac{\ell}{n} \mathrm{Ent}(f) \leq C_{\text{UBF}} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{V}{\ell}} \mu[\mathrm{Ent}_S(f)]. \tag{4.19}$$

In this section, we establish the following theorem.

Theorem 4.4.2. *For an arbitrary b-marginally bounded spin system on a graph of maximum degree Δ, if $\lceil \theta n \rceil$-UBF holds with constant C_{UBF} and $0 < \theta \leq \frac{b^{2(\Delta+1)}}{4e\Delta^2}$, then GBF holds with constant $C_{\text{GBF}} = C_{\text{UBF}} \times O\left((\theta\, b^2)^{-1} \log(1/b) \Delta^3\right)$.*

In this section we prove Theorem 4.4.2 by establishing general block factorization of entropy from uniform block factorization.

As it will become apparent in our proof of this theorem, there is a trade-off between the upper bound for θ and the value we can deduce for C_{GBF}; in particular, we could allow for UBF to hold for larger $\theta < 1$ (i.e., with a better dependence on Δ) at the expense of an additional factor depending on Δ in C_{GBF}.

We turn to the proof of Theorem 4.4.2. Recall that a graph G of maximum degree Δ is k-partite, with $k \leq \Delta + 1$. Let $\{V_1, ..., V_k\}$ denote the independent sets $V_i \subseteq V$ corresponding to a k-partition of G. A key step in the proof of Theorem 4.4.2 is to establish the following factorization statement.

Lemma 4.4.3. *Suppose that for an arbitrary b-marginally bounded spin system on a graph of maximum degree Δ, $\lceil \theta n \rceil$-UBF holds with constant C_{UBF} and $\theta \leq \frac{b^{2(\Delta+1)}}{4e\Delta^2}$. Then,*

$$\text{Ent}(f) \leq KC_{\text{UBF}} \sum_{i=1}^{k} \mu[\text{Ent}_{V_i}(f)], \tag{4.20}$$

where the constant K satisfies $K = O(\Delta^2(\theta\, b^2)^{-1} \log(1/b))$.

We refer to Eq. (4.20) as a k-partite factorization of entropy with constant KC_{UBF}. Once we have Lemma 4.4.3, Theorem 4.4.2 is implied by the following lemma.

Lemma 4.4.4. *Suppose that for an arbitrary spin system on a graph of maximum degree Δ, k-partite factorization of entropy holds with constant C. Then, GBF holds with constant Ck.*

We provide next the proofs of Lemmas 4.4.3 and 4.4.4.

Proof of Lemma 4.4.4. Let $\alpha = (\alpha_B)_{B \subseteq V}$ be a probability distribution over the subsets of V. Observe that for all $j = 1, ..., k$ and all $\tau \in \Omega(V \setminus V_j)$, $\mu^{\tau}_{V_j}$ is a product measure on $\Omega^{\tau}_{V_j}$. Therefore, we can apply Lemma 4.3.3 with $\Lambda = V_j$ and $\hat{\alpha} = (\hat{\alpha}_U)_{U \subseteq V_j}$, where

$\hat{\alpha}_U = \omega^{-1} \sum_{B \subseteq V} \alpha_B \mathbf{1}(V_j \cap B = U)$ and $\omega = \sum_{B \subseteq V} \alpha_B \mathbf{1}(V_j \cap B \neq \emptyset)$. We get

$$\delta(\hat{\alpha}) \operatorname{Ent}^{\tau}_{V_j}(f) \leq \sum_{U \subseteq V_j} \hat{\alpha_U} \, \mu^{\tau}_{V_j}[\operatorname{Ent}_U(f)] = \omega^{-1} \sum_{B \subseteq V} \alpha_B \, \mu^{\tau}_{V_j}[\operatorname{Ent}_{V_j \cap B}(f)]. \tag{4.21}$$

Observe that

$$\omega \delta(\hat{\alpha}) = \min_{x \in V_j} \sum\nolimits_{U \subseteq V_j : U \ni x} \hat{\alpha}_U = \min_{x \in V_j} \sum\nolimits_{B \subseteq V : B \ni x} \alpha_B \geq \delta(\alpha),$$

and from Eq. (4.13) we have $\mu[\operatorname{Ent}_{V_j \cap B}(f)] \leq \mu[\operatorname{Ent}_B(f)]$. Hence, taking expectation in Eq. (4.21) with respect to μ we obtain

$$\delta(\alpha) \, \mu[\operatorname{Ent}_{V_j}(f)] \leq \sum_{B \subseteq V} \alpha_B \, \mu[\operatorname{Ent}_B(f)].$$

Summing over j we have, for all $f : \Omega \to \mathbb{R}_+$,

$$\delta(\alpha) \sum_{j=1}^{k} \mu[\operatorname{Ent}_{V_j}(f)] \leq \sum_{j=1}^{k} \sum_{B \subseteq V} \alpha_B \, \mu[\operatorname{Ent}_B(f)],$$

and since by assumption k-partite factorization of entropy holds with constant C, we have

$$\delta(\alpha) \operatorname{Ent}(f) \leq C \sum_{j=1}^{k} \sum_{B \subseteq V} \alpha_B \, \mu[\operatorname{Ent}_B(f)] \leq C \, k \sum_{B \subseteq V} \alpha_B \, \mu[\operatorname{Ent}_B(f)].$$

Hence, GBF holds with constant Ck. □

The main idea behind the proof of Lemma 4.4.3 can be roughly explained as follows. The ℓ-UBF assumption with $\ell \sim \theta n$ is the factorization statement Eq. (4.19). If the set S in Eq. (4.19) were an independent set, then suitable applications of Lemma 4.3.1 and Lemma 4.3.4 would yield the desired conclusion. Moreover, the same conclusion would continue to hold if S were made of bounded connected components. The delicate part of the argument consists in exploiting the fact that if θ is sufficiently small then one can

effectively reduce the problem to case of bounded connected components.

Proof of Lemma 4.4.3. Since $\lceil \theta n \rceil$-UBF holds by assumption, setting $C = C_{\text{UBF}}$ one has

$$\text{Ent}(f) \leq \frac{C}{\theta} \mathbb{E} \left[\mu \left[\text{Ent}_S(f) \right] \right], \tag{4.22}$$

where S is a random set with uniform distribution over all subsets of V of cardinality $\lceil \theta n \rceil$, and $\mathbb{E}$ denotes the corresponding expectation.

Let $S_1, S_2, \ldots$ denote the connected components of S in G (taken in some arbitrary order) and for $i > 1$ let $S_{<i} = \cup_{j=1}^{i-1} S_j$. Then $\mu_{S_{<i+1}}$ has the product structure $\mu_{S_{<i+1}} = \otimes_{j=1}^{i} \mu_{S_j}$. By Lemmas 4.3.1 and 4.3.4, one has the decomposition

$$\mu \left[\text{Ent}_S(f) \right] = \sum_{i \geq 1} \mu \left[\text{Ent}_{S_{<i+1}}(\mu_{S_{<i}}(f)) \right] = \sum_{i \geq 1} \mu [\text{Ent}_{S_i}(\mu_{S_{<i}}(f))], \tag{4.23}$$

where we have used Eq. (4.16) with $A = S_i$ and $B = S_{<i}$. For $\tau \in \Omega(V \setminus S_i)$, let $\Gamma(S_i, \tau)$ be the optimal constant so that

$$\text{Ent}_{S_i}^{\tau}(\mu_{S_{<i}}(f)) \leq \Gamma(S_i, \tau) \sum_{j=1}^{k} \mu_{S_i}^{\tau} \left[\text{Ent}_{V_j \cap S_i}(\mu_{S_{<i}}(f)) \right].$$

Let $\Gamma(S_i) = \max_{\tau \in \Omega(V \setminus S_i)} \Gamma(S_i, \tau)$. Then,

$$\mu \left[\text{Ent}_S(f) \right] \leq \sum_{i \geq 1} \Gamma(S_i) \sum_{j=1}^{k} \mu \left[\text{Ent}_{V_j \cap S_i}(\mu_{S_{<i}}(f)) \right].$$

We observe next that for all $j = 1, ..., k$ one has

$$\mu \left[\text{Ent}_{V_j \cap S_i}(\mu_{S_{<i}}(f)) \right] \leq \mu \left[\text{Ent}_{V_j \cap S_i}(\mu_{V_j \cap S_{<i}}(f)) \right]. \tag{4.24}$$

To see this, we apply Lemma 4.3.4 with $A = V_j \cap S_i$, $B = S_{<i}$ and $U = V_j \cap S_{<i}$. Since $\mu_{S_{<i+1}} = \otimes_{j=1}^{i} \mu_{S_j}$ the assumptions for that lemma are satisfied and we obtain Eq. (4.24)

from Eq. (4.17).

Summarizing, we have obtained

$$\mathrm{Ent}(f) \leq \frac{C}{\theta} \sum_{j=1}^{k} \mathbb{E}\left[\sum_{i\geq 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right]\right]. \tag{4.25}$$

We show next that for all $j = 1, ..., k$

$$\mathbb{E}\left[\sum_{i\geq 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right]\right] \leq C'\mu\left[\mathrm{Ent}_{V_j}(f)\right], \tag{4.26}$$

with $C' = O\left(\frac{\log(1/b)}{b^2}\Delta^2\right)$. Combined with Eq. (4.25), this concludes the proof of the lemma.

Let us fix j and let $v_1, v_2, \ldots$ denote an ordering of the sites in $V_j \cap S$ such that $v_1, ..., v_{|V_j\cap S_1|}$ is an ordering of $V_j \cap S_1$, $v_{|V_j\cap S_1|+1}, ..., v_{|V_j\cap S_1|+|V_j\cap S_2|}$ is an ordering of $V_j \cap S_2$ and so on. Since, for all $i \geq 1$, $\mu_{V_j\cap S_i}$ is a product measure, Lemmas 4.3.1 and 4.3.4 (as in Eq. (4.23)) imply

$$\mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right] = \sum_{h=|V_j\cap S_1|+\cdots+|V_j\cap S_{i-1}|+1}^{|V_j\cap S_1|+\cdots+|V_j\cap S_i|} \mu\left[\mathrm{Ent}_{v_h}(\rho_{v_h}(f))\right],$$

where ρ_{v_h} is the conditional distribution obtained from μ by freezing the spins at all the sites outside V_j, together with all the sites $v_h, v_{h+1}, \ldots, v_{|V_j\cap S|}$.

Using this decomposition and rearranging one finds

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i\geq 1} \Gamma(S_i)\, \mu\left[\mathrm{Ent}_{V_j\cap S_i}(\mu_{V_j\cap S_{<i}}(f))\right]\right] \\
&= \mathbb{E}\left[\sum_{i\geq 1} \Gamma(S_i) \sum_{h=|V_j\cap S_1|+\cdots+|V_j\cap S_{i-1}|+1}^{|V_j\cap S_1|+\cdots+|V_j\cap S_i|} \mu\left[\mathrm{Ent}_{v_h}(\rho_{v_h}(f))\right]\right] \\
&= \mathbb{E}\left[\sum_{h} \mu\left[\mathrm{Ent}_{v_h}(\rho_{v_h}(f))\right] \Gamma(S(v_h))\right],
\end{aligned} \tag{4.27}$$

where $S(v_h)$ denotes the (unique) connected component of S containing v_h. Notice that for each realization of S, $\mu_{V_j \cap S}$ is a product measure and so one has from Lemmas 4.3.1 and 4.3.4 that

$$\sum_h \mu \left[\mathrm{Ent}_{v_h}(\rho_{v_h}(f)) \right] = \mu \left[\mathrm{Ent}_{V_j \cap S}(f)) \right] \leq \mu \left[\mathrm{Ent}_{V_j}(f) \right];$$

the inequality follows from Eq. (4.13).

Observe that each term $\mu[\mathrm{Ent}_{v_h}(\rho_{v_h}(f))]$, as well as the sequence $\{v_h\}$, depends on the realization S only through $V_j \cap S$. Therefore,

$$\mathbb{E}\left[\sum_h \mu \left[\mathrm{Ent}_{v_h}(\rho_{v_h}(f)) \right] \Gamma(S(v_h)) \right] = \mathbb{E}\left[\sum_h \mu \left[\mathrm{Ent}_{v_h}(\rho_{v_h}(f)) \right] \mathbb{E}\left[\Gamma(S(v_h)) \,|\, V_j \cap S \right] \right],$$

where $\mathbb{E}\left[\Gamma(S(v_h)) \,|\, V_j \cap S \right]$ is the conditional expectation of $\Gamma(S(v_h))$ given the realization $V_j \cap S$. Therefore, Eq. (4.26) follows if we prove that

$$\max_{W \subseteq V_j} \max_{v \in W} \mathbb{E}\left[\Gamma(S(v)) \,|\, V_j \cap S = W \right] \leq C'. \tag{4.28}$$

Now, for a b marginally bounded spin system, it follows from Lemma 4.2 in [45] and Eq. (4.13) that

$$\Gamma(S(v)) \leq \zeta |S(v)|^3 z^{|S(v)|},$$

where $\zeta = \zeta(b) = \frac{3 \log(1/b)}{2b^2}$ and $z = 1/b^2$. Thus,

$$\max_{W \subseteq V_j} \max_{v \in W} \mathbb{E}\left[\Gamma(S(v)) \,|\, V_j \cap S = W \right] \leq \zeta \cdot \max_{W \subseteq V_j} \max_{v \in W} \mathbb{E}\left[|S(v)|^3 z^{|S(v)|} \;\middle|\; V_j \cap S = W \right]. \tag{4.29}$$

To bound the expectation on the right-hand-side of Eq. (4.29), we consider the graph G_2 with vertex set V and edge set $E \cup E_2$, where E is the edge set of G and E_2 is the set of all pairs of vertices with a common neighbor in G. Note that G_2 has maximum degree Δ^2. Let $\mathcal{A}_v(a)$ be the collection of subsets of vertices $U \subseteq V$ such that $|U| \geq a$, $v \in U$ and the

induced subgraph $G_2[U]$ of U in G_2 is connected.

Now, let us fix the set $W = V_j \cap S$ and the vertex $v \in W$ and let $S_2 := (S(v) \cap V_{\neq j}) \subseteq S$, where $V_{\neq j} := \bigcup_{i:i\neq j} V_i$. We claim that when the event $\{|S(v)| = a\}$ occurs for some $a \in \mathbb{N}$, then $S_2 \in \mathcal{A}_v(\frac{a}{\Delta+1})$. Indeed, $G_2[S_2]$ is connected, since $S(v)$ is connected in G and removing the vertices in $V_j \setminus \{v\}$ from $S(v)$ will not disconnect S_2 in G_2. Moreover, $\Delta|S(v) \cap V_{\neq j}| \geq |S(v) \cap V_j|$, and so

$$a = |S(v) \cap V_j| + |S(v) \cap V_{\neq j}| \leq (\Delta + 1)|S(v) \cap V_{\neq j}|,$$

which implies that $|S_2| = |S(v) \cap V_{\neq j}| \geq a/(\Delta + 1)$. Given S, let $T_2(v)$ denote the connected component of S in G_2 containing v, and note that $S_2 \subseteq T_2(v)$. Then, for any $W \subseteq V_j$, $v \in W$ and integer $a \geq 1$ we get

$$\begin{aligned}\mathbb{P}\left(|S(v)| = a \,|\, V_j \cap S = W\right) &\leq \mathbb{P}\left(\exists\, S_2 \in \mathcal{A}_v\left(\frac{a}{\Delta+1}\right); S_2 \subseteq S\right) \\ &\leq \mathbb{P}\left(|T_2(v)| \geq \frac{a}{\Delta+1}\right). \end{aligned} \tag{4.30}$$

Next we use Lemma 4.3 from [45], which implies that for any integer $m \geq 1$,

$$\mathbb{P}\left(|T_2(v)| = m\right) \leq \frac{\ell}{n}(2e\Delta^2\theta)^{m-1}. \tag{4.31}$$

Indeed, the only difference with respect to Lemma 4.3 from [45] is that we have maximum degree Δ^2 here instead of Δ. In particular, if $2e\Delta^2\theta \leq 1/2$, using $\frac{\ell}{n} \leq 2\theta$,

$$\mathbb{P}\left(|T_2(v)| \geq \frac{a}{\Delta+1}\right) \leq 4\theta(2e\Delta^2\theta)^{\lfloor\frac{a}{\Delta+1}\rfloor-1} \leq \Delta^{-2}(2e\Delta^2\theta)^{\lfloor\frac{a}{\Delta+1}\rfloor}. \tag{4.32}$$

It follows that

$$\mathbb{E}\left[|S(v)|^3 z^{|S(v)|} \,|\, V_j \cap S = W\right] = \sum_{a\geq 1} a^3 z^a \cdot \mathbb{P}\left(|S(v)| = a \,|\, V_j \cap S = W\right) \tag{4.33}$$

$$\leq \Delta^{-2} \sum_{a \geq 1} a^3 (2e\Delta^2 \theta z^{\Delta+1})^{\lfloor \frac{a}{\Delta+1} \rfloor} \leq C_1 \Delta^2, \qquad (4.34)$$

for some absolute constant C_1 provided that $2e\Delta^2\theta z^{\Delta+1} \leq 1/2$. This implies that

$$\max_{W \subseteq V_j} \max_{v \in W} \mathbb{E}\left[|S(v)|^3 z^{|S(v)|} \mid V_j \cap S = W\right] \leq C_1 \Delta^2.$$

Hence, Eqs. (4.26) and (4.28) hold with $C' = C_1 \zeta \Delta^2$, and so k-partite factorization holds with constant $C_{\text{UBF}} C_1 \zeta \Delta^2 / \theta$. □

4.5 Optimal Mixing of Swendsen-Wang Dynamics

In this section, we show that for ferromagnetic Potts models, the k-partite factorization of entropy, as defined in Eq. (4.20), implies optimal mixing of the Swendsen-Wang (SW) dynamics. Since we have already established that, for any spin system, k-partite factorization is implied by spectral independence, we then deduce Theorem 4.1.2 from the introduction.

We again take $G = (V, E)$ to be an n-vertex graph of maximum degree Δ and μ to be the Potts distribution on G with configuration space $\Omega = [q]^V$. The SW dynamics takes a spin configuration, transforms it into a "joint" spin-edge configuration, performs a step in the joint space, and then drops the edges to obtain a new Potts configuration. Formally, from a Potts configuration $\sigma_t \in [q]^V$, a transition $\sigma_t \to \sigma_{t+1}$ of the SW dynamics is defined as follows:

1. Let $M_t = M(\sigma_t)$ denote the set of monochromatic edges in σ_t.

2. Independently for each edge $e \in M_t$, keep e with probability $p = 1 - \exp(-\beta)$ and remove e with probability $1 - p$. Let $A_t \subseteq M_t$ denote the resulting subset.

3. In the subgraph (V, A_t), independently for each connected component C (including isolated vertices), choose a spin s_C uniformly at random from $[q]$ and assign to each vertex in C the spin s_C. This spin assignment defines σ_{t+1}.

It will be useful for us to consider the "joint" Edwards-Sokal distribution for G with parameters $p \in [0,1]$ and integer $q \geq 2$. Let $\Omega_{\text{J}} = \Omega \times \{0,1\}^E$ be the set of "joint" spin-edge configurations (σ, A) consisting of a spin assignment to the vertices $\sigma \in \Omega$ and a subset of edges $A \subseteq E$. The Edwards-Sokal measure assigns to each $(\sigma, A) \in \Omega_{\text{J}}$ a probability given by

$$\nu(\sigma, A) = \frac{1}{Z_{\text{J}}} p^{|A|}(1-p)^{|E|-|A|}\mathbf{1}(\sigma \sim A), \tag{4.35}$$

where $\sigma \sim A$ means that $A \subseteq M(\sigma)$ (i.e., every edge in A is monochromatic in σ) and Z_{J} is the corresponding normalizing constant or partition function. When $p = 1 - e^{-\beta}$, the "spin marginal" of ν is precisely the Potts distribution μ and $Z_G = Z_{\text{J}}$, and the "edge marginal" of ν corresponds to the random-cluster measure; see, e.g., [56, 63] for extensive background on these measures.

A key concept in our strategy to prove optimal mixing results for the SW dynamics is the *spin/edge factorization* of entropy in the joint space Ω_{J}. This spin/edge factorization was shown in [24, Lemma 1.8] to imply $O(\log n)$ mixing of the SW dynamics on any graph. Moreover, in [24] it was proved that for bipartite graphs, even/odd factorization of entropy for μ implies the desired spin/edge factorization of entropy in the joint space Ω_{J}. We will generalize the argument from [24] to general graphs, and show that a k-partite factorization of entropy for μ implies spin/edge factorization of entropy in the joint space Ω_{J}, and thus combined with [24, Lemma 1.8] this will complete the proof of $O(\log n)$ mixing of the SW dynamics. Note, this $O(\log n)$ bound is optimal as there are graphs of bounded degree where the SW dynamics requires $\Omega(\log n)$ steps to mix.

Before stating our results, we stipulate some notation. For a function $f : \Omega_{\text{J}} \to \mathbb{R}_+$, we write $\mathrm{Ent}_\nu(f) = \nu\left[f \log\left(\frac{f}{\nu(f)}\right)\right]$ for the entropy of f with respect to ν. For a fixed configuration $\sigma \in \Omega$ and subset of edges $A \subseteq E$, $\mathrm{Ent}_\nu(f \,|\, \sigma)$ and $\mathrm{Ent}_\nu(f \,|\, A)$ denote the entropy of f with respect to the conditional measures $\nu(\cdot \,|\, \sigma)$ and $\nu(\cdot \,|\, A)$, respectively.

More precisely, for a given $\sigma \in \Omega$, $\nu(\cdot \,|\, \sigma)$ is the measure ν conditioned on the event that the spin configuration is equal to σ, and for a given $A \subseteq E$, $\nu(\cdot \,|\, A)$ is the measure ν conditioned on the event that the edge configuration is equal to A. In this way, $\mathrm{Ent}_\nu(f \,|\, \sigma)$ and $\mathrm{Ent}_\nu(f \,|\, A)$ are functions of σ and A, respectively, and $\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma)\right]$, $\nu\left[\mathrm{Ent}_\nu(f \,|\, A)\right]$ denote the corresponding expectations with respect to ν.

Theorem 4.5.1. *Suppose μ satisfies the k-partite factorization of entropy with constant C_{par}; see Eq.* (4.20)*. Then, there exists a constant $C = C(C_{\mathrm{par}}, \beta, \Delta)$ such that for all $f : \Omega_{\text{J}} \mapsto \mathbb{R}_+$*

$$\mathrm{Ent}_\nu(f) \leq C\ \left(\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma)\right] + \nu[\mathrm{Ent}_\nu(f \,|\, A)]\right). \tag{4.36}$$

The constant C satisfies $C = C_{\mathrm{par}} \times O(\beta\Delta^2 e^{\beta\Delta})$.

We call Eq. (4.36) the *spin/edge factorization of entropy* with constant C for the joint measure ν. The main motivation for this inequality is the result established in [24, Lemma 1.8] that on any n-vertex graph, approximate spin/edge factorization with constant C implies that the SW dynamics has discrete time entropy decay with rate $\delta = 1/C$, and therefore, by Lemma 2.2.4, satisfies $T_{\mathrm{mix}} = O(\log n)$. Theorem 4.1.2 from the introduction now follows immediately.

Proof of Theorem 4.1.2. For the Potts model one has $e^{\beta\Delta} = O(1/b)$. Therefore, the results follows from Theorem 4.6.1, Lemma 4.4.3, Theorem 4.5.1 and [24, Lemma 1.8]. □

Let $\{V_1, ..., V_k\}$ be the k-partition of G, where $k \leq \Delta + 1$, as in Section 4.4. For all $j \in [k]$ let $\nu(\cdot \,|\, \overline{\sigma_{V_j}}, A)$ denote the measure ν conditioned on $\overline{\sigma_{V_j}} = \{\sigma_v,\, v \notin V_j\}$ and $A \subseteq E$. We use $\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}, A)$ to denote the corresponding conditional entropy and $\nu\left[\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}, A)\right]$ for its expectation with respect to ν. Theorem 4.5.1 will follow from the following lemmas.

Lemma 4.5.2. *For all $f : \Omega_{\mathfrak{J}} \mapsto \mathbb{R}_+$ and all $j \in [k]$ we have*

$$\nu\left[\mathrm{Ent}_\nu(f \mid A)\right] \geq \nu\left[\mathrm{Ent}_\nu(f \mid \overline{\sigma_{V_j}}, A)\right].$$

Lemma 4.5.3. *There exists a constant $\delta_1 > 0$ such that, for all $f : \Omega_{\mathfrak{J}} \mapsto \mathbb{R}_+$ and all $j \in [k]$,*

$$\nu\left[\mathrm{Ent}_\nu(f \mid \sigma)\right] + \nu\left[\mathrm{Ent}_\nu(f \mid \overline{\sigma_{V_j}}, A)\right] \geq \delta_1\, \nu\left[\mathrm{Ent}_\nu(f \mid \overline{\sigma_{V_j}})\right].$$

The constant δ_1 satisfies $1/\delta_1 = O(\beta\Delta e^{\beta\Delta})$.

Lemma 4.5.4. *If μ satisfies the k-partite factorization with constant C_{par}, then for all $f : \Omega_{\mathfrak{J}} \mapsto \mathbb{R}_+$,*

$$\sum_{j=1}^{k} \nu\left[\mathrm{Ent}_\nu(f \mid \overline{\sigma_{V_j}})\right] \geq \delta_2 \mathrm{Ent}_\nu(f),$$

where $\delta_2 = \frac{1}{C_{\mathrm{par}}}$.

Proof of Theorem 4.5.1. By combining the bounds from Lemmas 4.5.2 to 4.5.4 we get

$$\nu\left[\mathrm{Ent}_\nu(f \mid \sigma) + \mathrm{Ent}_\nu(f \mid A)\right] \geq \frac{\delta_1\delta_2}{k}\mathrm{Ent}_\nu(f), \tag{4.37}$$

and so, using also $k \leq \Delta + 1$, the spin/edge factorization holds with constant

$$C = \frac{k}{\delta_1\delta_2} = C_{\mathrm{par}} \times O(\beta\Delta^2 e^{\beta\Delta}). \qquad \square$$

We briefly discuss next the proof of Lemmas 4.5.2 to 4.5.4, which are the respective counterparts of Lemmas 4.3, 4.4 and 4.5 in [24] for the bipartite setting.

Proof of Lemma 4.5.2. This is an instance of the same monotonicity that has already been seen in Lemma 4.3.2. In this particular case, it follows from the argument in the proof of

Lemma 4.3 in [24] by simply substituting σ_O with $\overline{\sigma_{V_j}}$ in that proof. □

Proof of Lemma 4.5.3. Let us fix $j \in [k]$. To simplify the notation, we shall use xy to denote the edge $\{x, y\}$, and view the edge configuration A as a vector in $\{0,1\}^E$. For any fixed configuration $\overline{\sigma_{V_j}}$ of spins, the conditional probability $\nu(\cdot \,|\, \overline{\sigma_{V_j}})$ is a product measure. That is,

$$\nu(\cdot \,|\, \overline{\sigma_{V_j}}) = \bigotimes_{x \in V_j} \nu_x(\cdot \,|\, \overline{\sigma_{V_j}}), \tag{4.38}$$

where, for each $x \in V_j$, $\nu_x(\cdot \,|\, \overline{\sigma_{V_j}})$ is the probability measure on $\{1, \ldots, q\} \times \{0,1\}^{\deg(x)}$ with $\deg(x)$ denoting the degree of x. The distribution $\nu_x(\cdot \,|\, \overline{\sigma_{V_j}})$ can be described s follows: pick the spin of site x according to the Potts measure on x conditioned on the spin of its neighbors in $V \setminus V_j$; then, independently for every edge $xy \in E$ incident to the vertex x, if $\sigma_x = \sigma_y$ set $A_{xy} = 1$ with probability p and set $A_{xy} = 0$ otherwise; if $\sigma_x \neq \sigma_y$, set $A_{xy} = 0$.

The measure $\nu(\cdot \,|\, \overline{\sigma_{V_j}}, A)$, obtained by further conditioning on a valid configuration of all edge variables A compatible with the fixed spins $\overline{\sigma_{V_j}}$, is again a product measure:

$$\nu(\cdot \,|\, \overline{\sigma_{V_j}}, A) = \bigotimes_{x \in V_j} \nu_x(\cdot \,|\, \overline{\sigma_{V_j}}, A), \tag{4.39}$$

where $\nu_x(\cdot \,|\, \overline{\sigma_{V_j}}, A)$ is the probability measure on $\{1, \ldots, q\}$ that is uniform if x has no incident edges in A, and is concentrated on the unique admissible value given $\overline{\sigma_{V_j}}$ and A otherwise.

Next, we note that $\nu(\cdot \,|\, \sigma)$ is a product of Bernoulli(p) random variables over the set of all monochromatic edges in σ, while it is concentrated on $A_e = 0$ on all remaining edges. Therefore we may write

$$\nu(\cdot \,|\, \sigma) = \bigotimes_{x \in V} \nu_x(\cdot \,|\, \sigma), \tag{4.40}$$

where $\nu_x(\cdot \,|\, \sigma)$ is the probability measure on $\{0,1\}^{\deg(x)}$, which is given by the product of Bernoulli(p) variables on all edges xy incident to x such that $\sigma_x = \sigma_y$ and is concentrated on $A_{xy} = 0$ if $\sigma_x \neq \sigma_y$.

We write $\mathrm{Ent}_x(\cdot \,|\, \overline{\sigma_{V_j}})$, $\mathrm{Ent}_x(\cdot \,|\, \overline{\sigma_{V_j}}, A)$, $\mathrm{Ent}_x(\cdot \,|\, \sigma)$ for the entropies with respect to the distributions $\nu_x(\cdot \,|\, \overline{\sigma_{V_j}})$, $\nu_x(\cdot \,|\, \overline{\sigma_{V_j}}, A)$, $\nu_x(\cdot \,|\, \sigma)$ respectively. The first observation is that, for every site x, there is a local factorization of entropies in the following sense. There exists a constant $\delta_1 > 0$ such that $1/\delta_1 = O(\beta\Delta e^{\beta\Delta})$, and such that for all functions $f \geq 0$ and all $x \in V_j$,

$$\nu_x \left[\mathrm{Ent}_x(f \,|\, \sigma) \,|\, \overline{\sigma_{V_j}} \right] + \nu_x \left[\mathrm{Ent}_x(f \,|\, \overline{\sigma_{V_j}}, A) \,|\, \overline{\sigma_{V_j}} \right] \geq \delta_1 \, \mathrm{Ent}_x(f \,|\, \overline{\sigma_{V_j}}); \tag{4.41}$$

this follows from a direct generalization of Lemma 4.7 from [24] for bipartite graphs; the proof of such generalization to the k-partite setting is the same as that of Lemma 4.7 and is thus omitted.

Next, we want to lift Eq. (4.41) to the product measure $\nu(\cdot \,|\, \overline{\sigma_{V_j}}) = \otimes_{x \in V_j} \nu_x(\cdot \,|\, \overline{\sigma_{V_j}})$. Let $x = 1, \ldots, n$ denote an arbitrary ordering of the sites $x \in V_j$. Let $A_x \in \{0,1\}^{\deg(x)}$ be the random variable corresponding to the state of the edges incident to x. We write $\xi_x = (\sigma_x, A_x)$ for the pair of variables corresponding to x. We first observe that

$$\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}) = \sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \overline{\sigma_{V_j}}) \,|\, \overline{\sigma_{V_j}} \right], \tag{4.42}$$

where $g_x = \nu \left[f \,|\, \overline{\sigma_{V_j}}, \xi_{x+1}, \ldots, \xi_n \right]$, $g_0 = f$ and $g_n = \nu \left[f \,|\, \overline{\sigma_{V_j}} \right]$. This identity is an instance of the decomposition in Lemma 4.3.1.

Putting together Eq. (4.41) and Eq. (4.42) yields

$$\begin{aligned} \delta_1 \, \mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}) &\leq \sum_{x=1}^{n} \nu \left[\nu_x \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \overline{\sigma_{V_j}} \right] + \nu_x \left[\mathrm{Ent}_x(g_{x-1} \,|\, \overline{\sigma_{V_j}}, A) \,|\, \overline{\sigma_{V_j}} \right] \,|\, \overline{\sigma_{V_j}} \right] \\ &= \sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) + \mathrm{Ent}_x(g_{x-1} \,|\, \overline{\sigma_{V_j}}, A) \,|\, \overline{\sigma_{V_j}} \right]. \end{aligned} \tag{4.43}$$

Proceeding as in the proof of Lemma 4.8 from [24], we obtain the following two inequalities:

$$\sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \sigma) \,|\, \overline{\sigma_{V_j}} \right] \leq \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \overline{\sigma_{V_j}} \right],$$
$$\sum_{x=1}^{n} \nu \left[\mathrm{Ent}_x(g_{x-1} \,|\, \overline{\sigma_{V_j}}, A) \,|\, \overline{\sigma_{V_j}} \right] \leq \nu \left[\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}, A) \,|\, \overline{\sigma_{V_j}} \right].$$

These two inequalities combined with Eq. (4.43) yields that

$$\delta_1 \, \mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}) \leq \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \overline{\sigma_{V_j}} \right] + \nu \left[\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}, A) \,|\, \overline{\sigma_{V_j}} \right]. \tag{4.44}$$

The results follows by taking expectations with respect to ν in Eq. (4.44). □

Proof of Lemma 4.5.4. From the definition of conditional entropy as well as the fact that $\nu(\cdot \,|\, \sigma_{V_j}, \overline{\sigma_{V_j}}) = \nu(\cdot \,|\, \sigma)$ we get

$$\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}}) = \mathrm{Ent}_\nu \left(\nu \left[f \,|\, \sigma \right] \,|\, \overline{\sigma_{V_j}} \right) + \nu \left[\mathrm{Ent}_\nu(f \,|\, \sigma) \,|\, \overline{\sigma_{V_j}} \right]. \tag{4.45}$$

(see eq. (4.5), (4.6) from Lemma 4.5 in [24]). Now, since the function $\nu \left[f \,|\, \sigma \right]$ depends only on the spin configuration σ, one has the identity

$$\sum_{j=1}^{k} \nu \left[\mathrm{Ent}_\nu(\nu[f \,|\, \sigma] \,|\, \overline{\sigma_{V_j}}) \right] = \sum_{j=1}^{k} \mu \left[\mathrm{Ent}(\nu[f \,|\, \sigma] \,|\, \overline{\sigma_{V_j}}) \right], \tag{4.46}$$

where the entropy in the right hand side is with respect to μ and not with respect to ν. Since k-partite factorization holds by assumption,

$$\sum_{j=1}^{k} \mu \left[\mathrm{Ent}(\nu[f \,|\, \sigma] \,|\, \overline{\sigma_{V_j}}) \right] \geq \delta_2 \, \mathrm{Ent} \left(\nu \left[f \,|\, \sigma \right] \right), \tag{4.47}$$

where $\delta_2 = 1/C_{\mathrm{par}}$. By taking functions depending only on σ_{V_j} for a single V_j one easily sees that C_{par} must be at least 1. Then, taking expectation and summing over j in Eq. (4.45),

and combining with Eq. (4.46) and Eq. (4.47), we get

$$\sum_{j=1}^{k} \nu\left[\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}})\right] \geq \delta_2 \,\mathrm{Ent}_\nu\left(\nu\left[f \,|\, \sigma\right]\right) + k\,\nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma)\right].$$

Using the simple decomposition $\mathrm{Ent}_\nu(f) = \mathrm{Ent}_\nu\left(\nu\left[f \,|\, \sigma\right]\right) + \nu\left[\mathrm{Ent}_\nu(f \,|\, \sigma)\right]$, and the fact that $\delta_2 \leq 1 \leq k$, we conclude that

$$\sum_{j=1}^{k} \nu\left[\mathrm{Ent}_\nu(f \,|\, \overline{\sigma_{V_j}})\right] \geq \delta_2 \,\mathrm{Ent}_\nu(f). \qquad \square$$

4.6 Uniform Block Factorization via Spectral Independence: A Direct Approach

In this section, we present a more direct proof of Theorem 3.2.4 which can be interesting to certain groups. The proof in this section uses a recursive arguments of approximate tensorization and subadditivity of entropy and does not utilize abstract simplicial complexes.

The goal of this section is to reformulate in the setting of spin systems some of the key facts that were derived in Section 3.4 and the references therein in the more general framework of simplicial complexes. This specialization yields some minor simplification in the main proofs, and may be of use for later reference. The approach consists in exploiting a recursive scheme which allows one to derive a global contraction estimate by analysing the spectral norm of a local operator. This is reminiscent of the recursive approach developed in [38, 35, 34], where similar ideas were used to derive spectral gap estimates for a class of conservative spin systems. The argument here is more robust and, unlike the one in [38, 35, 34], it does not rely on symmetries of the underlying measures.

We first introduce some notation. Let f be a function of the full spin configuration σ, and $U \subseteq V = [n]$ a subset of vertices. We use the notation $\mu^U = \mu_{V\setminus U}$ for the conditional distribution given the spins in U, and write $\mathrm{Av}_{|U|=\ell}$ for the uniform average over all sets $U \subseteq [n]$ with ℓ vertices. We are going to prove the following result which is a reformulation of Theorem 3.2.4.

Theorem 4.6.1. *If the spin system is η-spectrally independent and b-marginally bounded then there exists a constant $C = O(1 + \frac{\eta}{b})$ such that for any $\ell = \{1, \dots, n-1\}$ and for all $f \geq 0$:*

$$\frac{n}{\ell} \operatorname{Av}_{|U|=\ell} \operatorname{Ent}(\mu^U f) \leq C \operatorname{Ent} f. \tag{4.48}$$

Moreover, for any $\theta \in (0, 1]$, there exists $C = \left(\frac{1}{\theta}\right)^{O(\frac{\eta}{b})}$ such that for $\ell = \lceil \theta n \rceil$:

$$\frac{\ell}{n} \operatorname{Ent} f \leq C \operatorname{Av}_{|\Lambda|=\ell} \mu \left[\operatorname{Ent}_\Lambda f\right]. \tag{4.49}$$

We remark that Eq. (4.48) is an approximate subadditivity statement, which coincides with Eq. (4.7) when $\ell = 1$. On the other hand Eq. (4.49) is the uniform block factorization statement ℓ-UBF with $\ell = \lceil \theta n \rceil$; see Definition 4.4.1. We articulate the proof in two steps. The first is a recursive scheme which allows one to go from a local inequality to a global one; see Lemma 4.6.3. The second step is a control of the local inequality; see Lemma 4.6.4.

4.6.1 Setting up the Recursion

If $U \subseteq V$, and $\tau = \tau_U$ a configuration of spins on U, recall that we use notation $\mu^\tau = \mu(\cdot \,|\, \tau)$ for the conditional distribution μ^U when the spins on U are given by τ. Moreover, we write $\mu^{\tau,x} = \mu(\cdot \,|\, \tau \cup \sigma_x)$ if we additionally condition on the spin σ_x at vertex $x \notin U$ and similarly for $\mu^{\tau,x,y} = \mu(\cdot \,|\, \tau \cup \sigma_x \cup \sigma_y)$ for $x, y \notin U$, so that e.g. the expression $\mu^\tau \left[\operatorname{Ent}_{\mu^{\tau,x,y}} f\right]$ indicates the entropy of f with respect to $\mu(\cdot \,|\, \tau \cup \sigma_x \cup \sigma_y)$,

$$\operatorname{Ent}_{\mu^{\tau,x,y}} f = \mu^{\tau,x,y}[f \log(f/\mu^{\tau,x,y}(f))]$$

averaged over the two spins σ_x, σ_y sampled according to μ^τ. Define the constants α_k, $k = 0, \ldots, n-2$, as the largest numbers such that the inequalities

$$(1+\alpha_k)\mathrm{Av}_{x\notin U}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \le \mathrm{Av}_{x,y\notin U}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f))\,, \tag{4.50}$$

hold for all $k = 0, \ldots, n-2$, for all $U \subseteq [n]$ with $|U| = k$, for all configurations τ on U and for all functions $f \ge 0$. The symbol $\mathrm{Av}_{x\notin U}$ denotes the uniform average over all $n-k$ vertices $x \notin U$, and $\mathrm{Av}_{x,y\notin U}$ stands for the uniform average over all $(n-k)(n-k-1)$ pairs (x, y) with $x, y \notin U$ and $x \neq y$. We refer to Eq. (4.50) as the local inequality, since for each choice of x, y, the distributions involved are concerned with the spins at two vertices only.

Remark 4.6.2. Fix $x, y \notin U$. Using $\mu^{\tau,x} f = \mu^{\tau,x}\mu^{\tau,x,y} f$, from Lemma 4.3.1 we have the decomposition

$$\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) = \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) + \mu^\tau\left[\mathrm{Ent}_{\mu^{\tau,x}}(\mu^{\tau,x,y}(f)\right].$$

In particular, $\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \ge \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))$ and therefore Eq. (4.50) is always true with $\alpha_k = 0$. If μ is a product measure then the subadditivity of entropy for product measures gives

$$\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \ge \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) + \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,y}(f)),$$

which implies the validity of Eq. (4.50) with $\alpha_k = 1$ for all $k = 0, \ldots, n-2$.

The recursion is based on the following statement, which rephrases [45, Theorem 5.4].

Lemma 4.6.3. *Let α_k, $k = 0, \ldots, n-2$, be defined by Eq.* (4.50)*. Then, for all functions $f \ge 0$,*

$$\mathrm{Av}_{|U|=j}\mathrm{Ent}(\mu^U f) \le (1-\kappa_j)\mathrm{Ent}(f), \qquad j = 1, \ldots, n-1, \tag{4.51}$$

where

$$\kappa_j = \frac{\sum_{i=j}^{n-1} \Gamma_i}{\sum_{i=0}^{n-1} \Gamma_i}\,, \qquad \Gamma_i = \prod_{k=0}^{i-1} \alpha_k\,, \quad \Gamma_0 = 1.$$

Proof. The claim Eq. (4.51) follows from the fact that for all $k = 1, \dots, n-1$:

$$\mathrm{Av}_{|U|=k} \mathrm{Ent}(\mu^U f) \le \delta_k \mathrm{Av}_{|U|=k+1} \mathrm{Ent}(\mu^U f)\,, \qquad \delta_k = \frac{\sum_{i=0}^{k-1} \Gamma_i}{\sum_{i=0}^{k} \Gamma_i}, \tag{4.52}$$

since $\mathrm{Av}_{|U|=n} \mathrm{Ent}(\mu^U f) = \mathrm{Ent}(f)$, and $\delta_j \delta_{j+1} \cdots \delta_{n-1} = (1 - \kappa_j)$.

To prove Eq. (4.52), note that it holds for $k = 1$ with $\delta_1 = 1/(1+\alpha_0) = \Gamma_0/(\Gamma_0 + \Gamma_1)$ by the assumption Eq. (4.50) at $\tau = \emptyset$. Next, we suppose it holds for $0 < k-1 < n-1$ and show it for k. For any $|U| = k+1$ and $U' \subseteq U$ with $|U'| = k-1$, setting $\{x, y\} = U \setminus U'$ and letting $\tau = \tau_{U'}$ be the configuration on U', as in Lemma 4.3.1 we have the decomposition

$$\begin{aligned} \mathrm{Ent}(\mu^U f) &= \mathrm{Ent}(\mu(\mu^U f \,|\, \tau_{U'})) + \mu\left[\mathrm{Ent}(\mu^U f \,|\, \tau_{U'})\right] \\ &= \mathrm{Ent}(\mu^{U'} f) + \mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y} f) \,|\, \tau_{U'}\right]. \end{aligned} \tag{4.53}$$

Averaging we obtain

$$\mathrm{Av}_{|U|=k+1} \mathrm{Ent}(\mu^U f) = \mathrm{Av}_{|U'|=k-1} \mathrm{Ent}(\mu^{U'} f) \tag{4.54}$$

$$+ \mathrm{Av}_{|U'|=k-1} \mathrm{Av}_{x,y \notin U'} \mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y} f) \,|\, \tau_{U'}\right]. \tag{4.55}$$

In the same way

$$\mathrm{Av}_{|U|=k} \mathrm{Ent}(\mu^U f) = \mathrm{Av}_{|U'|=k-1} \mathrm{Ent}(\mu^{U'} f) \tag{4.56}$$

$$+ \mathrm{Av}_{|U'|=k-1} \mathrm{Av}_{x \notin U'} \mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x} f) \,|\, \tau_{U'}\right]. \tag{4.57}$$

From Eq. (4.50),

$$\mathrm{Av}_{|U|=k+1} \mathrm{Ent}(\mu^U f) - \mathrm{Av}_{|U'|=k-1} \mathrm{Ent}(\mu^{U'} f) \tag{4.58}$$

$$\geq (1+\alpha_{k-1})\mathrm{Av}_{|U'|=k-1}\mathrm{Av}_{x\notin U'}\mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}f)\,|\,\tau_{U'}\right] \tag{4.59}$$

$$= (1+\alpha_{k-1})\left[\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu^U f) - \mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu^{U'}f)\right]. \tag{4.60}$$

Therefore,

$$\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu^U f) \geq (1+\alpha_{k-1})\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu^U f) - \alpha_{k-1}\mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu^{U'}f). \tag{4.61}$$

By the inductive assumption Eq. (4.52) at $k-1$ we have

$$\begin{aligned}\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu^U f) &\geq (1+\alpha_{k-1}-\alpha_{k-1}\delta_{k-1})\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu^U f)\\ &= \delta_k^{-1}\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu^U f).\end{aligned}$$

□

4.6.2 Estimating Local Coefficients

The next step is an estimate on the coefficients α_k appearing in Eq. (4.50).

Lemma 4.6.4. *If the spin system is η-spectrally independent and b-marginally bounded then the local inequality Eq.* (4.50) *holds with*

$$\alpha_k \geq 1 - \frac{2\eta}{b(n-k-1)}. \tag{4.62}$$

Proof. Fix $U \subseteq V$, $|U| = k \leq n-2$ and $\tau = \tau_U$. We may assume $\mu^\tau(f) = 1$, which implies $\mu^\tau(\mu^{\tau,x,y}(f)) = \mu^\tau(\mu^{\tau,x}(f)) = 1$ for all $x, y \notin U$. For simplicity, we write $\mathrm{Av}_{x,y}$ and Av_x for the averages $\mathrm{Av}_{x,y\notin U}$ and $\mathrm{Av}_{x\notin U}$. Observe that

$$\begin{aligned}&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))\\ &\quad= \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x,y}(f)\log\mu^{\tau,x,y}(f) - \mu^{\tau,x}(f)\log\mu^{\tau,x}(f) - \mu^{\tau,y}(f)\log\mu^{\tau,y}(f)\right]\end{aligned} \tag{4.63}$$

$$= \mathrm{Av}_{x,y}\, \mu^{\tau} \left[\mu^{\tau,x,y}(f) \log \frac{\mu^{\tau,x,y}(f)}{\mu^{\tau,x}(f)\mu^{\tau,y}(f)} \right]. \tag{4.64}$$

Using $a \log(a/b) \geq a - b$ for all $a, b \geq 0$,

$$\begin{aligned} &\mathrm{Av}_{x,y}\, \mathrm{Ent}_{\mu^{\tau}}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_{x}\, \mathrm{Ent}_{\mu^{\tau}}(\mu^{\tau,x}(f)) \\ &\quad \geq 1 - \mathrm{Av}_{x,y}\, \mu^{\tau} \left[\mu^{\tau,x}(f)\mu^{\tau,y}(f) \right] \\ &\quad = -\mathrm{Av}_{x,y}\, \mu^{\tau} \left[(\mu^{\tau,x}(f) - 1)(\mu^{\tau,y}(f) - 1) \right]. \end{aligned} \tag{4.65}$$

We may rewrite

$$\begin{aligned} &\mathrm{Av}_{x,y}\, \mu^{\tau} \left[(\mu^{\tau,x}(f) - 1)(\mu^{\tau,y}(f) - 1) \right] \\ &\qquad = \frac{1}{n-k-1} \sum_{(x,a) \in \mathcal{P}} \nu(x,a)\varphi(x,a)[J^{\tau}\varphi](x,a), \end{aligned} \tag{4.66}$$

where

$$\varphi(x,a) = \mu^{\tau}(f \,|\, \sigma_x = a) - 1 = [\mu^{\tau,x}(f)](a) - 1,$$

$\mathcal{P}$ is the set of all pairs (x,a) where $x \in V \setminus U$ (if U is the set where $\tau = \tau_U$ is specified) and $a \in [q]$, ν denotes the probability measure on $\mathcal{P}$ obtained by setting

$$\nu(x,a) = \frac{1}{n-k}\, \mu^{\tau}(\sigma_x = a),$$

and $J^{\tau} : \mathcal{P} \times \mathcal{P} \mapsto \mathbb{R}$ denotes the influence matrix from Definition 2.1.4. Note that in the derivation of Eq. (4.66) we have used the fact that for each fixed $y \notin U$ one has

$$\sum_{a' \in [q]} \nu(y,a')\varphi(y,a') = \frac{1}{n-k}\, \mu^{\tau}(\mu^{\tau,y}(f) - 1) = 0.$$

Observe that J^τ is self-adjoint in $L^2(\mathcal{P},\nu)$:

$$\nu(x,a)J^\tau(x,a;y,a') = \nu(y,a')J^\tau(y,a';x,a). \tag{4.67}$$

In particular, its eigenvalues are real. Let $\eta \geq 0$ denote its largest eigenvalue (the eigenvalue zero always exists since all row sums of J^τ vanish). Letting $\langle\cdot,\cdot\rangle$ denote the scalar product in $L^2(\mathcal{P},\nu)$ we have $\langle\psi, J^\tau\psi\rangle \leq \eta\langle\psi,\psi\rangle$ for all $\psi \in L^2(\mathcal{P},\nu)$. Therefore,

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right] \\
&\quad= \frac{1}{n-k-1}\langle\varphi, J^\tau\varphi\rangle \leq \frac{\eta}{n-k-1}\langle\varphi,\varphi\rangle \\
&\quad= \frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)^2\right] = \frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)).
\end{aligned} \tag{4.68}$$

Recalling Eq. (4.65) we have shown

$$\begin{aligned}
&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \\
&\qquad\qquad \geq -\frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)).
\end{aligned} \tag{4.69}$$

Next, observe that for every fixed $x \notin U$, setting $h^\tau(\sigma_x) = [\mu^{\tau,x}(f)](\sigma_x)$:

$$\begin{aligned}
\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)) &= \sum_a \mu^\tau(\sigma_x = a)(h^\tau(a)-1)^2 \\
&\leq \frac{1}{b}\left(\sum_a \mu^\tau(\sigma_x = a)|h^\tau(a)-1|\right)^2
\end{aligned}$$

where $b = \min_{x\notin U}\min_a \mu^\tau(\sigma_x = a)$, as in Definition 2.1.1, with the minimum over a restricted to spin values that are allowed at x, that is such that $\mu^\tau(\sigma_x = a) > 0$, and we have used $\sum_i a_i^2 \leq (\sum_i a_i)^2$ for all $a_i \geq 0$. Pinsker's inequality shows that

$$\sum_a \mu^\tau(\sigma_x = a)|h^\tau(a)-1| \leq \sqrt{2\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))}.$$

It follows that

$$\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)) \leq \frac{2}{b}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)). \tag{4.70}$$

Inserting Eq. (4.70) into Eq. (4.69) concludes the proof. □

4.6.3 Proof of Theorem 4.6.1

From Lemma 4.6.3, we see that Eq. (4.48) holds with $C = \frac{n}{\ell}(1-\kappa_\ell)$. From Lemma 4.6.4 if follows that

$$\alpha_k \geq \max\{1 - R/(n-k-1), 0\}, \qquad R = \lceil 2\eta/b \rceil.$$

Using this bound in the definition of the coefficients κ_ℓ and rearranging, see Section 2.2 of [45], it is not hard to see that for any $1 \leq \ell \leq n-1$:

$$\kappa_\ell \geq \frac{(n-\ell-1)\cdots(n-\ell-R)}{(n-1)\cdots(n-R)}. \tag{4.71}$$

In particular,

$$\frac{n}{\ell}(1-\kappa_\ell) \leq \frac{n}{\ell}\left(1 - \frac{(n-\ell-1)\cdots(n-\ell-R)}{(n-1)\cdots(n-R)}\right).$$

Remarkably, the expression in the right hand side above is decreasing with ℓ, and therefore it is always less than $R+1$, its value at $\ell = 1$. This shows that Eq. (4.48) holds with $C \leq R+1 = O(1+\frac{\eta}{b})$.

To prove Eq. (4.49), we start with the decomposition

$$\mathrm{Av}_{|\Lambda|=\ell}\,\mu\left[\mathrm{Ent}_\Lambda f\right] = \mathrm{Ent}(f) - \mathrm{Av}_{|U|=n-\ell}\,\mathrm{Ent}\left[\mu^U f\right],$$

which follows from Lemma 4.3.1. Therefore Lemma 4.6.3 implies that Eq. (4.49) holds

with $C = \frac{\ell}{n\,\kappa_{n-\ell}}$. Using Eq. (4.71) we see that

$$\frac{\ell}{n\,\kappa_{n-\ell}} \leq \frac{(n-1)\cdots(n-R)}{(\ell-1)\cdots(\ell-R)}.$$

In particular, if $\ell = \lceil \theta n \rceil$ with $\theta \in (0,1]$ fixed, then for all sufficiently large n one has $\frac{\ell}{n\,\kappa_{n-\ell}} \leq (\frac{1}{\theta})^{O(R)}$. This ends the proof of Theorem 4.6.1.

CHAPTER 5

SPECTRAL INDEPENDENCE VIA STRONG SPATIAL MIXING APPROACH: 2-SPIN SYSTEMS

In this chapter, we establish spectral independence for general 2-spin systems. For the antiferromagnetic case we show that spectral independence holds when the parameters lie in the so-called tree uniqueness region; this matches the parameter regime of the strong spatial mixing properties. This chapter is based on [44].

5.1 Optimal Mixing Results for 2-Spin Systems

Given an n-vertex graph $G = (V, E)$, configurations of the 2-spin model are the 2^n assignments of spins $0, 1$ to the vertices. A 2-spin system is defined by three parameters: edge weights $\beta, \gamma > 0$ and a vertex weight $\lambda > 0$. Edge parameter β controls the (relative) strength of interaction between neighboring 1-spins, γ corresponds to neighboring 0-spins, and λ is the external field applied to vertices with 1-spins.

Every spin configuration $\sigma \in \{0, 1\}^V$ is assigned a weight

$$w_G(\sigma) = \beta^{m_1(\sigma)}\gamma^{m_0(\sigma)}\lambda^{n_1(\sigma)},$$

where, for spin $s \in \{0, 1\}$, $m_s(\sigma) = \#\{uv \in E : \sigma_u = \sigma_v = s\}$ is the number of monochromatic edges with spin s, and $n_1(\sigma) = \#\{v \in V : \sigma_v = 1\}$ is the number of vertices with spin 1 (as is standard, the parameters are normalized so we can avoid two additional parameters). The Gibbs distribution over spin configurations is given by $\mu_G(\sigma) = \frac{w_G(\sigma)}{Z_G(\beta,\gamma,\lambda)}$, where $Z_G(\beta, \gamma, \lambda) = \sum_{\sigma\in\{0,1\}^V} \beta^{m_1(\sigma)}\gamma^{m_0(\sigma)}\lambda^{n_1(\sigma)}$ is the partition function.

There are two examples of particular interest: the hardcore model and the Ising model.

When $\beta = 0$ and $\gamma = 1$ then the only configurations with non-zero weight are independent sets of G and the weight of an independent set σ is $w(\sigma) = \lambda^{|\sigma|}$; this example is known as the *hardcore model* where the parameter λ corresponds to the fugacity.

In the case $\beta = \gamma$ then the important quantity is the total number of monochromatic edges $m(\sigma) = m_0(\sigma) + m_1(\sigma)$ and the weight of a configuration σ is $w(\sigma) = \beta^{m(\sigma)}\lambda^{n_1(\sigma)}$; this is the classical *Ising model* where the parameter β corresponds to the inverse temperature and λ is the external field ($\lambda = 1$ means no external field). Note, when $\beta > 1$ then the model is *ferromagnetic* as neighboring vertices prefer to have the same spin, and $\beta < 1$ is the *antiferromagnetic* Ising model. In the general 2-spin system, the model is ferromagnetic when $\beta\gamma > 1$ and antiferromagnetic when $\beta\gamma < 1$. (When $\beta\gamma = 1$ we get a trivial product distribution.)

It was long conjectured that the simple Glauber dynamics is rapidly mixing in the tree uniqueness region. This was recently proved by Anari, Liu, and Oveis Gharan [4] for the hardcore model; they proved, for all $\delta \in (0, 1)$, the mixing time is $n^{O(\exp(1/\delta))}$ whenever $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ where $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$ is the tree uniqueness/non-uniqueness phase transition threshold. We improve this result and establish optimal mixing time of the Glauber dynamics in the uniqueness region.

Theorem 1.1.1 (Hard-core Model)**.** *Let $\Delta \geq 3$ be an integer and let $\delta \in (0, 1)$ be a real. For every n-vertex graph G of maximum degree Δ and every $0 < \lambda \leq (1 - \delta)\lambda_c(\Delta)$, the mixing time of the Glauber dynamics for the hardcore model on G with fugacity λ is at most $Cn \log n$ where $C = C(\Delta, \delta)$ is a constant independent of n.*

The spectral independence bound we obtain here is better than the previous result [4]. Our improved result follows from a simpler, cleaner proof approach which enables us to extend our result to a wide variety of 2-spin models, matching the key results for the correlation decay algorithm with vastly improved running times.

Our proof approach unifies the three major algorithmic tools for approximate counting: correlation decay, polynomial interpolation, and MCMC. Most known results for both cor-

relation decay and polynomial interpolation approach are proved by showing contraction of a suitably defined potential function on the so-called tree recursions; the tree recursions arise as a result of Weitz's self-avoiding walk tree that we will describe in more detail later in this chapter. A recent work of Shao and Sun [118] unifies these two approaches by showing that the contraction which is normally used to prove efficiency of the correlation decay algorithm, also implies (under some additional analytic conditions) that the polynomial interpolation approach is efficient.

Here we prove that this same contraction of a potential function also implies optimal mixing time of the Glauber dynamics; see Definition 5.2.1 and Theorem 5.2.2 for a detailed statement. Our proof utilizes several new tools concerning Weitz's self-avoiding walk tree, which are detailed in Section 5.4. In particular, we show that the partition function of a graph G divides the partition function of Weitz's self-avoiding walk tree; see Lemma 5.4.1. This result is potentially of independent interest for establishing absence of zeros for the partition function with complex parameters, as it enables one to consider the self-avoiding walk tree. This result also yields a new, useful equivalence for bounding the influence in a graph in terms of the self-avoiding tree, which strengthens the previously known connection by Weitz [127]; see Lemma 5.4.1 for details.

As an easy consequence we obtain rapid mixing for the Glauber dynamics for the antiferromagnetic Ising model in the tree uniqueness region. In terms of the edge activity, the two critical points for the Ising model on the Δ-regular tree are at $\beta_c(\Delta) = \frac{\Delta-2}{\Delta}$ and $\overline{\beta}_c(\Delta) = \frac{1}{\beta_c(\Delta)} = \frac{\Delta}{\Delta-2}$; the first lies in the antiferromagnetic regime, while the second lies in the ferromagnetic regime. If $\beta_c(\Delta) < \beta < \overline{\beta}_c(\Delta)$, then uniqueness holds for all external field λ on the Δ-regular tree.

As mentioned earlier, for the ferromagnetic Ising model, an FPRAS was known for general graphs [77]. Furthermore, Mossel and Sly [106] proved $O(n \log n)$ mixing time of the Glauber dynamics for the ferromagnetic Ising model when $1 < \beta < \overline{\beta}_c(\Delta)$. However, rapid mixing for the antiferromagnetic Ising model in the tree uniqueness region was not

known.

We provide the following optimal mixing result for the case $\beta > \beta_c(\Delta)$. Note, when $\beta \leq \beta_c$ there is an additional uniqueness region for certain values of the external field λ; this region is covered by Theorem 1.1.3.

Theorem 1.1.2 (Ising model)**.** *Let $\Delta \geq 3$ be an integer and let $\delta \in (0,1)$ be a real. For every n-vertex graph G of maximum degree Δ, every $\beta \in [\frac{\Delta-2+\delta}{\Delta-\delta}, \frac{\Delta-\delta}{\Delta-2+\delta}]$, and every $\lambda > 0$, the mixing time of the Glauber dynamics for the Ising model on G with edge activity β and external field λ is at most $Cn \log n$ where $C = C(\Delta, \delta)$ is a constant independent of n.*

Our results for the hardcore and Ising models fit within a larger framework of general antiferromagnetic 2-spin systems. Recall that the antiferromagnetic case is when $\beta\gamma < 1$.

For general 2-spin systems the appropriate tree phase transition is more complicated as there are models where the tree uniqueness threshold is not monotone in Δ. Hence the appropriate notion is "up-to-Δ uniqueness" as considered by [89]. Roughly speaking, we say uniqueness with gap $\delta \in (0,1)$ holds on the d-regular tree if for every integer $\ell \geq 1$, all vertices at distance ℓ from the root have total "influence" $\lesssim (1-\delta)^\ell$ on the marginal of the root. We say up-to-Δ uniqueness with gap δ holds if uniqueness with gap δ holds on the d-regular tree for all $1 \leq d \leq \Delta$; see Section 5.3 for the precise definition.

Both Theorems 1.1.1 and 1.1.2 are corollaries of the following general optimal mixing result which holds for general antiferromagnetic 2-spin systems in the entire tree uniqueness region.

Theorem 1.1.3 (Antiferromagnetic 2-Spin Systems)**.** *Let $\Delta \geq 3$ be an integer and let $\delta \in (0,1)$ be a real. Let (β, γ, λ) with $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$ be parameters specifying an antiferromagnetic 2-spin system which is up-to-Δ unique with gap δ. For every n-vertex graph G of maximum degree Δ, the mixing time of the Glauber dynamics for the antiferromagnetic 2-spin system on G with parameters (β, γ, λ) is at most $Cn \log n$ where $C = C(\Delta, \delta, \beta, \gamma, \lambda)$ is a constant independent of n.*

We also match existing correlation decay results [67, 118] for ferromagnetic 2-spin models; see Section 5.10 for results and proofs.

5.2 Establishing Spectral Independence by the Potential Method

The tree recursion is very useful in the study of approximating counting. Consider a tree rooted at r. Suppose that r has d children, denoted by $v_1, \ldots, v_d$. For $1 \leq i \leq \Delta_i$ we define T_{v_i} to be the subtree of T rooted at v_i that contains all descendant of v_i. Let $R_r = \mu_T(\sigma_r = 1)/\mu_T(\sigma_r = 0)$ denote the marginal ratio of the root, and $R_{v_i} = \frac{\mu_{T_{v_i}}(\sigma_{v_i}=1)}{\mu_{T_{v_i}}(\sigma_{v_i}=0)}$ for each subtree. The *tree recursion* is a formula that computes R_r given $R_{v_1}, \ldots, R_{v_d}$, due to the independence of T_{v_i}'s. More specifically, we can write $R_r = F_d(R_{v_1}, \ldots, R_{v_d})$ where $F_d : [0, +\infty]^d \to [0, +\infty]$ is a multivariate function such that for $(x_1, \ldots, x_d) \in [0, \infty]^d$,

$$F_d(x_1, \ldots, x_d) = \lambda \prod_{i=1}^{d} \frac{\beta x_i + 1}{x_i + \gamma}.$$

In this chapter, however, we pay particular interest in the log of marginal ratios. The reason is that we will carefully study the *pairwise influence matrix* Ψ_G of the Gibbs distribution μ_G, introduced in [4] and defined as for every $r, v \in V$

$$\Psi_G(r \to v) = \mu_G(\sigma_v = 1 \mid \sigma_r = 1) - \mu_G(\sigma_v = 1 \mid \sigma_r = 0).$$

One crucial observation we make in this chapter is that the influence $\Psi_G(r \to v)$ of r on v can be viewed as the derivative of $\log R_r$ with respect to the log external field at v (see Lemma 5.5.3). Thus, it is more convenient for us to work with the log ratios. To this end, we rewrite the tree recursion as $\log R_v = H_d(\log R_{v_1}, \ldots, \log R_{v_d})$ where $H_d : [-\infty, +\infty]^d \to [-\infty, +\infty]$ is a function such that for $(y_1, \ldots, y_d) \in [-\infty, +\infty]^d$,

$$H_d(y_1, \ldots, y_d) = \log \lambda + \sum_{i=1}^{d} \log \left(\frac{\beta e^{y_i} + 1}{e^{y_i} + \gamma} \right).$$

Observe that $H = \log \circ F \circ \exp$. Moreover, we define

$$h(y) = -\frac{(1-\beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}$$

for $y \in [-\infty, +\infty]$, so that $\frac{\partial}{\partial y_i} H_d(y_1, \ldots, y_d) = h(y_i)$ for each i.

To prove our main results, we use the potential method, which has been widely used to establish the decay of correlation. By choosing a suitable potential function for the log ratios, we show that the total influence from a given vertex decays exponentially with the distance, and thus establish rapid mixing of the Glauber dynamics. Let us first specify our requirements on the potential. For every integer $d \geq 0$, we define a bounded interval J_d which contains all log ratios at a vertex of degree d. More specifically, we let $J_d = \big[\log(\lambda\beta^d), \log(\lambda/\gamma^d)\big]$ when $\beta\gamma < 1$, and $J_d = \big[\log(\lambda/\gamma^d), \log(\lambda\beta^d)\big]$ when $\beta\gamma > 1$. Furthermore, define $J = \bigcup_{d=0}^{\Delta-1} J_d$ to be the interval containing all log ratios with degree less than Δ.

Definition 5.2.1 ((α, c)-Potential function)**.** Let $\Delta \geq 3$ be an integer. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$ and $\lambda > 0$. Let $\Xi : [-\infty, +\infty] \to (-\infty, +\infty)$ be a differentiable and increasing function with image $S = \Xi[-\infty, +\infty]$ and derivative $\psi = \Xi'$. For any $\alpha \in (0, 1)$ and $c > 0$, we say Ξ is an (α, c)-*potential function* with respect to Δ and (β, γ, λ) if it satisfies the following conditions:

1. (Contraction) For every integer d such that $1 \leq d < \Delta$ and every $(\tilde{y}_1, \ldots, \tilde{y}_d) \in S^d$, we have
$$\big\|\nabla H_d^{\Xi}(\tilde{y}_1, \ldots, \tilde{y}_d)\big\|_1 = \sum_{i=1}^{d} \frac{\psi(y)}{\psi(y_i)} \cdot |h(y_i)| \leq 1 - \alpha$$
where $H_d^{\Xi} = \Xi \circ H_d \circ \Xi^{-1}$, $y_i = \Xi^{-1}(\tilde{y}_i)$ for $1 \leq i \leq d$, and $y = H_d(y_1, \ldots, y_d)$.

2. (Boundedness) For every $y_1, y_2 \in J$, we have
$$\frac{\psi(y_2)}{\psi(y_1)} \cdot |h(y_1)| \leq \frac{c}{\Delta}.$$

In the definition of (α, c)-potential, one should think of y as the log marginal ratio at a vertex and the potential function is of $\log R$. The following theorem establishes rapid mixing of the Glauber dynamics given an (α, c)-potential function.

Theorem 5.2.2. *Let $\Delta \geq 3$ be an integer. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$ and $\lambda > 0$. Suppose that there is an (α, c)-potential with respect to Δ and (β, γ, λ) for some $\alpha \in (0, 1)$ and $c > 0$. Then for every n-vertex graph G of maximum degree Δ, the Gibbs distribution of the 2-spin system on G with parameters (β, γ, λ) is spectrally independent with constant*

$$\eta = \frac{c}{\alpha}.$$

We outline our proofs in Section 5.4. Note that in both Definition 5.2.1 and Theorem 5.2.2, the constant c is allowed to depend on the maximum degree Δ and parameters (β, γ, λ) in general. For example, a straightforward black-box application of the potential in [89] would give $c = \Theta(\Delta)$ for the Boundedness condition. However, this is undesirable for graphs with potentially unbounded degrees. One of our contributions is that we show the Boundedness condition holds for a universal constant c *independent* of Δ and (β, γ, λ).

In Section 5.8, we give a slightly more general definition of (α, c)-potentials, which relaxes the Boundedness condition, and is necessary for our analysis of antiferromagnetic 2-spin systems with $0 \leq \beta < 1 < \gamma$. Theorem 5.2.2 still holds for this larger class of potentials.

We remark that in all previous works of the potential method, results and proofs are always presented in terms of F_d, the tree recursion of R, and Φ, a potential function of R. In fact, our results can also be translated into the language of (F_d, Φ). To see this, since $H_d = \log \circ F_d \circ \exp$, it is straightforward to check that $H_d^{\Xi} = \Xi \circ H_d \circ \Xi^{-1} = \Phi \circ F_d \circ \Phi^{-1} = F_d^{\Phi}$ if we pick $\Phi = \Xi \circ \log$, and thereby $\nabla H_d^{\Xi} = \nabla F_d^{\Phi}$. This implies that the Contraction condition in Definition 5.2.1 holds for (H_d, Ξ) if and only if the corresponding contraction condition holds for (F_d, Φ). The Boundedness condition can also be stated equivalently for

(F_d, Φ). Nevertheless, in this chapter we choose to work with (H_d, Ξ) for the following two reasons. First, as mentioned earlier, the fact that $\Psi_G(r \to v)$ is a derivative of $\log R_r$ makes it natural to consider the tree recursion for the log ratios. Indeed, it is easier and cleaner to present our results and proofs using (H_d, Ξ) directly rather than switching to (F_d, Φ). Second, the potential function Ξ we will use is obtained from the exact potential Φ in [89], by the transformation $\Xi = \Phi \circ \exp$. (To be more precise, we also multiply a constant factor which only simplifies our calculation and does not matter much; also notice that [89] denotes the potential function by φ and its derivative by $\Phi = \varphi'$.) It is intriguing to notice that the derivative of this potential is simply $\psi = \sqrt{|h|}$. Then the Contraction condition has a nice form: $\sum_{i=1}^{d} \sqrt{h(y)h(y_i)} \leq 1 - \alpha$; and the Boundedness condition only involves an upper bound on $h(y)$. This seems to shed some light on the mysterious potential function Φ from [89], and also indicates that H_d is a meaningful variant of the tree recursion to consider. To add one more evidence, for a lot of cases (e.g., $\frac{\Delta-2}{\Delta} < \sqrt{\beta\gamma} < \frac{\Delta}{\Delta-2}$) where the potential $\Phi = \log$ is picked, that just means we can pick Ξ to be the identity function and H_d itself is contracting without any nontrivial potential.

5.3 Preliminaries for 2-Spin Systems

Here we give relevant definitions for 2-spin systems that are used in this chapter.

Uniqueness Let $\Delta \geq 3$ be an integer or $\Delta = \infty$. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$. For $1 \leq d < \Delta$, define

$$f_d(R) = \lambda \left(\frac{\beta R + 1}{R + \gamma} \right)^d$$

and denote the unique fixed point of f_d by R_d^*. For $\delta \in (0, 1)$, we say the parameters (β, γ, λ) are *up-to-Δ unique with gap δ* if $|f_d'(R_d^*)| < 1 - \delta$ for all $1 \leq d < \Delta$.

Ratio and Influence Consider the 2-spin system on a graph $G = (V, E)$. Let $\Lambda \subseteq V$ and $\sigma_\Lambda \in \{0,1\}^\Lambda$. For all $v \in V \backslash \Lambda$, we define the *marginal ratio* at v to be

$$R_G^{\sigma_\Lambda}(v) = \frac{\mu_G(\sigma_v = 1 \mid \sigma_\Lambda)}{\mu_G(\sigma_v = 0 \mid \sigma_\Lambda)}.$$

For all $u, v \in V \backslash \Lambda$, recall that we define the *(pairwise) influence* of u on v by

$$\Psi_G^{\sigma_\Lambda}(u \to v) = \mu_G(\sigma_v = 1 \mid \sigma_u = 1,\, \sigma_\Lambda) - \mu_G(\sigma_v = 1 \mid \sigma_u = 0,\, \sigma_\Lambda).$$

Write $\Psi_G^{\sigma_\Lambda}$ for the *(pairwise) influence matrix* whose entries are given by $\Psi_G^{\sigma_\Lambda}(u \to v)$.

Weitz's Self-Avoiding Walk Tree Let $G = (V, E)$ be a connected graph and $r \in V$ be a vertex of G. The *self-avoiding walk (SAW) tree* is defined as follows. Suppose that there is a total ordering of the vertex set V. A self-avoiding walk from r is a path $r = v_0 - v_1 - \cdots - v_\ell$ such that $v_i \neq v_j$ for all $0 \leq i < j \leq \ell$. The SAW tree $T_{\text{SAW}}(G, r)$ is a tree rooted at r, consisting of all self-avoiding walks $r = v_0 - v_1 - \cdots - v_\ell$ with $\deg(v_\ell) = 1$, and those appended with one more vertex that closes the cycle (i.e., $r = v_0 - v_1 - \cdots - v_\ell - v_i$ for some $0 \leq i \leq \ell - 2$ such that $\{v_\ell, v_i\} \in E$). Note that a vertex of G might have many copies in the SAW tree, and the degrees of vertices are preserved except for leaves. See Fig. 5.1 for an example.

We can define a 2-spin system on $T_{\text{SAW}}(G, r)$ with the same parameters (β, γ, λ), in which some of the leaves are fixed to a particular spin. More specifically, for a self-avoiding walk $r = v_0 - v_1 - \cdots - v_\ell$ appended with v_i, we fix v_i to be spin 1 if $v_{i+1} < v_\ell$ with respect to the total ordering on V, and spin 0 if $v_{i+1} > v_\ell$. For each $v \in V$ we denote the set of all free (unfixed) copies of v in $T_{\text{SAW}}(G, r)$ by $\mathcal{C}_v$. For $\Lambda \subseteq V$ and a partial configuration $\sigma_\Lambda \in \{0,1\}^\Lambda$, we define the SAW tree with conditioning σ_Λ by assigning the spin σ_v to every copy $\hat{v}$ of v from $\mathcal{C}_v$ and removing all descendants of $\hat{v}$, for each $v \in \Lambda$. Note that in general, different copies of v from $\mathcal{C}_v$ can receive different spin assignments. Finally, in the

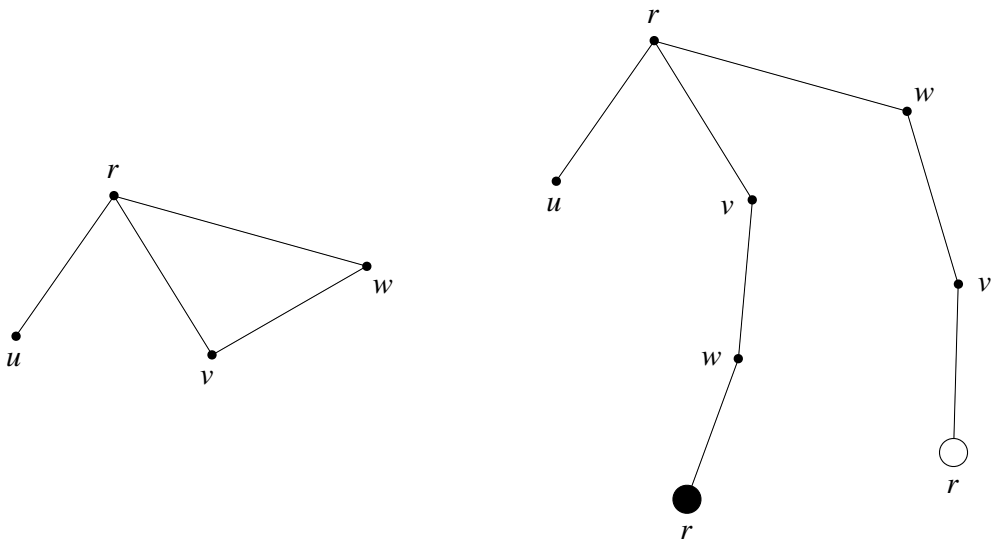

Figure 5.1: A graph G and the self-avoiding walk tree $T_{\text{SAW}}(G, r)$ rooted at r. Vertices with the same label in $T_{\text{SAW}}(G, r)$ are copies of the same vertex from G. (●/○: fixed to spin 1/0.)

case that every vertex v has a distinct field λ_v, all copies of v from $\mathcal{C}_v$ will have the same field λ_v in the SAW tree.

5.4 Proof Outline

In this section, we give an overview of our proof approach.

Self-Avoiding Walk Trees Preserve Influences From standard linear algebra, we know that the maximum eigenvalue of $\Psi_G^{\sigma_\Lambda}$ is upper bounded by both the 1-norm $\|\Psi_G^{\sigma_\Lambda}\|_1 = \max_{r \in V} \sum_{v \in V} |\Psi_G^{\sigma_\Lambda}(v \to r)|$, which corresponds to total influences on a vertex r, and the infinity-norm $\|\Psi_G^{\sigma_\Lambda}\|_\infty = \max_{r \in V} \sum_{v \in V} |\Psi_G^{\sigma_\Lambda}(r \to v)|$, corresponding to total influences of r. In [4] the authors use $\|\Psi_G^{\sigma_\Lambda}\|_1$ as an upper bound on $\lambda_{\max}(\Psi_G^{\sigma_\Lambda})$. Roughly speaking, they show that the sum of absolute influences on a fixed vertex r, is upper bounded by the maximum absolute influences on r in the self-avoiding walk tree rooted at r, over all boundary conditions. Here in this chapter, we will use $\|\Psi_G^{\sigma_\Lambda}\|_\infty$ to upper bound $\lambda_{\max}(\Psi_G^{\sigma_\Lambda})$ instead. In fact, much more is true if we look at the influences from r in the self-avoiding

tree. We show that for every vertex $v \in V$, the influence $\Psi_G^{\sigma_\Lambda}(r \to v)$ in G is preserved in the self-avoiding walk tree $T = T_{\text{SAW}}(G, r)$ rooted at r, in the form of sum of influences $\Psi_T^{\sigma_\Lambda}(r \to \hat{v})$ over all copies $\hat{v}$ of v.

The way we establish this fact is by viewing the partition function as a polynomial in λ. In fact, it will be useful to consider the more general case with an arbitrary external field λ_v for every $v \in V$. Let $\boldsymbol{\lambda} = \{\lambda_v : v \in V\}$ denote the fields. For $\Lambda \subseteq V$ and $\sigma_\Lambda \in \{0,1\}^\Lambda$, the weight of $\sigma \in \{0,1\}^{V\setminus\Lambda}$ conditional on σ_Λ is defined to be $w_G(\sigma \mid \sigma_\Lambda) = \beta^{m_1(\sigma|\sigma_\Lambda)}\gamma^{m_0(\sigma|\sigma_\Lambda)}\prod_{v\in V\setminus\Lambda}\lambda_v^{\sigma_v}$ where $m_i(\cdot \mid \sigma_\Lambda)$ is the number of i-i edges with at least one endpoint in $V\setminus\Lambda$ for $i = 0, 1$. Furthermore, $Z_G^{\sigma_\Lambda} = \sum_{\sigma\in\{0,1\}^{V\setminus\Lambda}} w_G(\sigma \mid \sigma_\Lambda)$ is the partition function conditioned on σ_Λ. We shall view β and γ as some fixed constants and think of $\boldsymbol{\lambda}$ as $n = |V|$ variables. In this sense, we regard the weights $w_G(\sigma \mid \sigma_\Lambda)$ as monomials in $\boldsymbol{\lambda}$ and the partition function $Z_G^{\sigma_\Lambda}$ as a polynomial in $\boldsymbol{\lambda}$. Moreover, the marginal ratios $R_G^{\sigma_\Lambda}(v)$ and the influences $\Psi_G^{\sigma_\Lambda}(r \to v)$ for $r, v \in V$ are all functions in $\boldsymbol{\lambda}$. Our main result is that the partition function of G divides that of $T_{\text{SAW}}(G, r)$ for each $r \in V$. From that, we show that the SAW tree preserves influences of the root, as well as re-establishing Weitz's celebrated result [127], see Lemma 5.5.4.

Lemma 5.4.1. *Let $G = (V, E)$ be a connected graph, $r \in V$ be a vertex and $\Lambda \subseteq V\setminus\{r\}$ such that $G\setminus\Lambda$ is connected. Let $T = T_{\text{SAW}}(G, r)$ be the self-avoiding walk tree of G rooted at r. Then for every $\sigma_\Lambda \in \{0,1\}^\Lambda$, $Z_G^{\sigma_\Lambda}$ divides $Z_T^{\sigma_\Lambda}$. More precisely, there exists a polynomial $P_{G,r}^{\sigma_\Lambda} = P_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$ independent of λ_r such that*

$$Z_T^{\sigma_\Lambda} = Z_G^{\sigma_\Lambda} \cdot P_{G,r}^{\sigma_\Lambda}. \tag{5.1}$$

As a corollary, for each vertex $v \in V$,

$$\mathcal{I}_G^{\sigma_\Lambda}(r \to v) = \sum_{\hat{v}\in\mathcal{C}_v} \mathcal{I}_T^{\sigma_\Lambda}(r \to \hat{v}), \tag{5.2}$$

where $\mathcal{C}_v$ is the set of all free (unfixed) copies of v in T.

Remark 5.4.2. We emphasize that for the purposes of bounding the total influence of a vertex in G, only Eq. (5.2) of Lemma 5.4.1 is needed, which can be proved in a purely combinatorial fashion. However, we believe the divisibility property Eq. (5.1) of the multivariate partition function of G and its self-avoiding walk tree may be of independent interest.

We note that a univariate version of the divisibility statement Eq. (5.1) has already appeared in [20] for the hardcore model and [93] for the zero-field Ising model in the study of complex roots of the partition function. From Lemma 5.4.1, we can get

$$\sum_{v \in V} |\Psi_G^{\sigma_\Lambda}(r \to v)| \leq \sum_{v \in V_T} |\Psi_T^{\sigma_\Lambda}(r \to v)|$$

for any fixed r. That means, we only need to upper bound the sum of all influences for trees, in order to get an upper bound on $\lambda_{\max}(\Psi_G^{\sigma_\Lambda})$.

Decay of Influences Given a Good Potential The tree recursion provides us a great tool for computing the (log) ratios of vertices recursively for trees. As we show in Lemma 5.5.3, the influence $\Psi_G^{\sigma_\Lambda}(r \to v)$ is in fact a version of derivative of the log marginal ratio at r. Thus, the tree recursion can be used naturally to relate these influences. We then apply the potential method, which has been widely used in literature to establish the decay of correlations (strong spatial mixing). The following lemma shows that the sum of absolute influences to distance k has exponential decay with k, which can be thought of as the decay of pairwise influences.

Lemma 5.4.3. *If there exists an (α, c)-potential function Ξ with respect to Δ and (β, γ, λ) where $\alpha \in (0, 1)$ and $c > 0$, then for every $\Lambda \subseteq V_T \backslash \{r\}$, $\sigma_\Lambda \in \{0, 1\}^\Lambda$ and all integers $k \geq 1$,*

$$\sum_{v \in L_r(k)} |\mathcal{I}_T^{\sigma_\Lambda}(r \to v)| \leq c \cdot (1 - \alpha)^{k-1}$$

where $L_r(k)$ denote the set of all free vertices at distance k away from r.

Theorem 5.2.2 is then proved by combining Lemma 5.4.1 and Lemma 5.4.3. We leave

its proof to Section 5.11.

Find a Good Potential As our final step, we need to find an (α, c)-potential function as defined in Definition 5.2.1. The potential Ξ we choose is exactly the one from [89], adapted to the log marginal ratios and the tree recursion H (see Section 5.7 for more details). We show that if the parameters (β, γ, λ) are up-to-Δ unique with gap $\delta \in (0, 1)$ and either $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$ or $\gamma \leq 1$, then Ξ is an (α, c)-potential.

Lemma 5.4.4. *Let $\Delta \geq 3$ be an integer. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$. Assume that (β, γ, λ) is up-to-Δ unique with gap $\delta \in (0, 1)$. Define the function Ξ implicitly by*

$$\Xi'(y) = \psi(y) = \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}} = \sqrt{|h(y)|}, \qquad \Xi(0) = 0. \tag{5.3}$$

If $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$, then Ξ is an (α, c)-potential function with $\alpha \geq \delta/2$ and $c \leq 1.5$. If $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma \leq 1$, then Ξ is an (α, c)-potential with $\alpha \geq \delta/2$ and $c \leq 18$; we can further take $c \leq 4$ if $\beta = 0$.

We deduce Theorem 1.1.3 for the case $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$ or $\gamma \leq 1$ from Theorem 5.2.2 and Lemma 5.4.4. The proof of it can be found in Section 5.11. The case that $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$ is trickier. As discussed in Section 5 of [89], when $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$, for some $\lambda > 0$ the spin system lies in the uniqueness region for arbitrary graphs, even with unbounded degrees (i.e., up-to-∞ unique). Thus, in this case the total influences of a vertex can be as large as $\Theta(\Delta/\delta)$, resulting in $n^{\Theta(\Delta/\delta)}$ mixing time. To deal with this, we consider a suitably weighted sum of absolute influences of a fixed vertex, which also upper bounds the maximum eigenvalue of the influence matrix. Definition 5.2.1 and Theorem 5.2.2 are then modified to a slightly stronger version. The statements and proofs for this case are presented in Section 5.8.

The rest of the chapter is organized as follows. In Section 5.5 we prove Lemma 5.4.1

about properties of the SAW tree. In Section 5.6 we establish Lemma 5.4.3 regarding the decay of influences by the potential method. We verify the Contraction condition in Section 5.7 for our choice of potential. Section 5.8 is devoted to the case that $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$, where a more general version of Definition 5.2.1 and Theorem 5.2.2 is required. In Section 5.9 we verify the Boundedness condition and its generalization for our potential in all cases. We consider ferromagnetic spin systems in Section 5.10. Finally, we prove all of our main results in Section 5.11.

5.5 Preservation of Influences for Self-Avoiding Walk Trees

In this section we show that the self-avoiding walk (SAW) tree, introduced in [127] (see also [117]), maintains all the influence of the root, and thus establishes Lemma 5.4.1. To do this, we show that the partition function of G, viewed as a polynomial of the external fields $\boldsymbol{\lambda}$, divides that of the SAW tree. From there we prove that the influence of the root vertex r on another vertex v in G, is exactly equal to that on all copies of v in the SAW tree. Using our proof approach, we show that the marginal of the root is maintained in the SAW tree, re-establishing Weitz's celebrated result [127], and also all pairwise covariances concerned with v are preserved.

Theorem 5.5.1. *Let $G = (V, E)$ be a connected graph, $r \in V$ be a vertex and $\Lambda \subseteq V \backslash \{r\}$ such that $G \backslash \Lambda$ is connected. Let $T = T_{\text{SAW}}(G, r)$ be the self-avoiding walk tree of G rooted at r. Then for every $\sigma_\Lambda \in \{0,1\}^\Lambda$, $Z_G^{\sigma_\Lambda}$ divides $Z_T^{\sigma_\Lambda}$. More precisely, there exists a polynomial $P_{G,r}^{\sigma_\Lambda} = P_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$ such that*

$$Z_T^{\sigma_\Lambda} = Z_G^{\sigma_\Lambda} \cdot P_{G,r}^{\sigma_\Lambda}.$$

Moreover, the polynomial $P_{G,r}^{\sigma_\Lambda}$ is independent of λ_r.

Remark 5.5.2. The proof of Theorem 5.5.1 can be adapted to give a purely combinatorial proof of Eq. (5.2) in Lemma 5.4.1. Like in the proof of [127, Theorem 3.1], one can

proceed via vertex splitting and telescoping, where instead of telescoping a product of marginal ratios, one instead telescopes a sum of single-vertex influences.

We remark that [20] proved a univariate version of Theorem 5.5.1 for the hardcore model, and [93] showed a similar result for the zero-field Ising model with a uniform edge weight. Our result holds for all 2-spin systems and arbitrary fields for each vertex. We can also generalize it to arbitrary edge weights for each edge in a straightforward fashion. It is crucial that the quotient polynomial $P_{G,r}^{\sigma_\Lambda}$ is independent of the field λ_r at the root, from which we can deduce the preservation of marginal and influences of the root immediately.

Before proving Theorem 5.5.1, we first give a few consequences of it. For all $u, v \in V \backslash \Lambda$, we define the *marginal* at v as $M_G^{\sigma_\Lambda}(v) = \mu_G(v = 1 \mid \sigma_\Lambda)$ (henceforth we write $v = i$ for the event $\sigma_v = i$ for convenience), and the *covariance* of u and v as

$$K_G^{\sigma_\Lambda}(u, v) = \mu_G(u = v = 1 \mid \sigma_\Lambda) - \mu_G(u = 1 \mid \sigma_\Lambda)\mu_G(v = 1 \mid \sigma_\Lambda).$$

The following lemma relates the quantities we are interested in with appropriate derivatives of the (log) partition function. Parts 1 and 2 of the lemma are folklore.

Lemma 5.5.3. *For every graph $G = (V, E)$, $\Lambda \subseteq V$ and $\sigma_\Lambda \in \{0, 1\}^\Lambda$, the following holds:*

1. *For all $v \in V$,*
$$\left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \log Z_G^{\sigma_\Lambda} = M_G^{\sigma_\Lambda}(v);$$

2. *For all $u, v \in V$,*
$$\left(\lambda_v \frac{\partial}{\partial \lambda_v}\right)\left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) \log Z_G^{\sigma_\Lambda} = \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) M_G^{\sigma_\Lambda}(u) = K_G^{\sigma_\Lambda}(u, v);$$

3. *For all $u, v \in V$,*
$$\left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \log R_G^{\sigma_\Lambda}(u) = \Psi_G^{\sigma_\Lambda}(u \to v).$$

Proof. The first two parts are standard. We include the proofs of these two facts for completeness. To see the first equality, we compute directly and get

$$\begin{aligned}
\left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \log Z_G^{\sigma_\Lambda} &= \frac{1}{Z_G^{\sigma_\Lambda}} \cdot \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) Z_G^{\sigma_\Lambda} \\
&= \frac{1}{Z_G^{\sigma_\Lambda}} \sum_{\sigma \in \{0,1\}^{V \setminus \Lambda}} \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \left(\beta^{m_1(\sigma)} \gamma^{m_0(\sigma)} \prod_{w \in V} \lambda_w^{\sigma_w}\right) \\
&= \frac{1}{Z_G^{\sigma_\Lambda}} \sum_{\sigma \in \{0,1\}^{V \setminus \Lambda}} \sigma_v \left(\beta^{m_1(\sigma)} \gamma^{m_0(\sigma)} \prod_{w \in V} \lambda_w^{\sigma_w}\right) \\
&= \sum_{\sigma \in \{0,1\}^{V \setminus \Lambda}} \sigma_v \cdot \mu_G(\sigma \mid \sigma_\Lambda) = M_G^{\sigma_\Lambda}(v).
\end{aligned}$$

For Part 2, using the result above, we can also get

$$\begin{aligned}
&\left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) \log Z_G^{\sigma_\Lambda} \\
&= \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \left(\frac{1}{Z_G^{\sigma_\Lambda}} \cdot \left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) Z_G^{\sigma_\Lambda}\right) \\
&= \frac{1}{Z_G^{\sigma_\Lambda}} \cdot \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) Z_G^{\sigma_\Lambda} - \frac{1}{(Z_G^{\sigma_\Lambda})^2} \cdot \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) Z_G^{\sigma_\Lambda} \cdot \left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) Z_G^{\sigma_\Lambda} \\
&= \frac{1}{Z_G^{\sigma_\Lambda}} \cdot \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \left(\sum_{\sigma \in \{0,1\}^{V \setminus \Lambda}} \sigma_u \left(\beta^{m_1(\sigma)} \gamma^{m_0(\sigma)} \prod_{w \in V} \lambda_w^{\sigma_w}\right)\right) - M_G^{\sigma_\Lambda}(u) \cdot M_G^{\sigma_\Lambda}(v) \\
&= \frac{1}{Z_G^{\sigma_\Lambda}} \sum_{\sigma \in \{0,1\}^{V \setminus \Lambda}} \sigma_u \cdot \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \left(\beta^{m_1(\sigma)} \gamma^{m_0(\sigma)} \prod_{w \in V} \lambda_w^{\sigma_w}\right) - M_G^{\sigma_\Lambda}(u) \cdot M_G^{\sigma_\Lambda}(v) \\
&= \frac{1}{Z_G^{\sigma_\Lambda}} \sum_{\sigma \in \{0,1\}^{V \setminus \Lambda}} \sigma_u \cdot \sigma_v \left(\beta^{m_1(\sigma)} \gamma^{m_0(\sigma)} \prod_{w \in V} \lambda_w^{\sigma_w}\right) - M_G^{\sigma_\Lambda}(u) \cdot M_G^{\sigma_\Lambda}(v) \\
&= \sum_{\sigma \in \{0,1\}^{V \setminus \Lambda}} \sigma_u \cdot \sigma_v \cdot \mu_G(\sigma \mid \sigma_\Lambda) - M_G^{\sigma_\Lambda}(u) \cdot M_G^{\sigma_\Lambda}(v) \\
&= K_G^{\sigma_\Lambda}(u, v).
\end{aligned}$$

□

For Part 3, we deduce from Part 2 that

$$\begin{aligned}\left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \log R_G^{\sigma_\Lambda}(u) &= \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) \log \left(\frac{M_G^{\sigma_\Lambda}(u)}{1 - M_G^{\sigma_\Lambda}(u)}\right) \\ &= \frac{\left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) M_G^{\sigma_\Lambda}(u)}{M_G^{\sigma_\Lambda}(u)\left(1 - M_G^{\sigma_\Lambda}(u)\right)} \\ &= \frac{K_G^{\sigma_\Lambda}(u,v)}{K_G^{\sigma_\Lambda}(u,u)}.\end{aligned}$$

It remains to show that

$$\Psi_G^{\sigma_\Lambda}(u \to v) = \frac{K_G^{\sigma_\Lambda}(u,v)}{K_G^{\sigma_\Lambda}(u,u)},$$

which actually holds for any two binary random variables. To see this, we first compute $K_G^{\sigma_\Lambda}(u,u) \cdot \Psi_G^{\sigma_\Lambda}(u \to v)$ by definition:

$$\begin{aligned}&K_G^{\sigma_\Lambda}(u,u) \cdot \Psi_G^{\sigma_\Lambda}(u \to v) \\ &= \mu_G(u=1 \mid \sigma_\Lambda) \cdot \mu_G(u=0 \mid \sigma_\Lambda) \cdot [\mu_G(v=1 \mid u=1,\, \sigma_\Lambda) - \mu_G(v=1 \mid u=0,\, \sigma_\Lambda)] \\ &= \mu_G(u=1, v=1 \mid \sigma_\Lambda) \cdot \mu_G(u=0 \mid \sigma_\Lambda) - \mu_G(u=1 \mid \sigma_\Lambda) \cdot \mu_G(u=0, v=1 \mid \sigma_\Lambda) \\ &= \mu_G(u=1, v=1 \mid \sigma_\Lambda) \cdot \mu_G(u=0, v=0 \mid \sigma_\Lambda) \\ &\qquad - \mu_G(u=1, v=0 \mid \sigma_\Lambda) \cdot \mu_G(u=0, v=1 \mid \sigma_\Lambda).\end{aligned}$$

Meanwhile, the covariance can be written as

$$\begin{aligned}&K_G^{\sigma_\Lambda}(u,v) = \mu_G(u=1, v=1 \mid \sigma_\Lambda) - \mu_G(u=1 \mid \sigma_\Lambda) \cdot \mu_G(v=1 \mid \sigma_\Lambda) \\ &= \mu_G(u=1, v=1 \mid \sigma_\Lambda) \cdot \mu_G(u=0, v=0 \mid \sigma_\Lambda) \\ &\qquad - \mu_G(u=1, v=0 \mid \sigma_\Lambda) \cdot \mu_G(u=0, v=1 \mid \sigma_\Lambda).\end{aligned}$$

This shows that $\Psi_G^{\sigma_\Lambda}(u \to v) = K_G^{\sigma_\Lambda}(u,v)/K_G^{\sigma_\Lambda}(u,u)$ and thus establishes Part 3.

We deduce Lemma 5.4.1 from Theorem 5.5.1 and the second item of the following lemma. The proof of Theorem 5.5.1 will be presented soon.

Lemma 5.5.4. *Let $G = (V, E)$ be a connected graph, $r \in V$ be a vertex and $\Lambda \subseteq V \backslash \{r\}$ such that $G \backslash \Lambda$ is connected. Let $T = T_{\text{SAW}}(G, r)$ be the self-avoiding walk tree of G rooted at r. Then for every $\sigma_\Lambda \in \{0,1\}^\Lambda$ we have:*

1. *([127, Theorem 3.1]) Preservation of marginal of the root r:*

$$M_G^{\sigma_\Lambda}(r) = M_T^{\sigma_\Lambda}(r) \qquad \textit{and} \qquad R_G^{\sigma_\Lambda}(r) = R_T^{\sigma_\Lambda}(r);$$

2. *Preservation of covariances and influences of r: for every $v \in V$,*

$$K_G^{\sigma_\Lambda}(r, v) = \sum_{\hat{v} \in \mathcal{C}_v} K_T^{\sigma_\Lambda}(r, \hat{v}) \qquad \textit{and} \qquad \Psi_G^{\sigma_\Lambda}(r \to v) = \sum_{\hat{v} \in \mathcal{C}_v} \Psi_T^{\sigma_\Lambda}(r \to \hat{v}).$$

where $\mathcal{C}_v$ is the set of all free (unfixed) copies of v in T.

Proof. By Theorem 5.5.1, there exists a polynomial $P_{G,r}^{\sigma_\Lambda} = P_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$ such that $Z_T^{\sigma_\Lambda} = Z_G^{\sigma_\Lambda} \cdot P_{G,r}^{\sigma_\Lambda}$ and $P_{G,r}^{\sigma_\Lambda}$ is independent of λ_r. Then it follows from Lemma 5.5.3 that

$$\begin{aligned} M_T^{\sigma_\Lambda}(r) &= \left(\lambda_r \frac{\partial}{\partial \lambda_r}\right) \log Z_T^{\sigma_\Lambda} \\ &= \left(\lambda_r \frac{\partial}{\partial \lambda_r}\right) \left(\log Z_G^{\sigma_\Lambda} + \log P_{G,r}^{\sigma_\Lambda}\right) \\ &= \left(\lambda_r \frac{\partial}{\partial \lambda_r}\right) \log Z_G^{\sigma_\Lambda} \\ &= M_G^{\sigma_\Lambda}(r), \end{aligned}$$

and therefore $R_T^{\sigma_\Lambda}(r) = R_G^{\sigma_\Lambda}(r)$. For the second item, again from Lemma 5.5.3 we get

$$K_G^{\sigma_\Lambda}(r, v) = \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) M_G^{\sigma_\Lambda}(r) = \left(\lambda_v \frac{\partial}{\partial \lambda_v}\right) M_T^{\sigma_\Lambda}(r).$$

Recall that for the spin system on the SAW tree T, every free copy $\hat{v}$ of v from $\mathcal{C}_v$ has the same external field $\lambda_{\hat{v}} = \lambda_v$. Then, by the chain rule of derivatives and Lemma 5.5.3, we

deduce that

$$K_G^{\sigma_\Lambda}(r,v) = \sum_{\hat{v}\in\mathcal{C}_v} \left(\lambda_{\hat{v}}\frac{\partial}{\partial\lambda_{\hat{v}}}\right) M_T^{\sigma_\Lambda}(r)\cdot\frac{\partial\lambda_{\hat{v}}}{\partial\lambda_v}\cdot\frac{\lambda_v}{\lambda_{\hat{v}}} = \sum_{\hat{v}\in\mathcal{C}_v} K_T^{\sigma_\Lambda}(r,\hat{v}).$$

Finally, we have

$$\Psi_G^{\sigma_\Lambda}(r\to v) = \left(\lambda_v\frac{\partial}{\partial\lambda_v}\right)\log R_G^{\sigma_\Lambda}(r) = \left(\lambda_v\frac{\partial}{\partial\lambda_v}\right)\log R_T^{\sigma_\Lambda}(r) = \sum_{\hat{v}\in\mathcal{C}_v}\Psi_T^{\sigma_\Lambda}(r\to\hat{v}),$$

where the last equality follows as above. □

We finish this section with the proof of Theorem 5.5.1. Before presenting our proof, let us first review the notations and definitions introduced earlier. Denote the set of fields at all vertices by $\boldsymbol{\lambda} = \{\lambda_v : v\in V\}$. For $\Lambda\subseteq V$ and $\sigma_\Lambda\in\{0,1\}^\Lambda$, the weight of $\sigma\in\{0,1\}^{V\setminus\Lambda}$ conditional on σ_Λ is given by

$$w_G(\sigma\mid\sigma_\Lambda) = \beta^{m_1(\sigma|\sigma_\Lambda)}\gamma^{m_0(\sigma|\sigma_\Lambda)}\prod_{v\in V\setminus\Lambda}\lambda_v^{\sigma_v},$$

where for $i=0,1$, $m_i(\cdot\mid\sigma_\Lambda)$ denotes the number of edges such that both endpoints receive the spin i and at least one of them is in $V\setminus\Lambda$. The partition function conditional on σ_Λ is defined as $Z_G^{\sigma_\Lambda} = \sum_{\sigma\in\{0,1\}^{V\setminus\Lambda}} w_G(\sigma\mid\sigma_\Lambda)$. For the SAW tree, we define the conditional weights and partition function in the same way. In particular, recall that when we fix a conditioning σ_Λ on the SAW tree, we also remove all descendants of $\hat{v}\in\mathcal{C}_v$ for each $v\in\Lambda$.

For every $v\in V\setminus\Lambda$ and $i\in\{0,1\}$, we shall write $v=i$ to represent the set of configurations such that $\sigma_v = i$ (i.e., $\{\sigma\in\{0,1\}^{V\setminus\Lambda} : \sigma_v = i\}$) and let $Z_G^{\sigma_\Lambda}(v=i)$ be sum of weights of all configurations with $v=i$. We further extend this notation and write $Z_G^{\sigma_\Lambda}(U=\sigma_U)$ for every $U\subseteq V\setminus\Lambda$ and $\sigma_U\in\{0,1\}^U$. For the SAW tree we adopt the same notations as well.

Proof of Theorem 5.5.1. We will show that there exists a polynomial $P_{G,r}^{\sigma_\Lambda} = P_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$, independent of λ_r, such that

$$Z_T^{\sigma_\Lambda}(r=1) = Z_G^{\sigma_\Lambda}(r=1) \cdot P_{G,r}^{\sigma_\Lambda} \quad \text{and} \quad Z_T^{\sigma_\Lambda}(r=0) = Z_G^{\sigma_\Lambda}(r=0) \cdot P_{G,r}^{\sigma_\Lambda}. \tag{5.4}$$

The high-level proof idea of Eq. (5.4) is similar to the corresponding result in [127, Theorem 3.1]. Let m be the number of edges with at least one endpoint in $V \backslash \Lambda$. We use induction on m. When $m = 0$ the statement is trivial since $T = G$. Assume that Eq. (5.4) holds for all graphs and all conditioning with less than m edges. Suppose that the root r has d neighbors $v_1, \ldots, v_d$. Define G' to be the graph obtained by replacing the vertex r with d vertices $r_1, \ldots, r_d$ and then connecting $\{r_i, d_i\}$ for $1 \le i \le d$.

Consider first the case where $(G \backslash \{r\}) \backslash \Lambda$ is still connected. For each i, let $G_i = G' - r_i$. Define the 2-spin system on G_i with the same parameters $(\beta, \gamma, \boldsymbol{\lambda})$, plus an additional conditioning that the vertices $r_1, \ldots, r_{i-1}$ are fixed to spin 0 while $r_{i+1}, \ldots, r_d$ are fixed to spin 1; we denote this conditioning by σ_{U_i} with $U_i = \{v_1, \ldots, v_d\} \backslash \{v_i\}$. Then, $T = T_{\text{SAW}}(G, r)$ can be generated by the following recursive procedure. Also see Fig. 5.2 for an illustration.

Algorithm: $T_{\text{SAW}}(G, r)$

1. For each i, let $T_i = T_{\text{SAW}}(G_i, v_i)$ plus the conditioning σ_{U_i};
2. Let $T = T_{\text{SAW}}(G, r)$ be the union of r and $T_1, \ldots, T_d$ by connecting $\{r, v_i\}$ for $1 \le i \le d$; output T.

For the purpose of proof, we also consider the 2-spin system on G' with the same parameters $(\beta, \gamma, \boldsymbol{\lambda})$, with an exception that we let the vertices $r_1, \ldots, r_d$ have no fields (i.e., setting $\lambda_{r_i} = 1$ for $1 \le i \le d$ instead of λ_r). We then observe that

$$Z_G^{\sigma_\Lambda}(r=1) = \lambda_r \cdot Z_{G'}^{\sigma_\Lambda}(r_1 = 1, \ldots, r_d = 1),$$

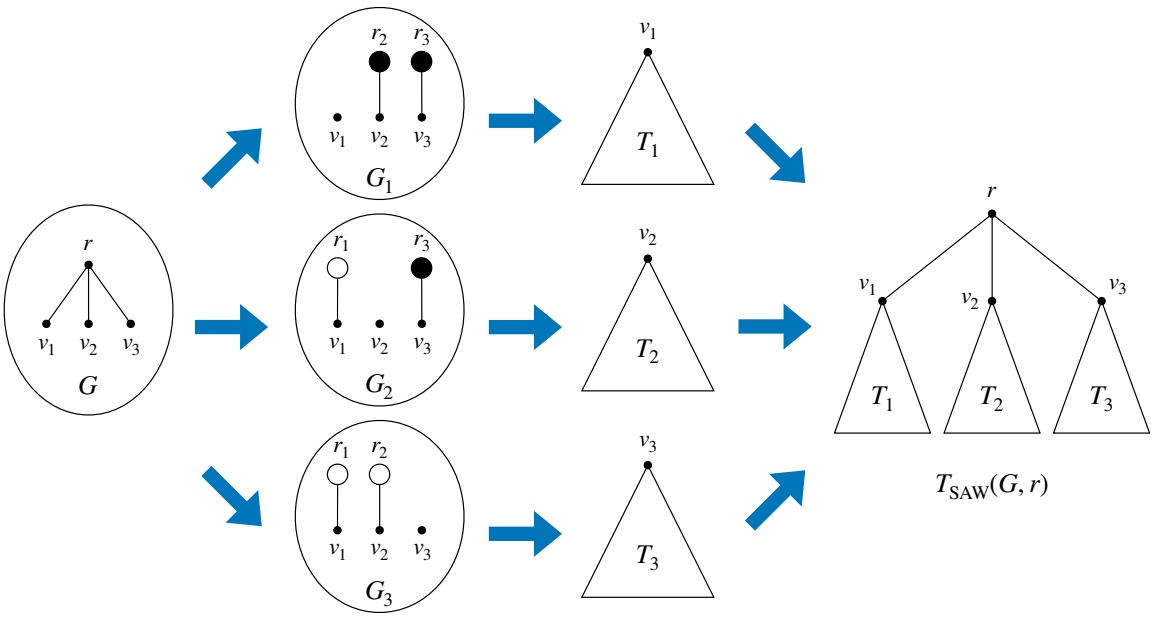

Figure 5.2: A recursive construction of the self-avoiding walk (SAW) tree. Here T_i is the SAW tree of G_i rooted at v_i for $i = 1, 2, 3$. (●/○: fixed to spin 1/0.)

and the same holds with spin 1 replaced by 0. For $1 \le i \le d$, let σ_{Λ_i} denote the union of the conditioning σ_Λ and σ_{U_i}, where $\Lambda_i = \Lambda \cup U_i$. Then for every $1 \le i \le d$ we have

$$Z_{G'}^{\sigma_\Lambda}(r_1 = 0, \ldots, r_{i-1} = 0, r_i = 1, \ldots, r_d = 1) = \beta \cdot Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 1) + Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 0).$$

Notice that both sides are independent of the field λ_r: for the left side, all r_i's do not have a field for the spin system on G'; for the right side, recall that we do not count the weight of fixed vertices for the conditional partition function for each G_i. Now define $Q_{G,r}^{\sigma_\Lambda} = Q_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$ by

$$Q_{G,r}^{\sigma_\Lambda} = \prod_{i=2}^{d} Z_{G'}^{\sigma_\Lambda}(r_1 = 0, \ldots, r_{i-1} = 0, r_i = 1, \ldots, r_d = 1),$$

which is independent of λ_r. Then we get

$$Z_G^{\sigma_\Lambda}(r = 1) \cdot Q_{G,r}^{\sigma_\Lambda} = \lambda_r \cdot \prod_{i=1}^{d} Z_{G'}^{\sigma_\Lambda}(r_1 = 0, \ldots, r_{i-1} = 0, r_i = 1, \ldots, r_d = 1)$$

$$= \lambda_r \cdot \prod_{i=1}^{d} \left(\beta \cdot Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 1) + Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 0)\right).$$

Using a similar argument, we also have

$$\begin{aligned} Z_G^{\sigma_\Lambda}(r = 0) \cdot Q_{G,r}^{\sigma_\Lambda} &= \prod_{i=1}^{d} Z_{G'}^{\sigma_\Lambda}(r_1 = 0, \ldots, r_i = 0, r_{i+1} = 1, \ldots, r_d = 1) \\ &= \prod_{i=1}^{d} \left(Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 1) + \gamma \cdot Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 0)\right). \end{aligned}$$

Since we assume that $(G\backslash\{r\})\backslash\Lambda$ is connected, the graph $G_i\backslash\Lambda$ is also connected for each i. Then, by the induction hypothesis, for each i there exists a polynomial $P_{G_i,v_i}^{\sigma_{\Lambda_i}} = P_{G_i,v_i}^{\sigma_{\Lambda_i}}(\boldsymbol{\lambda})$ such that

$$Z_{T_i}^{\sigma_{\Lambda_i}}(r = 1) = Z_{G_i}^{\sigma_{\Lambda_i}}(r = 1) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}} \quad \text{and} \quad Z_{T_i}^{\sigma_{\Lambda_i}}(r = 0) = Z_{G_i}^{\sigma_{\Lambda_i}}(r = 0) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}};$$

these polynomials are independent of λ_r since the conditional partition functions for G_i's do not involve λ_r. Now if we let

$$P_{G,r}^{\sigma_\Lambda} = Q_{G,r}^{\sigma_\Lambda} \cdot \prod_{i=1}^{d} P_{G_i,v_i}^{\sigma_{\Lambda_i}},$$

then it follows from the tree recursion that

$$\begin{aligned} Z_T^{\sigma_\Lambda}(r = 1) &= \lambda_r \cdot \prod_{i=1}^{d} \left(\beta \cdot Z_{T_i}^{\sigma_{\Lambda_i}}(v_i = 1) + Z_{T_i}^{\sigma_{\Lambda_i}}(v_i = 0)\right) \\ &= \lambda_r \cdot \prod_{i=1}^{d} \left(\beta \cdot Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 1) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}} + Z_{G_i}^{\sigma_{\Lambda_i}}(v_i = 0) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}}\right) \\ &= Z_G^{\sigma_\Lambda}(r = 1) \cdot Q_{G,r}^{\sigma_\Lambda} \cdot \prod_{i=1}^{d} P_{G_i,v_i}^{\sigma_{\Lambda_i}} \\ &= Z_G^{\sigma_\Lambda}(r = 1) \cdot P_{G,r}^{\sigma_\Lambda}. \end{aligned}$$

The other equality $Z_T^{\sigma_\Lambda}(r = 0) = Z_G^{\sigma_\Lambda}(r = 0) \cdot P_{G,r}^{\sigma_\Lambda}$ is established in the same way. This

completes the proof for the case that $(G\backslash\{r\})\backslash\Lambda$ is connected.

If $(G\backslash\{r\})\backslash\Lambda$ has two or more connected components, we can construct $T_{\text{SAW}}(G,r)$ by the SAW tree of each component. Recall that G' is defined by splitting the vertex r into d copies in the graph G. Suppose that $G'\backslash\Lambda$ has k connected component for an integer $k \geq 2$. Let $G'_{(1)},\dots,G'_{(k)}$ be the subgraphs induced by each component, along with vertices from Λ that are adjacent to it. For each j, let $G_{(j)}$ be the graph obtained from $G'_{(j)}$ by contracting all copies of r into one vertex $r_{(j)}$, and let $T_{(j)} = T_{\text{SAW}}(G'_{(j)}, r_{(j)})$. Observe that once we contract the roots $r_{(1)},\dots,r_{(k)}$ of $T_{(1)},\dots,T_{(k)}$, the resulting tree is $T_{\text{SAW}}(G,r)$.

We define the 2-spin system on each $G_{(j)}$ with the same parameters $(\beta,\gamma,\boldsymbol{\lambda})$, except that the vertex $r_{(j)}$ does not have a field (i.e., $\lambda_{r_{(j)}} = 1$ instead of λ_r). For $1 \leq j \leq k$, let $\Lambda_{(j)} = \Lambda \cap V(G_{(j)})$ and $\sigma_{\Lambda_{(j)}}$ be the configuration σ_Λ restricted on $\Lambda_{(j)}$. Then $G_{(j)}\backslash\Lambda_{(j)}$ is connected for every j and, since $k \geq 2$, each $G_{(j)}$ with conditioning $\sigma_{\Lambda_{(j)}}$ has fewer than m edges. Thus, we can apply the induction hypothesis; namely, for $1 \leq j \leq k$ there exists a polynomial $P^{\sigma_{\Lambda_{(j)}}}_{G_{(i)},r_{(i)}} = P^{\sigma_{\Lambda_{(j)}}}_{G_{(i)},r_{(i)}}(\boldsymbol{\lambda})$, which is independent of λ_r, such that

$$Z^{\sigma_{\Lambda_{(j)}}}_{T_{(j)}}(r_{(j)}=1) = Z^{\sigma_{\Lambda_{(j)}}}_{G_{(j)}}(r_{(j)}=1)\cdot P^{\sigma_{\Lambda_{(j)}}}_{G_{(j)},r_{(j)}}$$

and

$$Z^{\sigma_{\Lambda_{(j)}}}_{T_{(j)}}(r_{(j)}=0) = Z^{\sigma_{\Lambda_{(j)}}}_{G_{(j)}}(r_{(j)}=0)\cdot P^{\sigma_{\Lambda_{(j)}}}_{G_{(j)},r_{(j)}}.$$

We define the polynomial $P^{\sigma_\Lambda}_{G,r} = P^{\sigma_\Lambda}_{G,r}(\boldsymbol{\lambda})$ to be

$$P^{\sigma_\Lambda}_{G,r} = \prod_{j=1}^{k} P^{\sigma_{\Lambda_{(j)}}}_{G_{(j)},r_{(j)}}.$$

It is then easy to check that

$$Z^{\sigma_\Lambda}_{T}(r=1) = \lambda_r \cdot \prod_{j=1}^{k} Z^{\sigma_{\Lambda_{(j)}}}_{T_{(j)}}(r_{(j)}=1) = \lambda_r \cdot \prod_{j=1}^{k}\left(Z^{\sigma_{\Lambda_{(j)}}}_{G_{(j)}}(r_{(j)}=1)\cdot P^{\sigma_{\Lambda_{(j)}}}_{G_{(j)},r_{(j)}}\right)$$

$$= Z_G^{\sigma_\Lambda}(r=1) \cdot \prod_{j=1}^{k} P^{\sigma_{\Lambda(j)}}_{G_{(j)},r_{(j)}} = Z_G^{\sigma_\Lambda}(r=1) \cdot P^{\sigma_\Lambda}_{G,r},$$

and similarly $Z_T^{\sigma_\Lambda}(r=0) = Z_G^{\sigma_\Lambda}(r=0) \cdot P^{\sigma_\Lambda}_{G,r}$. The theorem then follows. □

5.6 Influence Bounds for Trees

In this section, we study the influences of the root on other vertices in a tree. We give an upper bound on the total influences of the root on all vertices at a fixed distance away. To do this, we apply the potential method, which has been used to establish the correlation decay property (see, e.g., [88, 89, 67]). Given an arbitrary potential function Ξ, our upper bound is in terms of properties of Ξ, involving bounds on $\left\|\nabla H_d^{\Xi}\right\|_1$ and $|\psi|$ where $\psi = \Xi'$. We then deduce Lemma 5.4.3 in the case that Ξ an (α, c)-potential.

Assume that $T = (V_T, E_T)$ is a tree rooted at r of maximum degree at most Δ. Let $\Lambda \subseteq V_T \backslash \{r\}$ and $\sigma_\Lambda \in \{0,1\}^\Lambda$ be arbitrary and fixed. Consider the 2-spin system on T with parameters (β, γ, λ), conditioned on σ_Λ. We need to bound the influence $\Psi_T^{\sigma_\Lambda}(r \to v)$ from the root r to another vertex $v \in V_T$. Notice that if v is disconnected from r when Λ is removed, then $\Psi_T^{\sigma_\Lambda}(r \to v) = 0$ by the Markov property of spin systems. Therefore, we may assume that, by removing all such vertices, Λ contains only leaves of T.

For a vertex $v \in V_T$, let $T_v = (V_{T_v}, E_{T_v})$ be the subtree of T rooted at v that contains all descendant of v; note that $T_r = T$. We will write $L_v(k) \subseteq V_T \backslash \Lambda$ for the set of all free vertices at distance k away from v in T_v. We pay particular interest in the marginal ratio at v in the subtree T_v, and write $R_v = R_{T_v}^{\sigma_\Lambda}(v)$ for simplicity. The $\log R_v$'s are related by the tree recursion H. If a vertex v has d children, denoted by $u_1, \ldots, u_d$, then the tree recursion is given by

$$\log R_v = H_d(\log R_{u_1}, \ldots, \log R_{u_d}),$$

where for $1 \leq d \leq \Delta$ and $(y_1, \ldots, y_d) \in [-\infty, \infty]^d$,

$$H_d(y_1, \ldots, y_d) = \log \lambda + \sum_{i=1}^{d} \log \left(\frac{\beta e^{y_i} + 1}{e^{y_i} + \gamma} \right).$$

Also recall that for $y \in [-\infty, +\infty]$, we define

$$h(y) = -\frac{(1 - \beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}$$

and $\frac{\partial}{\partial y_i} H_d(y_1, \ldots, y_d) = h(y_i)$ for all $1 \leq i \leq d \leq \Delta$.

The following lemma allows us to bound the sum of all influences from the root to distance k, using an arbitrary potential function.

Lemma 5.6.1. *Let $\Xi : [-\infty, +\infty] \to (-\infty, +\infty)$ be a differentiable and increasing (potential) function with image $S = \Xi[-\infty, +\infty]$ and derivative $\psi = \Xi'$. Denote the degree of the root r by Δ_r. Then for every integer $k \geq 1$,*

$$\sum_{v \in L_r(k)} |\mathcal{I}_T^{\sigma_\Lambda}(r \to v)| \leq \Delta_r A_\Xi B_\Xi \left(\max_{1 \leq d < \Delta} \sup_{\tilde{\boldsymbol{y}} \in S^d} \left\| \nabla H_d^\Xi(\tilde{\boldsymbol{y}}) \right\|_1 \right)^{k-1}$$

where

$$A_\Xi = \max_{u \in L_r(1)} \left\{ \frac{|h(\log R_u)|}{\psi(\log R_u)} \right\} \quad \textit{and} \quad B_\Xi = \max_{v \in L_r(k)} \{\psi(\log R_v)\}.$$

Before proving Lemma 5.6.1, we first present two useful properties of the influences on trees. Firstly, it was shown in [4] that the influences satisfy the following form of chain rule on trees.

Lemma 5.6.2 ([4, Lemma B.2])**.** *Suppose that $u, v, w \in V_T$ are three distinct vertices such that u is on the unique path from v to w. Then*

$$\Psi_T^{\sigma_\Lambda}(v \to w) = \Psi_T^{\sigma_\Lambda}(v \to u) \cdot \Psi_T^{\sigma_\Lambda}(u \to w).$$

Secondly, for two adjacent vertices on a tree, the influence from one to the other is given by the function h.

Lemma 5.6.3. *Let $v \in V_T$ and u be a child of v in the subtree T_v. Then*

$$\Psi_T^{\sigma_\Lambda}(v \to u) = h(\log R_u).$$

Proof. The lemma can be proved through an explicit computation of the influence. Here we present a more delicate proof utilizing Lemma 5.5.3, which gives some insights into the relation between the influence and the function h. We assume that v has d children in the subtree T_v, denoted by $u_1 = u$ and $u_2, \dots, u_d$ respectively. We also assume, as a more general setting than uniform fields, that each vertex w is attached to a field λ_w of its own. Then Lemma 5.5.3 and the tree recursion imply that

$$\begin{aligned}
\Psi_T^{\sigma_\Lambda}(v \to u) = \Psi_{T_v}^{\sigma_\Lambda}(v \to u) &= \left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) \log R_v \\
&= \left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) H_d(\log R_{u_1}, \dots, \log R_{u_d}) \\
&= \sum_{i=1}^{d} \frac{\partial}{\partial \log R_{u_i}} H_d(\log R_{u_1}, \dots, \log R_{u_d}) \cdot \left(\lambda_u \frac{\partial}{\partial \lambda_u}\right) \log R_{u_i} \\
&= \sum_{i=1}^{d} h(\log R_{u_i}) \cdot \Psi_{T_{u_i}}^{\sigma_\Lambda}(u_i \to u) = h(\log R_u),
\end{aligned}$$

where the last equality is because $\Psi_{T_{u_i}}^{\sigma_\Lambda}(u_i \to u) = 0$ for $u_i \neq u$ and $\Psi_{T_u}^{\sigma_\Lambda}(u \to u) = 1$. □

We are now ready to prove Lemma 5.6.1.

Proof of Lemma 5.6.1. For a vertex $v \in V_T$, denote the number of its children by d_v; note that $d_r = \Delta_r$. Let $u_1, \dots, u_{\Delta_r}$ be the children of the root r. We may assume that all these children of r are free, since if u_i is fixed then $\Psi_T^{\sigma_\Lambda}(r \to u_i) = 0$ by definition. Then by

Lemma 5.6.2 and Lemma 5.6.3, we get

$$\begin{aligned}\sum_{v\in L_r(k)}|\Psi_T^{\sigma_\Lambda}(r\to v)| &= \sum_{i=1}^{\Delta_r}|\Psi_T^{\sigma_\Lambda}(r\to u_i)|\sum_{v\in L_{u_i}(k-1)}|\Psi_T^{\sigma_\Lambda}(u_i\to v)|\\ &= \sum_{i=1}^{\Delta_r}|h(\log R_{u_i})|\sum_{v\in L_{u_i}(k-1)}|\Psi_T^{\sigma_\Lambda}(u_i\to v)|\\ &= \sum_{i=1}^{\Delta_r}\frac{|h(\log R_{u_i})|}{\psi(\log R_{u_i})}\sum_{v\in L_{u_i}(k-1)}\psi(\log R_{u_i})\,|\Psi_T^{\sigma_\Lambda}(u_i\to v)|\,.\end{aligned}$$

Hence, we obtain that

$$\begin{aligned}\sum_{v\in L_r(k)}|\Psi_T^{\sigma_\Lambda}(r\to v)| \le \Delta_r\cdot\max_{1\le i\le\Delta_r}\left\{\frac{|h(\log R_{u_i})|}{\psi(\log R_{u_i})}\right\}\\ \cdot\max_{1\le i\le\Delta_r}\left\{\sum_{v\in L_{u_i}(k-1)}\psi(\log R_{u_i})\,|\Psi_T^{\sigma_\Lambda}(u_i\to v)|\right\}.\end{aligned}\tag{5.5}$$

Next, we show by induction that for every vertex $u\in V_T\backslash\{r\}$ and every integer $k\ge 0$ we have

$$\sum_{v\in L_u(k)}\psi(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to v)| \le \max_{v\in L_u(k)}\{\psi(\log R_v)\}\cdot\left(\max_{w\in V_{T_u}}\sup_{\tilde{\boldsymbol{y}}\in S^{d_w}}\left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^k.\tag{5.6}$$

Observe that once we establish Eq. (5.6), the lemma follows immediately by plugging Eq. (5.6) into Eq. (5.5). We will use induction on k to prove Eq. (5.6). When $k=0$, if $u\in\Lambda$ is fixed then $L_u(0)=\emptyset$ and there is nothing to show; otherwise, Eq. (5.6) becomes

$$\psi(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to u)|\le\psi(\log R_u),$$

which holds with equality since $\Psi_T^{\sigma_\Lambda}(u\to u)=1$. Now suppose that Eq. (5.6) holds for some integer $k-1\ge 0$ (and for every vertex $u\in V_T\backslash\{r\}$). Let $u\in V_T\backslash\{r\}$ be arbitrary and denote the children of u by $w_1,\ldots,w_d$, where $1\le d<\Delta$ (if $d=0$ then $L_u(k)=\emptyset$

and Eq. (5.6) holds trivially). Again by Lemma 5.6.2 and Lemma 5.6.3 we have

$$\begin{aligned}
&\sum_{v\in L_u(k)} \psi(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to v)| \\
&= \sum_{i=1}^{d} \psi(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to w_i)| \sum_{v\in L_{w_i}(k-1)} |\Psi_T^{\sigma_\Lambda}(w_i\to v)| \\
&= \sum_{i=1}^{d} \frac{\psi(\log R_u)}{\psi(\log R_{w_i})}\,|h(\log R_{w_i})| \sum_{v\in L_{w_i}(k-1)} \psi(\log R_{w_i})\,|\Psi_T^{\sigma_\Lambda}(w_i\to v)|\,.
\end{aligned}$$

Using the induction hypothesis, we get

$$\begin{aligned}
&\sum_{v\in L_u(k)} \psi(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to v)| \\
&\le \sum_{i=1}^{d} \frac{\psi(\log R_u)}{\psi(\log R_{w_i})}\,|h(\log R_{w_i})| \cdot \max_{v\in L_{w_i}(k-1)} \{\psi(\log R_v)\} \\
&\qquad\qquad \cdot \left(\max_{w\in V_{T_{w_i}}} \sup_{\tilde{\boldsymbol{y}}\in S^{d_w}} \left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^{k-1} \\
&\le \max_{v\in L_u(k)} \{\psi(\log R_v)\} \cdot \left(\max_{w\in V_{T_u}\setminus\{u\}} \sup_{\tilde{\boldsymbol{y}}\in S^{d_w}} \left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^{k-1} \\
&\qquad\qquad \cdot \sum_{i=1}^{d} \frac{\psi(\log R_u)}{\psi(\log R_{w_i})}\,|h(\log R_{w_i})| \\
&\le \max_{v\in L_u(k)} \{\psi(\log R_v)\} \cdot \left(\max_{w\in V_{T_u}} \sup_{\tilde{\boldsymbol{y}}\in S^{d_w}} \left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^{k},
\end{aligned}$$

where the last inequality follows from that

$$\begin{aligned}
\sum_{i=1}^{d} \frac{\psi(\log R_u)}{\psi(\log R_{w_i})}\,|h(\log R_{w_i})| &= \sum_{i=1}^{d} \left|\frac{\partial}{\partial \Xi(\log R_{w_i})}\, H_d^{\Xi}\left(\Xi(\log R_{w_1}),\ldots,\Xi(\log R_{w_d})\right)\right| \\
&= \left\|\nabla H_d^{\Xi}\left(\Xi(\log R_{w_1}),\ldots,\Xi(\log R_{w_d})\right)\right\|_1.
\end{aligned}$$

This establishes Eq. (5.6), and thus completes the proof of the lemma. □

We then derive Lemma 5.4.3 as a corollary.

Proof of Lemma 5.4.3. Since Ξ is an (α, c)-potential, the Contraction condition implies that

$$\max_{1\le d<\Delta} \sup_{\tilde{\boldsymbol{y}}\in S^d} \left\|\nabla H_d^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1 \le 1-\alpha.$$

Meanwhile, since the degree of a vertex $v \in V_T\backslash\{r\}$ in the subtree T_v is less than Δ, we have $\log R_v \in J$. Then the Boundedness condition implies that for all $u \in L_r(1)$ and $v \in L_r(k)$,

$$\frac{\psi(\log R_v)}{\psi(\log R_u)} \cdot |h(\log R_u)| \le \frac{c}{\Delta}.$$

Therefore, we get

$$\Delta_r A_\Xi B_\Xi = \Delta_r \cdot \max_{u\in L_r(1)} \left\{\frac{|h(\log R_u)|}{\psi(\log R_u)}\right\} \cdot \max_{v\in L_r(k)} \{\psi(\log R_v)\} \le c.$$

The lemma then follows immediately from Lemma 5.6.1. □

5.7 Verifying a Good Potential: Contraction

In this section, we make a first step for proving Lemma 5.4.4. Let $\Delta \ge 3$ be an integer. Let β, γ, λ be reals such that $0 \le \beta \le \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$. Recall that define our potential function $\Xi : [-\infty, +\infty] \to (-\infty, +\infty)$ through its derivative by

$$\Xi'(y) = \psi(y) = \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}}, \qquad \Xi(0) = 0. \tag{5.7}$$

The following lemma implies that the potential Ξ given by Eq. (5.3) is well-defined.

Lemma 5.7.1. *For all $\beta, \gamma > 0$ such that $\beta\gamma < 1$, we have*

$$\int_{-\infty}^{+\infty} \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}} < +\infty.$$

Proof. For the $+\infty$ side we have

$$\int_0^{+\infty} \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}} = \int_0^{+\infty} \sqrt{\frac{1-\beta\gamma}{\beta e^y+\gamma e^{-y}+\beta\gamma+1}} < \int_0^{+\infty} \frac{1}{\sqrt{\beta e^y}} < +\infty.$$

Similarly, for the $-\infty$ side we have

$$\int_{-\infty}^0 \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}} < \int_{-\infty}^0 \frac{1}{\sqrt{\gamma e^{-y}}} < +\infty. \qquad \square$$

If (β, γ, λ) is up-to-Δ unique with gap $\delta \in (0, 1)$, then we show that Ξ satisfies the Contraction condition for $\alpha = \delta/2$. This holds for all parameters (β, γ, λ) in the uniqueness region, *without* requiring that $\gamma \leq 1$. Later in Section 5.9, we establish the Boundedness condition for Ξ when $\gamma \leq 1$, completing the proof of Lemma 5.4.4. The case of $\gamma > 1$ is more complicated and is left to Section 5.8.

Before giving our proof, we first point out that the potential function Ξ is essentially the same potential function Φ used in [89] (notice that [89] uses φ as the notation of the potential function and $\Phi = \varphi'$ for its derivative). Recall that the tree recursion for the marginal ratios is given by the function $F_d : [0, +\infty]^d \to [0, +\infty]$ where $1 \leq d \leq \Delta$ such that for all $(x_1, \ldots, x_d) \in [0, +\infty]^d$,

$$F_d(x_1, \ldots, x_d) = \lambda \prod_{i=1}^{d} \frac{\beta x_i + 1}{x_i + \gamma}.$$

The potential function $\Phi : [0, +\infty] \to (-\infty, +\infty)$ from [89] is defined implicitly via its derivative as

$$\Phi'(x) = \varphi(x) = \frac{1}{\sqrt{x(\beta x+1)(x+\gamma)}}, \qquad \Phi(1) = 0.$$

The follows lemma explains how we obtain our potential Ξ from Φ.

Lemma 5.7.2. *We have* $\Xi = \sqrt{1-\beta\gamma} \cdot (\Phi \circ \exp)$*; namely,* $\Xi(y) = \sqrt{1-\beta\gamma} \cdot \Phi(e^y)$ *for all* $y \in [-\infty, +\infty]$.

Proof. It is straightforward to check that

$$\begin{aligned}\psi(y) &= \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}} \\ &= \sqrt{1-\beta\gamma}\cdot e^y \cdot \sqrt{\frac{1}{e^y(\beta e^y+1)(e^y+\gamma)}} \\ &= \sqrt{1-\beta\gamma}\cdot e^y \varphi(e^y).\end{aligned}$$

Therefore,

$$\begin{aligned}\Xi(y) &= \int_0^y \psi(t)\,\mathrm{d}t = \sqrt{1-\beta\gamma}\cdot\int_0^y e^t\varphi(e^t)\,\mathrm{d}t \\ &= \sqrt{1-\beta\gamma}\cdot\int_1^{e^y}\varphi(s)\,\mathrm{d}s = \sqrt{1-\beta\gamma}\cdot\Phi(e^y).\end{aligned}$$ □

Combining the results of Lemmas 12, 13 and 14 from [89], we get that the potential function Φ satisfies the following gradient bound when (β,γ,λ) is in the uniqueness region. Note that this can be regarded as the Contraction condition but for Φ and F_d.

Theorem 5.7.3 ([89])**.** *Let $S_\Phi = \Phi[0,+\infty]$ be the image of Φ. If the parameters (β,γ,λ) are up-to-Δ unique with gap $\delta\in(0,1)$, then for every integer d such that $1\le d<\Delta$ and every $(\tilde{x}_1,\ldots,\tilde{x}_d)\in S_\Phi^d$,*

$$\left\|\nabla F_d^\Phi(\tilde{x}_1,\ldots,\tilde{x}_d)\right\|_1 \le \sqrt{1-\delta}$$

where $F_d^\Phi = \Phi\circ F_d\circ\Phi^{-1}$.

Recall our definition from Section 5.2. The tree recursion, in terms of the log marginal ratios, is described by the function $H_d:[-\infty,+\infty]^d\to[-\infty,+\infty]$ where $1\le d\le\Delta$ such that for every $(y_1,\ldots,y_d)\in[-\infty,+\infty]^d$,

$$H_d(y_1,\ldots,y_d) = \log\lambda + \sum_{i=1}^d \log\left(\frac{\beta e^{y_i}+1}{e^{y_i}+\gamma}\right).$$

Observe that $H_d = \log \circ F_d \circ \exp$, since we move from ratios to log ratios. We are now ready to establish the Contraction condition for Ξ.

Lemma 5.7.4. *Let $S_\Xi = \Xi[-\infty, +\infty]$ be the image of Ξ. If the parameters (β, γ, λ) are up-to-Δ unique with gap $\delta \in (0, 1)$, then for every integer d such that $1 \leq d < \Delta$ and every $(\tilde{y}_1, \ldots, \tilde{y}_d) \in S_\Xi^d$,*

$$\left\| \nabla H_d^\Xi(\tilde{y}_1, \ldots, \tilde{y}_d) \right\|_1 \leq \sqrt{1 - \delta}$$

where $H_d^\Xi = \Xi \circ H_d \circ \Xi^{-1}$.

Proof. Define the linear function $a : \mathbb{R} \to \mathbb{R}$ to be $a(x) = \sqrt{1 - \beta\gamma} \cdot x$ for $x \in \mathbb{R}$. Then Lemma 5.7.2 gives $\Xi = a \circ \Phi \circ \exp$, and thereby $\Xi \circ \log = a \circ \Phi$. It follows that for every $1 \leq d < \Delta$,

$$H_d^\Xi = \Xi \circ H_d \circ \Xi^{-1} = \Xi \circ \log \circ F_d \circ \exp \circ \Xi^{-1} = a \circ \Phi \circ F_d \circ \Phi^{-1} \circ a^{-1} = a \circ F_d^\Phi \circ a^{-1}.$$

That means, for every $(\tilde{y}_1, \ldots, \tilde{y}_d) \in S_\Xi^d$ we have

$$H_d^\Xi(\tilde{y}_1, \ldots, \tilde{y}_d) = \sqrt{1 - \beta\gamma} \cdot F_d^\Phi(\tilde{x}_1, \ldots, \tilde{x}_d)$$

where $\tilde{x}_i = \tilde{y}_i / \sqrt{1 - \beta\gamma}$ for $1 \leq i \leq d$. Then, for each i,

$$\frac{\partial}{\partial \tilde{y}_i} H_d^\Xi(\tilde{y}_1, \ldots, \tilde{y}_d) = \sqrt{1 - \beta\gamma} \cdot \frac{\partial}{\partial \tilde{x}_i} F_d^\Phi(\tilde{x}_1, \ldots, \tilde{x}_d) \cdot \frac{\mathrm{d}\tilde{x}_i}{\mathrm{d}\tilde{y}_i} = \frac{\partial}{\partial \tilde{x}_i} F_d^\Phi(\tilde{x}_1, \ldots, \tilde{x}_d).$$

This implies that $\nabla H_d^\Xi(\tilde{y}_1, \ldots, \tilde{y}_d) = \nabla F_d^\Phi(\tilde{x}_1, \ldots, \tilde{x}_d)$ for all $(\tilde{y}_1, \ldots, \tilde{y}_d) \in S_\Xi^d$, and the lemma then follows from Theorem 5.7.3. □

5.8 Remaining Antiferromagnetic Cases

In this section, we discuss the case where $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$. As studied in [89], in this case the uniqueness region is more complicated. For example, there exists a critical $\lambda_c^* > 0$ such that the 2-spin system with $\lambda < \lambda_c^*$ is in the uniqueness region for arbitrary graphs; namely, (β, γ, λ) is up-to-∞ unique. To deal with large degrees, we need to relax the Boundedness condition in Definition 5.2.1 and define a more general version of (α, c)-potentials. We shall see that Theorem 5.2.2 still holds for this general (α, c)-potential. The reason behind it is that in order to bound the maximum eigenvalue of the influence matrix, it suffices to consider a vertex-weighted sum of absolute influences of a vertex with large degree.

Remark 5.8.1. We give more background on the uniqueness region in Section 5.9.1. Note that in a recent revision of [89], the authors updated the descriptions of the uniqueness region for the case $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$, fixing a small error in the previous version. Statements and proofs in this section and Section 5.9 of this thesis are also adjusted accordingly based on the new version of [89].

Recall that our goal is to bound the maximum eigenvalue of the matrix $\Psi_G^{\sigma_\Lambda}$. We can do this by upper bounding the absolute row sum $\sum_{v \in V \setminus \Lambda} |\Psi_G^{\sigma_\Lambda}(r \to v)|$ for fixed r, thereby giving us a valid upper bound on $\lambda_{\max}(\Psi_G^{\sigma_\Lambda})$. However, this approach does not work when $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$. In this case, the potential Ξ fails to be an (α, c)-potential for a universal constant c independent of Δ. In fact, no such (α, c)-potentials exist as the absolute row sum $\sum_{v \in V \setminus \Lambda} |\Psi_G^{\sigma_\Lambda}(r \to v)|$ can be as large as $\Theta(\Delta)$. Especially, if the parameters (β, γ, λ) are up-to-∞ unique, which means the spin system has uniqueness for arbitrary graphs, then the absolute row sum $\sum_{v \in V \setminus \Lambda} |\Psi_G^{\sigma_\Lambda}(r \to v)|$ can be $\Theta(n)$ where $n = |V|$. We give a specific example where this is the case.

Example 5.8.2. Consider the antiferromagnetic 2-spin system specified by parameters $\beta = 0$, $\gamma > 1$ and $\lambda > 0$ on the star graph centered at r with Δ leaves. A simple calculation

reveals that $|\mathcal{I}_G(r \to v)| = \frac{\lambda}{\lambda+\gamma}$ for any leaf vertex $v \neq r$. Hence, $\sum_{v \neq r} |\mathcal{I}_G(r \to v)| = \Delta \cdot \frac{\lambda}{\lambda+\gamma}$. Now, since $\gamma > 1$, we have

$$\lambda_c = \lambda_c(\gamma, \Delta) = \min_{1<d<\Delta} \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}} = \Theta_\gamma(1),$$

forcing $\sum_{v \neq r} |\mathcal{I}_G(r \to v)| = \Theta_\gamma(\Delta)$ even when $\lambda < \lambda_c$ lies in the uniqueness region. However, we still have $\lambda_{\max}(\Psi_G) = O(1)$ since $\sum_{v \neq r} |\Psi_G(v \to r)| = O(1)$.

To solve this issue, one might want to consider the absolute column sum, involving the sum of absolute influences on a fixed vertex. However, this will not allow us to use the beautiful connection between graphs and SAW trees as showed in Lemma 5.4.1. Instead, we consider here a vertex-weighted version of the absolute row sum of $\Psi_G^{\sigma_\Lambda}$, which also upper bounds the maximum eigenvalue.

Lemma 5.8.3. *Let $\rho : V \to \mathbb{R}^+$ be a positive weight function of vertices. If there is a constant $\xi > 0$ such that for every $r \in V$ we have*

$$\sum_{v \in V \setminus \Lambda} \rho_v \cdot |\Psi_G^{\sigma_\Lambda}(r \to v)| \leq \xi \cdot \rho_r, \tag{5.8}$$

then $\lambda_{\max}(\Psi_G^{\sigma_\Lambda}) \leq \xi$.

Proof. Let $\mathcal{P} = \mathrm{diag}\{\rho_v : v \in V \setminus \Lambda\}$. The assumption is equivalent to $\|\mathcal{P}^{-1} \Psi_G^{\sigma_\Lambda} \mathcal{P}\|_\infty \leq \xi$. It follows that $\lambda_{\max}(\Psi_G^{\sigma_\Lambda}) = \lambda_{\max}(\mathcal{P}^{-1} \Psi_G^{\sigma_\Lambda} \mathcal{P}) \leq \xi$. □

We then modify our definition of (α, c)-potentials from Definition 5.2.1 which allows a weaker Boundedness condition. We remark that the only two differences between Definition 5.8.4 and Definition 5.2.1 is that: we allow $\Delta = \infty$; and the Boundedness condition is relaxed to what we call General Boundedness. Recall that for every $0 \leq d < \Delta$, we let $J_d = \big[\log(\lambda \beta^d), \log(\lambda/\gamma^d)\big]$ when $\beta\gamma < 1$, and $J_d = \big[\log(\lambda/\gamma^d), \log(\lambda \beta^d)\big]$ when $\beta\gamma > 1$.

Definition 5.8.4 (General (α, c)-potential function)**.** Let $\Delta \geq 3$ be an integer or $\Delta = \infty$. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$ and $\lambda > 0$. Let $\Xi : [-\infty, +\infty] \to (-\infty, +\infty)$ be a differentiable and increasing function with image $S = \Xi[-\infty, +\infty]$ and derivative $\psi = \Xi'$. For any $\alpha \in (0, 1)$ and $c > 0$, we say Ξ is a *general* (α, c)*-potential function* with respect to Δ and (β, γ, λ) if it satisfies the following conditions:

1. (Contraction) For every integer d such that $1 \leq d < \Delta$ and every $(\tilde{y}_1, \ldots, \tilde{y}_d) \in S^d$, we have
$$\left\|\nabla H_d^{\Xi}(\tilde{y}_1, \ldots, \tilde{y}_d)\right\|_1 = \sum_{i=1}^{d} \frac{\psi(y)}{\psi(y_i)} \cdot |h(y_i)| \leq 1 - \alpha$$
where $H_d^{\Xi} = \Xi \circ H_d \circ \Xi^{-1}$, $y_i = \Xi^{-1}(\tilde{y}_i)$ for $1 \leq i \leq d$, and $y = H_d(y_1, \ldots, y_d)$.

2. (General Boundedness) For all integers d_1, d_2 such that $0 \leq d_1, d_2 < \Delta$, and all reals $y_1 \in J_{d_1}, y_2 \in J_{d_2}$, we have
$$\frac{\psi(y_2)}{\psi(y_1)} \cdot |h(y_1)| \leq \frac{2c}{d_1 + d_2 + 2}.$$

Notice that General Boundedness is a weaker condition than Boundedness. To see this, if a potential function Ξ satisfies Boundedness with parameter c, then for every $0 \leq d_i < \Delta$ and every $y_i \in J_{d_i}$ where $i = 1, 2$ we have
$$\frac{\psi(y_2)}{\psi(y_1)} \cdot |h(y_1)| \leq \frac{c}{\Delta} \leq \frac{2c}{d_1 + d_2 + 2}.$$

The following theorem generalizes Theorem 5.2.2 and shows that a general (α, c)-potential function is sufficient to establish rapid mixing of the Glauber dynamics.

Theorem 5.8.5. *Let $\Delta \geq 3$ be an integer or $\Delta = +\infty$. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$ and $\lambda > 0$. Suppose that there is a general (α, c)-potential with respect to Δ and (β, γ, λ) for some $\alpha \in (0, 1)$ and $c > 0$. Then for every n-vertex graph G of maximum degree Δ, the Gibbs distribution μ of the 2-spin system on G with parameters*

(β, γ, λ) *is spectrally independent with constant*

$$\eta = \frac{2c}{\alpha}.$$

We then give a counterpart of Lemma 5.4.4, showing that Ξ is a general (α, c)-potential when $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$. Theorem 1.1.3 for this case is then obtained from Theorem 5.8.5 and Lemma 5.8.6.

Lemma 5.8.6. *Let $\Delta \geq 3$ be an integer. Let β, γ, λ be reals such that $0 \leq \beta < 1 < \gamma$ and $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$. Assume that (β, γ, λ) is up-to-Δ unique with gap $\delta \in (0, 1)$. Then the function Ξ defined implicitly by Eq.* (5.3) *is a general (α, c)-potential function with $\alpha \geq \delta/2$ and $c \leq 18$; we can further take $c \leq 4$ if $\beta = 0$.*

For Lemma 5.8.6, the Contraction condition of Ξ follows from Lemma 5.7.4, and General Boundedness is proved in Section 5.9 together with all other cases.

In the rest of this section, we prove Theorem 5.8.5 in the same way of Theorem 5.2.2, as outlined in Section 5.4. The major difference here is that we consider a weighted sum of absolute influences $\sum_{v \in V \setminus \Lambda} \rho_v \cdot |\Psi_G^{\sigma_\Lambda}(r \to v)|$ where $\rho : V \to \mathbb{R}^+$ is a weight function. This is sufficient for us to bound the eigenvalue of the influence matrix, as indicated by Lemma 5.8.3. We will choose the weight of a vertex v to be $\rho_v = \Delta_v$, the degree of v. The following lemma provides us an upper bound on the weighted sum of absolute influences to distance k, given a general (α, c)-potential. In particular, it generalizes Lemma 5.4.3.

Lemma 5.8.7. *If there exists a general (α, c)-potential function Ξ with respect to Δ and (β, γ, λ) where $\alpha \in (0, 1)$ and $c > 0$, then for every $\Lambda \subseteq V_T \setminus \{r\}$, $\sigma_\Lambda \in \{0, 1\}^\Lambda$ and all integers $k \geq 1$,*

$$\sum_{v \in L_r(k)} \Delta_v \cdot |\mathcal{I}_T^{\sigma_\Lambda}(r \to v)| \leq 2c \cdot (1 - \alpha)^{k-1} \cdot \Delta_r$$

where $L_r(k)$ denote the set of all free vertices at distance k away from r.

To prove Lemma 5.8.7, we first state the following generalization of Lemma 5.6.1 for

any weight function ρ. The proof of Lemma 5.8.8 is identical to Lemma 5.6.1 and we omit here.

Lemma 5.8.8. *Let $\Xi : [-\infty, +\infty] \to (-\infty, +\infty)$ be a differentiable and increasing (potential) function with image $S = \Xi[-\infty, +\infty]$ and derivative $\psi = \Xi'$. Denote the degree of the root r by Δ_r. Then for every integer $k \geq 1$,*

$$\sum_{v \in L_r(k)} \rho_v \cdot |\mathcal{I}_T^{\sigma_\Lambda}(r \to v)| \leq \Delta_r A_\Xi B_\Xi^\rho \left(\max_{1 \leq d < \Delta} \sup_{\tilde{\boldsymbol{y}} \in S^d} \left\| \nabla H_d^\Xi(\tilde{\boldsymbol{y}}) \right\|_1 \right)^{k-1}$$

where

$$A_\Xi = \max_{u \in L_r(1)} \left\{ \frac{|h(\log R_u)|}{\psi(\log R_u)} \right\} \quad \textit{and} \quad B_\Xi^\rho = \max_{v \in L_r(k)} \{\rho_v \cdot \psi(\log R_v)\}.$$

We then prove Lemma 5.8.7 and Theorem 5.8.5.

Proof of Lemma 5.8.7. Denote the degree of a vertex $v \in V_T \backslash \{r\}$ by Δ_v, and the degree of v in the subtree T_v by $d_v = \Delta_v - 1$. Pick the weights of vertices to be $\rho_v = \Delta_v$ for all $v \in V_T$. Since Ξ is a general (α, c)-potential, the Contraction condition implies that

$$\max_{1 \leq d < \Delta} \sup_{\tilde{\boldsymbol{y}} \in S^d} \left\| \nabla H_d^\Xi(\tilde{\boldsymbol{y}}) \right\|_1 \leq 1 - \alpha.$$

Since $\log R_v \in J_{d_v}$ by the definition of J_d, the General Boundedness condition implies that for all $u \in L_r(1)$ and $v \in L_r(k)$,

$$\frac{\psi(\log R_v)}{\psi(\log R_u)} \cdot |h(\log R_u)| \leq \frac{2c}{\Delta_u + \Delta_v}.$$

Therefore, we get

$$\Delta_r A_\Xi B_\Xi^\rho = \Delta_r \cdot \max_{u \in L_r(1)} \left\{ \frac{|h(\log R_u)|}{\psi(\log R_u)} \right\} \cdot \max_{v \in L_r(k)} \{\Delta_v \cdot \psi(\log R_v)\} \leq 2c \cdot \Delta_r.$$

The lemma then follows immediately from Lemma 5.8.8. □

Proof of Theorem 5.8.5. The proof of Theorem 5.8.5 is almost identical to Theorem 5.2.2. We point out that the only difference here is that we consider the weighted sum of absolute influences of a given vertex. Since the SAW tree preserve degrees of vertices, we can still apply Lemma 5.4.1. Then, combining Lemmas 5.4.1, 5.8.3 and 5.8.7, we complete the proof of the theorem. □

5.9 Verifying a Good Potential: Boundedness

In this subsection, we show the Boundedness or General Boundedness condition for our potential function Ξ defined by Eq. (5.3) in different ranges of parameters. Combining Lemma 5.7.4, we complete the proofs of Lemma 5.4.4 and Lemma 5.8.6.

In Section 5.9.1 we give background on the uniqueness region of parameters (β, γ, λ), based on the work of [89]. We then show Boundedness and General Boundedness in Section 5.9.2. Proofs of technical lemmas are left to Section 5.9.3.

5.9.1 Preliminaries for the Uniqueness Region

In this section we give a brief description of the uniqueness region of parameters (β, γ, λ). All the results here, and also their proofs, can be found in Lemma 21 from the latest version of [89].

Let $\Delta \geq 3$ be an integer and β, γ, λ be reals. We assume that $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$. For $1 \leq d \leq \Delta$ define

$$f_d(R) = \lambda \left(\frac{\beta R + 1}{R + \gamma} \right)^d$$

and denote the unique fixed point of f_d by R_d^*. Recall that the parameters (β, γ, λ) are up-to-Δ unique with gap $\delta \in (0, 1)$ if $|f_d'(R_d^*)| < 1 - \delta$ for all $1 \leq d < \Delta$.

When $\beta = 0$, the spin system is called a *hard-constraint model*. In this case, there exists a critical threshold for the external field defined as

$$\lambda_c = \lambda_c(\gamma, \Delta) = \min_{1<d<\Delta} \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}},$$

such that the parameters $(0, \gamma, \lambda)$ are up-to-Δ unique if and only if $\lambda < \lambda_c$. In particular, when $\gamma \leq 1$ the critical field is given by

$$\lambda_c = \lambda_c(\gamma, \Delta) = \frac{\gamma^{\Delta}(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}.$$

When $\beta > 0$, the spin system is called a *soft-constraint model*. If $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$, then (β, γ, λ) is up-to-Δ unique for all $\lambda > 0$. If $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ the uniqueness region is more complicated which we now describe. Let

$$\overline{\Delta} = \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}},$$

so that for every $1 \leq d < \overline{\Delta}$ we have $d \cdot \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} < 1$, and for every $d \geq \overline{\Delta}$ we have $d \cdot \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} \geq 1$. For every $\overline{\Delta} \leq d < \Delta$, we define $x_1(d) \leq x_2(d)$ to be the two positive roots of the quadratic equation

$$\frac{d(1-\beta\gamma)x}{(\beta x+1)(x+\gamma)} = 1.$$

More specifically, $x_1(d)$ and $x_2(d)$ are given by

$$x_1(d) = \frac{\theta(d) - \sqrt{\theta(d)^2 - 4\beta\gamma}}{2\beta} \qquad \text{and} \qquad x_2(d) = \frac{\theta(d) + \sqrt{\theta(d)^2 - 4\beta\gamma}}{2\beta}$$

where

$$\theta(d) = d(1-\beta\gamma) - (1+\beta\gamma).$$

Notice that $\theta(d) \geq 2\sqrt{\beta\gamma}$ for all $d \geq \overline{\Delta}$. For $i = 1, 2$ we let

$$\lambda_i(d) = x_i(d) \left(\frac{x_i(d) + \gamma}{\beta x_i(d) + 1} \right)^d.$$

Then, the parameters (β, γ, λ) are up-to-Δ unique if and only if λ belongs to the following regime

$$\mathcal{A} = \bigcap_{\overline{\Delta} \leq d < \Delta} \Big[(0, \lambda_1(d)) \cup (\lambda_2(d), \infty) \Big]. \tag{5.9}$$

In particular, when $\gamma \leq 1$ there are two critical thresholds $0 < \lambda_c < \overline{\lambda}_c$ such that the parameters (β, γ, λ) are up-to-Δ unique if and only if $\lambda < \lambda_c$ or $\lambda > \overline{\lambda}_c$ (i.e., $\mathcal{A} = (0, \lambda_c) \cup (\overline{\lambda}_c, \infty)$), where

$$\lambda_c = \lambda_c(\beta, \gamma, \Delta) = \min_{\overline{\Delta} \leq d < \Delta} \lambda_1(d)$$

and

$$\overline{\lambda}_c = \overline{\lambda}_c(\beta, \gamma, \Delta) = \max_{\overline{\Delta} \leq d < \Delta} \lambda_2(d) = \lambda_2(\Delta - 1).$$

The following bounds on the critical fields are helpful for our proofs later.

Lemma 5.9.1. *1. If $\beta = 0$, then for every integer d such that $1 < d < \Delta$ we have*

$$\lambda_c \leq \frac{4\gamma^{d+1}}{d-1}.$$

2. If $\beta > 0$ and $\sqrt{\beta\gamma} \leq \frac{\Delta - 2}{\Delta}$, then for every integer d such that $\overline{\Delta} \leq d < \Delta$ we have

$$\lambda_1(d) \leq \frac{18\gamma^{d+1}}{\theta(d)} \qquad \textit{and} \qquad \lambda_2(d) \geq \frac{\theta(d)}{18\beta^{d+1}}$$

where $\theta(d) = d(1 - \beta\gamma) - (1 + \beta\gamma)$.

The proof of Lemma 5.9.1 is postponed to Section 5.9.3.

5.9.2 Proofs of Boundedness

In this section we complete the proofs of Lemma 5.4.4 and Lemma 5.8.6 by establishing Boundedness and General Boundedness in the corresponding range of parameters.

Let $\Delta \geq 3$ be an integer. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$. Recall that the potential function Ξ is defined by

$$\Xi'(y) = \psi(y) = \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}} = \sqrt{|h(y)|}, \qquad \Xi(0) = 0. \tag{1}$$

It is surprising to find out that $\psi = \sqrt{|h|}$, as the potential Ξ is exactly the one from [89] as indicated by Lemma 5.7.2. This seems not to be a coincidence, and it provides some intuition why the potential from [89] works. More importantly, the fact that $\psi = \sqrt{|h|}$ is helpful in our proof of Boundedness and General Boundedness. Recall that for $0 \leq d < \Delta$ and $\beta\gamma < 1$ we let $J_d = \left[\log(\lambda\beta^d), \log(\lambda/\gamma^d)\right]$ to be the range of log marginal ratios of a vertex with d children. Then for every $0 \leq d_i < \Delta$ and $y_i \in J_{d_i}$ where $i = 1, 2$, we have

$$\frac{\psi(y_2)}{\psi(y_1)} \cdot |h(y_1)| = \sqrt{|h(y_1)| \cdot |h(y_2)|}. \tag{5.10}$$

The following lemma gives upper bounds on $\sqrt{|h(y_1)| \cdot |h(y_2)|}$, from which and Eq. (5.10) we deduce Boundedness and General Boundedness immediately. The brackets in the lemma indicate which lemma the bound is applied to.

Lemma 5.9.2. *Let $\Delta \geq 3$ be an integer. Let β, γ, λ be reals such that $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$. Assume that the parameters (β, γ, λ) are up-to-Δ unique with gap $\delta \in (0, 1)$. Then for all integers d_1, d_2 such that $0 \leq d_1, d_2 < \Delta$, and all reals $y_i \in J_{d_i}$ where $i = 1, 2$, the following holds:*

H. Hard-constraint models: $\beta = 0$ and $\lambda < \lambda_c$.

H.1. (Lemma 5.4.4) If $\gamma \leq 1$, then

$$|h(y_1)| \leq \frac{4}{\Delta}.$$

H.2. (Lemma 5.8.6) If $\gamma > 1$, then

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \leq \frac{8}{d_1 + d_2 + 2}.$$

S. Soft-constraint models: $\beta > 0$ and $\lambda \in \mathcal{A}$.

S.1. (Lemma 5.4.4) If $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$, then

$$|h(y_1)| \leq \frac{1.5}{\Delta}.$$

S.2. (Lemma 5.4.4) If $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma \leq 1$, then

$$|h(y_1)| \leq \frac{18}{\Delta}.$$

S.3. (Lemma 5.8.6) If $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ and $\gamma > 1$, then

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \leq \frac{36}{d_1 + d_2 + 2}.$$

The following lemma, whose proof can be found in Section 5.9.3, is helpful.

Lemma 5.9.3. *For every $y \in [-\infty, +\infty]$ we have*

$$|h(y)| = \frac{|1 - \beta\gamma|e^y}{(\beta e^y + 1)(e^y + \gamma)} \leq \frac{|1 - \sqrt{\beta\gamma}|}{1 + \sqrt{\beta\gamma}}.$$

We present here the proof of Lemma 5.9.2.

Proof of Lemma 5.9.2. We use notations and results from Section 5.9.1.

H. Hard-constraint models: $\beta = 0$ and $\lambda < \lambda_c$.

H.1. $\gamma \leq 1$.

For every $y_1 \in J_{d_1}$ we deduce from Lemma 5.9.1 that

$$e^{y_1} \leq \frac{\lambda}{\gamma^{d_1}} \leq \frac{\lambda_c}{\gamma^{\Delta-1}} \leq \frac{4\gamma}{\Delta - 2}.$$

Hence,

$$|h(y_1)| = \frac{e^{y_1}}{e^{y_1} + \gamma} \leq \frac{\frac{4\gamma}{\Delta-2}}{\frac{4\gamma}{\Delta-2} + \gamma} = \frac{4}{\Delta + 2} \leq \frac{4}{\Delta}.$$

H.2. $\gamma > 1$.

Let $\bar{y} = \frac{y_1+y_2}{2}$ and $\bar{d} = \frac{d_1+d_2}{2}$. Then we get

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} = \sqrt{\frac{e^{y_1}}{e^{y_1} + \gamma}} \cdot \sqrt{\frac{e^{y_2}}{e^{y_2} + \gamma}} = \frac{1}{\sqrt{(1 + \gamma e^{-y_1})(1 + \gamma e^{-y_2})}} \leq \frac{1}{1 + \gamma e^{-\bar{y}}},$$

where the last inequality follows from the AM–GM inequality by

$$(1+\gamma e^{-y_1})(1+\gamma e^{-y_2}) = 1+\gamma(e^{-y_1}+e^{-y_2})+\gamma^2 e^{-2\bar{y}} \geq 1+2\gamma e^{-\bar{y}}+\gamma^2 e^{-2\bar{y}} = (1+\gamma e^{-\bar{y}})^2.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$, we have

$$e^{\bar{y}} = \sqrt{e^{y_1} \cdot e^{y_2}} \leq \sqrt{\frac{\lambda}{\gamma^{d_1}} \cdot \frac{\lambda}{\gamma^{d_2}}} = \frac{\lambda}{\gamma^{\bar{d}}}.$$

If $\bar{d} \geq 2$, then we deduce from Lemma 5.9.1 and $\gamma > 1$ that

$$e^{\bar{y}} \leq \frac{\lambda_c}{\gamma^{\lfloor \bar{d} \rfloor}} \leq \frac{4\gamma}{\lfloor \bar{d} \rfloor - 1}.$$

It follows that

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \leq \frac{1}{1+\gamma e^{-\bar{y}}} \leq \frac{1}{1+\frac{\lfloor \bar{d} \rfloor - 1}{4}} = \frac{4}{\lfloor \bar{d} \rfloor + 3} \leq \frac{8}{d_1 + d_2 + 2}.$$

If $\bar{d} < 2$, then it is easy to see that

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \leq 1 \leq \frac{8}{d_1 + d_2 + 2}.$$

S. Soft-constraint models: $\beta > 0$ *and* $\lambda \in \mathcal{A}$.

S.1. $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$.

For every $y_1 \in J$ we deduce from Lemma 5.9.3 that

$$|h(y_1)| \leq \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} \leq \frac{1}{\Delta - 1} \leq \frac{1.5}{\Delta}.$$

S.2. $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ *and* $\gamma \leq 1$.

In this case, we have either $\lambda < \lambda_c$ or $\lambda > \overline{\lambda}_c$ where $\lambda_c, \overline{\lambda}_c$ are the two critical fields. Consider first $\lambda > \overline{\lambda}_c$. For every $y_1 \in J_{d_1}$ we deduce from Lemma 5.9.1 and $\beta < 1$ that

$$e^{y_1} \geq \lambda \beta^{d_1} \geq \overline{\lambda}_c \beta^{\Delta - 1} \geq \frac{\theta(\Delta - 1)}{18\beta}$$

where $\theta(d) = d(1-\beta\gamma) - (1+\beta\gamma)$. Hence,

$$\begin{aligned} |h(y_1)| &= \frac{(1-\beta\gamma)e^{y_1}}{(\beta e^{y_1} + 1)(e^{y_1} + \gamma)} = \frac{1-\beta\gamma}{\beta e^{y_1} + \gamma e^{-y_1} + (1+\beta\gamma)} \\ &\leq \frac{1-\beta\gamma}{\frac{\theta(\Delta-1)}{18} + (1+\beta\gamma)} = \frac{18(1-\beta\gamma)}{(\Delta-1)(1-\beta\gamma) + 17(1+\beta\gamma)} \leq \frac{18}{\Delta}. \end{aligned}$$

Next we consider $\lambda < \lambda_c$. For every $y_1 \in J_{d_1}$ we deduce from Lemma 5.9.1 and $\gamma \leq 1$

that

$$e^{y_1} \leq \frac{\lambda}{\gamma^{d_1}} \leq \frac{\lambda_c}{\gamma^{\Delta-1}} \leq \frac{18\gamma}{\theta(\Delta-1)}.$$

Hence,

$$|h(y_1)| = \frac{1-\beta\gamma}{\beta e^{y_1} + \gamma e^{-y_1} + (1+\beta\gamma)} \leq \frac{1-\beta\gamma}{\frac{\theta(\Delta-1)}{18} + (1+\beta\gamma)} \leq \frac{18}{\Delta}.$$

S.3. $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$ *and* $\gamma > 1$.

Let $\bar{y} = \frac{y_1+y_2}{2}$, $\bar{d} = \frac{d_1+d_2}{2}$, $d_L = \lfloor \bar{d} \rfloor$, and $d_R = \lceil \bar{d} \rceil$. We first consider some trivial cases. If $\bar{d} \leq 2$ then it is easy to see that

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \leq 1 \leq \frac{6}{d_1 + d_2 + 2}.$$

If $\bar{d} > 2$ and $d_L \leq \overline{\Delta}$, then we deduce from Lemma 5.9.3 that

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \leq \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} = \frac{1}{\overline{\Delta}} \leq \frac{2}{d_1 + d_2 - 2} \leq \frac{6}{d_1 + d_2 + 2}.$$

Hence, in the following we may assume that $\bar{d} > 2$ and $d_L > \overline{\Delta}$.

Since the parameters (β, γ, λ) are up-to-Δ unique, we have $\lambda \in \mathcal{A}$ where the regime $\mathcal{A}$ is given by Eq. (5.9). Observe that

$$\mathcal{A} \subseteq (0, \lambda_1(d_L)) \cup (\lambda_2(d_R), \infty) \cup (\lambda_2(d_L), \lambda_1(d_R))$$

where the last interval is nonempty only when $\lambda_2(d_L) < \lambda_1(d_R)$. This means that λ is contained in at least one of the three intervals. We establish the bound by considering these three cases separately.

Case 1: $\lambda < \lambda_1(d_L)$. By the Cauchy-Schwarz inequality, we have

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} = \sqrt{\frac{1-\beta\gamma}{\beta e^{y_1} + \gamma e^{-y_1} + (1+\beta\gamma)}} \cdot \sqrt{\frac{1-\beta\gamma}{\beta e^{y_2} + \gamma e^{-y_2} + (1+\beta\gamma)}}$$

$$\leq \frac{1-\beta\gamma}{\sqrt{(\beta e^{y_1}+\gamma e^{-y_1})(\beta e^{y_2}+\gamma e^{-y_2})}+(1+\beta\gamma)}. \tag{5.11}$$

Therefore, we get

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \leq \frac{1-\beta\gamma}{\gamma e^{-\bar{y}}+(1+\beta\gamma)}.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$ and $\gamma > 1$, we deduce from Lemma 5.9.1 that

$$e^{\bar{y}} \leq \frac{\lambda}{\gamma^{\bar{d}}} \leq \frac{\lambda_1(d_L)}{\gamma^{d_L}} \leq \frac{18\gamma}{\theta(d_L)},$$

where $\theta(d_L) = d_L(1-\beta\gamma)-(1+\beta\gamma)$. It follows that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \leq \frac{1-\beta\gamma}{\gamma e^{-\bar{y}}+(1+\beta\gamma)} \leq \frac{1-\beta\gamma}{\frac{\theta(d_L)}{18}+(1+\beta\gamma)} \leq \frac{36}{d_1+d_2+2}.$$

Case 2: $\lambda > \lambda_2(d_R)$. Similarly, we obtain from Eq. (5.11) that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \leq \frac{1-\beta\gamma}{\beta e^{\bar{y}}+(1+\beta\gamma)}.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$ and $\beta < 1$, we deduce from Lemma 5.9.1 that

$$e^{\bar{y}} \geq \lambda\beta^{\bar{d}} \geq \lambda_2(d_R)\beta^{d_R} \geq \frac{\theta(d_R)}{18\beta},$$

where $\theta(d_R) = d_R(1-\beta\gamma)-(1+\beta\gamma)$. It follows that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \leq \frac{1-\beta\gamma}{\beta e^{\bar{y}}+(1+\beta\gamma)} \leq \frac{1-\beta\gamma}{\frac{\theta(d_R)}{18}+(1+\beta\gamma)} \leq \frac{36}{d_1+d_2+2}.$$

Case 3: $\lambda_2(d_L) < \lambda < \lambda_1(d_R)$. We may assume that $d_1 \geq d_2$. By Eq. (5.11), we obtain

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \leq \frac{1-\beta\gamma}{\sqrt{\beta\gamma}e^{\frac{y_2-y_1}{2}}+(1+\beta\gamma)}.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$ and $\beta < 1 < \gamma$, we have

$$e^{y_2 - y_1} \geq \beta^{d_2} \gamma^{d_1} \geq \beta^{d_L} \gamma^{d_R}.$$

Meanwhile, we deduce from Lemma 5.9.1 that

$$\frac{\theta(d_L)}{18\beta^{d_L+1}} \leq \lambda_2(d_L) < \lambda < \lambda_1(d_R) \leq \frac{18\gamma^{d_R+1}}{\theta(d_R)},$$

which implies

$$\sqrt{\beta\gamma} e^{\frac{y_2 - y_1}{2}} \geq \sqrt{\beta^{d_L+1}\gamma^{d_R+1}} \geq \frac{\sqrt{\theta(d_L)\theta(d_R)}}{18} \geq \frac{\theta(d_L)}{18}.$$

It follows that

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \leq \frac{1 - \beta\gamma}{\sqrt{\beta\gamma} e^{\frac{y_2 - y_1}{2}} + (1 + \beta\gamma)} \leq \frac{1 - \beta\gamma}{\frac{\theta(d_L)}{18} + (1 + \beta\gamma)} \leq \frac{36}{d_1 + d_2 + 2}. \quad \square$$

5.9.3 Proofs of Technical Lemmas

Proof of Lemma 5.9.1. 1. For every $1 < d < \Delta$ we have

$$\lambda_c \leq \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}} = \frac{\gamma^{d+1}}{d-1} \left(\frac{d}{d-1} \right)^d \leq \frac{4\gamma^{d+1}}{d-1},$$

where the last inequality follows from that $(\frac{d}{d-1})^d \leq 4$ for all integer $d > 1$.

2. For every $\overline{\Delta} \leq d < \Delta$ we have

$$x_1(d) = \frac{2\gamma}{\theta(d) + \sqrt{\theta(d)^2 - 4\beta\gamma}} \leq \frac{2\gamma}{\theta(d)}.$$

Observe that the function $\frac{x+\gamma}{\beta x+1}$ is monotone increasing in x when $\beta\gamma < 1$, and thus we

deduce that

$$\frac{x_1(d)+\gamma}{\beta x_1(d)+1} \le \frac{\frac{2\gamma}{\theta(d)}+\gamma}{\frac{2\beta\gamma}{\theta(d)}+1} = \gamma \cdot \frac{2+d(1-\beta\gamma)-(1+\beta\gamma)}{2\beta\gamma+d(1-\beta\gamma)-(1+\beta\gamma)} = \gamma \cdot \frac{d+1}{d-1}.$$

Therefore,

$$\lambda_1(d) = x_1(d)\left(\frac{x_1(d)+\gamma}{\beta x_1(d)+1}\right)^d \le \frac{2\gamma}{\theta(d)} \cdot \gamma^d \cdot \left(\frac{d+1}{d-1}\right)^d \le \frac{18\gamma^{d+1}}{\theta(d)}$$

where the last inequality follows from that $(\frac{d+1}{d-1})^d \le 9$ for all integer $d > 1$.

The second part can be proved similarly. For every $\overline{\Delta} \le d < \Delta$ we have

$$x_2(d) = \frac{\theta(d)+\sqrt{\theta(d)^2-4\beta\gamma}}{2\beta} \ge \frac{\theta(d)}{2\beta},$$

and hence,

$$\frac{x_2(d)+\gamma}{\beta x_2(d)+1} \ge \frac{\frac{\theta(d)}{2\beta}+\gamma}{\frac{\theta(d)}{2}+1} = \frac{1}{\beta} \cdot \frac{d(1-\beta\gamma)-(1+\beta\gamma)+2\beta\gamma}{d(1-\beta\gamma)-(1+\beta\gamma)+2} = \frac{1}{\beta} \cdot \frac{d-1}{d+1}.$$

We then conclude that

$$\lambda_2(d) = x_2(d)\left(\frac{x_2(d)+\gamma}{\beta x_2(d)+1}\right)^d \ge \frac{\theta(d)}{2\beta} \cdot \frac{1}{\beta^d} \cdot \left(\frac{d-1}{d+1}\right)^d \ge \frac{\theta(d)}{18\beta^{d+1}},$$

where the last inequality again follows from that $(\frac{d+1}{d-1})^d \le 9$ for all integer $d > 1$. □

Proof of Lemma 5.9.3. We deduce from the AM–GM inequality that

$$|h(y)| = \frac{|1-\beta\gamma|}{\beta e^y + \gamma e^{-y} + 1 + \beta} \le \frac{|1-\beta\gamma|}{2\sqrt{\beta\gamma}+1+\beta} = \frac{|1-\sqrt{\beta\gamma}|}{1+\sqrt{\beta\gamma}},$$

as claimed. □

5.10 Ferromagnetic Cases

In the ferromagnetic case, the best known correlation decay results are given in [67, 118]. Using the potential functions in [67] and [118], we show the following two results, which match the known correlation decay results. In fact, the potential function from [118] turns out to be an (α, c)-potential function for constants $\alpha = \Theta(\delta)$ and $c \leq O(1)$.

Theorem 5.10.1. *Fix an integer $\Delta \geq 3$, positive real numbers β, γ, λ and $0 < \delta < 1$, and assume (β, γ, λ) satisfies one of the following three conditions:*

1. $\frac{\Delta-2+\delta}{\Delta-\delta} \leq \sqrt{\beta\gamma} \leq \frac{\Delta-\delta}{\Delta-2+\delta}$*, and λ is arbitrary;*
2. $\sqrt{\beta\gamma} \geq \frac{\Delta}{\Delta-2}$ *and* $\lambda \leq (1-\delta)\frac{\gamma}{\max\{1,\beta^{\Delta-1}\}\cdot((\Delta-2)\beta\gamma-\Delta)}$*;*
3. $\sqrt{\beta\gamma} \geq \frac{\Delta}{\Delta-2}$ *and* $\lambda \geq \frac{1}{1-\delta} \cdot \frac{(\Delta-2)\beta\gamma-\Delta}{\beta\cdot\min\{1,1/\gamma^{\Delta-1}\}}$*.*

Then the identity function $\Xi(y) = y$ (based on the potential given in [118]) is an (α, c)-potential function for $\alpha = \Theta(\delta)$ and $c = O(1)$. Furthermore, for every n-vertex graph G of maximum degree Δ, the mixing time of the Glauber dynamics for the 2-spin system on G with parameters (β, γ, λ) is $O(n \log n)$.

Remark 5.10.2. Condition 1 includes both the ferromagnetic case $1 < \sqrt{\beta\gamma} \leq \frac{\Delta-\delta}{\Delta-2+\delta}$ and the antiferromagnetic case $\frac{\Delta-2+\delta}{\Delta-\delta} \leq \sqrt{\beta\gamma} < 1$. Note that in both cases (β, γ, λ) is up-to-Δ unique with gap δ. For the antiferromagnetic case, the identity function Ξ is an (α, c)-potential with $c \leq 1.5$ and a better contraction rate $\alpha \geq \delta$, compared with the bound $\alpha \geq \delta/2$ of the potential Ξ given by Eq. (5.3) in Lemma 5.4.4. For the ferromagnetic case with $\beta = \gamma > 1$ (Ising model), $O(n \log n)$ mixing was previously known in [106].

The potential function from [67] is indeed an (α, c)-potential, but c must, unfortunately, depend on Δ. We have the following result, which is weaker than the correlation decay algorithm in [67] for unbounded-degree graphs.

Theorem 5.10.3. *Fix an integer $\Delta \geq 3$, and nonnegative real numbers β, γ, λ satisfying $\beta \leq 1 \leq \gamma$, $\sqrt{\beta\gamma} \geq \frac{\Delta}{\Delta-2}$, and $\lambda < \left(\frac{\gamma}{\beta}\right)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. Then for every n-vertex graph G with maximum degree Δ, the mixing time of the Glauber dynamics for the ferromagnetic 2-spin system on G with parameters (β, γ, λ) is $O(n \log n)$.*

Proofs of these theorems are provided below.

Proof of Theorem 5.10.1. Throughout, we use the "trivial" potential function $\Xi(y) = y$. Note that then, $\psi(y) = 1$ is a constant function. Now, we prove Contraction and Boundedness. The mixing result then follows from Theorem 5.2.2 and Theorem 1.2.1. We split into the three cases.

1. We first prove the Contraction part. By Lemma 5.9.3, for all $y \in [-\infty, +\infty]$ we have

$$|h(y)| \leq \frac{|1 - \sqrt{\beta\gamma}|}{1 + \sqrt{\beta\gamma}} \leq \frac{1 - \delta}{\Delta - 1}.$$

   Now let us prove the Boundedness condition. From the above inequality we have

$$|h(y)| \leq \frac{1}{\Delta - 1} \leq \frac{1.5}{\Delta}$$

   for $\Delta \geq 3$.

2. For the Contraction part, by $\log(\lambda \max\{1, 1/\gamma^{\Delta-1}\}) \leq y_i \leq \log(\lambda \max\{1, \beta^{\Delta-1}\})$, we have

$$\left|\frac{\partial H_d(\boldsymbol{y})}{\partial y_i}\right| = |h(y_i)| = \frac{\beta\gamma - 1}{1 + \beta\gamma + \gamma e^{-y_i} + \beta e^{y_i}} \leq \frac{\beta\gamma - 1}{1 + \beta\gamma + \gamma e^{-y_i}}$$
$$\leq \frac{\beta\gamma - 1}{1 + \beta\gamma + \frac{\gamma}{\lambda \max\{1, \beta^{\Delta-1}\}}}.$$

   Since we assumed $\lambda \leq (1 - \delta)\frac{\gamma}{\max\{1, \beta^{\Delta-1}\} \cdot ((\Delta-2)\beta\gamma - \Delta)}$, it follows that we have the

upper bound

$$\begin{aligned}\frac{\beta\gamma-1}{1+\beta\gamma+\frac{(\Delta-2)\beta\gamma-\Delta}{1-\delta}} &= (1-\delta)\frac{\beta\gamma-1}{(\Delta-1-\delta)\beta\gamma-(\Delta-1+\delta)}\\ &= (1-\delta)\frac{\beta\gamma-1}{(\Delta-1-\delta)(\beta\gamma-1)+2\delta}\\ &\le \frac{1-\delta}{\Delta-1-\delta} \le (1-\Theta(\delta))\frac{1}{\Delta-1}.\end{aligned}$$

Now, we prove the Boundedness condition. Since $\lambda \le \frac{\gamma}{\max\{1,\beta^{\Delta-1}\}\cdot((\Delta-2)\beta\gamma-\Delta)}$, it follows that $y \le \log(\lambda\max\{1,\beta^{\Delta-1}\}) \le \log\left(\frac{\gamma}{(\Delta-2)\beta\gamma-\Delta}\right)$. A simple calculation reveals that $\frac{\gamma}{(\Delta-2)\beta\gamma-\Delta} \le \sqrt{\frac{\gamma}{\beta}}$ and so by Lemma 5.9.3, we have

$$\begin{aligned}|h(y)| &\le \left|h\left(\log\left(\frac{\gamma}{(\Delta-2)\beta\gamma-\Delta}\right)\right)\right| \le \frac{(\beta\gamma-1)e^{\log\left(\frac{\gamma}{(\Delta-2)\beta\gamma-\Delta}\right)}}{e^{\log\left(\frac{\gamma}{(\Delta-2)\beta\gamma-\Delta}\right)}+\gamma}\\ &= (\beta\gamma-1)\frac{1}{1+(\Delta-2)\beta\gamma-\Delta} = \frac{\beta\gamma-1}{(\Delta-2)(\beta\gamma-1)-1} \le O(1/\Delta).\end{aligned}$$

3. For the Contraction part, by $\log(\lambda\max\{1,1/\gamma^{\Delta-1}\}) \le y_i \le \log(\lambda\max\{1,\beta^{\Delta-1}\})$, we have

$$\begin{aligned}\left|\frac{\partial H_d(\boldsymbol{y})}{\partial y_i}\right| = |h(y_i)| &= \frac{\beta\gamma-1}{1+\beta\gamma+\gamma e^{-y_i}+\beta e^{y_i}} \le \frac{\beta\gamma-1}{1+\beta\gamma+\beta e^{y_i}}\\ &\le \frac{\beta\gamma-1}{1+\beta\gamma+\beta\lambda\max\{1,1/\gamma^{\Delta-1}\}}.\end{aligned}$$

Since we assumed $\lambda \ge \frac{1}{1-\delta}\cdot\frac{(\Delta-2)\beta\gamma-\Delta}{\beta\cdot\min\{1,1/\gamma^{\Delta-1}\}}$, it follows that we have the upper bound

$$\frac{\beta\gamma-1}{1+\beta\gamma+\frac{(\Delta-2)\beta\gamma-\Delta}{1-\delta}}$$

which is again is upper bounded by $(1-\Theta(\delta))\frac{1}{\Delta-1}$ as we calculated in case 2 above.

Now, we prove the Boundedness condition. Note that since $\lambda \ge \frac{(\Delta-2)\beta\gamma-\Delta}{\beta\min\{1,1/\gamma^{\Delta-2}}$, it follows that $y \ge \log(\lambda\min\{1,1/\gamma^{\Delta-1}\} \ge \log\left(\frac{(\Delta-2)\beta\gamma-\Delta}{\beta}\right)$. A simple calculation

reveals that $\frac{(\Delta-2)\beta\gamma-\Delta}{\beta} \geq \sqrt{\frac{\gamma}{\beta}}$ and so by Lemma 5.9.3, we have

$$|h(y)| \leq \left|h\left(\log\left(\frac{(\Delta-2)\beta\gamma-\Delta}{\beta}\right)\right)\right| \leq (\beta\gamma-1)\frac{1}{\beta\cdot\frac{(\Delta-2)\beta\gamma-\Delta}{\beta}+1}$$
$$= \frac{\beta\gamma-1}{(\Delta-2)(\beta\gamma-1)-1} \leq O(1/\Delta). \qquad \square$$

Next, we use results from [67] to prove Theorem 5.10.3. Their potential function is implicitly defined by its derivative for the marginal ratios as

$$\Phi'(R) = \phi(R) = \min\left\{\frac{\beta\gamma-1}{\alpha\gamma\log\frac{\lambda+\gamma}{\beta\lambda+1}}, \frac{1}{R\log\frac{\lambda}{R}}\right\}$$

for a constant $0 \leq \alpha \leq 1$ depending only on β, γ, λ (see [67] for a precise definition). In our context, the corresponding potential for the log ratios is

$$\Xi'(y) = \psi(y) = e^y\phi(e^y) = \min\left\{\frac{\beta\gamma-1}{\alpha\gamma\log\frac{\lambda+\gamma}{\beta\lambda+1}}e^y, \frac{1}{\log\frac{\lambda}{e^y}}\right\}$$

and is bounded by constants depending on $\beta, \gamma, \lambda, \Delta$ for $\log(\lambda/\gamma^{\Delta-1}) \leq y \leq \log\lambda$.

One of the main technical results in [67] is showing that the tree recursion is contracting with the potential function Φ, and the derivative ϕ is bounded in the sense that there exist positive constants C_1, C_2 depending only on β, γ, λ such that $C_1 \leq \phi(R) \leq C_2$ for all $0 \leq R \leq \lambda$. [67] refer to such a function as a *universal potential function*.

In our context, we get that Ξ is an (α, c)-potential function which satisfies Definition 5.2.1, but with a constant c that depends on γ, Δ. Indeed, in the worst case, we have

$$\max_{y_1,y_2}\frac{\psi(y_2)}{\psi(y_1)} \geq \frac{\psi(\log\lambda)}{\psi(\log(\lambda/\gamma^{\Delta-1}))} = \frac{\lambda\frac{\beta\gamma-1}{\alpha\gamma\log\frac{\lambda+\gamma}{\beta\lambda+1}}}{\frac{\beta\gamma-1}{\alpha\log\frac{\lambda+\gamma}{\beta\lambda+1}}\cdot\frac{\lambda}{\gamma^\Delta}} = \gamma^{\Delta-1}.$$

More precisely, we have the following result from [67], stated in terms of the log marginal ratios.

Theorem 5.10.4. *Assume β, γ, λ are nonnegative real numbers satisfying $\beta \leq 1 \leq \gamma$, $\sqrt{\beta\gamma} \geq 1$, and $\lambda < \left(\frac{\gamma}{\beta}\right)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. Then the function Ξ is an (α, c)-potential function for a constant $0 < \alpha < 1$ depending on β, γ, λ, and a constant $c > 0$ depending on $\beta, \gamma, \lambda, \Delta$.*

Combined with Theorem 5.2.2 and Theorem 1.2.1, this gives $Cn \log n$ mixing with a constant C depending only on $\beta, \gamma, \lambda, \Delta$. We note this is weaker than the correlation decay result in [67] when the maximum degree Δ is unbounded.

5.11 Proofs of Mixing Results

In this section we give the proofs of Theorem 1.1.1, Theorem 1.1.2, Theorem 1.1.3 and Theorem 5.2.2.

Proof of Theorem 5.2.2. It suffices to bound $\sum_{v \in V \setminus \{r\}} |\mathcal{I}_G^{\sigma_\Lambda}(r \to v)|$ for all graphs $G = (V, E)$ with $n = |V|$ vertices and all boundary conditions σ_Λ on a subset Λ of i vertices. We deduce that

$$\begin{aligned}
\sum_{v \in V \setminus \{r\}} |\mathcal{I}_G^{\sigma_\Lambda}(r \to v)| &\leq \sum_{v \in V_T \setminus \{r\}} |\mathcal{I}_T^{\sigma_\Lambda}(r \to v)| && \text{(Lemma 5.4.1; } T = T_{\text{SAW}}(G, r)) \\
&= \sum_{k=1}^{\infty} \sum_{v \in L_r(k)} \left|\mathcal{I}_T^{\sigma_{\Lambda_\Lambda}}(r \to v)\right| && \text{(split the sum by levels)} \\
&\leq c \sum_{k=1}^{\infty} (1-\alpha)^{k-1} && \text{(Lemma 5.4.3)} \\
&= \frac{c}{\alpha}.
\end{aligned}$$

The theorem then follows. □

Proof of Theorem 1.1.3. We leverage Theorem 5.2.2 and Theorem 5.8.5, which shows $\frac{c}{\alpha}$-spectral independence as long as there is an (α, c)-potential, or $\frac{2c}{\alpha}$-spectral independence if there is a general (α, c)-potential. We use the potential given by Eq. (5.3), which is an adaptation of the potential function in [89] to the log marginal ratios. When (β, γ, λ) is

up-to-Δ unique with gap $\delta \in (0,1)$, it is an (α, c)-potential or a general (α, c)-potential by Lemma 5.4.4 and Lemma 5.8.6, with $\alpha \geq \delta/2$ and c a universal constant specified by the range of parameters. The theorem then follows from Theorem 1.2.1. □

Proof of Theorem 1.1.1. By Claim 5.11.1 given below, $\lambda \leq (1-\delta)\lambda_c(\Delta)$ implies up-to-Δ uniqueness with gap $\geq \delta/4$. Since $\gamma \leq 1$, we can again appeal to Lemma 5.4.4 to obtain an (α, c)-potential with $\alpha \geq \delta/8$ and $c \leq 4$. Theorem 1.1.1 then follows by Theorem 5.2.2 with $\frac{32}{\delta}$-spectral independence and Theorem 1.2.1. □

Proof of Theorem 1.1.2. By Claim 5.11.2 given below, $\beta \geq \beta_c(\Delta) + \delta(1-\beta_c(\Delta))$ implies up-to-Δ uniqueness with gap δ. Again, appealing to Lemma 5.4.4, we obtain an (α, c)-potential with $\alpha \geq \delta/2$ and $c \leq 1.5$. Theorem 1.1.2 then follows by Theorem 5.2.2 with $\frac{3}{\delta}$-spectral independence and Theorem 1.2.1.

Though we technically get $\frac{3}{\delta}$ by using the [89] potential, we can improve it to $\frac{1.5}{\delta}$-spectral independence by using the trivial identity function as the potential. See the first case of Theorem 5.10.1 and Remark 5.10.2. □

We next state and prove Claim 5.11.1 and Claim 5.11.2, which relate the parameter gaps with the uniqueness gaps.

Claim 5.11.1 (Hardcore Model; Lemma C.1 from [4])**.** *Fix an integer $\Delta \geq 3$, $0 < \delta < 1$, and $\beta = 0, \gamma > 0$. If $\lambda \leq (1-\delta)\lambda_c(\gamma, \Delta)$, then (β, γ, λ) is up-to-Δ unique with gap $\delta/4$.*

Claim 5.11.2 (Large $\sqrt{\beta\gamma}$)**.** *Fix an integer $\Delta \geq 3$, and $0 < \delta < 1$. If $\sqrt{\beta\gamma} \geq \frac{\Delta-2}{\Delta} + \delta\left(1 - \frac{\Delta-2}{\Delta}\right) = \frac{\Delta-2(1-\delta)}{\Delta}$, then (β, γ, λ) is up-to-Δ unique with gap $0 < \delta < 1$ for all λ. Note if $\beta = \gamma$, this is precisely the condition $\beta \geq \beta_c(\Delta) + \delta(1-\beta_c(\Delta))$.*

Proof. Consider the univariate recursion for the marginal ratios with $d < \Delta$ children $f_d(R) = \lambda\left(\frac{\beta R+1}{R+\gamma}\right)^d$. Differentiating, we have

$$f_d'(R) = d\lambda\left(\frac{\beta R+1}{R+\gamma}\right)^{d-1} \cdot \left(\frac{\beta}{R+\gamma} - \frac{\beta R+1}{(R+\gamma)^2}\right)$$

$$= -d(1-\beta\gamma)\lambda\left(\frac{\beta R+1}{R+\gamma}\right)^d \cdot \frac{1}{(\beta R+1)(R+\gamma)}$$
$$= -d(1-\beta\gamma)\cdot\frac{f_d(R)}{(\beta R+1)(R+\gamma)}.$$

At the unique fixed point R_d^*, we have $f_d(R_d^*) = R_d^*$ so

$$|f_d'(R_d^*)| = d(1-\beta\gamma)\frac{R_d^*}{(\beta R_d^*+1)(R_d^*+\gamma)}.$$

By Lemma 5.9.3, we have the upper bound

$$|f_d'(R_d^*)| \le d\cdot\frac{1-\beta\gamma}{(1+\sqrt{\beta\gamma})^2} = d\cdot\frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}}.$$

Since we assumed $\sqrt{\beta\gamma} \ge \frac{\Delta-2(1-\delta)}{\Delta}$, we obtain

$$d\cdot\frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} \le d\cdot\frac{\Delta-(\Delta-2(1-\delta))}{\Delta+(\Delta-2(1-\delta))} = d\cdot\frac{1-\delta}{\Delta-1+\delta} \le (1-\delta)\frac{d}{\Delta-1}.$$

As this is at most $1-\delta$ for all $d<\Delta$, we have up-to-Δ uniqueness with gap δ. □

CHAPTER 6

SPECTRAL INDEPENDENCE VIA STRONG SPATIAL MIXING APPROACH: MATCHINGS AND COLORINGS

In this chapter we establish spectral independence for the monomer-dimer model on arbitrary bounded-degree graphs and for random colorings on bounded-degree triangle-free graphs. By Theorem 1.2.1 we obtain optimal mixing times of the Glauber dynamics for these models. Our results match the current best known parameter regimes for strong spatial mixing. We consider the monomer-dimer model for matchings in Section 6.1, which is based on [45]. Section 6.2, based on [43], is devoted to colorings on triangle-free graphs.

6.1 Optimal Mixing Results for Monomer-Dimer Model

We prove optimal mixing time bounds for the monomer-dimer model on all matchings of a graph with constant maximum degree. Given a graph $G = (V, E)$ and a fugacity $\lambda > 0$, the Gibbs distribution μ for the monomer-dimer model is defined on the collection $\mathcal{M}$ of all matchings of G where $\mu(M) = w(M)/Z$ for $w(M) = \lambda^{|M|}$. The Glauber dynamics for the monomer-dimer model adds or deletes a random edge in each step. In particular, from $X_t \in \mathcal{M}$, choose an edge e uniformly at random from E and let $X' = X_t \oplus e$. If $X' \in \mathcal{M}$ then let $X_{t+1} = X'$ with probability $w(X')/(w(X') + w(X_t))$ and otherwise let $X_{t+1} = X_t$.

We prove $O(n \log n)$ mixing time for the Glauber dynamics for sampling matchings on bounded-degree graphs with n vertices. A classical result of Jerrum and Sinclair [76] yields rapid mixing of the Glauber dynamics for any graph even with unbounded degrees, but the best mixing time bound was $O(n^2 m \log n)$ [75] where m is the number of edges.

Theorem 1.1.5 (Monomer-Dimer Model)**.** *Let $\Delta \geq 3$ be an integer and let $\lambda > 0$ be a*

real. For every n-vertex graph G of maximum degree Δ, the mixing time of the Glauber dynamics for the monomer-dimer model on G with fugacity λ is at most $Cn\log n$ where $C = C(\Delta, \lambda)$ is a constant independent of n.

For the monomer-dimer model, spectral independence was not known previously. Following the proof strategy of Chapter 5 and utilizing the two-step recursion from [19], we show the following.

Theorem 6.1.1. *Let $\Delta \geq 3$ be an integer and $\lambda > 0$ be a real. Then for every graph $G = (V, E)$ of maximum degree at most Δ with $m = |E|$, the Gibbs distribution μ of the monomer-dimer model on G with fugacity λ is η-spectrally independent for $\eta = \min\left\{2\lambda\Delta, 2\sqrt{1+\lambda\Delta}\right\}$.*

Our goal in the rest of this section is to obtain spectral independence bounds for the monomer-dimer model; in particular, we prove Theorem 6.1.1. Fix a graph $G = (V_G, E_G)$ and a positive real number $\lambda > 0$. We define the monomer-dimer model on G to be the distribution μ_G on 2^{E_G} supported on all matchings of G, where $\mu_G(M) \propto \lambda^{|M|}$. We note this model may be identified with the hardcore model on the line graph $L(G)$ with parameter λ, and hence, may also be viewed as a 2-spin system with spins in $\{0, 1\}$. As above, we also think of the states as being assignments $\sigma : E \to \{0, 1\}$ such that $\{e \in E : \sigma(e) = 0\}$ is a matching.

Recall that we define the pairwise influence as

$$\Psi_G^\tau(e \to f) = \mu_G(\sigma_f = 0 \mid \sigma_e = 0, \sigma_\Lambda = \tau) - \mu_G(\sigma_f = 0 \mid \sigma_e = 1, \sigma_\Lambda = \tau)$$

for every $\Lambda \subseteq E_G$, every feasible boundary condition $\tau : \Lambda \to \{0, 1\}$, and every pair of distinct edges $e, f \notin \Lambda$. We say the Gibbs distribution μ_G of the monomer-dimer model is η-spectrally independent if $\lambda_1(\Psi_G^\tau) \leq \eta$ for all $\Lambda \subseteq E$ and all feasible $\tau : \Lambda \to \{0, 1\}$, where Ψ_G is the associated matrix of pairwise influences. We note this notion of spectral independence is equivalent to Definition 2.1.4; see [4].

We prove the following. Note that Theorem 6.1.1 is a consequence of it as the maximum eigenvalue of a matrix is upper bounded by the maximum absolute row sum.

Theorem 6.1.2. *Fix an integer $\Delta \geq 3$, and a positive real number $\lambda > 0$. Then for every graph $G = (V_G, E_G)$ of maximum degree at most Δ with $|E| = m$, every $\Lambda \subseteq E_G$, and every feasible boundary condition $\tau : \Lambda \to \{0, 1\}$, the Gibbs distribution μ of the monomer-dimer model on G with fugacity λ satisfies the inequality*

$$\sum_{f \in E_G : f \neq e, f \notin \Lambda} |\Psi_G^\tau(e \to f)| \leq \min\left\{2\lambda\Delta, 2\sqrt{1+\lambda\Delta}\right\}$$

for all edges $e \in E_G$. In particular, the Gibbs distribution μ_G is η-spectrally independent for

$$\eta = \min\left\{2\lambda\Delta, 2\sqrt{1+\lambda\Delta}\right\}.$$

Remark 6.1.3. We note [3] independently gave an $O(1)$-spectral independence bound for vertex-to-vertex influences using Hurwitz stability which is incomparable to our result here.

Our proof follows the strategy used in Chapter 5. Specifically, we prove Theorem 6.1.2 in two steps. In the first step, we prove a reduction for bounding the total influence of an edge in G to the total influence of an edge in the associated tree of self-avoiding walks in G. To do this, we extend known results [60] on the univariate matching polynomial, following a similar but simpler argument used in [44]. In the second step, we bound the total influence of an edge in any tree of maximum degree at most Δ by leveraging the associated tree recursions. We formalize these in the following two intermediate theorems.

Theorem 6.1.4 (Reduction from Graphs to Trees)**.** *Fix a graph $G = (V_G, E_G)$, $r \in V_G$, $e \sim r$ incident to r, $f \in E_G$, and $\lambda > 0$. Then there exists a tree $T = T_{\mathrm{SAW}}(G, r) =$*

(V_T, E_T) *such that the following inequality holds:*

$$\sum_{f \in E_G} |\Psi_G(e \to f)| \leq \sum_{g \in E_T} |\Psi_T(e \to g)| .$$

Theorem 6.1.5 (Total Edge Influence in Trees)**.** *Let $T = (V_T, E_T)$ be any tree of maximum degree $\leq \Delta$, $e \in E_T$ be any edge, and fix $\lambda > 0$. Then we have the bound*

$$\sum_{f \in E_T : f \neq e} |\Psi_T(e \to f)| \leq \min\left\{2\lambda\Delta, 2\sqrt{1+\lambda\Delta}\right\}.$$

Remark 6.1.6. We note that this bound on the total influence of an edge is tight for the infinite Δ-regular tree. However, it turns out that bounding the maximum eigenvalue of the influence matrix using the total influence of an edge is not tight for the infinite Δ-regular tree.

Assuming the truth of these two theorems, we now give a straightforward proof of Theorem 6.1.2.

Proof of Theorem 6.1.2. Fix G, $\Lambda \subseteq E_G$ and $\tau : \Lambda \to \{0, 1\}$. Let $H = (V_H, E_H)$ be the graph obtained from G by deleting all edges $e \in \Lambda$ such that $\tau(e) = 1$, and deleting all edges $f \in \Lambda$ along with edges incident to them such that $\tau(f) = 0$. Observe that H is a subgraph of G with maximum degree at most Δ, and crucially, the conditional distribution μ_G^τ is precisely μ_H. By Theorems 6.1.4 and 6.1.5, we have the bound

$$\lambda_1(\Psi_G^\tau) = \lambda_1(\Psi_H) \leq \min\{2\lambda\Delta, 2\sqrt{1+\lambda\Delta}\}.$$

As G, Λ, τ were arbitrary, the claim follows. □

All that remains is to prove Theorems 6.1.4 and 6.1.5. We do this in Sections 6.1.1 and 6.1.2, respectively, noting that the arguments are completely independent of one another.

6.1.1 Reducing Influences in Graphs to Influences in Trees: Proof of Theorem 6.1.4

Fix a graph $G = (V_G, E_G)$ with maximum degree $\leq \Delta$, and a vertex $r \in V_G$. Let $T_{\mathrm{SAW}}(G, r)$ denote the self-avoiding walk tree in G rooted at r; in the context of matchings, this is known as the "path tree" [60], and we refer to Section 5.3 and [60] for formal definitions. Note that we do not impose any boundary conditions on $T_{\mathrm{SAW}}(G, r)$ like in [127]. For every vertex $u \in V_G$, we write $C(u)$ to be the set of copies of u in T. Similarly, for every edge $e \in E_G$, we write $C(e)$ to be the set of copies of e in T.

We prove the following more fine-grained relationship between pairwise influences in G and pairwise influences in $T = T_{\mathrm{SAW}}(G, r)$.

Proposition 6.1.7 (Influence in G to Influence in $T_{\mathrm{SAW}}(G, r)$)**.** *For every graph $G = (V_G, E_G)$, $r \in V_G$, $e \sim r$, $f \in E_G$, and edge activities $x_e \geq 0$, if we let $T = T_{\mathrm{SAW}}(G, r) = (V_T, E_T)$, then we have the identity*

$$\Psi_G(e \to f) = \sum_{f' \in C(f)} \Psi_T(e \to f').$$

We note that by the Triangle Inequality, Proposition 6.1.7 immediately implies Theorem 6.1.4. Hence, it suffices to prove Proposition 6.1.7, which we do by generalizing properties of the univariate matching polynomial.

Define the following multivariate edge-matching polynomial.

$$\mathcal{M}_G(x_e : e \in E_T) = \sum_{M \subseteq E \text{ matching}} \prod_{e \in M} x_e.$$

$\mathcal{M}_G$ is also the partition function of the monomer-dimer model on G with edge activities $x_e \geq 0$. Furthermore, if $r \in V_G$ is arbitrary, and we denote $T = T_{\mathrm{SAW}}(G, r)$, then define

$$\overline{\mathcal{M}}_T(x_e : e \in E_G) \stackrel{\mathrm{def}}{=} \mathcal{M}_T(\overline{x}_f : f \in E_T)$$

where $\overline{x}_f = x_e$ for all $f \in C(e)$ and all $e \in E_G$. We note that while $\mathcal{M}_G$ is always multiaffine, $\overline{\mathcal{M}}_T$ is not. Furthermore, $\mathcal{M}_G$ is not homogeneous. Finally, note that the degree of any edge $e \sim r$ incident to r is 1 in $\overline{\mathcal{M}}_T$ since no self-avoiding walk can reuse e after using e to leave r. We will crucially need the following decomposition of $\mathcal{M}_G$.

Lemma 6.1.8. *For every graph $G = (V_G, E_G)$ and any vertex $v \in V_G$, we have the identity*

$$\mathcal{M}_G(x) = \mathcal{M}_{G-r}(x) + \sum_{v \sim r} x_{rv}\, \mathcal{M}_{G-r-v}(x).$$

Proof. Group the matchings for which r is not saturated in the term $\mathcal{M}_{G-r}(x)$. Similarly, group the matchings for which a fixed edge $e = \{r, v\}$ incident to r is selected in $\mathcal{M}_{G-r-v}(x)$. □

We prove the following, a univariate analog of which was already proved in [60].

Lemma 6.1.9. *For every graph $G = (V_G, E_G)$ and $r \in V_G$, taking $T = T_{\mathrm{SAW}}(G, r)$, we have the identity*

$$\frac{\mathcal{M}_G(x)}{\mathcal{M}_{G-r}(x)} = \frac{\overline{\mathcal{M}}_T(x)}{\overline{\mathcal{M}}_{T-r}(x)}.$$

Furthermore, we may write $\overline{\mathcal{M}}_T(x) = \mathcal{M}_G(x) \cdot q(x)$ for some polynomial q which does not depend on x_e for any $e \sim r$.

First, let us see how to use Lemma 6.1.9 to prove Proposition 6.1.7.

Proof of Proposition 6.1.7. Fix $r \in V_G$, and write $T = T_{\mathrm{SAW}}(G, r)$. By Lemma 6.1.9, we have that $\overline{\mathcal{M}}_T(x) = \mathcal{M}_G(x) \cdot q(x)$ for a polynomial q which does not depend on x_e for all $e \sim r$. It follows that if $e \sim r$, then

$$\mu_T(\sigma_e = 0) = (x_e \partial_{x_e} \log \overline{\mathcal{M}}_T)(x) = (x_e \partial_{x_e} \log \mathcal{M}_G)(x) = \mu_G(\sigma_e = 0).$$

It then also follows that for any $e \sim r$ and any edge $f \in E_G$, we have the identity

$$(x_f x_e \cdot \partial_{x_f} \partial_{x_e} \log \overline{\mathcal{M}}_T)(x) = (x_f x_e \cdot \partial_{x_f} \partial_{x_e} \log \mathcal{M}_G)(x).$$

Now, let us understand the left-hand and right-hand sides separately as influences. For the right-hand side, we have that

$$\begin{aligned}
(x_f x_e \cdot \partial_{x_f} \partial_{x_e} \log \mathcal{M}_G)(x) &= x_e x_f \cdot \partial_{x_f} \frac{(\partial_{x_e} \mathcal{M}_G)(x)}{\mathcal{M}_G(x)} \\
&= x_f x_e \cdot \left(\frac{(\partial_{x_f} \partial_{x_e} \mathcal{M}_G)(x)}{\mathcal{M}_G(x)} - \frac{(\partial_{x_f} \mathcal{M}_G)(x) \cdot (\partial_{x_e} \mathcal{M}_G)(f)}{\mathcal{M}_G(x)^2} \right) \\
&= \mu_G(\sigma_e = 0, \sigma_f = 0) - \mu_G(\sigma_e = 0) \cdot \mu_G(\sigma_f = 0) \\
&= \mu_G(\sigma_e = 0) \cdot (\mu_G(\sigma_f = 0 \mid \sigma_e = 0) - \mu_G(\sigma_f = 0)) \\
&= \mu_G(\sigma_e = 0) \cdot \mu_G(\sigma_e = 1) \cdot \Psi_G(e \to f).
\end{aligned}$$

For the left-hand side, we have by the Chain Rule that

$$\begin{aligned}
(x_f x_e \cdot \partial_{x_f} \partial_{x_e} \log \overline{\mathcal{M}}_T)(x) &= x_f x_e \cdot \partial_{x_f} \frac{(\partial_{x_e} \overline{\mathcal{M}}_T)(x)}{\overline{\mathcal{M}}_T(x)} \\
&= \sum_{f' \in C(f)} x_f x_e \cdot \partial_{\overline{x}_{f'}} \frac{(\partial_{x_e} \mathcal{M}_T)(\overline{x})}{\mathcal{M}_T(\overline{x})} \Bigg|_{\overline{x}=x} \\
&= \sum_{f' \in C(f)} \mu_T(\sigma_e = 0) \cdot \mu_T(\sigma_e = 1) \cdot \Psi_T(e \to f').
\end{aligned}$$

Since $\mu_G(\sigma_e = 0) = \mu_T(\sigma_e = 0)$ and $\mu_G(\sigma_e = 1) = \mu_T(\sigma_e = 1)$, the claim follows. □

All that remains is to prove Lemma 6.1.9.

Proof of Lemma 6.1.9. We go by induction on the graph. First, we note that the claim is trivial in the case where G itself is a tree, since then $T = G$ and $\mathcal{M}_G = \overline{\mathcal{M}}_T$ (i.e. q is

identically 1). This forms our base case. Now, by Lemma 6.1.8 we may write

$$\mathcal{M}_G(x) = \mathcal{M}_{G-r}(x) + \sum_{v \sim r} x_{rv}\, \mathcal{M}_{G-r-v}(x)$$

where we note the polynomials $\mathcal{M}_{G-r}(x), \mathcal{M}_{G-r-v}(x)$ do not depend on any x_e for $e \sim r$. Therefore, we deduce that

$$\begin{aligned}
\frac{\mathcal{M}_G(x)}{\mathcal{M}_{G-r}(x)} &= 1 + \sum_{v \sim r} x_{rv} \cdot \frac{\mathcal{M}_{G-r-v}(x)}{\mathcal{M}_{G-r}(x)} \\
&= 1 + \sum_{v \sim r} x_{rv} \cdot \frac{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G-r,v)-v}(x)}{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G-r,v)}(x)} && \text{(Induction)} \\
&= 1 + \sum_{v \sim r} x_{rv} \cdot \frac{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)-r-v}(x)}{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)-r}(x)} \\
&\qquad\qquad (T_{\mathrm{SAW}}(G-r,v) \cong \text{subtree of } T_{\mathrm{SAW}}(G,r) \text{ rooted at } v) \\
&= \frac{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)-r}(x) + \sum_{v \sim r} x_{rv} \overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)-r-v}(x)}{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)-r}(x)} \\
&= \frac{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)}(x)}{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)-r}(x)}.
\end{aligned}$$

This proves the first claim. For the second claim, we go by induction again. The base case where G is a tree is again immediate. For the inductive step, we have

$$\begin{aligned}
\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)}(x) &= \mathcal{M}_G(x) \cdot \frac{\overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G,r)-r}(x)}{\mathcal{M}_{G-r}(x)} \\
&= \mathcal{M}_G(x) \cdot \frac{\prod_{v \sim u} \overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G-r,v)}(x)}{\mathcal{M}_{G-r}(x)} \\
&\qquad (\text{Deleting } r \text{ in } T_{\mathrm{SAW}}(G,r) \text{ disconnects the subtrees } T_{\mathrm{SAW}}(G-r,v).) \\
&= \mathcal{M}_G(x) \cdot q(x). \\
&\qquad (\text{Induction: } \mathcal{M}_{G-r}(x) \text{ divides } \overline{\mathcal{M}}_{T_{\mathrm{SAW}}(G-r,v)}(x) \text{ for any } v)
\end{aligned}$$

This shows the lemma. □

6.1.2 The Total Influence in a Tree: Proof of Theorem 6.1.5

At this point, we can forget self-avoiding walk trees, and just focus on the special case where G itself is a tree T. Throughout, we assume our tree T has maximum degree at most Δ. If $e \in T$ with endpoints r_1, r_2, then we may view T as two trees $T(r_1), T(r_2)$ on disjoint sets of vertices which are connected by the edge e, with $T(r_1)$ being rooted at r_1 and $T(r_2)$ being rooted at r_2. If v is a vertex in $T(r_1)$ (resp. $T(r_2)$), we write $T(v)$ for the subtree of $T(r_1)$ (resp. $T(r_2)$) rooted at v. We also write $L_v(k)$ for the set of descendants of v (in $T(v)$) at distance exactly k from v. We will let $\mu_T(r)$ to denote the probability that the vertex $r \in V_T$ is saturated in a random matching drawn from the Gibbs distribution. Similarly, we will write $\mu_T(\overline{r}) = 1 - \mu_T(r)$.

We now prove several intermediate technical results which we will use to deduce Theorem 6.1.5. To state them, we will need the following recursion for the probabilities $\mu_G(\overline{r})$ in the monomer-dimer model on G with fugacity $\lambda \geq 0$:

$$\mu_G(\overline{r}) = \frac{1}{1 + \lambda \sum_{v \sim r} \mu_{G-r}(\overline{v})} \stackrel{\text{def}}{=} F_\lambda(\mu_{G-r}(\overline{v}) : v \sim r).$$

We note this is immediate from Lemma 6.1.8 and using that $\mu_G(\overline{r}) = \frac{\mathcal{M}_{G-r}(x)}{\mathcal{M}_G(x)}$ where we take $x = \lambda \mathbf{1}$.

Proposition 6.1.10. *Fix a tree T and an edge $e \in E_T$ with endpoints r_1, r_2. Then we have the bound*

$$\sum_{f \in E_T : f \neq e} |\Psi_T(e \to f)| \leq 2 \sum_{k=1}^{\infty} \max \left\{ \prod_{i=1}^{k} \mu_{T(u_i)}(u_i) \;\middle|\; u_1 \in e,\, u_{i+1} \in L_{u_i}(1),\, \forall i = 1, \ldots, k-1 \right\}$$

where we write $e = e_1, u_1, e_2, u_2, \ldots, u_k, e_{k+1} = f$ for the unique path from e to f such that edge e_i connects vertices u_{i-1} and u_i.

Proposition 6.1.11. *Consider the potential function $\Phi(x) = \log x$ with derivative $\varphi(x) = \frac{1}{x}$. Then for every tree T rooted at r and every sequence of vertices $r = u_1, \ldots, u_k$, we have the inequality*

$$\prod_{i=1}^{k} \mu_{T(u_i)}(u_i) \leq \min\left\{ \left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k, \sup_y \left\|\nabla(\Phi \circ F_\lambda^{\circ 2} \circ \Phi^{-1})(y)\right\|_1^{\lfloor k/2 \rfloor} \right\}$$

for every k.

Given Proposition 6.1.11, we will also need a bound on the gradient norm. Conveniently, this gradient norm was already analyzed in [19] to establish the correlation decay property.

Lemma 6.1.12 ([19, Lemma 3.3])**.** *Consider the potential function $\Phi(x) = \log x$. Then we have the following bound on the norm of the gradient for the two-step log-marginal recursion.*

$$\sup_y \left\|\nabla(\Phi \circ F_\lambda^{\circ 2} \circ \Phi^{-1})(y)\right\|_1 \leq 1 - \frac{2}{\sqrt{1+\lambda\Delta}+1}.$$

We now show how to use intermediate technical results to prove Theorem 6.1.5. Immediately following, we will prove Propositions 6.1.10 and 6.1.11.

Proof of Theorem 6.1.5. By combining Propositions 6.1.10 and 6.1.11 and Lemma 6.1.12, we have the following two bounds

$$\begin{aligned}
\sum_{f \in E_T : f \neq e} |\Psi_T(e \to f)| &\leq 2\sum_{k=1}^{\infty} \left(1 - \frac{2}{\sqrt{1+\lambda\Delta}+1}\right)^{\lfloor k/2 \rfloor} \\
&= -2 + 4\sum_{k=0}^{\infty} \left(1 - \frac{2}{\sqrt{1+\lambda\Delta}+1}\right)^{k} \\
&= 2\sqrt{1+\lambda\Delta}; \\
\sum_{f \in E_T : f \neq e} |\Psi_T(e \to f)| &\leq 2\sum_{k=1}^{\infty} \left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k = 2\frac{\lambda\Delta}{1+\lambda\Delta} \cdot \frac{1}{1 - \frac{\lambda\Delta}{1+\lambda\Delta}} = 2\lambda\Delta.
\end{aligned}$$

This proves the theorem. □

Proof of Proposition 6.1.10

The crux of the proof rests on the following factorization of the pairwise influence in trees, which was already observed in prior work [4, 44] in the context of vertex-spin systems.

Lemma 6.1.13 (Factorization of Pairwise Influence in Trees)**.** *Fix two edges $e, f \in E_T$. Let $e = e_1, u_1, e_2, u_2, \ldots, u_k, e_{k+1} = f$ be the unique path in T from e to f, where edge e_i connects vertices u_{i-1} and u_i. Then we have*

$$\Psi_T(e \to f) = \prod_{i=1}^{k} \Psi_T(e_i \to e_{i+1}).$$

Proof of Lemma 6.1.13. It suffices to show that if g is any edge on the unique path from e to f, then $\Psi_T(e \to f) = \Psi_T(e \to g) \cdot \Psi_T(g \to f)$; the full claim then follows by induction. This simpler identity follows immediately from the fact that conditioning on g disconnects e from f so that they become independent. □

Now, we finish the proof of Proposition 6.1.10. Observe that

$$\begin{aligned}
\sum_{f \in E_T : f \neq e} |\Psi_T(e \to f)| &= \sum_{k=1}^{\infty} \left(\sum_{\substack{f \in E_{T(r_1)} \\ \operatorname{dist}(e,f)=k}} + \sum_{\substack{f \in E_{T(r_2)} \\ \operatorname{dist}(e,f)=k}} \right) \prod_{i=1}^{k} |\Psi_T(e_i \to e_{i+1})| \\
&= \sum_{k=1}^{\infty} \left(\sum_{\substack{f \in E_{T(r_1)} \\ \operatorname{dist}(e,f)=k}} + \sum_{\substack{f \in E_{T(r_2)} \\ \operatorname{dist}(e,f)=k}} \right) \prod_{i=1}^{k} \mu_{T-e_i}(\sigma_{e_{i+1}} = 0) \\
&\leq 2 \max_{r \in e} \sum_{k=1}^{\infty} \sum_{\substack{f \in E_{T(r)} \\ \operatorname{dist}(e,f)=k}} \prod_{i=1}^{k} \mu_{T-e_i}(\sigma_{e_{i+1}} = 0)
\end{aligned}$$

and hence, to prove the claim, it suffices to show that for each $r \in e = \{r_1, r_2\}$ and each

positive integer k, we have

$$\sum_{\substack{f\in E_{T(r)}\\ \mathrm{dist}(e,f)=k}} \prod_{i=1}^{k} \mu_{T-e_i}(\sigma_{e_{i+1}}=0) \leq \max\left\{ \prod_{i=1}^{k} \mu_{T(u_i)}(u_i) \;\middle|\; u_1 = r,\, u_{i+1}\in L_{u_i}(1),\, \forall i = 1,\dots,k-1 \right\}.$$

To show this, we prove the following more general inequality

$$\sum_{\substack{f\in E_{T(r)}\\ \mathrm{dist}(e,f)=k}} \prod_{i=1}^{k} \mu_{T-e_i}(\sigma_{e_{i+1}}=0) \leq \sum_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall 1\leq i\leq j}} \prod_{i=1}^{j} \mu_{T(u_i)}(\sigma_{\{u_i,u_{i+1}\}}=0)\cdot \max_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall j+1\leq i\leq k}} \left\{ \prod_{i=j+1}^{k} \mu_{T(u_i)}(u_i) \right\}$$

for all j, by induction. The desired inequality is the special case $j = 0$ and the base case $j = k$ holds trivially with equality. Assume the inequality holds for some $j + 1 \leq k$. We now prove that it holds for j. This follows by:

$$\begin{aligned}
&\sum_{\substack{f\in E_{T(r)}\\ \mathrm{dist}(e,f)=k}} \prod_{i=1}^{k} \mu_{T-e_i}(\sigma_{e_{i+1}}=0)\\
&\leq \sum_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall 1\leq i\leq j+1}} \prod_{i=1}^{j+1} \mu_{T(u_i)}(\sigma_{\{u_i,u_{i+1}\}}=0) \max_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall j+2\leq i\leq k}} \left\{ \prod_{i=j+2}^{k} \mu_{T(u_i)}(u_i) \right\}\\
&= \sum_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall 1\leq i\leq j}} \prod_{i=1}^{j} \mu_{T(u_i)}(\sigma_{\{u_i,u_{i+1}\}}=0)\\
&\qquad \times \sum_{u_{j+2}\in L_{u_{j+1}}(1)} \mu_{T(u_{j+1})}(\sigma_{\{u_{j+1},u_{j+2}\}}=0) \max_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall j+2\leq i\leq k}} \left\{ \prod_{i=j+2}^{k} \mu_{T(u_i)}(u_i) \right\}\\
&\leq \sum_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall 1\leq i\leq j}} \prod_{i=1}^{j} \mu_{T(u_i)}(\sigma_{\{u_i,u_{i+1}\}}=0)
\end{aligned}$$

$$\times \max_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall j+1\le i\le k}} \left\{\prod_{i=j+2}^{k} \mu_{T(u_i)}(u_i)\right\} \underbrace{\sum_{u_{j+2}\in L_{u_{j+1}}(1)} \mu_{T(u_{j+1})}(\sigma_{\{u_{j+1},u_{j+2}\}}=0)}_{=\mu_{T(u_{j+1})}(u_{j+1})}$$

$$= \sum_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall 1\le i\le j}} \prod_{i=1}^{j} \mu_{T(u_i)}(\sigma_{\{u_i,u_{i+1}\}}=0) \max_{\substack{u_{i+1}\in L_{u_i}(1)\\ \forall j+1\le i\le k}} \left\{\prod_{i=j+1}^{k} \mu_{T(u_i)}(u_i)\right\}.$$

This completes the proof of the proposition.

Proof of Proposition 6.1.11

We first prove the upper bound of $\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k$ as it is simpler and does not require the use of the potential function Φ. To see this, observe that

$$\mu_G(\overline{r}) = \frac{1}{1+\lambda\sum_{v\sim r}\mu_{G-r}(\overline{v})} \ge \frac{1}{1+\lambda\Delta}$$

and hence $\mu_G(r) \le \frac{\lambda\Delta}{1+\lambda\Delta}$. It follows that

$$\prod_{i=1}^{k} \mu_{T(u_i)}(u_i) \le \left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k$$

for any path $u_1,\ldots,u_k$ in T starting from the root. Applying Proposition 6.1.10, we have that the total influence is upper bounded by

$$2\sum_{k=1}^{\infty}\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k = \frac{2\lambda\Delta}{1+\lambda\Delta}\cdot\frac{1}{1-\frac{\lambda\Delta}{1+\lambda\Delta}} = 2\lambda\Delta.$$

It turns out, one may view this simple analysis as a "one-step" version of the proof of Proposition 6.1.11.

Composing with the recursion F_λ yields the following recursion for the log-marginals

$y_i = \log p_i = \Phi(p_i)$:

$$(\Phi \circ F_\lambda \circ \Phi^{-1})(y) = -\log\left(1 + \lambda \sum_{i=1}^{d} \exp(y_i)\right);$$
$$(\Phi \circ F_\lambda^{\circ 2} \circ \Phi^{-1})(y) = -\log\left(1 + \lambda \sum_{i=1}^{d} \frac{1}{1 + \lambda \sum_{j=1}^{d} \exp(y_{ij})}\right).$$

Differentiating and applying the Inverse Function Theorem, we obtain that

$$(\partial_{p_i} F_\lambda)(p) = -\frac{\lambda}{\left(1 + \lambda \sum_{j=1}^{d} p_j\right)^2} = -\lambda F_\lambda(p)^2;$$

and

$$\begin{aligned}\partial_{y_i}(\Phi \circ F_\lambda \circ \Phi^{-1})(y) &= \frac{(\Phi \circ F_\lambda)(p)}{\Phi(p_i)} \cdot (\partial_{p_i} F_\lambda)(p) \\ &= -\frac{p_i}{F_\lambda(p)} \cdot \lambda F_\lambda(p)^2 = -\lambda \cdot p_i \cdot F_\lambda(p);\end{aligned}$$

and

$$\begin{aligned}\partial_{y_{ij}}(\Phi \circ F_\lambda^{\circ 2} \circ \Phi^{-1})(y) &= \frac{(\Phi \circ F_\lambda^{\circ 2})(p)}{\Phi(p_{ij})} \cdot (\partial_{p_{ij}} F_\lambda^{\circ 2})(y) \\ &= \frac{p_{ij}}{F_\lambda^{\circ 2}(p)} \cdot ((\partial_{p_j} F_\lambda) \circ F_\lambda)(p) \cdot (\partial_{p_i} F_\lambda)(p) \\ &= \frac{p_{ij}}{F_\lambda^{\circ 2}(p)} \cdot \left(-\lambda (F_\lambda \circ F_\lambda)(p)^2\right) \cdot \left(-\lambda F_\lambda(p)^2\right) \\ &= (\lambda \cdot p_{ij} \cdot F_\lambda(p)) \cdot \left(\lambda \cdot F_\lambda(p) \cdot F_\lambda^{\circ 2}(p)\right).\end{aligned}$$

With this, we now write

$$\begin{aligned}&\prod_{i=1}^{k} \mu_{T(u_i)}(u_i) \\ &= \prod_{i=1}^{k} \sum_{u_{i+1} \in L_{u_i}(1)} \lambda \cdot \mu_{T(u_i)}(\overline{u_i}) \cdot \mu_{T(u_{i+1})}(\overline{u_{i+1}})\end{aligned}$$

$$
\begin{aligned}
&\leq \prod_{i=1}^{\lfloor k/2 \rfloor} \left(\sum_{v \in L_{u_{2i-1}}(1)} \lambda \cdot \mu_{T(u_{2i-1})}(\overline{u_{2i-1}}) \cdot \mu_{T(v)}(\overline{v}) \right) \\
&\qquad\qquad \cdot \left(\sum_{w \in L_{u_{2i}}(1)} \lambda \cdot \mu_{T(u_{2i})}(\overline{u_{2i}}) \cdot \mu_{T(w)}(\overline{w}) \right) \\
&\leq \prod_{i=1}^{\lfloor k/2 \rfloor} \left(\sum_{v \in L_{u_{2i-1}}(1)} \sum_{w \in L_v(1)} \left(\lambda \cdot \mu_{T(u_{2i-1})}(\overline{u_{2i-1}}) \cdot \mu_{T(v)}(\overline{v}) \right) \cdot \left(\lambda \cdot \mu_{T(v)}(\overline{v}) \cdot \mu_{T(w)}(\overline{w}) \right) \right) \\
&\leq \sup_y \left\| \nabla (\Phi \circ F_\lambda^{\circ 2} \circ \Phi^{-1})(y) \right\|_1^{\lfloor k/2 \rfloor},
\end{aligned}
$$

as claimed.

6.2 Optimal Mixing Results for Colorings on Triangle-free Graphs

Let $\alpha^* \approx 1.763$ be the solution to $\exp(1/x) = x$; this threshold has appeared in several related results for colorings, though obtaining corresponding algorithms has been challenging. For example, for $\alpha > \alpha^*$, Gamarnik, Katz, and Misra [59] proved SSM on triangle-free graphs when $q > \alpha\Delta + \beta$ for some constant $\beta = \beta(\alpha)$; see also [61] for a related result on amenable graphs. It was not until recently that the SSM result of [59] was converted to an algorithm for triangle-free graphs by Liu, Sinclair, and Srivastava [92] utilizing the complex zeros approach; however, just as for 2-spin systems, the polynomial exponent in the running time depends exponentially on Δ and the distance of α from α^*.

Our main contribution is to develop the spectral independence approach for colorings, and analyze Glauber dynamics in the regime $q \geq \alpha\Delta + 1$ for all $\alpha > \alpha^*$ on triangle-free graphs. Our result applies for all Δ and we show that the bound on spectral independence does not depend on Δ and q, yielding substantially faster randomized algorithms for sampling/counting colorings than the previous deterministic ones (at the expense of using randomness). The following result shows the second part of Theorem 1.1.4.

Theorem 6.2.1 (Colorings of Triangle-Free Graphs)**.** *Let $\alpha^* \approx 1.763$ denote the solution to $\exp(1/x) = x$. For all $\alpha > \alpha^*$, for any triangle-free graph $G = (V, E)$ with maximum*

degree $\Delta \geq 3$ and any integer $q \geq \alpha\Delta + 1$, the mixing time of the Glauber dynamics on G with q colors is at most $Cn \log n$, where $C = C(\alpha, \Delta)$ is a constant independent of n.

We remark that the constant C is a function of the gap $\alpha - \alpha^*$ and is independent of q.

6.2.1 Preliminaries for Colorings and List-Colorings

Let $q \geq 3$ be an integer and denote by $[q] := \{1, \ldots, q\}$. To apply Theorem 1.2.1, we need to establish the following theorem; Theorem 6.2.1 then follows immediately.

Theorem 6.2.2. *Let $\alpha^* \approx 1.763$ denote the solution to $\exp(1/x) = x$. For all $\alpha > \alpha^*$, there exists $\eta = \eta(\alpha) > 0$ such that, for any triangle-free graph $G = (V, E)$ with maximum degree Δ and any integer $q \geq \alpha\Delta + 1$, the uniform distribution μ over all proper q-colorings of G is η-spectrally independent.*

To establish spectral independence for colorings we also need the more general notion of list-colorings. A list-coloring instance is a pair (G, L) where $G = (V, E)$ is a graph and $L = \{L(v)\}_{v \in V}$ prescribes a list $L(v) \subseteq [q]$ of available colors for each $v \in V$; it will also be convenient to assume that the vertices of G are ordered by some relation $<$ (the ordering itself does not matter). A proper list-coloring for the instance (G, L) is an assignment $\sigma : V \to [q]$ such that $\sigma_v \in L(v)$ for each $v \in V$ and $\sigma_v \neq \sigma_w$ for each $\{v, w\} \in E$. The instance is satisfiable iff such a proper list-coloring exists. Note, q-colorings corresponds to the special case where $L(v) = [q]$ for each $v \in V$. For a satisfiable list-coloring instance (G, L), we will denote by $U_{G,L}$ the set $\{(v, i) \mid v \in V, i \in L(v)\}$, by $\Omega_{G,L}$ the set of all proper list-colorings, and by $\mathbb{P}_{G,L}$ the uniform distribution over $\Omega_{G,L}$; we will omit G from notations when it is clear from context. We typically use σ to denote a random list-coloring that is distributed according to $\mathbb{P}_{G,L}$.

We will be interested in analyzing the Glauber dynamics on $\Omega_{G,L}$. This is a Markov chain $(Z_t)_{t \geq 0}$ of list-colorings which starts from an arbitrary $Z_0 \in \Omega_{G,L}$ and at each time $t \geq 0$ updates the current list-coloring Z_t to Z_{t+1} by selecting a vertex $v \in V$ u.a.r. and

setting $Z_{t+1}(v) = c$, where c is a color chosen u.a.r. from the set $L(v)\backslash Z_t(N_G(v))$; for a vertex $w \neq v$, the color of w is unchanged, i.e., $Z_{t+1}(w) = Z_t(w)$. The transition matrix of the Glauber dynamics will be denoted by $\mathcal{P} = \mathcal{P}_{G,L}$.

To ensure satisfiability of (G, L) as well as ergodicity of the Glauber dynamics, we will henceforth assume the well-known condition that $|L(v)| \geq \Delta_G(v) + 2$ for all $v \in V$, where $\Delta_G(v) = |N_G(v)|$ and $N_G(v)$ is the set of neighbors of v in G. Then, Glauber dynamics converges to the uniform distribution over $\Omega_{G,L}$.

Remark 6.2.3. To ensure satisfiability, it suffices to have the assumption $|L(v)| \geq \Delta_G(v)+1$ for all $v \in V$; in fact, for every $v \in V$ and $i \in L(v)$ there exists a list-coloring σ of (G, L) with $\sigma_v = i$. The slightly stronger condition $|L(v)| \geq \Delta_G(v) + 2$ for every $v \in V$ ensures that any two list-colorings σ, τ are "connected" by a sequence of list-colorings where consecutive list-colorings differ at the color of a single vertex. (A clique with $q + 1$ vertices gives a counterexample to this latter property for q-colorings).

For the spectral independence approach, we will need to consider conditional distributions of $\mathbb{P}_{G,L}$ given a partial list-coloring on a subset of vertices. For a partial list-coloring τ on a subset $S \subseteq V$, let (G_τ, L_τ) be the list-coloring instance on the induced subgraph $G[V\backslash S]$ with lists obtained from L by removing the unavailable colors that have been assigned by τ for each vertex in $V\backslash S$, i.e., $L_\tau = \{L_\tau(v)\}_{v \in V\backslash S}$ where for $v \in V\backslash S$ we have $L_\tau(v) = L(v)\backslash\tau(N_G(v) \cap S)$.

To capture those instances of list-colorings obtained from an instance of q-colorings by assigning fixed colors to a subset of vertices, the following notion of (Δ, q)-list-colorings will be useful.

Definition 6.2.4. Let Δ, q be positive integers with $\Delta \geq 3$ and $q \geq \Delta + 2$. We say that (G, L) is a (Δ, q)-list-coloring instance if $G = (V, E)$ has maximum degree Δ and for each $v \in V$ it holds that $L(v) \subseteq [q]$ and $|L(v)| \geq q - \Delta + \Delta_G(v)$.

In our proofs henceforth, it will be convenient to define the following slightly more accurate form of the region of (Δ, q) where our results apply to.

Definition 6.2.5 (Parameter Region Λ_ε)**.** Let $\alpha^* \approx 1.763$ denote the solution to $\exp(1/x) = x$. For $\varepsilon > 0$, define $\Lambda_\varepsilon = \{(\Delta, q) \in \mathbb{N}^2 \mid \Delta \geq 3,\ q \geq \alpha\Delta + \beta\}$ where $\alpha = (1+\varepsilon)\alpha^*$ and $\beta = 2 - \alpha + \frac{\alpha}{2(\alpha^2-1)} < 0.655$.

The core of our argument behind the proof of Theorem 6.2.2 is to establish the following bound on $\lambda_1(\Psi)$ by studying the list-coloring distribution.

Theorem 6.2.6. *Let $\varepsilon > 0$ be arbitrary, and suppose that (G, L) is a (Δ, q)-list-coloring instance with $q \geq (1+\varepsilon)\alpha^*\Delta + 1$ and G a triangle-free graph. Then, $\lambda_1(\Psi) \leq 8\left(\frac{1}{\varepsilon} + 1\right)\frac{\Delta}{q}$ where $\Psi = \Psi_{G,L}$ is the influence matrix.*

Theorem 6.2.2 is a consequence of Theorem 6.2.6 by definitions of spectral independence.

6.2.2 Establishing Spectral Independence: Proof of Theorem 6.2.6

Let (G, L) be a (Δ, q)-list-coloring instance as in Theorem 6.2.6. Let $\Psi = \Psi_{G,L}$ be the square matrix with indices from the set $U_{G,L}$, where the entry indexed by $(v, i), (w, k) \in U_{G,L}$ is 0 if $v = w$, and

$$\Psi\big((v,i),(w,k)\big) = \mathbb{P}_{G,L}(\sigma_w = k \mid \sigma_v = i) - \mathbb{P}_{G,L}(\sigma_w = k), \quad \text{if } v \neq w.$$

Our goal is to bound the spectral radius of the influence matrix $\Psi_{G,L}$. In this section, we will show that for $\alpha > \alpha^*$ there exists a constant $C = C(\alpha)$ such that whenever $q \geq \alpha\Delta + 1$ it holds that $\lambda_1(\Psi_{G,L}) \leq C\Delta/q$.

Henceforth, it will be convenient to extend $\Psi_{G,L}$ by setting $\Psi_{G,L}((v,i),(w,k)) = 0$ when $k \notin L(w)$ or $i \notin L(v)$. It is well-known that, for any square matrix the spectral radius is bounded by the maximum of the L_1-norms of the rows. In our setting, the bound on $\lambda_1(\Psi_{G,L})$ will therefore be obtained by showing that, for an arbitrary vertex v of G and

a color $i \in L(v)$, it holds that

$$\sum_{w \in V \setminus \{v\}} \sum_{k \in [q]} \left| \Psi_{G,L}\big((v,i),(w,k)\big) \right| \leq 8 \left(\frac{1}{\varepsilon} + 1 \right) \frac{\Delta}{q}. \tag{6.1}$$

To bound the sum in Eq. (6.1), we introduce the *maximum influence*, which describes the maximum difference of the marginal probability at w under all color choices of v.

Definition 6.2.7 (Maximum Influences)**.** Let (G, L) be a (Δ, q)-list-coloring instance. Let v, w be two vertices of G. The maximum influence of v on w is defined to be

$$\Psi_{G,L}[v \to w] = \max_{i,j \in L(v)} \max_{Q \subseteq [q]} \left| \mathbb{P}_{G,L}(\sigma_w \in Q \mid \sigma_v = i) - \mathbb{P}_{G,L}(\sigma_w \in Q \mid \sigma_v = j) \right|.$$

Lemma 6.2.8. *For all distinct $v, w \in V$ and $i \in L(v)$ we have*

$$\sum_{k \in [q]} |\Psi_{G,L}((v,i),(w,k))| \leq 2 \cdot \Psi_{G,L}[v \to w].$$

Proof. Let $\pi \in \mathbb{R}^q$ such that $\pi(k) = \mathbb{P}_{G,L}(\sigma_w = k)$ for each $k \in [q]$. For $i \in L(v)$, let $\pi_i \in \mathbb{R}^q$ such that $\pi_i(k) = \mathbb{P}_{G,L}(\sigma_w = k \mid \sigma_v = i)$ for each $k \in [q]$. Finally, let $\nu(i) = \mathbb{P}_{G,L}(\sigma_v = i)$ for each $i \in L(v)$. It then follows that

$$\sum_{k \in [q]} |\Psi_{G,L}((v,i),(w,k))| = \sum_{k \in [q]} |\pi_i(k) - \pi(k)| = \|\pi_i - \pi\|_1 .$$

Meanwhile, by the law of total probability we have $\pi(k) = \sum_{j \in L(v)} \nu(j)\pi_j(k)$ for all $k \in [q]$. Hence, $\pi = \sum_{j \in L(v)} \nu(j)\pi_j$. We then deduce from the triangle inequality that

$$\|\pi_i - \pi\|_1 = \left\| \sum_{j \in L(v)} \nu(j)(\pi_i - \pi_j) \right\|_1 \leq \sum_{j \in L(v)} \nu(j) \|\pi_i - \pi_j\|_1 \leq \max_{i,j \in L(v)} \|\pi_i - \pi_j\|_1 .$$

Notice that

$$\frac{1}{2}\left\|\pi_i - \pi_j\right\|_1 = d_{\mathrm{TV}}(\pi_i, \pi_j) = \max_{Q \subseteq [q]} \left|\mathbb{P}_{G,L}(\sigma_w \in Q \mid \sigma_v = i) - \mathbb{P}_{G,L}(\sigma_w \in Q \mid \sigma_v = j)\right|.$$

The lemma then follows immediately. □

Hence, to bound the sum in Eq. (6.1), it suffices to bound the sum $\sum_{w \in V \setminus \{v\}} \Psi_{G,L}[v \to w]$ instead. Our ultimate goal is to write a recursion for this latter sum, bounding by an analogous sum for the neighbors of v (in the graph where v is deleted). To get on the right track, we start by writing a recursion for influences.

Definition 6.2.9. Let (G, L) be a list-coloring instance where $G = (V, E)$ is a graph and $L = \{L(v)\}_{v \in V}$ is a collection of color lists. Let $v \in V$. For $u \in N_G(v)$ and colors $i, j \in L(v)$ with $i \neq j$, we denote by (G_v, L_u^{ij}) the list-coloring instance with $G_v = G \setminus v$ and lists $L_u^{ij} = \{L_u^{ij}(w)\}_{w \in V \setminus \{v\}}$ obtained from L by:

- removing the color i from the lists $L(u')$ for $u' \in N_G(v)$ with $u' < u$,
- removing the color j from the lists $L(u')$ for $u' \in N_G(v)$ with $u' > u$, and
- keeping the remaining lists unchanged.

The following lemma will be crucial in our recursive approach to bound influences, and follows by adapting suitably ideas from [59]. The proof of it can be found in Section 6.2.3.

Lemma 6.2.10. *Let (G, L) be a (Δ, q)-list-coloring instance with $G = (V, E)$ and $L = \{L(v)\}_{v \in V}$. Then, for $v \in V$ and arbitrary colors $i, j \in L(v)$ with $i \neq j$, for all $w \in V \setminus \{v\}$ and $k \in [q]$, we have*

$$\mathbb{P}(\sigma_w = k \mid \sigma_v = i) - \mathbb{P}(\sigma_w = k \mid \sigma_v = j) = \sum_{u \in N_G(v)} \frac{\mathbb{P}_u^{ij}(\sigma_u = j)}{\mathbb{P}_u^{ij}(\sigma_u \neq j)} \cdot \Psi_u^{ij}\big((u, j), (w, k)\big) - \frac{\mathbb{P}_u^{ij}(\sigma_u = i)}{\mathbb{P}_u^{ij}(\sigma_u \neq i)} \cdot \Psi_u^{ij}\big((u, i), (w, k)\big),$$

where $\mathbb{P} := \mathbb{P}_{G,L}$ *and, for* $u \in N_G(v)$, $\mathbb{P}_u^{ij} := \mathbb{P}_{G_v, L_u^{ij}}$ *and* $\Psi_u^{ij} := \Psi_{G_v, L_u^{ij}}$.

Recall that we set $\Psi_u^{ij}\big((u,c),(w,k)\big) = 0$ for $c \notin L_u^{ij}(u)$. To apply Lemma 6.2.10 recursively, it will be helpful to consider multiple list-coloring instances on the same graph G. For a collection of lists $\mathcal{L} = \{L_1, \ldots, L_t\}$, where each $L \in \mathcal{L}$ is a set of lists of all vertices for G, we use $(G, \mathcal{L})$ to denote the collection of $|\mathcal{L}|$ list-coloring instances $\{(G, L_1), \ldots, (G, L_t)\}$. When considering the pair $(G, \mathcal{L})$ or (G, L), we usually omit the graph G when it is clear from the context.

Definition 6.2.11. Let $(G, \mathcal{L})$ be a collection of list-colorings instances with $G = (V, E)$ and a collection of lists $\mathcal{L}$ on G. For $v \in V$, we define $\mathcal{L}_v$ to be the collection of lists for $G_v = G \setminus v$ obtained from $\mathcal{L}$ by setting

$$\mathcal{L}_v = \big\{L_u^{ij} \mid L \in \mathcal{L}, u \in N_G(v), i, j \in L(v) \text{ with } i \neq j\big\}.$$

Note that $(G_v, \mathcal{L}_v)$ consists of $|\mathcal{L}_v| = \sum_{L \in \mathcal{L}} \Delta_G(v) \cdot |L(v)| \cdot (|L(v)| - 1)$ list-coloring instances.

Lemma 6.2.12. *If* $(G, \mathcal{L})$ *is a collection of* (Δ, q)*-list-coloring instances, then for every vertex* v *of* G, $(G_v, \mathcal{L}_v)$ *is also a collection of* (Δ, q)*-list-coloring instances.*

Proof. Let $L_v \in \mathcal{L}_v$ be arbitrary, so that L_v is obtained from some $L \in \mathcal{L}$. Then, by definition, for $u \notin N_G(v)$ we have $|L_v(u)| = |L(u)|$ and $\Delta_{G \setminus v}(u) = \Delta_G(u)$, while for $u \in N_G(v)$ we have $|L_v(u)| \geq |L(u)| - 1$ and $\Delta_{G_v}(u) = \Delta_G(u) - 1$. This implies that $\mathcal{L}_v$ is (Δ, q)-induced. □

Definition 6.2.13. Let $(G, \mathcal{L})$ be a collection of (Δ, q)-list-coloring instances with $G = (V, E)$. Fix a vertex $v \in V$ and let $w \in V \setminus \{v\}$. The maximum influence of v on w with respect to $(G, \mathcal{L})$ is defined to be

$$\Psi_{G,\mathcal{L}}[v \to w] = \max_{L \in \mathcal{L}} \Psi_{G,L}[v \to w].$$

The *total* maximum influence of v with respect to $(G, \mathcal{L})$ is defined to be 0 if $\Delta_G(v) = 0$, and

$$\Psi^*_{G,\mathcal{L}}(v) = \frac{1}{\Delta_G(v)} \sum_{w \in V \setminus \{v\}} \Psi_{G,\mathcal{L}}[v \to w] \quad \text{if } \Delta_G(v) \geq 1.$$

The following lemma gives a recursive bound on the total maximum influence.

Lemma 6.2.14. *Let $(G, \mathcal{L})$ be a collection of list-coloring instances and v be a vertex of G with $\Delta_G(v) \geq 1$. Then, with $G_v, \mathcal{L}_v$ as in Definition 6.2.11,*

$$\Psi^*_{G,\mathcal{L}}(v) \leq \max_{u \in N_G(v)} \Big\{ R_{G_v,\mathcal{L}_v}(u) \big(\Delta_{G_v}(u) \cdot \Psi^*_{G_v,\mathcal{L}_v}(u) + 1 \big) \Big\},$$

where $R_{G_v,\mathcal{L}_v}(u) = \max_{L \in \mathcal{L}_v} \max_{c \in L(u)} \frac{\mathbb{P}_{G_v,L}(\sigma_u = c)}{\mathbb{P}_{G_v,L}(\sigma_u \neq c)}$ for $u \in N_G(v)$.

Proof. Suppose that $G = (V, E)$. For convenience, we will drop the subscripts $G, \mathcal{L}$ from influences and use the subscript v as a shorthand for the subscripts $G_v, \mathcal{L}_v$ of influences and the quantity R. We will soon show that for every $w \in V \setminus \{v\}$, we have

$$\Psi[v \to w] \leq \sum_{u \in N_G(v)} R_v(u) \cdot \Psi_v[u \to w]. \tag{6.2}$$

Assuming Eq. (6.2) for the moment, we have that

$$\begin{aligned}
\Psi^*(v) &= \frac{1}{\Delta_G(v)} \sum_{w \in V \setminus \{v\}} \Psi[v \to w] \leq \frac{1}{\Delta_G(v)} \sum_{w \in V \setminus \{v\}} \sum_{u \in N_G(v)} R_v(u) \cdot \Psi_v[u \to w] \\
&= \frac{1}{\Delta_G(v)} \sum_{u \in N_G(v)} R_v(u) \cdot \Bigg(\sum_{w \in V \setminus \{v,u\}} \Psi_v[u \to w] + \Psi_v[u \to u] \Bigg) \\
&\leq \max_{u \in N_G(v)} \Big\{ R_v(u) \big(\Delta_{G_v}(u) \cdot \Psi^*_v(u) + 1 \big) \Big\},
\end{aligned}$$

which is precisely the desired inequality (observe that $\Psi_v[u \to u] = 1$ by definition). To prove Eq. (6.2), consider $L \in \mathcal{L}$, $i, j \in L(v)$ with $i \neq j$, and $Q \subseteq [q]$. For simplicity, let $\mathbb{P} := \mathbb{P}_{G,L}$ and, for $u \in N_G(v)$, $\mathbb{P}^{ij}_u := \mathbb{P}_{G_v,L^{ij}_u}$, $\Psi^{ij}_u := \Psi_{G_v,L^{ij}_u}$, and $\Psi^{ij}_u = \Psi_{G_v,L^{ij}_u}$. Let also $P^{ij}_{w,Q} := \mathbb{P}(\sigma_w \in Q \mid \sigma_v = i) - \mathbb{P}(\sigma_w \in Q \mid \sigma_v = j)$ and $\Psi^{ij}_u\big((u,i),(w,Q)\big) := \mathbb{P}^{ij}_u(\sigma_w \in Q \mid$

$\sigma_u = i) - \mathbb{P}_u^{ij}(\sigma_w \in Q)$, so that from Lemma 6.2.10 we have

$$P_{w,Q}^{ij} = \sum_{u \in N_G(v)} \frac{\mathbb{P}_u^{ij}(\sigma_u = j)}{\mathbb{P}_u^{ij}(\sigma_u \neq j)} \cdot \Psi_u^{ij}\big((u,j),(w,Q)\big) - \frac{\mathbb{P}_u^{ij}(\sigma_u = i)}{\mathbb{P}_u^{ij}(\sigma_u \neq i)} \cdot \Psi_u^{ij}\big((u,i),(w,Q)\big). \quad (6.3)$$

By the law of total probability, we have

$$\begin{aligned} &\sum_{c \in L_u^{ij}(u)} \mathbb{P}_u^{ij}(\sigma_u = c) \cdot \Psi_u^{ij}\big((u,c),(w,Q)\big) \\ = &\sum_{c \in L_u^{ij}(u)} \mathbb{P}_u^{ij}(\sigma_u = c)\big(\mathbb{P}_u^{ij}(\sigma_w \in Q \mid \sigma_u = c) - \mathbb{P}_u^{ij}(\sigma_w \in Q)\big) \\ = &\, 0; \end{aligned}$$

so we conclude that

$$\begin{aligned} m_u^{ij}(Q) &:= \min_{i' \in L_u^{ij}(u)} \Psi_u^{ij}\big((u,i'),(w,Q)\big) \leq 0, \quad \text{and} \\ M_u^{ij}(Q) &:= \max_{j' \in L_u^{ij}(u)} \Psi_u^{ij}\big((u,j'),(w,Q)\big) \geq 0. \end{aligned} \quad (6.4)$$

Observe further that

$$\begin{aligned} \Psi_u^{ij}[u \to w] &= \max_{i',j' \in L_u^{ij}(u)} \max_{Q' \subseteq [q]} \big|\mathbb{P}_u^{ij}(\sigma_w \in Q' \mid \sigma_u = i') - \mathbb{P}_u^{ij}(\sigma_w \in Q' \mid \sigma_v = j')\big| \\ &\geq M_u^{ij}(Q) - m_u^{ij}(Q). \end{aligned} \quad (6.5)$$

Combining Eqs. (6.3) to (6.5) we obtain that

$$\begin{aligned} P_{w,Q}^{ij} &\leq \sum_{u \in N_G(v)} R_v(u)\big(M_u^{ij}(Q) - m_u^{ij}(Q)\big) \\ &\leq \sum_{u \in N_G(v)} R_v(u) \cdot \Psi_u^{ij}[u \to w] \\ &\leq \sum_{u \in N_G(v)} R_v(u) \cdot \Psi_v[u \to w]. \end{aligned}$$

Since $\Psi_{G,L}[v \to w] = \max_{i,j \in L(v)} \max_{Q \subseteq [q]} P^{ij}_{w,Q}$, by taking maximum over $i, j \in L(v)$ and $Q \subseteq [q]$ of the left-hand side, we obtain the same upper for $\Psi_{G,L}[v \to w]$. We then obtain Eq. (6.2) by taking maximum over $L \in \mathcal{L}$, and thus finish the proof. □

For the bound in Lemma 6.2.14 to be useful, we need to show that the ratio $R(u)$ defined there is strictly less than $1/\Delta_G(u)$. The following lemma does this for $(\Delta, q) \in \Lambda_\varepsilon$, building on ideas from [61, 59].

Lemma 6.2.15. *Let $\varepsilon > 0$ and $(\Delta, q) \in \Lambda_\varepsilon$. Let (G, L) be a (Δ, q)-list-coloring instance with G a triangle-free graph. Then for every vertex u of G with degree at most $\Delta - 1$ and every color $c \in L(u)$, we have*

$$\frac{\mathbb{P}_{G,L}(\sigma_u = c)}{\mathbb{P}_{G,L}(\sigma_u \neq c)} \leq \min\left\{\frac{1}{(1+\varepsilon)\Delta_G(u)},\ \frac{4}{q}\right\}.$$

Remark 6.2.16. We remark that our region Λ_ε is slightly smaller than that of [61], where similar bounds are shown for $q \geq \alpha\Delta - \gamma$ for $\gamma \approx 0.4703$. The difference is that the arguments in [61] upper-bound $\mathbb{P}_L(\sigma_u = c)$ instead of the ratio $\mathbb{P}_L(\sigma_u = c)/\mathbb{P}_L(\sigma_u \neq c)$ which is relevant here, and which is clearly larger than $\mathbb{P}_L(\sigma_u = c)$. See also the discussion before the upcoming Lemma 6.2.19.

Note that when $\Delta_G(u)$ is small, the bound $1/\Delta_G(u)$ is poor and we shall apply the simpler crude bound $4/q$. The proof of Lemma 6.2.15 can be found in Section 6.2.3. Combining Lemmas 6.2.14 and 6.2.15, we can now bound the total influence.

Theorem 6.2.17. *Let $\varepsilon > 0$ and $(\Delta, q) \in \Lambda_\varepsilon$. Suppose that $(G, \mathcal{L})$ is a collection of (Δ, q)-list-coloring instances where G is a triangle-free graph. Then for every vertex v of G we have $\Psi^*_{G,\mathcal{L}}(v) \leq \frac{4}{q}\left(\frac{1}{\varepsilon} + 1\right)$.*

Proof. Let $v_0 = v$, $G^0 = G$ and $\mathcal{L}^0 = \mathcal{L}$. For $\ell \geq 0$, we will define inductively a sequence of (Δ, q)-list-coloring instances $(G^\ell, \mathcal{L}^\ell)$ and a vertex v_ℓ in G^ℓ as follows. Let $G^{\ell+1}$ be the graph obtained from G^ℓ by deleting v_ℓ, i.e., $G^{\ell+1} = G^\ell \backslash v_\ell$ and $\mathcal{L}^{\ell+1} = \mathcal{L}^\ell_{v_\ell}$. Note that all

neighbors of v_ℓ in G_ℓ have degree at most $\Delta - 1$ in $G^{\ell+1}$. Moreover, since by induction $(G^\ell, \mathcal{L}^\ell)$ is a set of (Δ, q)-list-coloring instances, by Lemma 6.2.12 so is $(G^{\ell+1}, \mathcal{L}^{\ell+1})$. Since $q \geq (1+\varepsilon)\alpha\Delta + 1$, combining Lemmas 6.2.14 and 6.2.15, we obtain that

$$\Psi^*_{G^\ell,\mathcal{L}^\ell}(v_\ell) \leq \frac{1}{1+\varepsilon} \cdot \max_{u \in N_{G^\ell}(v_\ell)} \left\{ \Psi^*_{G^{\ell+1},\mathcal{L}^{\ell+1}}(u) \right\} + \frac{4}{q}. \tag{6.6}$$

We let $v_{\ell+1}$ be the vertex $u \in N_{G^\ell}(v_\ell)$ that attains the maximum of the right-hand side of Eq. (6.6), so

$$\Psi^*_{G^\ell,\mathcal{L}^\ell}(v_\ell) \leq \frac{1}{1+\varepsilon} \cdot \Psi^*_{G^{\ell+1},\mathcal{L}^{\ell+1}}(v_{\ell+1}) + \frac{4}{q}. \tag{6.7}$$

Hence, we obtain a sequence of vertices $v_0, v_1, \ldots, v_m$ and a sequence of collections of lists $\mathcal{L}^0, \mathcal{L}^1, \ldots, \mathcal{L}^m$, till when $\Delta_{G^m}(v_m) = 0$ and thus $\Psi^*_{G^m,\mathcal{L}^m}(v_m) = 0$. From this, and since Eq. (6.7) holds for all $0 \leq \ell \leq m-1$, we obtain by solving the recursion that $\Psi^*_{G,\mathcal{L}}(v) \leq \frac{4/q}{1-(1+\varepsilon)^{-1}} = \frac{4}{q}\left(\frac{1}{\varepsilon} + 1\right)$, as wanted. □

Combining Theorem 6.2.17 with Lemma 6.2.8 and Definition 6.2.13 of total maximum influence gives Eq. (6.1), which therefore yields the bound $\lambda_1(\Psi_{G,L}) \leq 8\left(\frac{1}{\varepsilon}+1\right)\frac{\Delta}{q}$ for any (Δ, q)-list-coloring instance (G, L) with $(\Delta, q) \in \Lambda_\varepsilon$, as claimed at the beginning of this section and which completes the proof of Theorem 6.2.6.

6.2.3 Proofs of Influence Recursion and Marginal Bounds

In this section, we give the proof of Lemma 6.2.10 and Lemma 6.2.15, which were used in the proof of Theorem 6.2.6.

Proof of Lemma 6.2.10. For convenience, set $P := \mathbb{P}(\sigma_w = k \mid \sigma_v = j) - \mathbb{P}(\sigma_w = k \mid \sigma_v = i)$.

Let $d = \Delta_G(v)$ and $u_1, \ldots, u_d$ be the neighbors of v in G in the order prescribed by the labelling on G. Let $N = N_G(v)$ and, for $t = 1, \ldots, d$, let $N_t = \{u_1, \ldots u_{t-1}\}$ be the set of

vertices preceding v_t. Then, with $G_v = G \backslash v$ and $L_v = \{L(u)\}_{u \in V \backslash \{v\}}$, we have

$$\begin{aligned} P &= \mathbb{P}_{G_v, L_v}\big(\sigma_w = k,\, j \notin \sigma_N\big) - \mathbb{P}_{G_v, L_v}\big(\sigma_w = k,\, i \notin \sigma_N\big) \\ &= \sum_{t=1}^{d} \mathbb{P}_{G_v, L_v}\big(\sigma_w = k,\, i \notin \sigma_{N_t},\, j \notin \sigma_{N \backslash N_t}\big) - \mathbb{P}_{G_v, L_v}\big(\sigma_w = k,\, i \notin \sigma_{N_{t+1}},\, j \notin \sigma_{N \backslash N_{t+1}}\big) \\ &= \sum_{u \in N_G(v)} \mathbb{P}_u^{ij}(\sigma_w = k \mid \sigma_u \neq j) - \mathbb{P}_u^{ij}(\sigma_w = k \mid \sigma_u \neq i). \end{aligned}$$

Now, for $u \in N_G(v)$, we have that

$$\begin{aligned} &\mathbb{P}_u^{ij}(\sigma_w = k \mid \sigma_u \neq i) - \mathbb{P}_u^{ij}(\sigma_w = k) \\ &= \begin{cases} 0, & \text{if } i \notin L(u), \\ -\dfrac{\mathbb{P}_u^{ij}(\sigma_u = i)}{\mathbb{P}_u^{ij}(\sigma_u \neq i)} \cdot \Psi_u^{ij}\big((u,i),(w,k)\big), & \text{if } i \in L(u). \end{cases} \end{aligned}$$

Summing this over $u \in N_G(v)$ yields the equality in the lemma. □

Next, we prove Lemma 6.2.15. For integers $\Delta, q \geq 3$ with $q \geq \Delta + 1$, the following function will be relevant for this section:

$$\Phi(\Delta, q) = \frac{q-2}{\Delta - 1} \cdot \left[\left(1 - \frac{1}{q - \Delta + 1}\right)^{q - \Delta + 1}\right]^{\frac{\Delta - 1}{q-2}}. \tag{6.8}$$

The following lemma is implicitly given in [61] in their proof of Lemma 15. Here we present a more direct proof, combining ideas from both [61] and [59].

Lemma 6.2.18. *Suppose that (G, L) is a (Δ, q)-list-coloring instance with $G = (V, E)$ a triangle-free graph. Then for every vertex $u \in V$ of degree at most $\Delta - 1$ and every color $c \in L(u)$, we have*

$$\frac{\mathbb{P}_{G,L}(\sigma_u = c)}{\mathbb{P}_{G,L}(\sigma_u \neq c)} \leq \frac{1}{\Phi(\Delta, q)} \cdot \frac{1}{\Delta_G(u)}.$$

Proof. By the law of total probability, it suffices to give an upper bound on $\frac{\mathbb{P}_{G_\tau, L_\tau}(\sigma_u = c)}{\mathbb{P}_{G_\tau, L_\tau}(\sigma_u \neq c)}$ for an arbitrary partial list-coloring τ on $V \backslash (u \cup N_G(u))$. In turn, since (G_τ, L_τ) is also

a (Δ, q)-list-coloring instance and G_τ is a star graph centered at u, it suffices to prove the lemma when G is a star graph centered at u. Henceforth, for convenience, we drop the subscript G from notation.

For $w \in N_G(u)$ and $c \in L(u)$ we define $\delta_c(w) = \mathbb{1}\{c \in L(w)\}$. For any $c, c' \in L(u)$, we have

$$\begin{aligned}\frac{\mathbb{P}_L(\sigma_u = c')}{\mathbb{P}_L(\sigma_u = c)} &= \prod_{w \in N_G(u)} \frac{|L(w)| - \delta_{c'}(w)}{|L(w)| - \delta_c(w)} \\ &\geq \prod_{w \in N_G(u)} \left(1 - \frac{\delta_{c'}(w)}{|L(w)|}\right) \\ &= \prod_{w \in N_G(u)} \left(1 - \frac{1}{|L(w)|}\right)^{\delta_{c'}(w)}.\end{aligned}$$

From this, and using the arithmetic-geometric mean inequality, it follows that

$$\begin{aligned}\frac{\mathbb{P}_L(\sigma_u \neq c)}{\mathbb{P}_L(\sigma_u = c)} &= \sum_{c' \in L(u)\setminus\{c\}} \prod_{w \in N_G(u)} \left(1 - \frac{1}{|L(w)|}\right)^{\delta_{c'}(w)} \\ &\geq \big(|L(u)| - 1\big)\Bigg(\prod_{c' \in L(u)\setminus\{c\}} \prod_{w \in N_G(u)} \left(1 - \frac{1}{|L(w)|}\right)^{\delta_{c'}(w)}\Bigg)^{\frac{1}{|L(u)|-1}} \\ &= \big(|L(u)| - 1\big)\Bigg(\prod_{w \in N_G(u)} \left(1 - \frac{1}{|L(w)|}\right)^{\sum_{c' \in L(u)\setminus\{c\}} \delta_{c'}(w)}\Bigg)^{\frac{1}{|L(u)|-1}} \\ &\geq \big(|L(u)| - 1\big)\Bigg(\prod_{w \in N_G(u)} \left(1 - \frac{1}{|L(w)|}\right)^{|L(w)|}\Bigg)^{\frac{1}{|L(u)|-1}}.\end{aligned}$$

Since $(1 - 1/m)^m$ is an increasing sequence in m and $|L(w)| \geq q - \Delta + 1$, we get

$$\frac{1}{\Delta_G(u)} \cdot \frac{\mathbb{P}_L(\sigma_u \neq c)}{\mathbb{P}_L(\sigma_u = c)} \geq \frac{|L(u)| - 1}{\Delta_G(u)} \cdot \left[\left(1 - \frac{1}{q - \Delta + 1}\right)^{q-\Delta+1}\right]^{\frac{\Delta_G(u)}{|L(u)|-1}}.$$

Since we have

$$\frac{|L(u)| - 1}{\Delta_G(u)} \geq \frac{q - \Delta - 1}{\Delta_G(u)} + 1 \geq \frac{q - \Delta - 1}{\Delta - 1} + 1 = \frac{q - 2}{\Delta - 1},$$

we deduce that

$$\frac{1}{\Delta_G(u)} \cdot \frac{\mathbb{P}_L(\sigma_u \neq c)}{\mathbb{P}_L(\sigma_u = c)} \geq \frac{q-2}{\Delta-1} \cdot \left[\left(1 - \frac{1}{q-\Delta+1} \right)^{q-\Delta+1} \right]^{\frac{\Delta-1}{q-2}} = \Phi(\Delta, q).$$

This shows the lemma. □

We then give a lower bound on the key function $\Phi(\Delta, q)$ defined in Lemma 6.2.18 when $(\Delta, q) \in \Lambda_\varepsilon$. Relevant to Remark 6.2.16, numerical experiments demonstrate that $\Phi(\Delta, q) < 1$ when $q = \alpha\Delta$ for α very close to α^*, indicating that the current proof approach cannot go beyond $q \geq \alpha\Delta$.

Lemma 6.2.19. *For every $\varepsilon > 0$ and $(\Delta, q) \in \Lambda_\varepsilon$, we have $\Phi(\Delta, q) \geq 1 + \left(1 + \frac{1}{\alpha^*}\right)\varepsilon$.*

Proof. Note that the condition $q \geq \alpha\Delta + \beta$ can be rewritten as

$$q - 2 \geq \alpha(\Delta - 1) + \frac{\alpha}{2(\alpha^2 - 1)}. \tag{6.9}$$

First by Lemma 17 (ii) of [61], which can be proved directly by comparing the power series expansions, we have

$$-(q-\Delta+1)\log\left(1 - \frac{1}{q-\Delta+1}\right) \leq 1 + \frac{1}{2(q-\Delta)}.$$

Since we have

$$q - \Delta = (q-2) - (\Delta - 1) + 1 > (\alpha - 1)(\Delta - 1),$$

it follows that

$$\Phi(\Delta, q) \geq \frac{q-2}{\Delta-1} \cdot \exp\left[-\left(1 + \frac{1}{2(\alpha-1)(\Delta-1)}\right) \cdot \frac{\Delta-1}{q-2} \right].$$

Notice that the right-hand side above is monotone increasing in $q-2$. Plugging in Eq. (6.9),

we deduce that

$$\begin{aligned}\Phi(\Delta, q) &\geq \alpha\left(1+\frac{1}{2(\alpha^2-1)(\Delta-1)}\right)\cdot\exp\left[-\frac{1}{\alpha}\cdot\frac{1+\frac{1}{2(\alpha-1)(\Delta-1)}}{1+\frac{1}{2(\alpha^2-1)(\Delta-1)}}\right]\\ &= \alpha\left(1+\frac{1}{2(\alpha^2-1)(\Delta-1)}\right)\cdot\exp\left(-\frac{1}{\alpha}-\frac{1}{2(\alpha^2-1)(\Delta-1)+1}\right)\\ &\geq \alpha e^{-\frac{1}{\alpha}}\cdot\left(1+\frac{1}{2(\alpha^2-1)(\Delta-1)}\right)\cdot\left(1-\frac{1}{2(\alpha^2-1)(\Delta-1)+1}\right)\\ &= \alpha e^{-\frac{1}{\alpha}}.\end{aligned}$$

Finally, since $\alpha = (1+\varepsilon)\alpha^*$ and $\alpha^* e^{-1/\alpha^*} = 1$, we obtain

$$\Phi(\Delta, q) \geq (1+\varepsilon)\alpha^* e^{-\frac{1}{\alpha^*}+\frac{\varepsilon}{\alpha}} \geq (1+\varepsilon)\left(1+\frac{\varepsilon}{\alpha}\right) = 1+\left(1+\frac{1}{\alpha^*}\right)\varepsilon. \qquad \square$$

We are now ready to prove Lemma 6.2.15.

Proof of Lemma 6.2.15. The first upper bound $\frac{1}{(1+\varepsilon)\Delta_G(u)}$ follows from Lemmas 6.2.18 and 6.2.19. For the second bound, first we have the following crude bound

$$\mathbb{P}_L(\sigma_u = c) \leq \frac{1}{|L(u)| - \Delta_G(u)} \leq \frac{1}{q-\Delta}.$$

Therefore,

$$\frac{\mathbb{P}_L(\sigma_u = c)}{\mathbb{P}_L(\sigma_u \neq c)} \leq \frac{1}{q-\Delta-1}.$$

Since $q - 2 \geq \alpha(\Delta - 1)$, we deduce that

$$\frac{q-\Delta-1}{q} \geq \frac{(q-2)-(\Delta-1)}{(q-2)+(\Delta-1)} \geq \frac{\alpha-1}{\alpha+1} \geq \frac{1}{4}.$$

It then follows that $\mathbb{P}_L(\sigma_u = c)/\mathbb{P}_L(\sigma_u \neq c) \leq 4/q$. $\square$

CHAPTER 7

SPECTRAL INDEPENDENCE VIA COUPLING METHODS

In this chapter, we show spectral independence for contractive distributions. As an application, we show that the uniform distribution of all q-colorings on a graph of maximum degree Δ is spectrally independent when $q > (11/6 - \varepsilon_0)\Delta$ for a universal constant $\varepsilon_0 \approx 10^{-5}$. This chapter is based on [23].

7.1 Optimal Mixing Results for Colorings and Potts Model

There are two broad approaches for establishing fast convergence of MCMC algorithms: probabilistic or analytic techniques. Probabilistic techniques primarily utilize the coupling method; a popular example is the path coupling method which has become a fundamental tool in theoretical computer science [30]. In contrast, analytic techniques establish decay to equilibrium by means of functional inequalities such as Poincaré or log-Sobolev inequalities, which correspond to decay of variance and relative entropy respectively. In particular, the so-called modified log-Sobolev inequality is often a powerful analytic tool in establishing tight bounds on the mixing time, while the weaker Poincaré inequality provides control on the spectral gap; see, e.g., [48, 98, 25].

These two approaches—probabilistic or analytic—appeared disparate. While coupling techniques have been used to prove Poincaré inequalities, there are no clear relations between the probabilistic approach and log-Sobolev inequalities. We establish a strong connection by proving that coupling inequalities in the form of bounds on the Ollivier-Ricci curvature of the Markov chain imply entropy decay, and hence the associated modified log-Sobolev inequality holds. In the context of spin systems on bounded-degree graphs, this settles a remarkable conjecture of Peres and Tetali (see Conjecture 3.1 in [54] and Remark 7.2.6).

For q-colorings of graphs with maximum degree Δ, Jerrum [74] proved that the Glauber dynamics has $O(n \log n)$ mixing time when $q > 2\Delta$. Jerrum's result was improved to $q > \frac{11}{6}\Delta$ in [125] and further improved to $q > (\frac{11}{6} - \varepsilon_0)\Delta$ for some small $\varepsilon_0 \approx 10^{-5} > 0$ by Chen et al. [42] by analyzing a Markov chain referred to as the flip dynamics; this implied $O(n^2)$ mixing time of the Glauber dynamics. We obtain $O(n \log n)$ mixing time of the Glauber dynamics, which is asymptotically optimal [71], and also obtain optimal bounds on the log-Sobolev and modified log-Sobolev constants. The following result shows the first part of Theorem 1.1.4.

Theorem 7.1.1 (Colorings)**.** *For q-colorings on an n-vertex graph of maximum degree Δ, when $q > (\frac{11}{6} - \varepsilon_0)\Delta$, where $\varepsilon_0 \approx 10^{-5} > 0$ is a fixed constant, the Glauber dynamics has mixing time $Cn \log n$ where $C = C(\Delta, q)$ is a constant independent of n.*

Moreover, we obtain improved results for the ferromagnetic Potts model. Unlike the Ising model, for the ferromagnetic Potts model known rapid mixing results for the Glauber dynamics do not reach the tree uniqueness threshold. The best known results [70, 124, 27] imply that the Glauber dynamics mixes in $O(n \log n)$ steps when $\beta < \beta_0$ where $\beta_0 = \max\left\{\frac{2}{\Delta}, \frac{1}{\Delta}\ln(\frac{q-1}{\Delta})\right\}$. In addition, [27] showed $\mathrm{poly}(n)$ mixing of the Glauber dynamics for $\beta < \beta_1$ where $\beta_1 = (1 - o(1))\frac{\ln q}{\Delta - 1}$, the $o(1)$ term tends to 0 as $q \to \infty$; see Remark 7.5.10 for more details. These results yield polynomial mixing time bounds for the Swendsen-Wang dynamics in the corresponding regimes of β. Note the critical point for the uniqueness threshold on the tree was established by Häggström [69] and it behaves as $\beta_u = \frac{\ln q}{\Delta - 1} + O(1)$; see [27]. In both regimes, we prove optimal bounds for the mixing time and (modified) log-Sobolev constant of the Glauber dynamics and also for the Swendsen-Wang dynamics.

Theorem 7.1.2 (Ferromagnetic Ising/Potts Model)**.** *For the* ferromagnetic Ising model *with $\beta < \beta_c(\Delta)$ on any n-vertex graph of maximum degree $\Delta \geq 3$, the Glauber dynamics has mixing time $Cn \log n$ and the Swendsen-Wang dynamics has mixing time $C' \log n$, where C*

and C' are constants independent of n. For the ferromagnetic Potts model *the same results hold when $\beta < \max\{\beta_0, \beta_1\}$.*

7.2 Establishing Spectral Independence for Contractive Distributions

We present a series of results showing in a general context that when there exists a contractive coupling then spectral independence holds.

Let d denote an arbitrary metric on Ω. A simple example is the *Hamming metric*, which for configurations $\sigma, \tau \in \Omega$ is defined to be $d_{\mathrm{H}}(\sigma, \tau) = |\{x \in V : \sigma_x \neq \tau_x\}|$. There are two types of more general metrics that we will consider: those within a constant factor of the Hamming metric and vertex-weighted Hamming metric for arbitrary weights. For $\gamma \geq 1$, a metric d on Ω is said to be γ-equivalent to the Hamming metric (or γ-equivalent for simplicity) if for all $\sigma, \tau \in \Omega$,

$$\frac{1}{\gamma} d_{\mathrm{H}}(\sigma, \tau) \leq d(\sigma, \tau) \leq \gamma d_{\mathrm{H}}(\sigma, \tau);$$

that is, a γ-equivalent metric is an arbitrary metric where every distance is within a factor γ of the Hamming distance. In contrast, we can generalize the Hamming distance by considering arbitrary weights for the vertices. Let $w : V \to \mathbb{R}_+$ be an arbitrary positive weight function. The *w-weighted Hamming metric* between two configurations $\sigma, \tau \in \Omega$ is defined to be

$$d_w(\sigma, \tau) = \sum_{x \in V} w(u) \mathbf{1}\{\sigma_u \neq \tau_u\}.$$

In particular, if $w_u = 1$ for all u then d_w is just the usual Hamming metric. Note there are no constraints on the weights except that they are positive; in particular, the weights can be a function of n.

We will often consider a class $\mathcal{P} = \{P^\tau : \tau \in \mathcal{T}\}$ of Markov chains associated with μ, where each P^τ is a Markov chain with stationary distribution μ^τ and $\tau \in \mathcal{T}$ is a pinning; for example, $\mathcal{P}$ can be the family of Glauber dynamics for all μ^τ's. In coupling proofs,

the goal is to design a coupling so that for an arbitrary pair of states the chains contract with respect to some distance metric after the coupled transition. Roughly speaking, for $\kappa \in (0,1)$, we say that μ is κ-contractive with respect to (w.r.t.) a collection $\mathcal{P}$ of Markov chains and a metric d if one step of every chain P^τ contracts the distance by a factor κ in expectation. This is formalized in the following definition.

Definition 7.2.1 (κ-Contraction)**.** Let $\mathcal{P}$ denote a collection of Markov chains associated with μ and let d be a metric on Ω. For $\kappa \in (0,1)$ we say that μ is κ-contractive w.r.t. $\mathcal{P}$ and d if for all $\tau \in \mathcal{T}$, all $X_0, Y_0 \in \Omega^\tau$, there exists a coupling $(X_0, Y_0) \to (X_1, Y_1)$ for P^τ such that:

$$\mathbb{E}[d(X_1, Y_1)|X_0, Y_0] \leq \kappa d(X_0, Y_0).$$

The following result shows that spectral independence holds if the Glauber dynamics has a contractive coupling.

Theorem 7.2.2.

(1) If μ is κ-contractive w.r.t. the Glauber dynamics and an arbitrary w-weighted Hamming metric, then μ is spectrally independent with constant $\eta = \frac{2}{(1-\kappa)n}$. In particular, if $\kappa \leq 1 - \frac{\varepsilon}{n}$, then $\eta \leq \frac{2}{\varepsilon}$.

(2) If the metric in (1) is not a weighted Hamming metric but instead an arbitrary γ-equivalent metric, then $\eta = \frac{2\gamma^2}{(1-\kappa)n}$. In particular, if $\kappa \leq 1 - \frac{\varepsilon}{n}$, then $\eta \leq \frac{2\gamma^2}{\varepsilon}$.

Notice that a κ-contractive coupling for the Hamming distance immediately implies $O(n \log n)$ mixing time of the Glauber dynamics (see, e.g., [30, 87]). But the above theorem offers two additional features. First, it allows arbitrary weights w and the resulting bound on the mixing time is independent of these weights, whereas a coupling argument, such as utilized in path coupling [30], yields a mixing time bound which depends on the ratio of $\max_u w(u)/\min_u w(u)$. Second, as discussed in the previous theorems, spectral independence (together with the easily satisfied marginal boundedness) implies optimal bounds on the mixing time and entropy decay rate for arbitrary heat-bath block dynamics.

We can extend Theorem 7.2.2 by replacing the Glauber dynamics with arbitrary Markov chains. In particular, we consider a general class of Markov chains which we call the *select-update dynamics*. In each step, the select-update dynamics picks a block $B \in \mathcal{B}$ randomly (with a distribution that may depend on the current configuration), and updates all vertices in B using the current configuration (and the pinning if there is one). Note that no assumptions are made on how to pick or update the blocks; the only requirement is that the dynamics converges to the correct stationary distribution. If the chain selects a block B from a fixed distribution over $\mathcal{B}$ and updates B using the conditional marginal distribution on B (under the pinning if applicable), then this is the standard heat-bath block dynamics that we introduced earlier; hence, the select-update dynamics is much more general than the weighted heat-bath block dynamics. Another example of the select-update dynamics is the flip dynamics for sampling random colorings of a graph; see Section 7.5.1.

We define $M = \max_{B \in \mathcal{B}} |B|$ to be the maximum block size and D to be the maximum probability of a vertex being selected in any step of the chain.

Theorem 7.2.3. *If μ is κ-contractive w.r.t. arbitrary select-update dynamics and an arbitrary γ-equivalent metric, then μ is spectrally independent with constant $\eta = \frac{2\gamma^2 DM}{1-\kappa}$.*

Theorem 7.2.3 generalizes Theorem 7.2.2(2) since $M = 1$ and $D = 1/n$ for the Glauber dynamics. If we further assume that the select-update dynamics updates each connected component of a block independently, then the maximum block size M can be replaced by the maximum component size of a block; see Remark 7.5.4. See also Theorem 7.5.2 for a stronger statement involving arbitrary Markov chains, where DM is replaced by the maximum expected distance of two chains when pinning a single vertex. This more general statement potentially applies to chains with unbounded block sizes, including the Swendsen-Wang dynamics.

It is worth remarking that, as a corollary of Theorem 7.2.3 we obtain that a coupling argument for the select-update dynamics where the maximum block size is constant (and $D/(1-\kappa) = O(1)$) implies $O(n \log n)$ mixing time of the Glauber dynamics, together

with the optimal mixing and optimal entropy decay for arbitrary heat-bath block dynamics.

Moreover, as a corollary of Theorem 7.2.2 we obtain that the Dobrushin uniqueness condition implies spectral independence. The Dobrushin uniqueness condition is a classical condition in statistical physics which considers the following dependency matrix.

Definition 7.2.4 (Dobrushin uniqueness condition)**.** The *Dobrushin dependency/influence matrix* $R \in \mathbb{R}^{V \times V}$ is defined by $R(x,x) = 0$ and

$$R(x,y) = \max\left\{d_{\mathrm{TV}}\left(\mu_y(\cdot \mid \sigma), \mu_y(\cdot \mid \tau)\right) : (\sigma,\tau) \in \mathcal{S}_{x,y}\right\} \qquad \text{for } x \neq y$$

where $\mathcal{S}_{x,y}$ is the set of all pairs of configurations on $V \setminus \{y\}$ that can differ only at x. The *Dobrushin uniqueness condition* holds if the maximum column sum of R is at most $1-\varepsilon$ for some $\varepsilon > 0$.

The Dobrushin dependency matrix for the entry $R(x,y)$ considers the worst case pair of configurations on the entire neighborhood of y which differ at x. If x is not a neighbor of y then $R(x,y) = 0$. Hence, the Dobrushin uniqueness condition states that for all y, $\sum_{x \in N(y)} R(x,y) < 1$. In contrast, the ALO influence matrix considers the influence of a disagreement at x on a vertex y (which is not necessarily a neighbor) and no other vertices are fixed, although one needs to consider all pinnings to establish spectral independence, so the notions are incomparable at first glance.

Using Theorem 7.2.2 we prove that the Dobrushin uniqueness condition implies spectral independence. Moreover, our result holds under generalizations of the Dobrushin uniqueness condition. Hayes [70] generalized it to the following spectral condition: if $\|R\|_2 \leq 1-\varepsilon$ for some $\varepsilon > 0$, then the mixing time of the Glauber dynamics is $O(n \log n)$. This was further generalized by Dyer et al. [52] to arbitrary matrix norms. We prove spectral independence when the spectral radius $\rho(R) < 1$, which is the strongest statement of this type as the spectral radius is no larger than any matrix norm; see Remark 7.4.1 for a more detailed discussion.

Theorem 7.2.5. *If the Dobrushin dependency matrix R satisfies $\rho(R) \leq 1 - \varepsilon$ for some $\varepsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\varepsilon$.*

Previously, Marton [101] (see also [62, 116]) showed that the spectral condition in Theorem 7.2.5 implies approximate tensorization of entropy and thus optimal bounds on the modified log-Sobolev constant for the Glauber dynamics. See also [18] for related results with an alternative technique. However, the approach in these works does not imply block factorization of entropy as in our case.

Remark 7.2.6. Our definition of κ-contraction is equivalent to the statement that the Markov chain has coarse Ollivier-Ricci curvature at least $1 - \kappa > 0$ with respect to the metric d [107]. Combining Theorem 7.2.2 with Theorem 4.1.1 we obtain a proof of the following version of the Peres-Tetali conjecture: if the Glauber dynamics has Ollivier-Ricci curvature at least $\varepsilon/n > 0$ then the Glauber dynamics has a modified log-Sobolev constant at least c/n and any α-weighted heat-bath block dynamics has a modified log-Sobolev constant at least $c\,\delta(\alpha)$, for some constant $c = c(\varepsilon, b, \Delta) > 0$, where $\delta(\alpha)$ is defined in Eq. (4.1). Replacing Theorem 7.2.2 with its generalization Theorem 7.2.3 we obtain the same conclusion under the much milder assumption that there exists some κ-contractive select-update dynamics satisfying $DM/(1 - \kappa) = O(1)$. The original Peres-Tetali conjecture in the setting of random walks on graphs is that if there exists a graph metric d such that the random walk has Ollivier-Ricci curvature at least $\lambda > 0$ with respect to d then the random walk has modified log-Sobolev constant at least $c\lambda > 0$, for some universal constant $c > 0$; see Conjecture 3.1 in Eldan et al. [54].

In the rest of this chapter we establish our main results that a contractive distribution is spectrally independent. These results in particular connect classic probabilistic approach for establishing fast mixing of Markov chains such as coupling with recent developments utilizing spectral independence. We first consider a special case of Theorem 7.2.2 concerned with Glauber dynamics and Hamming metric in Section 7.3; this will serve as a concrete example to illustrate our approach for establishing spectral independence. In

Section 7.4, we consider arbitrary metric and prove Theorem 7.2.2. Finally, we consider general Markov chains and metrics in Section 7.5 and prove Theorem 7.2.3.

7.3 Warm-up: Contraction for Glauber Dynamics and Hamming Metric

In this section, we prove a simpler version of Theorem 7.2.2, which already gives the main idea of our proof approach for establishing spectral independence. We show that, if the distribution μ is contractive w.r.t. the Glauber dynamics and the Hamming metric, then it is spectrally independent.

Theorem 7.3.1. *If μ is κ-contractive w.r.t. the Glauber dynamics and the Hamming metric for some $\kappa \in (0,1)$, then μ is spectrally independent with constant $\eta = \frac{2}{(1-\kappa)n}$. In particular, if $\kappa \leq 1 - \varepsilon/n$, then $\eta \leq 2/\varepsilon$.*

Remark 7.3.2. In this chapter, the Glauber dynamics P^{τ}_{GL} for the conditional distribution μ^{τ} with a pinning τ on $U \subseteq V$ is defined as follows: in each step the chain picks a vertex $u \in V$ u.a.r. and updates its spin conditioned on all other vertices and τ. In particular, all pinned vertices in U are allowed to be selected and when this happens the configuration will remain the same (no updates will be made). This setting can make our theorem statements and proofs easier to understand, and will not harm our results since we only consider these chains for the purpose of analysis rather than actually running them. Alternatively, we can define the Glauber dynamics $\tilde{P}^{\tau}_{\mathrm{GL}}$ for μ^{τ} in the following way: in each step an *unpinned* vertex $u \in V \setminus U$ is selected u.a.r. and updated accordingly. Note that $\tilde{P}^{\tau}_{\mathrm{GL}}$ is faster than P^{τ}_{GL} and the contraction rate of $\tilde{P}^{\tau}_{\mathrm{GL}}$ depends on the number of unpinned vertices. If we assume μ^{τ} is κ_ℓ-contractive w.r.t. $\tilde{P}^{\tau}_{\mathrm{GL}}$ and d_{H} where $\ell = |V \setminus U|$, then an analog of Theorem 7.3.1 can show that μ is spectrally independent with

$$\eta = \max_{\ell=1,\ldots,n} \left\{ \frac{2}{(1-\kappa_\ell)\ell} \right\}.$$

However, in actual applications such as under the Dobrushin uniqueness condition in Sec-

tion 7.4, the contraction rate satisfies $\kappa_\ell \le 1 - \varepsilon/\ell$, so we eventually get $\eta \le 2/\varepsilon$ just as from Theorem 7.3.1.

Recall that for any pinning $\tau \in \mathcal{T}$ we let μ^τ be the conditional distribution over Ω^τ given τ, and the ALO influence matrix Ψ^τ is a square matrix indexed by $\mathcal{P}^\tau$ and defined as $\Psi(u, j; u, k) = 0$ and

$$\Psi^\tau(u, j; v, k) = \mu^\tau(\sigma_v = k \mid \sigma_u = j) - \mu^\tau(\sigma_v = k) \text{ for } u \neq v.$$

The distribution μ is said to be η-spectrally independent if $\lambda_1(\Psi^\tau) \le \eta$ for all pinning τ.

Our goal is to upper bound the maximum eigenvalue of the ALO influence matrix Ψ^τ for a given pinning τ. In fact, to make notations simpler we will only consider the case where there is no pinning; the proof is identical by replacing Ω, μ, Ψ with $\Omega^\tau, \mu^\tau, \Psi^\tau$ when an arbitrary pinning τ is given. To upper bound $\lambda_1(\Psi)$, a standard approach that has been applied in previous works [4, 44, 43, 55, 45] is to upper bound the infinity norm of Ψ. More specifically, for each $(u, j) \in \mathcal{P}$ we define

$$S(u, j) = \sum_{(v,k)\in\mathcal{P}} |\Psi(u, j; v, k)| \tag{7.1}$$

to be the sum of absolute influences of a given pair (u, j). The quantity $S(u, j)$ can be thought of as the total influence of (u, j) on all other vertex-spin pairs. If one can show $S(u, j) \le \eta$ for all $(u, j) \in \mathcal{P}$, then it immediately follows that

$$\lambda_1(\Psi) \le \|\Psi\|_\infty = \max_{(u,j)\in\mathcal{P}} S(u, j) \le \eta.$$

Hence, it suffices to show that $S(u, j) = O(1)$.

Fix $(u, j) \in \mathcal{P}$, and define the distribution $\nu = \mu(\cdot \mid \sigma_u = j)$; namely, ν is the conditional distribution of μ with the pinning $\sigma_u = j$. The key observation we make here is that the quantity $S(u, j)$ can be viewed as the difference of the expectation of some function

f under the two measures μ and ν. More specifically, we define

$$f(\sigma) = \sum_{(v,k)\in\mathcal{P}} t(u,j;v,k)\,\mathbf{1}_{\{\sigma_v=k\}}, \tag{7.2}$$

where

$$t(u,j;v,k) = \operatorname{sgn}(\Psi(u,j;v,k)) = \begin{cases} +1, & \Psi(u,j;v,k) > 0; \\ -1, & \Psi(u,j;v,k) < 0; \\ 0, & \Psi(u,j;v,k) = 0. \end{cases}$$

With this definition it follows that

$$\begin{aligned} S(u,j) &= \sum_{(v,k)\in\mathcal{P}} t(u,j;v,k)\Psi(u,j;v,k) \\ &= \sum_{(v,k)\in\mathcal{P}} t(u,j;v,k)\mu(\sigma_v = k \mid \sigma_u = j) - t(u,j;v,k)\mu(\sigma_v = k) \\ &= \mathbb{E}_\nu f - \mathbb{E}_\mu f. \end{aligned}$$

Therefore, the absolute sum of influences $S(u,j)$ describes, in some sense, the "distance" of the two distributions ν and μ measured by f.

To be more precise about our last statement, we review some standard definitions about the Wasserstein distance. Let (Ω, d) be a finite metric space. We say a function $f : \Omega \to \mathbb{R}$ is *L-Lipschitz* w.r.t. the metric d if for all $\sigma, \tau \in \Omega$ we have

$$|f(\sigma) - f(\tau)| \le L d(\sigma,\tau).$$

For every function $f : \Omega \to \mathbb{R}$, we let $L_d(f)$ be the optimal Lipschitz constant of f w.r.t. the metric d; i.e., $L_d(f) = \inf\{L \ge 0 : f \text{ is } L\text{-Lipschitz w.r.t. } d\}$. For a pair of distributions μ

and ν on Ω, the 1-*Wasserstein distance* w.r.t. the metric d between μ and ν is defined as

$$W_{1,d}(\mu,\nu) = \inf\left\{\mathbb{E}_\pi[d(\sigma,\tau)] \mid \pi \in \mathcal{C}(\mu,\nu)\right\}$$

where $\mathcal{C}(\mu,\nu)$ denotes the set of all couplings of μ,ν (i.e., $\pi(\cdot,\cdot) \in \mathcal{C}(\mu,\nu)$ is a joint distribution over $\Omega\times\Omega$ with the marginals on the first and second coordinates being μ and ν respectively) and (σ,τ) is distributed as π; equivalently, the 1-Wasserstein distance can be represented as

$$W_{1,d}(\mu,\nu) = \sup\left\{\mathbb{E}_\mu f - \mathbb{E}_\nu f \mid f:\Omega\to\mathbb{R},\ L_d(f)\le 1\right\}. \tag{7.3}$$

Observe that, the function f defined by Eq. (7.2) is 2-Lipschitz w.r.t. the Hamming metric d_{H}; to see this, if $\sigma,\tau\in\Omega$ and $d_{\mathrm{H}}(\sigma,\tau) = k$ then by the definition of f we have $|f(\sigma)-f(\tau)|\le 2k$. Therefore, we deduce from Eq. (7.3) that

$$S(u,j) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \le L_{d_{\mathrm{H}}}(f)\, W_{1,d_{\mathrm{H}}}(\nu,\mu) \le 2W_{1,d_{\mathrm{H}}}(\nu,\mu).$$

That means, if one can show $W_{1,d_{\mathrm{H}}}(\nu,\mu) = O(1)$ for μ and $\nu = \mu(\cdot \mid \sigma_u = j)$ for any pair (u,j), then $\lambda_1(\Psi) = O(1)$ and spectral independence would follow.

The following lemma, which generalizes previous works [29, 113], will be used to bound the Wasserstein distance of two distributions and may be interesting of its own. Roughly speaking, it claims that if μ,ν are the stationary distributions of two Markov chains P,Q (e.g., Glauber dynamics) respectively, and if μ is contractive w.r.t. P and the two chains P,Q are "close" to each other in one step, then the Wasserstein distance between ν and μ is small. The special case where $\Omega = \{+,-\}^n$ and P,Q are both the Glauber dynamics appeared in [29, Theorem 3.1] and [113, Theorem 2.1], but here we do not make any assumption on the state space or the chains, which is crucial to our applications in Section 7.5.1.

Lemma 7.3.3. *Let (Ω, d) be a finite metric space. Let μ, ν be two distributions over Ω, and P, Q be two Markov chains on Ω with stationary distributions μ, ν respectively. If μ is κ-contractive w.r.t. the chain P and the metric d, then for every $f : \Omega \to \mathbb{R}$ we have*

$$|\mathbb{E}_\mu f - \mathbb{E}_\nu f| \le \frac{L_d(f)}{1-\kappa} \mathbb{E}_\nu \left[W_{1,d}(P(\sigma, \cdot), Q(\sigma, \cdot)) \right]$$

where $P(\sigma, \cdot)$ is the distribution after one step of the chain P when starting from σ and similarly for $Q(\sigma, \cdot)$. As a consequence,

$$W_{1,d}(\mu, \nu) \le \frac{1}{1-\kappa} \mathbb{E}_\nu \left[W_{1,d}(P(\sigma, \cdot), Q(\sigma, \cdot)) \right].$$

We remark that Lemma 7.3.3 holds in a very general setting, and (Ω, d) can be any finite metric space. It shows that if two Markov chains are close to each other, then their stationary distributions must be close to each other, under the assumption that one of the chains is contractive.

Proof of Lemma 7.3.3. The proof imitates the arguments from [29, 113]. Assume for now that P is irreducible; this is a conceptually easier case and we will consider general P later. Since P is irreducible, let h be the principal solution to the Poisson equation $(I - P)h = \bar{f}$ where $\bar{f} = f - \mathbb{E}_\mu f$; that is,

$$h = \sum_{t=0}^{\infty} P^t \bar{f}. \tag{7.4}$$

See Lemma 2.1 in [29] and the references in that paper for backgrounds on the Poisson equation. We then have

$$\mathbb{E}_\nu f - \mathbb{E}_\mu f = \mathbb{E}_\nu \bar{f} = \mathbb{E}_\nu[(I - P)h] = \mathbb{E}_\nu[(Q - P)h]$$

where the last equality is due to $\nu = \nu Q$. For each $\sigma \in \mathrm{supp}(\nu) \subseteq \Omega$, we deduce from

Eq. (7.3) that

$$((Q-P)h)(\sigma) = \mathbb{E}_{Q(\sigma,\cdot)}h - \mathbb{E}_{P(\sigma,\cdot)}h \leq L_d(h)\, W_{1,d}(Q(\sigma,\cdot), P(\sigma,\cdot)).$$

It remains to bound the Lipschitz constant of h. For $\sigma, \tau \in \Omega$,

$$\begin{aligned} |h(\sigma) - h(\tau)| &\leq \sum_{t=0}^{\infty} \left|(P^t \bar{f})(\sigma) - (P^t \bar{f})(\tau)\right| \\ &= \sum_{t=0}^{\infty} \left|\mathbb{E}_{P^t(\sigma,\cdot)}\bar{f} - \mathbb{E}_{P^t(\tau,\cdot)}\bar{f}\right| \\ &\leq L_d(f) \sum_{t=0}^{\infty} W_{1,d}(P^t(\sigma,\cdot), P^t(\tau,\cdot)) \end{aligned}$$

where the last inequality again follows from Eq. (7.3). Since μ is κ-contractive w.r.t. P and d, for all $\sigma, \tau \in \Omega$ and every integer $t \geq 1$ we have

$$W_{1,d}(P^t(\sigma,\cdot), P^t(\tau,\cdot)) \leq \kappa^t d(\sigma,\tau).$$

We then deduce that

$$|h(\sigma) - h(\tau)| \leq L_d(f) \sum_{t=0}^{\infty} \kappa^t d(\sigma,\tau) = \frac{L_d(f)}{1-\kappa}\, d(\sigma,\tau).$$

This implies that $L_d(h) \leq L_d(f)/(1-\kappa)$ and the lemma then follows.

Next, we show how to remove the assumption that P is irreducible. Observe that in the proof above we only need the irreducibility of P to guarantee that the function h given by Eq. (7.4) is well-defined; i.e., the series on the right-hand side of Eq. (7.4) is convergent. The rest of the proof does not require the irreducibility of P. In fact, one can deduce the convergence of Eq. (7.4) solely from the contraction of P. Note that for all $\sigma \in \Omega$,

$$\left|P^t \bar{f}(\sigma)\right| = \left|P^t \bar{f}(\sigma) - \mathbb{E}_\mu P^t \bar{f}\right|$$

$$= \left| P^t \bar{f}(\sigma) - \sum_{\tau \in \Omega} \mu(\tau) P^t \bar{f}(\tau) \right|$$
$$\leq \sum_{\tau \in \Omega} \mu(\tau) \left| P^t \bar{f}(\sigma) - P^t \bar{f}(\tau) \right|$$

where the first equality follows from $\mathbb{E}_\mu P^t \bar{f} = \mathbb{E}_\mu \bar{f} = 0$. Since Ω is finite, to show that Eq. (7.4) is convergent for all $\sigma \in \Omega$, it suffices to show that for all $\sigma, \tau \in \Omega$ the series $\sum_{t=0}^{\infty} \left| P^t \bar{f}(\sigma) - P^t \bar{f}(\tau) \right|$ is convergent. Actually, our proof before has already showed that

$$\sum_{t=0}^{\infty} \left| P^t \bar{f}(\sigma) - P^t \bar{f}(\tau) \right| \leq \frac{L_d(f)}{1-\kappa} d(\sigma, \tau) < \infty$$

using only the contraction of P, where we have $L_d(f) < \infty$ and $\sup_{\sigma,\tau \in \Omega} d(\sigma, \tau) < \infty$ because Ω is finite. Therefore, the lemma remains true without the assumption of irreducibility of P. □

Given Lemma 7.3.3, we can now complete the proof of Theorem 7.3.1.

Proof of Theorem 7.3.1. For every $(u, j) \in \mathcal{P}$, we deduce from Lemma 7.3.3 that

$$S(u, j) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \leq \frac{L_{d_{\mathrm{H}}}(f)}{1-\kappa} \mathbb{E}_\nu \left[W_{1,d_{\mathrm{H}}}(P(\sigma, \cdot), Q(\sigma, \cdot)) \right] \tag{7.5}$$

where $S(u, j)$ is given by Eq. (7.1), f is given by Eq. (7.2), P is the Glauber dynamics for μ, and Q is the Glauber dynamics for $\nu = \mu^{(u,j)} = \mu(\cdot \mid \sigma_u = j)$ (we use (u, j) to denote the pinning $\sigma_u = j$). We claim that for every $\sigma \in \Omega^{(u,j)}$,

$$W_{1,d_{\mathrm{H}}}(P(\sigma, \cdot), Q(\sigma, \cdot)) \leq \frac{1}{n}. \tag{7.6}$$

To see this, let σ_1 and σ_2 be the configurations after one step of P and Q respectively when starting from σ. We can couple σ_1 and σ_2 by picking the same vertex to update in the Glauber dynamics. If the picked vertex is not u, then we can make $\sigma_1 = \sigma_2$; meanwhile, if u is picked, which happens with probability $1/n$, then $d_{\mathrm{H}}(\sigma_1, \sigma_2) \leq 1$ where the discrepancy

is caused by the pinning $\sigma_u = j$. Therefore, the 1-Wasserstein distance between σ_1 and σ_2 is upper bounded by $1/n$; this justifies our claim. Combining $L_{d_{\mathrm{H}}}(f) \leq 2$ and Eq. (7.6), we obtain from Eq. (7.5) that $S(u,j) \leq \frac{2}{(1-\kappa)n}$ for each (u,j); consequently, $\lambda_1(\Psi) \leq \frac{2}{(1-\kappa)n}$. The same argument holds for μ^τ under any pinning τ as well, and spectral independence then follows. □

7.4 Contraction for Glauber Dynamics and General Metrics

In this section, we generalize the Hamming metric assumption in Theorem 7.3.1 to any weighted Hamming metric or any metric equivalent to Hamming, which establishes Theorem 7.2.2. We restate it here for convenience.

Theorem 7.2.2.

(1) If μ is κ-contractive w.r.t. the Glauber dynamics and an arbitrary w-weighted Hamming metric, then μ is spectrally independent with constant $\eta = \frac{2}{(1-\kappa)n}$. In particular, if $\kappa \leq 1 - \frac{\varepsilon}{n}$, then $\eta \leq \frac{2}{\varepsilon}$.

(2) If the metric in (1) is not a weighted Hamming metric but instead an arbitrary γ-equivalent metric, then $\eta = \frac{2\gamma^2}{(1-\kappa)n}$. In particular, if $\kappa \leq 1 - \frac{\varepsilon}{n}$, then $\eta \leq \frac{2\gamma^2}{\varepsilon}$.

We prove the two cases of Theorem 7.2.2 separately. We first consider the weighted Hamming metric. Recall that for a positive weight function $w : V \to \mathbb{R}_+$, the w-weighted Hamming metric $d = d_w$ is given by

$$d_w(\sigma,\tau) = \sum_{u \in V} w(u)\mathbf{1}\{\sigma_u \neq \tau_u\} \text{ for } \sigma,\tau \in \Omega.$$

In particular, if $w(u) = 1$ for all u then d is the usual Hamming metric.

Unfortunately, the proof of Theorem 7.3.1 does not work directly in this scenario. The reason is that the right-hand side of Eq. (7.5), with d_{H} replaced by $d = d_w$ now, can be as large as $O\left(\frac{w_{\max}}{w_{\min}}\right)$ (more specifically, $L_d(f) = O\left(\frac{1}{w_{\min}}\right)$ and $W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) =$

$O(w_{\max})$), which can be unbounded since we are not making any assumption on w. To deal with this, we need to take the vertex weights into account when defining the function f and, more importantly, when defining the absolute sum of influences $S(u, j)$.

Proof of Theorem 7.2.2(1). For ease of notation we may assume that there is no pinning; the proof remains the same with an arbitrary pinning τ. For fixed $(u, j) \in \mathcal{P}$, we define the w-weighted sum of absolute influences given by

$$S_w(u, j) = \sum_{(v,k)\in\mathcal{P}} w(v)\,|\Psi(u, j; v, k)|. \tag{7.7}$$

Such weighted sums were considered in [44, Lemma 22] to deduce spectral independence. We claim that if $S_w(u, j) \le \eta\, w(u)$ for all $(u, j) \in \mathcal{P}$ for some $\eta > 0$, then $\lambda_1(\Psi) \le \eta$. To see this, let $\tilde{w} \in \mathbb{R}_+^{|\mathcal{P}|}$ with $\tilde{w}(u, j) = w(u)$ and let $W = \mathrm{diag}(\tilde{w})$; the assumption of the claim then implies that $\|W^{-1}\Psi W\|_\infty \le \eta$ and thus $\lambda_1(\Psi) = \lambda_1(W^{-1}\Psi W) \le \eta$. Therefore, it suffices to upper bound the ratio $S_w(u, j)/w(u)$.

Let $\nu = \mu^{(u,j)} = \mu(\cdot \mid \sigma_u = j)$ be the conditional distribution with pinning $\sigma_u = j$, and define

$$f_w(\sigma) = \sum_{(v,k)\in\mathcal{P}} w(v)\, t(u, j; v, k)\, \mathbf{1}_{\{\sigma_v=k\}} \tag{7.8}$$

where $t(u, j; v, k) = \mathrm{sgn}(\Psi(u, j; v, k))$. Observe that $L_d(f_w) \le 2$ and

$$S_w(u, j) = \mathbb{E}_\nu f_w - \mathbb{E}_\mu f_w.$$

It then follows from Lemma 7.3.3 that

$$S_w(u, j) \le \frac{2}{1-\kappa}\, \mathbb{E}_\nu \left[W_{1,d}(P(\sigma, \cdot), Q(\sigma, \cdot))\right]$$

where P, Q are the Glauber dynamics for μ, ν respectively. For every $\sigma \in \Omega^{(u,j)}$ we have

$$W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) \leq \frac{w(u)}{n},$$

since if we couple the configurations σ_1, σ_2 after one step of P, Q respectively by picking the same vertex to update, then $d(\sigma_1, \sigma_2) = w(x)$ only when the site u is picked, and $\sigma_1 = \sigma_2$ otherwise. Therefore, we get $S_w(u,j) \leq \frac{2w(u)}{(1-\kappa)n}$ for every $(u,j) \in \mathcal{P}$, implying that $\lambda_1(\Psi) \leq \frac{2}{(1-\kappa)n}$. The same argument works for μ^τ under any pinning τ as well, which establishes spectral independence. □

Next we consider the second part of Theorem 7.2.2. Recall that a metric d on Ω is said to be γ-equivalent (to the Hamming metric) for some $\gamma > 1$ if for all $\sigma, \tau \in \Omega$

$$\frac{1}{\gamma} d_{\mathrm{H}}(\sigma,\tau) \leq d(\sigma,\tau) \leq \gamma d_{\mathrm{H}}(\sigma,\tau).$$

To prove the second part, we follow the proof approach for Theorem 7.3.1, and in particular the right-hand side of Eq. (7.9) below (analogous to Eq. (7.5)) can be upper bounded using the γ-equivalence.

Proof of Theorem 7.2.2(2). For every $(u,j) \in \mathcal{P}$, we deduce from Lemma 7.3.3 that

$$S(u,j) = \mathbb{E}_\nu f - \mathbb{E}_\mu f \leq \frac{L_d(f)}{1-\kappa} \mathbb{E}_\nu \left[W_{1,d}(P(\sigma,\cdot), Q(\sigma,\cdot)) \right] \tag{7.9}$$

where $S(u,j)$ and f are defined by Eq. (7.1), Eq. (7.2) respectively, and P, Q are the Glauber dynamics for μ and $\nu = \mu^{(u,j)} = \mu(\cdot \mid \sigma_u = j)$ respectively. Since d is γ-equivalent, for all $\sigma, \tau \in \Omega$ we have

$$|f(\sigma) - f(\tau)| \leq 2 d_{\mathrm{H}}(\sigma,\tau) \leq 2\gamma d(\sigma,\tau);$$

this shows $L_d(f) \leq 2\gamma$. Meanwhile, by the definition of 1-Wasserstein distance for every

$\sigma \in \Omega^{(u,j)}$ we have

$$W_{1,d}(P(\sigma,\cdot),Q(\sigma,\cdot)) = \inf\{\mathbb{E}_\pi[d(\sigma,\tau)] \mid \pi \in \mathcal{C}(P(\sigma,\cdot),Q(\sigma,\cdot))\}$$
$$\leq \gamma \inf\{\mathbb{E}_\pi[d_{\mathrm{H}}(\sigma,\tau)] \mid \pi \in \mathcal{C}(P(\sigma,\cdot),Q(\sigma,\cdot))\} = \gamma W_{1,d_{\mathrm{H}}}(P(\sigma,\cdot),Q(\sigma,\cdot)) \leq \frac{\gamma}{n}$$

where the last inequality is Eq. (7.6). Thus, we obtain from Eq. (7.9) that $S(u,j) \leq \frac{2\gamma^2}{(1-\kappa)n}$. The rest of the proof is the same as Theorem 7.3.1. □

As an application of Theorem 7.2.2(1), we show that the Dobrushin uniqueness condition, as well as its generalizations [70, 52], implies spectral independence. Recall that the Dobrushin dependency matrix R is a $|V| \times |V|$ matrix defined as $R(u,u) = 0$ and

$$R(u,v) = \max\{d_{\mathrm{TV}}(\mu_v(\cdot \mid \sigma), \mu_v(\cdot \mid \tau)) : (\sigma,\tau) \in \mathcal{S}_{u,v}\} \text{ for } x \neq y$$

where $\mathcal{S}_{u,v}$ is the set of pairs of configurations on $V \setminus \{v\}$ that differ at most at u. Denote the spectral radius of a square matrix M by $\rho(M)$. If M is nonnegative, then $\rho(M)$ is an eigenvalue of M by the Perron-Frobenius theorem. We prove Theorem 7.2.5 from the introduction.

Theorem 7.2.5. *If the Dobrushin dependency matrix R satisfies $\rho(R) \leq 1 - \varepsilon$ for some $\varepsilon > 0$, then μ is spectrally independent with constant $\eta = 2/\varepsilon$.*

Remark 7.4.1. If $\|R\|_\infty < 1$, then the Glauber dynamics mixes rapidly by a simple application of the path coupling method of Bubley and Dyer [30]. The same is true under the Dobrushin uniqueness condition, i.e., when $\|R\|_1 < 1$. Hayes [70] generalized the condition to the spectral norm $\|R\|_2 < 1$. Dyer, Goldberg, and Jerrum [52] further improved it to $\|R\| < 1$ for any matrix norm (where the mixing time depends logarithmly on the norm of the all-one matrix). Our condition $\rho(R) < 1$ in Theorem 7.2.5 is technically better than previous works since for a nonnegative matrix R one has $\rho(R) \leq \|R\|$ for any matrix norm,

and the inequality can be strict for all norms when R is not irreducible; see [52] for related discussions. Finally, we point out that if R is symmetric then $\rho(R) = \|R\|_2$.

It is known that the Glauber dynamics is contractive for some weighted Hamming metric if the weight vector satisfies a spectral condition related to R.

Lemma 7.4.2 ([52, Lemma 20]). *If $w \in \mathbb{R}_+^V$ is a positive vector such that $Rw \leq (1-\varepsilon)w$ entrywisely, then μ is $(1-\varepsilon/n)$-contractive w.r.t. the Glauber dynamics and the w-weighted Hamming metric $d = d_w$.*

The following fact about nonnegative matrices is helpful.

Lemma 7.4.3 ([103, Example 7.10.2]). *If $M, N \in \mathbb{R}_+^{n\times n}$ are two nonnegative square matrices such that $M \leq N$ entrywisely, then $\rho(M) \leq \rho(N)$.*

We give below the proof of Theorem 7.2.5.

Proof of Theorem 7.2.5. Consider first the case that there is no pinning. If the Dobrushin dependency matrix R is irreducible, then the right principal eigenvector w associated with the eigenvalue $\rho(R)$ satisfies $Rw = \rho(R)w \leq (1-\varepsilon)w$ and $w > 0$ by the Perron-Frobenius theorem. Hence, Lemma 7.4.2 and (the proof of) Theorem 7.2.2(1) immediately yield $\lambda_1(\Psi) \leq 2/\varepsilon$. However, if R is reducible, we cannot use the principal eigenvector directly since it may have zero entries. We instead consider the matrix $R_\delta = R + \delta O$ where O is the all-one matrix and $\delta > 0$ is a tiny constant. Let w_δ be the right principal eigenvector of R_δ associated with the eigenvalue $\rho(R_\delta)$. Since R_δ is irreducible, $w_\delta > 0$ by the Perron-Frobenius theorem. Moreover, $Rw_\delta \leq R_\delta w_\delta = \rho(R_\delta)w_\delta$. Since $\lim_{\delta\to 0} R_\delta = R$, we have $\lim_{\delta\to 0}\rho(R_\delta) = \rho(R)$; see, e.g., Remark 3.4 in [2]. Thus, $\rho(R_\delta) < 1$ for sufficiently small δ. Then by Lemma 7.4.2 and Theorem 7.2.2(1), for δ small enough, we have $\lambda_1(\Psi) \leq 2/(1-\rho(R_\delta))$. Taking $\delta \to 0$ and using the assumption that $\rho(R) \leq 1-\varepsilon$, we obtain $\lambda_1(\Psi) \leq 2/\varepsilon$.

Next, consider the conditional measure μ^τ with a pinning τ on a subset $U \subseteq V$. Let R^τ be the Dobrushin dependency matrix for μ^τ; note that by definition $R^\tau(u,v) = 0$ if $u \in U$

or $v \in U$, and $R^\tau(u,v) \le R(u,v)$ for all $u, v \in V$. We deduce from Lemma 7.4.3 that $\rho(R^\tau) \le \rho(R) \le 1-\varepsilon$ and thus this is reduced to the no-pinning case. Therefore, we get $\lambda_1(\Psi^\tau) \le 2/\varepsilon$ for all τ and spectral independence then follows. □

7.5 Contraction for General Markov Chains and General Metrics

In this section, we generalize Theorem 7.3.1 to arbitrary "local" Markov chains and arbitrary metrics close to the Hamming metric. In particular, we prove Theorem 7.2.3.

Consider a collection of Markov chains $\mathcal{P} = \{P^\tau : \tau \in \mathcal{T}\}$ associated with μ, where each P^τ is a Markov chain on Ω^τ with stationary distribution μ^τ. Intuitively, one can think of $\mathcal{P}$ as the same dynamics applied to all conditional distributions μ^τ; for example, $\mathcal{P}$ can be the collection of Glauber dynamics for all μ^τ's. We are particularly interested in local dynamics; these are Markov chains that make local updates on the configuration in each step, e.g., Glauber dynamics for spin systems or flip dynamics for colorings. Alternatively, we can describe local dynamics as those insensitive to pinnings; that is, if the dynamics is applied to both μ and $\mu^{(u,j)}$ with a pinning $\sigma_u = j$, then with high probability there is no difference in the two chains or the discrepancy caused by the pinning will not propagate. This motivates the following definition.

Definition 7.5.1. We say a collection $\mathcal{P}$ of Markov chains associated with μ is Φ-*local* if for any two adjacent pinnings $\tau \in \mathcal{T}$ and $\tau' = \tau \cup (u,j)$ where $(u,j) \in \mathcal{P}^\tau$ (i.e., τ' combines τ and the pinning $\sigma_u = j$), and for all $\sigma \in \Omega^{\tau'}$, we have

$$W_{1,d_{\mathrm{H}}}(P^\tau(\sigma,\cdot), P^{\tau'}(\sigma,\cdot)) \le \Phi.$$

We show that for such local dynamics contraction implies spectral independence.

Theorem 7.5.2. *If μ is κ-contractive w.r.t. a Φ-local collection $\mathcal{P}$ of Markov chains and a γ-equivalent metric d for some $\kappa \in (0,1)$, then μ is spectrally independent with constant $\eta = \frac{2\gamma^2\Phi}{1-\kappa}$.*

Proof. The proof is similar to that of Theorem 7.3.1 and Theorem 7.2.2(2). For an arbitrary pinning τ and $(u, j) \in \mathcal{P}^\tau$, we define

$$S^\tau(u, j) = \sum_{(v,k)\in\mathcal{P}^\tau} |\Psi^\tau(u, j; v, k)| \tag{7.10}$$

and

$$f^\tau(\sigma) = \sum_{(v,k)\in\mathcal{P}^\tau} t^\tau(u, j; v, k)\, \mathbf{1}_{\{\sigma_v=k\}} \tag{7.11}$$

where $t^\tau(u, j; v, k) = \operatorname{sgn}(\Psi^\tau(u, j; v, k))$; these definitions are analogous to Eq. (7.1) and Eq. (7.2) with pinning τ. Let $\tau' = \tau \cup (u, j)$. Then we deduce from Lemma 7.3.3 that

$$S^\tau(u, j) = \mathbb{E}_{\mu^{\tau'}} f^\tau - \mathbb{E}_{\mu^\tau} f^\tau \le \frac{L_d(f^\tau)}{1-\kappa}\, \mathbb{E}_{\mu^{\tau'}} \left[W_{1,d}(P^\tau(\sigma, \cdot), P^{\tau'}(\sigma, \cdot)) \right].$$

As shown in the proof of Theorem 7.2.2(2), since d is γ-equivalent to the Hamming metric we have $L_d(f^\tau) \le \gamma L_{d_{\mathrm{H}}}(f^\tau) \le 2\gamma$ and for all $\sigma \in \Omega^{\tau'}$ we have

$$W_{1,d}(P^\tau(\sigma, \cdot), P^{\tau'}(\sigma, \cdot)) \le \gamma W_{1,d_{\mathrm{H}}}(P^\tau(\sigma, \cdot), P^{\tau'}(\sigma, \cdot)) \le \gamma\Phi$$

using the Φ-locality of $\mathcal{P}$. Therefore, we obtain that $S^\tau(u, j) \le \frac{2\gamma^2\Phi}{1-\kappa}$ for all $(u, j) \in \mathcal{P}^\tau$. This yields $\lambda_1(\Psi^\tau) \le \frac{2\gamma^2\Phi}{1-\kappa}$ and spectral independence follows. □

To better understand local dynamics, we consider a very general type of Markov chains which we call *select-update dynamics*; examples include the Glauber dynamics, heat-bath block dynamics, and flip dynamics. Let $\mathcal{B}$ be a collection of blocks associated with the select-update dynamics and fix some pinning τ. Given the current configuration $\sigma^t \in \Omega^\tau$, the next configuration σ^{t+1} is generated as follows:

1. SELECT: Select a block $B \in \mathcal{B}$ from some distribution p_t over $\mathcal{B}$;

2. UPDATE: Resample the configuration on B from some distribution ν_B^t.

We try to make weakest assumptions on the selection rule p_t and the update rule ν_B^t: the selection distribution p_t is allowed to depend on the current configuration σ^t but is independent of the pinning τ, and the update distribution ν_B^t is allowed to depend on the whole current configuration σ^t and the part of the pinning τ contained in B. In particular, the heat-bath block dynamics is a special case of the select-update dynamics: the selection rule $p_t = \alpha$ is a fixed distribution over $\mathcal{B}$ and the update rule ν_B^t is the marginal distribution on B conditioned on σ^t outside B and the pinning τ in B.

Remark 7.5.3. The assumption that the selection rule p_t is independent of the pinning τ is not necessary, but it is helpful for stating and proving our theorems and does not weaken our results. Roughly speaking, we only require that the collection of the select-update dynamics is the same dynamics applied to all μ^τ's, and the selection rule p_t can be conditioned on containing at least one unpinned vertex. See the discussions in Remark 7.3.2 for the Glauber dynamics.

We write $\mathcal{P}_\mathcal{B}$ for a collection of select-update dynamics associated with μ. Denote the maximum block size of $\mathcal{B}$ by

$$M = \max_{B \in \mathcal{B}} |B|,$$

and the maximum probability of a vertex being selected in Step 1 by

$$D = \max_{p_t} \max_{u} \sum_{B \in \mathcal{B}: u \in B} p_t(B)$$

where we maximize over all selection rules p_t that can occur. We can show that the select-update dynamics $\mathcal{P}_\mathcal{B}$ is Φ-local with $\Phi = DM$; using this and Theorem 7.5.2 we establish Theorem 7.2.3, which we restate here for convenience.

Theorem 7.2.3. *If μ is κ-contractive w.r.t. arbitrary select-update dynamics and an arbitrary γ-equivalent metric, then μ is spectrally independent with constant $\eta = \frac{2\gamma^2 DM}{1-\kappa}$.*

Proof. It suffices to show that the select-update dynamics $\mathcal{P}_\mathcal{B}$ is Φ-local with $\Phi = DM$; the theorem would then follows immediately from Theorem 7.5.2. Consider two adjacent

pinnings τ and $\tau' = \tau \cup (u, j)$ where $(u, j) \in \mathcal{P}^\tau$. For $\sigma \in \Omega^{\tau'}$, let σ_1 and σ_2 be the two configurations obtained from σ after one step of P^τ and $P^{\tau'}$ respectively. We couple σ_1 and σ_2 by picking the same block $B \in \mathcal{B}$ in Step 1 of the select-update dynamics. If $u \notin B$, then we have $\sigma_1 = \sigma_2$. Meanwhile, if $u \in B$, which happens with probability at most D, we have $d_{\mathrm{H}}(\sigma_1, \sigma_2) \leq |B| \leq M$. Therefore,

$$W_{1,d_{\mathrm{H}}}(P^\tau(\sigma, \cdot), P^{\tau'}(\sigma, \cdot)) \leq DM.$$

This establishes the (DM)-locality for $\mathcal{P}_\mathcal{B}$. □

Remark 7.5.4. If we further assume that in Step 2 the select-update dynamics resamples a block independently for each of its components (i.e., the update rule ν_B^t is a product distribution over all components of the induced subgraph $G[B]$), then in Theorem 7.2.3 the maximum block size M can be replaced by the maximum component size of all blocks.

7.5.1 Application: Flip Dynamics for Colorings

In this section we establish spectral independence for colorings utilizing Theorem 7.2.3.

Theorem 7.5.5. *Let $\varepsilon_0 \approx 10^{-5} > 0$ be a fixed constant. Let $\Delta, q \geq 3$ be integers and $q > (\frac{11}{6} - \varepsilon_0)\Delta$. Then there exists $\eta = \eta(\Delta, q) > 0$ such that the following holds.*

Let μ be the uniform distribution over all proper q-colorings of a graph $G = (V, E)$ of maximum degree at most Δ. Then μ is spectrally independent with constant η.

To apply Theorem 7.2.3, we need a contractive Markov chain for sampling colorings of a graph. Vigoda considered the *flip dynamics* [125] and showed that it is contractive for the Hamming metric when the number of colors $q > \frac{11}{6}\Delta$. Recently, [42] improved the bound to $q > (\frac{11}{6} - \varepsilon_0)\Delta$ for a fixed tiny constant $\varepsilon_0 \approx 10^{-5}$, using variable-length coupling or an alternative metric. Our result on spectral independence builds upon contraction results for the flip dynamics.

We first describe the flip dynamics. Let Ω be the set of all proper q-colorings of G. Fix a pinning τ on $U \subseteq V$. For a coloring $\sigma \in \Omega$, a vertex $u \in V$, and a color $j \in [q]$, denote by $L_\sigma(u, j)$ the bicolored component containing u with colors j and σ_u; that is, the set of all vertices which can be reached from u through an alternating (σ_u, j)-colored path. Given the coloring σ^t at time t, the flip dynamics with pinning τ generates the next coloring σ^{t+1} as follows:

1. Pick a vertex $u \in V$ u.a.r. and a color $j \in [q]$ u.a.r.;
2. If $L_{\sigma^t}(u, j)$ contains a pinned vertex (i.e., $L_{\sigma^t}(u, j) \cap U \neq \emptyset$), then $\sigma^{t+1} = \sigma^t$;
3. If all vertices in $L_{\sigma^t}(u, j)$ are free (i.e., $L_{\sigma^t}(u, j) \cap U = \emptyset$), then flip the two colors of $L_{\sigma^t}(u, j)$ with probability p_s/s where $s = |L_{\sigma^t}(u, j)|$.

The flip dynamics is specified by the flip parameters $\{p_s\}_{s=1}^{\infty}$. In [125] and the recent improvement [42], the flip parameters are chosen in such a way that $p_s = 0$ for all $s \geq 7$; i.e., in each step at most six vertices change their colors. We set the flip parameters as in Observation 5.1 from [42], where the authors established contraction of the flip dynamics using the path coupling method.

Lemma 7.5.6 ([42])**.** *Under the assumptions of Theorem 7.5.5, there exists a constant $\varepsilon = \varepsilon(\Delta, q) > 0$ and a 2-equivalent metric d such that μ is $(1 - \varepsilon/n)$-contractive w.r.t. the flip dynamics and the metric d.*

We remark that the pinning τ induces a list coloring instance where each unpinned vertex has a color list to choose its color from, and the results of [42] generalize naturally to list colorings. Also, in this chapter we assume that the flip dynamics may pick a pinned vertex and stay at the current coloring. This does not weaken our results since we only consider the flip dynamics for analysis rather than actually running it; see Remark 7.3.2 addressing the same issue for the Glauber dynamics and also Remark 7.5.3 for general select-update dynamics.

We give below the proof of Theorem 7.5.5.

Proof of Theorem 7.5.5. Note that the flip dynamics belongs to the class of select-update dynamics, where the associated $\mathcal{B}$ is the collection of connected subsets of vertices. Since the flip parameters satisfy $p_s > 0$ only for $s \leq 6$, we have $M \leq 6$. Moreover, we have $D \leq \Delta^6/n$ since a vertex u is in the selected bicolored component $L_{\sigma^t}(v, j)$ only if $\mathrm{dist}(u, v) \leq 5$, which happens with probability at most Δ^6/n. The theorem then follows from Lemma 7.5.6 and Theorem 7.2.3. □

We conclude here with the proof of Theorem 7.1.1.

Proof of Theorem 7.1.1. By Theorem 7.5.5 the uniform distribution μ of proper colorings is spectrally independent. Then the results follows immediately from Theorem 4.1.1. □

7.5.2 Application: Block Dynamics for Potts Model

Here we apply Theorems 7.3.1 and 7.2.3 to the ferromagnetic Potts model to establish spectral independence.

Theorem 7.5.7. *Let $\Delta \geq 3$ and $q \geq 2$ be integers. Let μ be the Gibbs distribution of the q-state ferromagnetic Potts model with inverse temperature parameter β on a graph $G = (V, E)$ of maximum degree at most Δ. Then, the following holds:*

1. *If $\beta < \max\left\{\frac{2}{\Delta}, \frac{1}{\Delta}\ln(\frac{q-1}{\Delta})\right\}$, then μ is spectrally independent with constant $\eta = \eta(\beta, \Delta)$.*
2. *For any $\delta > 0$ there exists $c = c(\delta, \Delta) > 0$ such that, if $\beta \leq \frac{\ln q - c}{\Delta - 1 + \delta}$ then μ is spectrally independent with constant $\eta = \eta(\delta, \beta, \Delta)$.*

To prove this theorem, we need the following results from [124] and [27] regarding the contraction of the Glauber dynamics and of the heat-bath block dynamics with a specific choice of blocks.

Lemma 7.5.8 ([124, Corollary 2.14] & [27, Proposition 2.2])**.** *Under the assumptions in Part 1 of Theorem 7.5.7, there exists a constant $\varepsilon = \varepsilon(\beta, \Delta)$ such that μ is $(1 - \frac{\varepsilon}{n})$-contractive w.r.t. the Glauber dynamics and the Hamming metric.*

Lemma 7.5.9 ([27, Theorem 2.7])**.** *Under the assumptions in Part 2 of Theorem 7.5.7, there exists a collection of blocks $\mathcal{B} = \{B_u\}_{u \in V}$ satisfying $u \in B_u$, $|B_u| = O(1/\delta)$ and $G[B_u]$ connected for all u, such that μ is $(1 - \frac{1}{2n})$-contractive w.r.t. the α-weighted heat-bath block dynamics for $\mathcal{B}$ and the Hamming metric, where α is the uniform distribution over $\mathcal{B}$.*

Remark 7.5.10. To be more precise, [27] shows that the conclusion of Lemma 7.5.9 is true when β, q, and the maximum block size $M = \max_{u \in V} |B_u|$ satisfy

$$\beta \left(\Delta - 1 + \frac{1}{M} \right) + 3M(\ln \Delta + \ln M) \leq \ln q. \tag{7.12}$$

Thus, for any $\delta > 0$, by taking $M = \lceil \delta^{-1} \rceil$ and $c = 3M(\ln \Delta + \ln M)$, our assumption $\beta \leq \frac{\ln q - c}{\Delta - 1 + \delta}$ in Part 2 of Theorem 7.5.7 implies Eq. (7.12). Moreover, if we take, say, $M \approx \sqrt{\ln q}$ (namely, $\delta \approx 1/\sqrt{\ln q}$), then $c = o(\ln q)$ and our assumption becomes $\beta \leq (1 - o(1)) \frac{\ln q}{\Delta - 1}$ where $o(1)$ tends to 0 as $q \to \infty$; this gives the bound β_1 in Theorem 7.1.2 from the introduction.

Theorem 7.5.7 is an immediate consequence of Lemmas 7.5.8 and 7.5.9 and the results proved in this section.

Proof of Theorem 7.5.7. Part 1 follows directly from Lemma 7.5.8 and Theorem 7.3.1. For Part 2, we note that the block dynamics from Lemma 7.5.9 corresponds to a select-update dynamics with $M = O(1/\delta)$ and $D = \Delta^{O(1/\delta)}/n$; the reason of the latter is that each u is in at most $\Delta^{O(M)}$ blocks. The theorem then follows from Lemma 7.5.9 and Theorem 7.2.3. □

We end this section with the proof of Theorem 7.1.2.

Proof of Theorem 7.1.2. For Ising model, spectral independence is known in the whole uniqueness region [44]. For Potts model, Theorem 7.5.7 establishes spectral independence

in the corresponding parameter regimes. The theorem then follows from Theorems 4.1.1 and 4.1.2. □

CHAPTER 8

SPECTRAL INDEPENDENCE VIA STABILITY OF PARTITION FUNCTION

In this chapter we prove, in an almost black-box fashion, that methods for establishing large zero-free regions needed for the polynomial interpolation method also yield spectral independence. This chapter is based on [46].

8.1 Optimal Mixing Results for Holant Problems

The polynomial interpolation method is a mathematically elegant approach which works in the following manner. To approximate a partition function at a positive real value λ, one needs to prove there is a zero-free region around λ in the complex plane which means that the partition function has no roots in an open region (in the complex plane) containing the point λ. This implies that one can approximate the Taylor series of a simple transformation of the partition function using a logarithmic number of terms, which yields a polynomial-time algorithm to approximate the partition function at λ.

We prove that a zero-free region implies spectral independence. This immediately yields several new rapid mixing results for MCMC methods. We also obtain significantly improved running times in many instances. For a spin system on a graph with n vertices and constant maximum degree Δ, the polynomial interpolation method [110] yields a running time of $O(n^C)$ where the constant C depends on Δ and parameters of the model. In contrast, spectral independence implies an optimal mixing time bound of $O(n \log n)$ for the Glauber dynamics by Theorem 1.2.1 (and more generally optimal mixing for the block dynamics by Theorem 4.1.1).

We state here three sample applications of our techniques; further applications are stated later in this section.

For a graph $G = (V, E)$, we say a vertex v is covered by a subset $S \subseteq E$ of edges if v is

incident to at least one edge in S. The subset $S \subseteq E$ is called an *edge cover* if all vertices are covered by S. Note there is always a trivial edge cover by setting $S = E$. An FPRAS (fully polynomial randomized approximation scheme) was presented for counting the number of edge covers for 3-regular graphs [22]. In [90] an FPTAS (deterministic analog of an FPRAS) for counting edge covers was presented for all graphs using the correlation decay approach, and the running time was $O(m^{1+\log_2 6}n^2)$ where m is the number of edges and n is the number of vertices. An FPRAS for all graphs using MCMC was presented in [73].

The correlation decay algorithm of [90] was extended to weighted (partial) edge covers (with worse running time guarantees) in [91]. In the weighted version, each edge has a weight $\lambda > 0$ and each vertex receives a penalty $\rho \in [0, 1]$ for being uncovered. Every subset $S \subseteq E$ is associated with the weight $w(S) = \rho^{|\mathrm{unc}(S)|}\lambda^{|S|}$, where $\mathrm{unc}(S)$ denotes the set of vertices that are not covered by S. The Gibbs distribution over all subsets of edges is given by $\mu(S) \propto w(S)$. Note, the case $\lambda = 1$ and $\rho = 0$ corresponds to uniformly random exact edge covers.

Finally, an FPTAS using the polynomial interpolation algorithm was presented for graphs with constant maximum degree [66], see also [21]. Using the zero-free results in [66] with our new technical contributions we immediately obtain an FPRAS using a simple MCMC algorithm and with significantly faster running time guarantees.

Theorem 8.1.1 (Weighted Edge Covers)**.** *Let $\Delta \geq 3$ be an integer and let $\lambda > 0$, $\rho \in [0, 1]$ be reals. For every n-vertex graph G of maximum degree Δ, the Glauber dynamics for sampling random weighted edge covers of G with parameters (λ, ρ) mixes in $Cn \log n$ steps where $C = C(\Delta, \lambda, \rho)$ is a constant independent of n.*

One of the seminal results in the field of approximate counting is the work of Jerrum and Sinclair [77] presenting an FPRAS for the partition function of the ferromagnetic Ising model on any graph. Recall that, the Ising model on a graph $G = (V, E)$ is described by two parameters, the edge activity $\beta_{\mathrm{Ising}} > 0$ and the vertex activity $\lambda_{\mathrm{Ising}} > 0$. The Gibbs distribution of the Ising model is over all $\{+, -\}$ spin assignments to vertices. Every

configuration $\sigma : V \to \{+, -\}$ has density $\mu_{\text{Ising}}(\sigma) \propto \beta_{\text{Ising}}^{m(\sigma)} \lambda_{\text{Ising}}^{|\sigma^{-1}(+)|}$ where $m(\sigma)$ denotes the number of monochromatic edges in σ and $\sigma^{-1}(+)$ is the set of vertices assigned spin $+$. The model is ferromagnetic when $\beta_{\text{Ising}} > 1$, in which case neighboring vertices are more likely to have the same spin.

The central task of the Jerrum-Sinclair algorithm is sampling from the Gibbs distribution for the high-temperature expansion of the Ising model which is defined on all subsets of edges weighted to prefer subgraphs with more even degree vertices. For a graph $G = (V, E)$, an edge weight $\lambda > 0$, and a vertex penalty $\rho \in [0, 1]$, the Gibbs distribution μ for weighted (partial) even subgraphs is defined on all subsets of edges; a subset $S \subseteq E$ has weight $w(S) = \rho^{|\mathsf{odd}(S)|}\lambda^{|S|}$ where $\mathsf{odd}(S)$ is the set of odd-degree vertices in the subgraph (V, S), and $\mu(S) \propto w(S)$. The weighted even subgraphs model is related to the ferromagnetic Ising model by $\beta_{\text{Ising}} = \frac{1+\lambda}{1-\lambda}$ and $\lambda_{\text{Ising}} = \frac{1+\rho}{1-\rho}$, for which one can easily transform a subset of edges from μ to a sample from μ_{Ising} [64]. Note that if $\rho = 0$ then μ is the distribution over all weighted exact even subgraphs, corresponding to the ferromagnetic Ising model without external fields (i.e., $\lambda_{\text{Ising}} = 1$).

In [77], an MCMC algorithm is presented to sample weighted even subgraphs of an arbitrary (unbounded-degree) graph in time $O(m^3 \text{poly}(1/\rho))$ where m is the number of edges. In another direction, [94] presents an FPTAS for approximating the partition function of the ferromagnetic Ising model with nonzero fields on bounded-degree graphs, using Barvinok's polynomial interpolation method and the Lee-Yang theory. As is common for this type of approach, the running time of [94] is n^C for a constant C depending on the maximum degree of the graph and the parameters of the Ising model.

Here we use our results relating zero-free regions and spectral independence to obtain a faster MCMC algorithm for bounded-degree graphs when $\rho > 0$.

Theorem 1.1.6 (Weighted Even Subgraphs; Ferromagnetic Ising Model)**.** *Let $\Delta \geq 3$ be an integer and let $\lambda > 0$, $\rho \in (0, 1]$ be reals. For every n-vertex graph G of maximum degree Δ, the mixing time of the Glauber dynamics for sampling random weighted even*

subgraphs of G with parameters (λ, ρ) is at most $Cn \log n$ where $C = C(\Delta, \lambda, \rho)$ is a constant independent of n. In particular, we get an approximate sampling algorithm with running time $O(n \log n)$ for the ferromagnetic Ising model with edge activity $\beta_{\text{Ising}} = \frac{1+\lambda}{1-\lambda}$ and external field $\lambda_{\text{Ising}} = \frac{1+\rho}{1-\rho}$.

Remark 8.1.2. In [77], the MCMC method can actually be used to obtain a sampler for $\rho = 0$ corresponding to weighted exact even subgraphs. This is achieved by taking $\rho = 1/n$ and using rejection sampling. Notice that the running time of [77] is polynomial in $1/\rho$, and therefore this gives a $\text{poly}(n)$ time algorithm for sampling weighted exact even subgraphs and hence for the ferromagnetic Ising model without fields. Unfortunately, Theorem 1.1.6 cannot be used to obtain a sampler for $\rho = 0$, since our bound on the mixing time of the Glauber dynamics (the constant C from Theorem 1.1.6) depends exponentially on $1/\rho$.

Finally, we simultaneously generalize [76, 53, 21] to all antiferromagnetic two-spin edge models, i.e., antiferromagnetic two-spin models on the class of line graphs. Again, in the bounded-degree regime we obtain optimal mixing times. Before we state the result, let us recall the definitions of 2-spin systems from Chapter 5. For a graph $G = (V, E)$ and fixed parameters $\beta \geq 0$, $\gamma > 0$, $\lambda > 0$, the Gibbs distribution of the corresponding two-spin edge model on G is given by

$$\mu(\sigma) \propto \beta^{m_1(\sigma)} \gamma^{m_0(\sigma)} \lambda^{|\sigma^{-1}(1)|}, \quad \forall \sigma \in \{0,1\}^E \tag{8.1}$$

where $m_i(\sigma)$ denotes the number of pairs of edges e, f sharing a single endpoint such that $\sigma(e) = \sigma(f) = i$, for each $i = 0, 1$. We say the system is antiferromagnetic if $\beta\gamma < 1$ and ferromagnetic if $\beta\gamma > 1$ (note that $\beta\gamma = 1$ corresponds to a trivial product measure). The case $\beta = 0$ and $\gamma = 1$ recovers the monomer-dimer model for matchings weighted by λ, and the case $\beta = \gamma$ recovers the Ising model on the line graph of G.

Theorem 8.1.3 (Antiferromagnetic Two-Spin Edge Models)**.** *Let $\Delta \geq 3$ be an integer and let $\beta \geq 0$, $\gamma > 0$, $\lambda > 0$ be reals such that $\beta\gamma < 1$. For every n-vertex graph G of maximum*

degree Δ, the Glauber dynamics for sampling from the antiferromagnetic two-spin edge model on G with parameters (β, γ, λ) mixes in $Cn \log n$ steps where $C = C(\Delta, \beta, \gamma, \lambda)$ is a constant independent of n.

We present further applications of our methods in Section 8.6.

8.2 Establishing Spectral Independence in Zero-Free Regions

We need a few preliminary definitions before formally stating our technical results. Our results hold for an arbitrary distribution on a discrete product space; this general setup contains spin systems as a special case. Let V be a finite set and we refer to the elements in V as vertices. For an integer $q \geq 2$, the set of spins is $Q = \{0\} \cup Q_1$ where $Q_1 = \{1, \ldots, q-1\}$ and we treat 0 as a special spin. The state space is $\Omega = Q^V$, the collection of all spin assignments of vertices. Finally, let $w : \Omega \to \mathbb{R}_{\geq 0}$ be a nonnegative weight function that is not always zero; i.e., $w(\sigma) > 0$ for at least one $\sigma \in \Omega$.

Let $\boldsymbol{\lambda} : V \times Q_1 \to \mathbb{C}$ be a vector of (complex) external fields; each $\lambda_{v,k}$ represents the weight of vertex v receiving spin k. Spin 0 has no external field. Given w, the *partition function* is a multivariate polynomial of $\boldsymbol{\lambda}$ defined as:

$$Z_w(\boldsymbol{\lambda}) = \sum_{\sigma \in \Omega} w(\sigma) \boldsymbol{\lambda}^\sigma, \quad \text{where } \boldsymbol{\lambda}^\sigma = \prod_{v \in V \colon \sigma_v \neq 0} \lambda_{v,\sigma_v}. \tag{8.2}$$

If $\boldsymbol{\lambda}$ is real and positive (i.e., every $\lambda_{v,k} \in \mathbb{R}_+$), then the *Gibbs distribution* $\mu = \mu_{w,\boldsymbol{\lambda}}$ is given by:

$$\mu(\sigma) = \frac{w(\sigma)\boldsymbol{\lambda}^\sigma}{Z_w(\boldsymbol{\lambda})}, \quad \forall \sigma \in \Omega. \tag{8.3}$$

Note that $Z_w(\boldsymbol{\lambda}) > 0$ since w is not identically zero.

To establish spectral independence we need to consider the model with an arbitrary "pinning" which is a fixed configuration on an arbitrary subset of vertices. We formally define pinnings and the associated notions in Section 8.3, and introduce the relevant notation here.

A configuration $\sigma \in \Omega$ is said to be valid or feasible if $w(\sigma) > 0$. For $\Lambda \subseteq V$, let Ω_Λ denote the set of pinnings on Λ; this is the set of configurations on Λ which have a valid extension to the remaining vertices $V \setminus \Lambda$. For $\Lambda \subseteq V$ and $\tau \in \Omega_\Lambda$, let $V^\tau = V \setminus \Lambda$ denote the set of unpinned vertices, let $Z_w^\tau(\boldsymbol{\lambda})$ be the multivariate conditional partition function under the pinning τ, and let μ^τ be the corresponding conditional distribution.

We can now define the notion of spectral independence. Let $\mathcal{T} = \bigcup_{\Lambda \subseteq V} \Omega_\Lambda$ be the collection of all pinnings. For $\tau \in \mathcal{T}$ let $\mathcal{P}^\tau = \{(v,k) \in V \times Q : v \in V^\tau, k \in \Omega_v^\tau\}$ be the collection of feasible vertex-spin pairs under τ, where Ω_v^τ represents the set of feasible spins at v conditioned on τ. Recall that, for fixed $\tau \in \mathcal{T}$, for every $(u,j),(v,k) \in \mathcal{P}^\tau$, the (pairwise) influence of (u,j) on (v,k) under the pinning τ is given by $\Psi_\mu^\tau(u,j;v,k) = 0$ for $u = v$ and

$$\Psi_\mu^\tau(u,j;v,k) = \mu(\sigma_v = k \mid \sigma_u = j, \sigma_\Lambda = \tau) - \mu(\sigma_v = k \mid \sigma_\Lambda = \tau) \quad \text{for } u \neq v.$$

For a square matrix M with real eigenvalues, let $\mathsf{EigMax}(M)$ denote the maximum eigenvalue of M. We say μ is spectrally independent with constant η if for every pinning $\tau \in \mathcal{T}$ one has

$$\mathsf{EigMax}(\Psi_\mu^\tau) \leq \eta.$$

For a non-empty region Γ of the complex plane, we say a multivariate polynomial $P(z_1, \ldots, z_n)$ is Γ-*stable* if $P(z_1, \ldots, z_n) \neq 0$ whenever $z_j \in \Gamma$ for all j, see Definition 2.5.1. We present a sequence of results connecting spectral independence of the distribution with stability of the partition function. Our first result holds when the zero-free region of the partition function is sufficiently "large", e.g., containing the whole positive real axis. Below for $\Gamma \subseteq \mathbb{C}$ let $\overline{\Gamma}$ denote the closure of Γ and let $\partial\Gamma$ be the boundary of Γ; for $\lambda \in \mathbb{C}$ let $\operatorname{dist}(\lambda, \partial\Gamma) = \inf_{z \in \partial\Gamma} |z - \lambda|$; see Section 2.5.2.

Theorem 8.2.1. *Let $\Gamma \subseteq \mathbb{C}$ be a non-empty open connected region such that Γ is unbounded and $0 \in \overline{\Gamma}$. If the multivariate partition function Z_w is Γ-stable, then for any*

$\lambda \in \mathbb{R}_+ \cap \Gamma$ *the Gibbs distribution* $\mu = \mu_{w,\lambda}$ *with the uniform external field* λ *is spectrally independent with constant*

$$\eta = \frac{8}{\delta}$$

where $\delta = \frac{1}{\lambda}\mathrm{dist}(\lambda, \partial\Gamma)$.

In particular, the statement is true when Γ *is a non-empty open connected region containing the positive real axis; i.e.,* $\mathbb{R}_+ \subseteq \Gamma$.

If we only know that a part of the positive real axis is contained in Γ, then we need to further assume that all conditional partition functions under pinnings are stable. In this case our bound on spectral independence depends on the marginal bound of the distribution μ, which is defined as

$$b = \min_{\substack{\tau \in \mathcal{T} \\ (v,k) \in \mathcal{P}^\tau}} \mu^\tau(\sigma_v = k).$$

Note that $b > 0$ since $\mathcal{P}^\tau$ contains only feasible vertex-spin pairs.

Theorem 8.2.2. *Let* $\lambda^* \in \mathbb{R}_+$ *and let* $\Gamma \subseteq \mathbb{C}$ *be a non-empty open connected region such that* $(0, \lambda^*) \subseteq \Gamma$ *(respectively,* $(\lambda^*, \infty) \subseteq \Gamma$*). If for every pinning* $\tau \in \mathcal{T}$ *the multivariate conditional partition function* Z_w^τ *is* Γ*-stable, then for any* $\lambda \in (0, \lambda^*)$ *(respectively,* $\lambda \in (\lambda^*, \infty)$*) the Gibbs distribution* $\mu = \mu_{w,\lambda}$ *with the uniform external field* λ *is spectrally independent with constant*

$$\eta = \frac{8}{\delta} \min\left\{\frac{1-b}{b}, \frac{\lambda}{b(\lambda^* - \lambda)} + 1\right\}$$

$$\left(\textit{respectively, } \eta = \frac{8}{\delta} \min\left\{\frac{1-b}{b}, \frac{\lambda^*}{b(\lambda - \lambda^*)} + 1\right\}\right)$$

where b *is the marginal bound for* μ *and* $\delta = \frac{1}{\lambda}\mathrm{dist}(\lambda, \partial\Gamma)$.

Remark 8.2.3. The first term $(1-b)/b$ is better when λ is close to λ^*, while the second term is better when λ is close to 0 (respectively, ∞), because usually b/λ is bounded from below when $\lambda \to 0$ (respectively, $b\lambda$ is bounded from below when $\lambda \to \infty$).

Remark 8.2.4. We point out here that Theorem 8.2.2 does not apply to the ferromagnetic Ising model. The celebrated Lee-Yang theorem states that the partition function for the ferromagnetic Ising model is $\mathbb{D}(0,1)$-stable and $\overline{\mathbb{D}}(0,1)^{\mathsf{c}}$-stable where $\mathbb{D}(0,1)$ denotes the open unit ball centered at 0 on the complex plane and $\overline{\mathbb{D}}(0,1)^{\mathsf{c}}$ denotes the exterior of $\mathbb{D}(0,1)$. However, when a pinning is applied, particularly when some vertices are pinned to $+$ and some are $-$, we do not have either $\mathbb{D}(0,1)$-stability or $\overline{\mathbb{D}}(0,1)^{\mathsf{c}}$-stability for the conditional partition function. To see this, notice that such a pinning can result in inconsistent external fields; some fields are < 1 (hence in $\mathbb{D}(0,1)$) while others are > 1 (hence in $\overline{\mathbb{D}}(0,1)^{\mathsf{c}}$), and the Lee-Yang theorem does not apply.

Meanwhile, one should not expect spectral independence to hold for the ferromagnetic Ising model at all temperatures and for all external fields, since the Glauber dynamics is slow mixing when the parameters lie in the tree non-uniqueness region.

If limited information about the zero-free region is given, then spectral independence still holds with a worse bound.

Theorem 8.2.5. *Let $\Gamma \subseteq \mathbb{C}$ be a non-empty open connected region. If for every pinning $\tau \in \mathcal{T}$ the multivariate conditional partition function Z_w^τ is Γ-stable, then for any $\lambda \in \mathbb{R}_+ \cap \Gamma$ the Gibbs distribution $\mu = \mu_{w,\lambda}$ with the uniform external field λ is spectrally independent with constant*

$$\eta = \frac{2}{b\delta^2}$$

where b is the marginal bound for μ and $\delta = \frac{1}{\lambda}\mathrm{dist}(\lambda, \partial\Gamma)$.

Our results Theorems 8.2.1, 8.2.2 and 8.2.5 also hold for non-uniform external fields, i.e., each pair (v,k) is assigned a distinct field $\lambda_{v,k}$, and the zero-free regions are allowed to be distinct for different pairs. See Theorem 8.4.2 for a formal statement.

Our work builds upon the recent work of Anari et al. [3]. Theorem 16 of [3] established spectral independence for any distribution over $\{0,1\}^V$ assuming that the generating polynomial is sector-stable (that is, Γ-stable where $\Gamma = \{z \in \mathbb{C} : |\mathrm{Arg}(z)| \leq \theta\}$ is a sector for

some $\theta \in (0, \pi/2)$). Our results Theorems 8.2.1, 8.2.2 and 8.2.5 strengthen theirs in two ways. First we consider an arbitrary discrete product space $\Omega = Q^V$ for a finite label set Q, rather than binary domains. Second and more importantly, we do not have any restriction on the zero-free region Γ and the results hold for any open connected region. This allows us to apply our results in a much broader setting. See Section 8.4 for more details.

To establish zero-free regions for our main applications, we utilize the approach in [66], which reduces the problem via Asano-Ruelle contractions [7, 114] to showing a sufficiently large zero-free region for a collection of bounded-degree univariate polynomials, one for each vertex of the input graph. These univariate polynomials are referred to as the local polynomials, since they only depend on the configuration restricted to edges incident to the given vertex. We note a very similar idea was also used in [126, 21] to establish zero-free regions, although their methods do not go through Asano-Ruelle contractions. See Section 8.5 for more details.

It was also shown in a sequence of papers [11, 16, 17, 12, 13, 112] that one can establish large zero-free regions via an inductive approach based on conditioning the distribution. This method of establishing zero-free regions also works nicely for us, as spectral independence requires a bound on the pairwise influences for all conditional distributions. We show that one can deduce rapid mixing of the Glauber dynamics in a nearly black-box fashion from these zero-free methods for several problems in Section 8.6.

Algorithmically, our results have several advantages over prior works utilizing zero-free regions. In particular, the polynomial interpolation method pioneered by Barvinok [10] typically only yields quasi-polynomial time algorithms in general, and polynomial time algorithms with exponent depending on the maximum degree for problems arising from graphs [110]. In contrast, we obtain fast algorithms for sampling and counting. Another feature of our approach is that we only need the zero-free region to be sufficiently large. This is in contrast to the polynomial interpolation technique, which needs the zero-free region to also contain a point at which the partition function is easily computable.

8.3 Preliminaries for Pinnings

In this section we give more details on pinnings and conditional distributions. Note that our setting in this chapter is more general than previous chapters as described in Chapter 2, mostly because here we consider an arbitrary weight function on configurations which is independent of the external fields, and we view the fields as variables of the partition function.

Let $q \geq 2$ be an integer and write $q_1 = q - 1$. Let V be a finite set of vertices and let $Q = \{0\} \cup Q_1$ be the set of spins where $Q_1 = \{1, \ldots, q_1\}$. Every spin assignment $\sigma : V \to Q$ is called a configuration. The state space $\Omega = Q^V$ is the collection of all configurations and let $w : \Omega \to \mathbb{R}_{\geq 0}$ be a nonnegative weight function that is not identically zero. A configuration $\sigma \in \Omega$ is said to be valid or feasible if $w(\sigma) > 0$. For $\Lambda \subseteq V$, define set of pinnings on Λ by

$$\Omega_\Lambda = \left\{\tau \in Q^\Lambda : \exists \text{ valid } \sigma \in \Omega \text{ s.t. } \sigma_\Lambda = \tau\right\}.$$

Note Ω_V is the set of all valid configurations. Let $\mathcal{T} = \bigcup_{\Lambda \subseteq V} \Omega_\Lambda$ be the collection of all pinnings.

For a pinning $\tau \in \mathcal{T}$, let V^τ denote the set of unpinned vertices; so if $\tau \in \Omega_\Lambda$ then $V^\tau = V \setminus \Lambda$. For $v \in V^\tau$, let Ω_v^τ be the set of valid spins at v under τ:

$$\Omega_v^\tau = \{k \in Q : \exists \text{ valid } \sigma \in \Omega \text{ s.t. } \sigma_\Lambda = \tau \text{ and } \sigma_v = k\}.$$

Define the collection of feasible vertex-spin pairs under τ by $\mathcal{P}^\tau = \{(v, k) \in V \times Q : v \in V^\tau, k \in \Omega_v^\tau\}$, and the collection of pairs with nonzero spins by $\mathcal{P}_1^\tau = \{(v, k) \in \mathcal{P}^\tau : k \neq 0\}$. We write Ω_v, $\mathcal{P}$, and $\mathcal{P}_1$ when no pinning is applied.

Let $\boldsymbol{\lambda} : \mathcal{P}_1 \to \mathbb{C}$ be a vector of complex external fields. Given a pinning $\tau \in \mathcal{T}$ on $\Lambda \subseteq V$ with $U = V^\tau = V \setminus \Lambda$, the conditional partition function Z_w^τ under τ is a

multivariate polynomial of $\boldsymbol{\lambda}$ defined as

$$Z_w^\tau(\boldsymbol{\lambda}) = \sum_{\sigma\in\Omega:\,\sigma_\Lambda=\tau} w(\sigma)\boldsymbol{\lambda}^{\sigma_U}, \quad \text{where } \boldsymbol{\lambda}^{\sigma_U} = \prod_{v\in U:\,\sigma_v\neq 0} \lambda_{v,\sigma_v}.$$

When there is no pinning, this matches Eq. (8.2) from the introduction. Observe that Z_w^τ depends only on the variables $\{\lambda_{v,k} : (v,k) \in \mathcal{P}_1^\tau\}$, and that Z_w^τ is not identically zero since τ is a pinning. If $\boldsymbol{\lambda}$ is real and positive, then $Z_w^\tau(\boldsymbol{\lambda}) > 0$ and we obtain the conditional Gibbs distribution:

$$\mu^\tau(\sigma) = \mu(\sigma \mid \sigma_\Lambda = \tau) = \frac{w(\sigma)\boldsymbol{\lambda}^{\sigma_U}}{Z_w^\tau(\boldsymbol{\lambda})}, \quad \forall \sigma \in \Omega \text{ s.t. } \sigma_\Lambda = \tau.$$

Again this matches Eq. (8.3) when there is no pinning.

8.4 Proofs of Spectral Independence via Stability

In this section, we deduce spectral independence of a distribution from the stability of the associated partition function, and thus prove Theorems 8.2.1, 8.2.2 and 8.2.5.

We first show that pinning preserves stability of the partition function if the zero-free region is unbounded and contains 0 in its closure. Intuitively, the pinning $\sigma_v = k$ for $k \neq 0$ corresponds to taking $\lambda_{v,k} = \infty$ (which is achieved by taking derivative with respect to $\lambda_{v,k}$), and the pinning $\sigma_v = 0$ corresponds to taking $\lambda_{v,k} = 0$ for all $k \in Q_1$. Hence, under an arbitrary pinning the conditional partition function is just the original partition function specialized at ∞ and 0 for specific external fields, and the fact that the closure of the zero-free region contains ∞ (i.e., unboundedness) and 0 guarantees that after specialization the resulted partition function is still stable. This is formalized by the following lemma.

Lemma 8.4.1. *Let $\{\Gamma_{v,k} \subseteq \mathbb{C} : (v,k) \in \mathcal{P}_1\}$ be a collection of non-empty open connected regions such that for every $(v,k) \in \mathcal{P}_1$ the region $\Gamma_{v,k}$ is unbounded and $0 \subseteq \overline{\Gamma_{v,k}}$. If the multivariate partition function Z_w is $\left(\prod_{(v,k)\in\mathcal{P}_1} \Gamma_{v,k}\right)$-stable, then for every pinning $\tau \in \mathcal{T}$*

the multivariate conditional partition function Z_w^τ is $\left(\prod_{(v,k)\in\mathcal{P}_1^\tau}\Gamma_{v,k}\right)$-stable.

Next, we show the following grand theorem in the multivariate setting, from which one can deduce the bounds on spectral independence given in Theorems 8.2.1, 8.2.2 and 8.2.5.

Theorem 8.4.2. *Let $\{\Gamma_{v,k}\subseteq\mathbb{C}:(v,k)\in\mathcal{P}_1\}$ be a collection of non-empty open connected regions, and let $\boldsymbol{\lambda}:\mathcal{P}_1\to\mathbb{R}_+$ such that $\lambda_{v,k}\in\mathbb{R}_+\cap\Gamma_{v,k}$ for each $(v,k)\in\mathcal{P}_1$. Suppose that for every pinning $\tau\in\mathcal{T}$ the multivariate conditional partition function Z_w^τ is $\left(\prod_{(v,k)\in\mathcal{P}_1^\tau}\Gamma_{v,k}\right)$-stable. Then the Gibbs distribution $\mu=\mu_{w,\boldsymbol{\lambda}}$ with external fields $\boldsymbol{\lambda}$ is spectrally independent with constant*

$$\eta=\frac{2}{b\delta^2}, \tag{8.4}$$

where b is the marginal bound for μ and

$$\delta=\min_{(v,k)\in\mathcal{P}_1}\frac{1}{\lambda_{v,k}}\mathrm{dist}(\lambda_{v,k},\partial\Gamma_{v,k}).$$

Furthermore:

1. *For each $v\in V$, let $\Gamma_v\subseteq\mathbb{C}$ be the connected component of the intersection region $\bigcap_{k\in Q_1}\frac{1}{\lambda_{v,k}}\Gamma_{v,k}$ which contains 1 (note that $1\in\frac{1}{\lambda_{v,k}}\Gamma_{v,k}$ for all (v,k)). If for every $v\in V$ the region Γ_v is unbounded and $0\in\overline{\Gamma_v}$, then spectral independence holds with constant*

$$\eta=\frac{8}{\delta}.$$

   In particular, the statement is true if $\mathbb{R}_+\subseteq\Gamma_{v,k}$ for each $(v,k)\in\mathcal{P}_1$.

2. *If there exists $\lambda^*\in\mathbb{R}_+$ such that $\lambda_{v,k}\in(0,\lambda^*)\subseteq\Gamma_{v,k}$ for every $(v,k)\in\mathcal{P}_1$, then spectral independence holds with constant*

$$\eta=\frac{8}{\delta}\min\left\{\frac{1-b}{b},\frac{\lambda_{\max}}{b(\lambda^*-\lambda_{\max})}+1\right\},$$

where $\lambda_{\max} = \max_{(v,k)\in\mathcal{P}_1} \lambda_{v,k}$.

3. *If there exists* $\lambda^* \in \mathbb{R}_+$ *such that* $\lambda_{v,k} \in (\lambda^*, \infty) \subseteq \Gamma_{v,k}$ *for every* $(v,k) \in \mathcal{P}_1$*, then spectral independence holds with constant*

$$\eta = \frac{8}{\delta} \min\left\{ \frac{1-b}{b},\ \frac{\lambda^*}{b(\lambda_{\min} - \lambda^*)} + 1 \right\},$$

where $\lambda_{\min} = \min_{(v,k)\in\mathcal{P}_1} \lambda_{v,k}$.

We prove Theorem 8.4.2 by upper bounding the absolute row sum of the influence matrix Ψ_μ^τ for any pinning $\tau \in \mathcal{T}$; namely, for each $(u,j) \in \mathcal{P}^\tau$ we bound the sum of absolute values of the influences from (u,j) to all other pairs $(v,k) \in \mathcal{P}^\tau$, see Lemma 8.4.4. We accomplish this by greatly strengthening and generalizing the proof strategy in [3].

At a high level, the work [3] views the sum of absolute influences as the derivative of some function f produced by the conditional partition functions. The variables of f are just the external fields of the partition function which lie in some zero-free region Γ and the stability of the conditional partition functions guarantees that the image of f is contained in some nice region Γ'. In [3], the authors study sector-stability of the partition function for the binary state space $\{0,1\}^V$; in particular, both the zero-free region Γ and the region Γ' containing the image are sectors for their choice of f. Then, by applying conformal mappings between the sector and the unit disk, the derivative of f can be upper bounded using the Schwarz-Pick Theorem (Theorem 2.5.3).

However, here we are facing a more challenging situation since we try to establish spectral independence from an arbitrary zero-free region Γ for any discrete product space Q^V. In fact, for us the regions Γ and Γ' are in abstract forms and to apply the Schwarz-Pick Theorem we need to design good mappings from Γ and Γ' to the unit disk. This is achieved by both carefully describing these regions and utilizing tools from complex analysis, especially the Riemann Mapping Theorem (Theorem 2.5.4). See Section 8.4.3 for details of this part.

We now provide the proofs of Theorems 8.2.1, 8.2.2 and 8.2.5 from the introduction.

Proof of Theorem 8.2.1. Follows from Lemma 8.4.1 and Item 1 of Theorem 8.4.2. □

Proof of Theorem 8.2.2. Follows from Items 2 and 3 of Theorem 8.4.2. □

Proof of Theorem 8.2.5. Follows from Eq. (8.4) of Theorem 8.4.2. □

After proving Lemma 8.4.1 in Section 8.4.1, we prove Theorem 8.4.2 in Section 8.4.2 and prove Lemma 8.4.4, a central lemma for bounding the absolute sum of influences, in Section 8.4.3.

8.4.1 Preservation of Stability under Pinnings

In this subsection we present the proof of Lemma 8.4.1.

Let $\tau \in \mathcal{T}$ be an arbitrary pinning on $\Lambda \subseteq V$ and let $U = V^\tau = V \setminus \Lambda$ be the set of unpinned vertices. We consider the conditional partition function Z_w^τ under the pinning τ. As discussed earlier, one can view Z_w^τ as obtained from the original partition function Z_w by specializing at 0 and taking derivatives for certain variables provided by the pinning τ. To be more precise, we define $\Lambda_0 = \{v \in \Lambda : \tau_v = 0\}$ to be the set of vertices pinned to spin 0, and let $\Lambda_1 = \{v \in \Lambda : \tau_v \neq 0\}$ be those pinned to nonzero spins. We also define $\tau_1 = \tau_{\Lambda_1}$ to be the pinning restricted to vertices with nonzero spins. The key observation here is that

$$Z_w^\tau(\boldsymbol{\lambda}) = \left(\frac{\partial}{\partial \boldsymbol{\lambda}_{\Lambda_1,\tau_1}} Z_w(\boldsymbol{\lambda}) \right) \Bigg|_{\boldsymbol{\lambda}_{\Lambda_0}=0} \tag{8.5}$$

where $\boldsymbol{\lambda}_{\Lambda_0} = 0$ represents plugging in $\lambda_{v,k} = 0$ for all $v \in \Lambda_0$ and $k \in Q_1$, and $\frac{\partial}{\partial \boldsymbol{\lambda}_{\Lambda_1,\tau_1}}$ represents taking derivatives $\frac{\partial}{\partial \lambda_{v,\tau_v}}$ for all $v \in \Lambda_1$. Hence, to establish Lemma 8.4.1 it suffices to show that specialization at 0 and differentiation preserves Γ-stability if the zero-free region Γ is unbounded and $0 \in \overline{\Gamma}$. This is actually true for any multi-affine polynomial, which is a polynomial whose monomials are all square-free.

Lemma 8.4.3. *Let $n \geq 1$ be an integer and let $\Gamma_1, \ldots, \Gamma_n \subseteq \mathbb{C}$ be non-empty open connected regions. Let $P \in \mathbb{C}[z_1, \ldots, z_n]$ be a multi-affine polynomial and assume that P is $(\prod_{\ell=1}^{n} \Gamma_\ell)$-stable. Then:*

1. *(Inversion) The polynomial*

$$P_1(z_1, z_2, \ldots, z_n) = z_1 P(\frac{1}{z_1}, z_2, \ldots, z_n)$$

   is $(\Gamma_1^{-1} \times \prod_{\ell=2}^{n} \Gamma_\ell)$-stable;

2. *(Specialization) If $0 \in \overline{\Gamma_1}$, then the polynomial*

$$P_2(z_2, \ldots, z_n) = P(0, z_2, \ldots, z_n)$$

   is either $(\prod_{\ell=2}^{n} \Gamma_\ell)$-stable or identically zero;

3. *(Differentiation) If Γ_1 is unbounded, then the polynomial*

$$P_3(z_2, \ldots, z_n) = \frac{\partial}{\partial z_1} P(z_1, z_2, \ldots, z_n)$$

   is either $(\prod_{\ell=2}^{n} \Gamma_\ell)$-stable or identically zero.

Proof. Consider first the inversion. Suppose for sake of contradiction that P_1 is not $(\Gamma_1^{-1} \times \prod_{\ell=2}^{n} \Gamma_\ell)$-stable. Then there exists $w_1 \in \Gamma_1^{-1}$ and $z_\ell \in \Gamma_\ell$ for $2 \leq \ell \leq n$ such that $P_1(w_1, z_2, \ldots, z_n) = 0$. Note that $w_1 = 1/z_1$ for some $z_1 \in \Gamma_1 \setminus \{0\}$. It follows that

$$0 = z_1 P_1(w_1, z_2, \ldots, z_n) = z_1 w_1 P\left(\frac{1}{w_1}, z_2, \ldots, z_n\right) = P(z_1, z_2, \ldots, z_n),$$

contradicting to the stability of P. Hence, we have the desired stability for P_1.

Next consider specialization. Since Γ_1 is open and $0 \in \overline{\Gamma_1}$, there exists a sequence of complex numbers $\{\zeta_m\}_{m=1}^{\infty}$ such that $\zeta_m \in \Gamma_1$ and $\lim_{m\to\infty} \zeta_m = 0$. Let $f_m(z_2, \ldots, z_n) =$

$P(\zeta_m, z_2, \ldots, z_n)$ be a polynomial of degree $\leq \deg(P)$ for each m. Then f_m is $(\prod_{\ell=2}^n \Gamma_\ell)$-stable by the stability assumption of P. Furthermore, the sequence $\{f_m\}_{m=1}^\infty$ converges to P_2 coefficient-wisely, and hence uniformly on compact subsets; see, e.g., Lemma 33 in [3]. Hurwitz' Theorem (Theorem 2.5.2) then implies that P_2 is either $(\prod_{\ell=2}^n \Gamma_\ell)$-stable or identically zero, as claimed.

Last we consider differentiation. Since Γ_1 is open and unbounded, we deduce that the region $\Gamma_1^{-1} = \{1/z : z \in \Gamma_1 \setminus \{0\}\}$ is open and satisfies $0 \in \overline{\Gamma_1^{-1}}$. Recall that we have shown the inversion $P_1(z_1, z_2, \ldots, z_n) = z_1 P(\frac{1}{z_1}, z_2, \ldots, z_n)$ is $(\Gamma_1^{-1} \times \prod_{\ell=2}^n \Gamma_\ell)$-stable. Now observe that, for a multi-affine polynomial P, the derivative P_3 of P with respect to z_1 is the same as the specialization of P_1 at $z_1 = 0$:

$$P_3(z_2, \ldots, z_n) = \frac{\partial}{\partial z_1} P(z_1, z_2, \ldots, z_n) = P_1(0, z_2, \ldots, z_n).$$

Hence, we immediately conclude from previous results that P_3 is either $(\prod_{\ell=2}^n \Gamma_\ell)$-stable or identically zero. □

Lemma 8.4.1 is an immediate consequence of Lemma 8.4.3.

Proof of Lemma 8.4.1. Observe that the partition function Eq. (8.2) is multi-affine. The lemma then follows from Eq. (8.5) and Lemma 8.4.3. Notice that the conditional partition functions are never identically zero since pinnings are extendable to valid full configurations. □

8.4.2 Bounds on Spectral Independence

In this subsection we prove Theorem 8.4.2.

An important observation is that it is sufficient to assume $1 \in \Gamma_{v,k}$ for every $(v, k) \in \mathcal{P}_1$ and consider the Gibbs distribution with the all-one external fields $\mathbf{1}$. In general, given the

external field $\lambda_{v,k} \in \Gamma_{v,k}$ for each (v,k), we may reweight the configurations by

$$\widetilde{w}(\sigma) = w(\sigma) \prod_{v \in V:\, \sigma_v \neq 0} \lambda_{v,\sigma_v}, \quad \forall \sigma \in \Omega \tag{8.6}$$

and the new partition function is

$$Z_{\widetilde{w}}(\boldsymbol{\lambda}) = \sum_{\sigma \in \Omega} \widetilde{w}(\sigma) \boldsymbol{\lambda}^{\sigma}. \tag{8.7}$$

In particular, $Z_w(\boldsymbol{\lambda}) = Z_{\widetilde{w}}(\mathbf{1})$ and $\mu_{w,\boldsymbol{\lambda}} = \mu_{\widetilde{w},\mathbf{1}}$ for the given $\boldsymbol{\lambda} = (\lambda_{v,k})$. In other words, we hide the external fields into the weight of configurations and under the new weights we are interested in the all-one external fields. This will simplify the notations. Note that, if for $\tau \in \mathcal{T}$ the multivariate conditional partition function Z_w^τ is $\big(\prod_{(v,k)\in\mathcal{P}_1^\tau} \Gamma_{v,k}\big)$-stable, then the reweighted conditional partition function $Z_{\widetilde{w}}^\tau$ is $\big(\prod_{(v,k)\in\mathcal{P}_1^\tau} \widetilde{\Gamma}_{v,k}\big)$-stable where $\widetilde{\Gamma}_{v,k} = \frac{1}{\lambda_{v,k}}\Gamma_{v,k}$ for each $(v,k) \in \mathcal{P}_1$.

In the rest of this section, we assume that $1 \in \Gamma_{v,k}$ and consider the case of all-one external fields. The following lemma is an important step towards deducing Theorem 8.4.2; it builds upon the proof strategy of [3] while generalizing their result to a great extent.

Lemma 8.4.4. *Consider the Gibbs distribution $\mu = \mu_{w,\mathbf{1}}$ with the all-one external fields $\mathbf{1}$. Let $\tau \in \mathcal{T}$ be a fixed pinning and let $\{\Gamma_{v,k} \subseteq \mathbb{C} : (v,k) \in \mathcal{P}_1^\tau\}$ be a collection of non-empty open connected regions such that $1 \in \Gamma_{v,k}$ for each $(v,k) \in \mathcal{P}_1^\tau$. For every $v \in V^\tau$ let $\Gamma_v \subseteq \mathbb{C}$ be the connected component of the intersection $\bigcap_{0\neq k\in\Omega_v^\tau} \Gamma_{v,k}$ that contains 1. If the multivariate conditional partition function Z_w^τ is $\big(\prod_{(v,k)\in\mathcal{P}_1^\tau} \Gamma_{v,k}\big)$-stable, then the influence matrix Ψ_μ^τ under the pinning τ satisfies*

$$\mathsf{EigMax}(\Psi_\mu^\tau) \leq \big\|\Psi_\mu^\tau\big\|_\infty \leq \min\left\{\frac{2}{b\delta^2}, \frac{8}{\delta} \max_{\substack{\tau\in\mathcal{T}\\(v,k)\in\mathcal{P}^\tau}} \operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right)\right\},$$

where: b is the marginal bound for μ;

$$\delta = \min_{(v,k)\in\mathcal{P}_1^\tau} \operatorname{dist}(1, \partial\Gamma_{v,k}); \quad p_{v,k}^\tau = \mu^\tau(\sigma_v = k);$$
$$\mathcal{C}_{v,k}^\tau = -\frac{1}{p_{v,k}^\tau}(\Gamma_{v,k} - 1)^{-1} = \left\{ -\frac{1}{p_{v,k}^\tau(z-1)} : z \in \Gamma_{v,k} \setminus \{1\} \right\}, \quad \textit{for } k \neq 0;$$
$$\mathcal{C}_{v,0}^\tau = \frac{1}{p_{v,0}^\tau}\left((\Gamma_v - 1)^{-1} + 1\right) = \left\{ \frac{z}{p_{v,0}^\tau(z-1)} : z \in \Gamma_v \setminus \{1\} \right\}.$$

The following two lemmas are helpful for bounding the distance $\operatorname{dist}(1, \mathcal{C}_{v,k}^\tau)$ which is given in Lemma 8.4.4.

Lemma 8.4.5. *Let $\tau \in \mathcal{T}$ and $(v,k) \in \mathcal{P}_1^\tau$.*

1. *If $\Gamma_{v,k}$ is unbounded, then*
$$\operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right) \leq 1.$$

2. *Let $\alpha_{v,k} = \inf(\Gamma_{v,k} \cap \mathbb{R}_+)$ and $\beta_{v,k} = \sup(\Gamma_{v,k} \cap \mathbb{R}_+)$. Then*
$$\operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right) \leq \min\left\{ \frac{\alpha_{v,k}}{p_{v,k}^\tau(1-\alpha_{v,k})} + \frac{1-p_{v,k}^\tau}{p_{v,k}^\tau},\ \frac{1}{p_{v,k}^\tau(\beta_{v,k}-1)} + 1 \right\},$$

with the convention that $\frac{1}{\infty} = 0$ if $\beta_{v,k} = \infty$.

Lemma 8.4.6. *Let $\tau \in \mathcal{T}$ and $(v,0) \in \mathcal{P}^\tau$.*

1. *If $0 \in \overline{\Gamma_v}$, then*
$$\operatorname{dist}\left(1, \mathcal{C}_{v,0}^\tau\right) \leq 1.$$

2. *Let $\alpha_v = \inf(\Gamma_v \cap \mathbb{R}_+)$ and $\beta_v = \sup(\Gamma_v \cap \mathbb{R}_+)$. Then*
$$\operatorname{dist}\left(1, \mathcal{C}_{v,0}^\tau\right) \leq \min\left\{ \frac{\alpha_v}{p_{v,0}^\tau(1-\alpha_v)} + 1,\ \frac{1}{p_{v,0}^\tau(\beta_v-1)} + \frac{1-p_{v,0}^\tau}{p_{v,0}^\tau} \right\},$$

with the convention that $\frac{1}{\infty} = 0$ if $\beta_v = \infty$.

Combining Lemmas 8.4.4 to 8.4.6, we are able to establish Theorem 8.4.2.

Proof of Theorem 8.4.2. As discussed at the beginning of this subsection, we can reweight the configurations by Eq. (8.6) and consider the all-one external fields under the new weights, so that Lemma 8.4.4 applies. In particular, for an arbitrary pinning $\tau \in \mathcal{T}$ the reweighted conditional partition function $Z^{\tau}_{\widetilde{w}}$ given by Eq. (8.7) is $\big(\prod_{(v,k)\in\mathcal{P}_1^{\tau}} \widetilde{\Gamma}_{v,k}\big)$-stable where $\widetilde{\Gamma}_{v,k} = \frac{1}{\lambda_{v,k}}\Gamma_{v,k}$ for each $(v,k) \in \mathcal{P}_1$. Thus, the first upper bound in Lemma 8.4.4 implies that the Gibbs distribution $\mu = \mu_{w,\boldsymbol{\lambda}} = \mu_{\widetilde{w},\mathbf{1}}$ is spectrally independent with constant $\eta = 2/(b\delta^2)$ where b is the marginal bound for μ and

$$\delta = \min_{(v,k)\in\mathcal{P}_1} \operatorname{dist}\left(1, \partial\widetilde{\Gamma}_{v,k}\right) = \min_{(v,k)\in\mathcal{P}_1} \frac{1}{\lambda_{v,k}} \operatorname{dist}\left(\lambda_{v,k}, \partial\Gamma_{v,k}\right)$$

as claimed.

For each $v \in V$, let $\widetilde{\Gamma}_v = \Gamma_v \subseteq \mathbb{C}$ be the connected component of the intersection $\bigcap_{k\in Q_1} \widetilde{\Gamma}_{v,k} = \bigcap_{k\in Q_1} \frac{1}{\lambda_{v,k}}\Gamma_{v,k}$ that contains 1. Then we further have the following.

1. If for every $v \in V$ the region $\widetilde{\Gamma}_v$ is unbounded and 0 is contained in the closure of $\widetilde{\Gamma}_v$, then the first part of Lemmas 8.4.5 and 8.4.6 implies that $\operatorname{dist}(1, \mathcal{C}^{\tau}_{v,k}) \leq 1$ for all $\tau \in \mathcal{T}$ and $(v,k) \in \mathcal{P}^{\tau}$. Hence, by the second upper bound in Lemma 8.4.4 spectral independence holds with constant $\eta = 8/\delta$.

   If $\mathbb{R}_+ \subseteq \Gamma_{v,k}$ for each $(v,k) \in \mathcal{P}_1$, then by definition $\widetilde{\Gamma}_v$ is unbounded and 0 is contained in the closure of $\widetilde{\Gamma}_v$ for every $v \in V$; therefore, spectral independence holds with $\eta = 8/\delta$.

2. If there exists $\lambda^* \in \mathbb{R}_+$ such that $\lambda_{v,k} \in (0, \lambda^*) \subseteq \Gamma_{v,k}$ for every $(v,k) \in \mathcal{P}_1$, then one has

$$\alpha_{v,k} = \inf\left(\widetilde{\Gamma}_{v,k} \cap \mathbb{R}_+\right) = 0, \quad \beta_{v,k} = \sup\left(\widetilde{\Gamma}_{v,k} \cap \mathbb{R}_+\right) \geq \frac{\lambda^*}{\lambda_{v,k}},$$
$$\alpha_{v} = \inf\left(\widetilde{\Gamma}_{v} \cap \mathbb{R}_+\right) = 0, \quad \beta_{v} = \sup\left(\widetilde{\Gamma}_{v} \cap \mathbb{R}_+\right) \geq \frac{\lambda^*}{\lambda_{v,\max}},$$

where $\lambda_{v,\max} = \max_{k \in Q_1} \lambda_{v,k}$. Thus, we deduce from the second part of Lemmas 8.4.5 and 8.4.6 that for all $\tau \in \mathcal{T}$ and $(v,k) \in \mathcal{P}^\tau$,

$$\begin{aligned}
\operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right) &\leq \min\left\{\frac{1-p_{v,k}^\tau}{p_{v,k}^\tau}, \frac{\lambda_{v,k}}{p_{v,k}^\tau(\lambda^* - \lambda_{v,k})} + 1\right\} \\
&\leq \min\left\{\frac{1-b}{b}, \frac{\lambda_{\max}}{b(\lambda^* - \lambda_{\max})} + 1\right\}, \quad \text{for } k \neq 0; \\
\operatorname{dist}\left(1, \mathcal{C}_{v,0}^\tau\right) &\leq \min\left\{1, \frac{\lambda_{v,\max}}{p_{v,0}^\tau(\lambda^* - \lambda_{v,\max})} + \frac{1-p_{v,0}^\tau}{p_{v,0}^\tau}\right\} \leq 1.
\end{aligned}$$

The second bound in Lemma 8.4.4 then yields the desired bound on spectral independence. Note that we may assume μ is supported on at least two configurations so that $(1-b)/b \geq 1$, namely $b \leq 1/2$; otherwise μ is concentrated on a single configuration and spectral independence holds with constant 0.

3. If there exists $\lambda^* \in \mathbb{R}_+$ such that $\lambda_{v,k} \in (\lambda^*, \infty) \subseteq \Gamma_{v,k}$ for every $(v,k) \in \mathcal{P}_1$, then one has

$$\begin{aligned}
\alpha_{v,k} = \inf\left(\widetilde{\Gamma}_{v,k} \cap \mathbb{R}_+\right) \leq \frac{\lambda^*}{\lambda_{v,k}}, \quad \beta_{v,k} = \sup\left(\widetilde{\Gamma}_{v,k} \cap \mathbb{R}_+\right) = \infty, \\
\alpha_{v} = \inf\left(\widetilde{\Gamma}_{v} \cap \mathbb{R}_+\right) \leq \frac{\lambda^*}{\lambda_{v,\min}}, \quad \beta_{v} = \sup\left(\widetilde{\Gamma}_{v} \cap \mathbb{R}_+\right) = \infty,
\end{aligned}$$

where $\lambda_{v,\min} = \min_{k \in Q_1} \lambda_{v,k}$. Thus, we deduce from the second part of Lemmas 8.4.5 and 8.4.6 that for all $\tau \in \mathcal{T}$ and $(v,k) \in \mathcal{P}^\tau$,

$$\begin{aligned}
\operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right) &\leq \min\left\{\frac{\lambda^*}{p_{v,k}^\tau(\lambda_{v,k} - \lambda^*)} + \frac{1-p_{v,k}^\tau}{p_{v,k}^\tau}, 1\right\} \leq 1, \quad \text{for } k \neq 0; \\
\operatorname{dist}\left(1, \mathcal{C}_{v,0}^\tau\right) &\leq \min\left\{\frac{\lambda^*}{p_{v,0}^\tau(\lambda_{v,\min} - \lambda^*)} + 1, \frac{1-p_{v,0}^\tau}{p_{v,0}^\tau}\right\} \\
&\leq \min\left\{\frac{\lambda^*}{b(\lambda_{\min} - \lambda^*)} + 1, \frac{1-b}{b}\right\}.
\end{aligned}$$

The second bound in Lemma 8.4.4 then yields the desired bound on spectral independence. □

Remark 8.4.7. For Item 2 of Theorem 8.4.2 our proof actually yields a more complicated but stronger constant for spectral independence:

$$\eta = \frac{8}{\delta} \max \left\{ 1, \max_{\substack{\tau \in \mathcal{T} \\ (v,k) \in \mathcal{P}_1^\tau}} \left\{ \min \left\{ \frac{1 - p_{v,k}^\tau}{p_{v,k}^\tau}, \frac{\lambda_{v,k}}{p_{v,k}^\tau (\lambda^* - \lambda_{v,k})} + 1 \right\} \right\} \right\}.$$

For Item 3, the constant is:

$$\eta = \frac{8}{\delta} \max \left\{ 1, \max_{\substack{\tau \in \mathcal{T} \\ (v,0) \in \mathcal{P}^\tau}} \left\{ \min \left\{ \frac{1 - p_{v,0}^\tau}{p_{v,0}^\tau}, \frac{\lambda^*}{p_{v,0}^\tau (\lambda_{v,\min} - \lambda^*)} + 1 \right\} \right\} \right\}.$$

These two bounds are more robust in the sense of Remark 8.2.3, namely, when some external fields are close to λ^* while others are close to 0 (respectively, ∞).

We end this subsection with the proofs of Lemmas 8.4.5 and 8.4.6. The proof of Lemma 8.4.4 is presented in Section 8.4.3.

Proof of Lemma 8.4.5. By definition we have

$$\operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right) = \inf_{1 \neq z \in \Gamma_{v,k}} \left| -\frac{1}{p_{v,k}^\tau (z-1)} - 1 \right|.$$

If $\Gamma_{v,k}$ is unbounded, then there exists a sequence $\{z_n\}$ such that $1 \neq z_n \in \Gamma_{v,k}$ and $\lim_{n\to\infty} |z_n| = \infty$. Therefore,

$$\operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right) \leq \liminf_{n\to\infty} \left| -\frac{1}{p_{v,k}^\tau (z_n - 1)} - 1 \right| \leq 1 + \liminf_{n\to\infty} \frac{1}{p_{v,k}^\tau |z_n - 1|} = 1.$$

This shows the first part.

For the second part, observe that $\alpha_{v,k} < 1 < \beta_{v,k}$ since $\Gamma_{v,k}$ is open and $1 \in \Gamma_{v,k}$. Hence, we obtain

$$\operatorname{dist}\left(1, \mathcal{C}_{v,k}^\tau\right) \leq \inf_{x \in \Gamma_{v,k} \cap (0,1)} \left| -\frac{1}{p_{v,k}^\tau (x-1)} - 1 \right|$$

$$= \frac{1}{p^\tau_{v,k}(1-\alpha_{v,k})} - 1$$
$$= \frac{\alpha_{v,k}}{p^\tau_{v,k}(1-\alpha_{v,k})} + \frac{1-p^\tau_{v,k}}{p^\tau_{v,k}},$$

and also

$$\operatorname{dist}\left(1, \mathcal{C}^\tau_{v,k}\right) \leq \inf_{x \in \Gamma_{v,k} \cap (1,\infty)} \left| -\frac{1}{p^\tau_{v,k}(x-1)} - 1 \right| = \frac{1}{p^\tau_{v,k}(\beta_{v,k}-1)} + 1.$$

The second part follows. □

Proof of Lemma 8.4.6. By definition we have

$$\operatorname{dist}\left(1, \mathcal{C}^\tau_{v,0}\right) = \inf_{1 \neq z \in \Gamma_v} \left| \frac{z}{p^\tau_{v,0}(z-1)} - 1 \right|.$$

If $0 \in \overline{\Gamma_v}$, then there exists a sequence $\{z_n\}$ such that $1 \neq z_n \in \Gamma_{v,k}$ and $\lim_{n\to\infty} z_n = 0$. Therefore,

$$\operatorname{dist}\left(1, \mathcal{C}^\tau_{v,0}\right) \leq \liminf_{n\to\infty} \left| \frac{z_n}{p^\tau_{v,0}(z_n-1)} - 1 \right| \leq 1 + \liminf_{n\to\infty} \frac{|z_n|}{p^\tau_{v,0}|z_n-1|} = 1.$$

This shows the first part.

For the second part, observe that $\alpha_v < 1 < \beta_v$ since Γ_v is open and $1 \in \Gamma_{v,k}$. Hence, we obtain

$$\operatorname{dist}\left(1, \mathcal{C}^\tau_{v,0}\right) \leq \inf_{x \in \Gamma_v \cap (0,1)} \left| \frac{x}{p^\tau_{v,0}(x-1)} - 1 \right| = \frac{\alpha_v}{p^\tau_{v,0}(1-\alpha_v)} + 1,$$

and also

$$\operatorname{dist}\left(1, \mathcal{C}^\tau_{v,0}\right) \leq \inf_{x \in \Gamma_v \cap (1,\infty)} \left| \frac{x}{p^\tau_{v,0}(x-1)} - 1 \right|$$
$$= \frac{\beta_v}{p^\tau_{v,0}(\beta_v-1)} - 1$$

$$= \frac{1}{p^{\tau}_{v,0}(\beta_v - 1)} + \frac{1 - p^{\tau}_{v,0}}{p^{\tau}_{v,0}}.$$

The second part follows. □

8.4.3 Bounding the Absolute Sum of Influences: Proof of Lemma 8.4.4

Let $\tau \in \mathcal{T}$ be an arbitrary pinning and fix τ. We will give an upper bound on the absolute row sum of the associated influence matrix Ψ^{τ}_{μ} under τ, which then provides an upper bound on the maximum eigenvalue of Ψ^{τ}_{μ}. In particular, for $(u, j) \in \mathcal{P}^{\tau}$ we define

$$\widetilde{S}^{\tau}_{\mu}(u, j) = \sum_{(v,k)\in\mathcal{P}^{\tau}} \left|\Psi^{\tau}_{\mu}(u, j; v, k)\right| = \sum_{(v,k)\in\mathcal{P}^{\tau}: v\neq u} \left|\mu^{\tau}(\sigma_v = k \mid \sigma_u = j) - \mu^{\tau}(\sigma_v = k)\right|$$

to be the absolute sum of influences in the row (u, j), and define

$$S^{\tau}_{\mu}(u, j) = \sum_{(v,k)\in\mathcal{P}^{\tau}_1} \left|\Psi^{\tau}_{\mu}(u, j; v, k)\right| = \sum_{(v,k)\in\mathcal{P}^{\tau}_1: v\neq u} \left|\mu^{\tau}(\sigma_v = k \mid \sigma_u = j) - \mu^{\tau}(\sigma_v = k)\right|$$

to be the partial absolute sum of influences for pairs (v, k) with $k \neq 0$ in the row (u, j). Notice that one has $\widetilde{S}^{\tau}_{\mu}(u, j) \leq 2S^{\tau}_{\mu}(u, j)$, because for each $(v, 0) \in \mathcal{P}^{\tau}$ it follows from the triangle inequality that

$$\left|\mu^{\tau}(\sigma_v = 0 \mid \sigma_u = j) - \mu^{\tau}(\sigma_v = 0)\right| \leq \sum_{k\in\Omega^{\tau}_v\setminus\{0\}} \left|\mu^{\tau}(\sigma_v = k \mid \sigma_u = j) - \mu^{\tau}(\sigma_v = k)\right|.$$

Hence,

$$\mathsf{EigMax}(\Psi^{\tau}_{\mu}) \leq \left\|\Psi^{\tau}_{\mu}\right\|_{\infty} = \max_{(u,j)\in\mathcal{P}^{\tau}} \widetilde{S}^{\tau}_{\mu}(u, j) \leq 2 \max_{(u,j)\in\mathcal{P}^{\tau}} S^{\tau}_{\mu}(u, j). \tag{8.8}$$

The rest of the proof aims to bound $S^{\tau}_{\mu}(u, j)$ for a fixed $(u, j) \in \mathcal{P}^{\tau}$. We consider two cases $j \neq 0$ and $j = 0$ separately, and prove the following.

Lemma 8.4.8. *For $j \neq 0$ we have*

$$S_\mu^\tau(u,j) \leq \min\left\{\frac{1}{b\delta^2}, \frac{4}{\delta}\mathrm{dist}\left(1, \mathcal{C}_{u,j}^\tau\right)\right\}.$$

Lemma 8.4.9. *For $j = 0$ we have*

$$S_\mu^\tau(u,0) \leq \min\left\{\frac{1}{b\delta^2}, \frac{4}{\delta}\mathrm{dist}\left(1, \mathcal{C}_{u,0}^\tau\right)\right\}.$$

Lemma 8.4.4 follows immediately from these two lemmas.

Proof of Lemma 8.4.4. Combining Eq. (8.8) and Lemmas 8.4.8 and 8.4.9. □

Proof of Lemma 8.4.8

Fix $(u,j) \in \mathcal{P}_1^\tau$. We follow the proof approach of [3] and view $S_\mu^\tau(u,j)$ as the derivative of certain function related to the partition function; the lemma then follows from an application of the Schwarz-Pick Theorem (Theorem 2.5.3) for bounding the derivative. For ease of notation we write

$$\mathcal{P}' = \{(v,k) \in \mathcal{P}_1^\tau : v \neq u\} \quad \text{and} \quad \mathcal{K} = \prod_{(v,k)\in\mathcal{P}'} \Gamma_{v,k}.$$

Define the multivariate complex function $f : \mathcal{K} \to \mathbb{C}$ as

$$f(\boldsymbol{\lambda}) = \frac{1}{p_{u,j}^\tau} \frac{Z_w^{\tau\cup(u,j)}(\boldsymbol{\lambda})}{Z_w^\tau(\boldsymbol{\lambda}\cup\mathbf{1}_u)}, \quad \text{for } \boldsymbol{\lambda} \in \mathcal{K} \tag{8.9}$$

where $\tau \cup (u,j) \in \mathcal{T}$ is the pinning that combines τ and $\sigma_u = j$, and $\boldsymbol{\lambda} \cup \mathbf{1}_u$ is the vector of external fields that combines $\boldsymbol{\lambda}$ and the all-one fields $\mathbf{1}_u$ at u (i.e., $\lambda_{u,j'} = 1$ for all $0 \neq j' \in \Omega_u^\tau$); notice that

$$Z_w^\tau(\boldsymbol{\lambda}\cup\mathbf{1}_u) = \sum_{j'\in\Omega_u^\tau} Z_w^{\tau\cup(u,j')}(\boldsymbol{\lambda}).$$

Note that f is well-defined since by our assumption $Z^{\tau}_{w}(\boldsymbol{\lambda} \cup \mathbf{1}_u) \neq 0$ whenever each $\lambda_{v,k} \in \Gamma_{v,k}$. The following claim summarizes several important properties of the function f.

Claim 8.4.10. *Let $f : \mathcal{K} \to \mathbb{C}$ be the multivariate complex function defined by Eq.* (8.9)*.*

1. *The function f is well-defined and holomorphic on $\mathcal{K}$, and $f(\mathbf{1}) = 1$.*

2. *For every $(v,k) \in \mathcal{P}'$,*
$$\left.\frac{\partial f}{\partial \lambda_{v,k}}\right|_{\boldsymbol{\lambda}=\mathbf{1}} = \Psi^{\tau}_{\mu}(u,j;v,k).$$

3. *Suppose that $f \not\equiv 1$. Let $\mathcal{A} \subseteq \mathbb{C}$ be an open region defined as*
$$\mathcal{A} = -\frac{1}{p^{\tau}_{u,j}} \left(\Gamma_{u,j} - 1\right)^{-1}.$$
Then $1 \notin \overline{\mathcal{A}}$. Let $\mathcal{A}_1$ be the connected component of $\overline{\mathcal{A}}^{\mathsf{c}}$ which contains 1. Then $\mathcal{A}_1$ is open and simply connected, and
$$\mathrm{image}(f) \subseteq \mathcal{A}_1.$$

If $f \equiv 1$, then Item 2 of Claim 8.4.10 implies that $\Psi^{\tau}_{\mu}(u,j;v,k) = 0$ for all $(v,k) \in \mathcal{P}'$, and hence $S^{\tau}_{\mu}(u,j) = 0$. In the rest of the proof we assume that $f \not\equiv 1$.

Given Claim 8.4.10, in order to bound $S^{\tau}_{\mu}(u,j)$ it suffices to bound $\|\nabla f(\mathbf{1})\|_1$. We do this by taking holomorphic functions $\varphi : \mathbb{D}(0,1) \to \mathcal{K}$, $\psi : \mathcal{A}_1 \to \mathbb{D}(0,1)$ and considering their composition with the function f. The bound on $\|\nabla f(\mathbf{1})\|_1$ would then follow from the Schwarz-Pick Theorem which bounds the derivative of a holomorphic function from the open unit dist into itself.

We now formalize this idea. Let $\varphi : \mathbb{D}(0,1) \to \mathcal{K}$ be a holomorphic vector-valued function such that for every $(v,k) \in \mathcal{P}'$, the (v,k)-coordinate function $\varphi_{v,k} : \mathbb{D}(0,1) \to \Gamma_{v,k}$ is holomorphic and satisfies $\varphi_{v,k}(0) = 1$ and $\varphi'_{v,k}(0) \in \mathbb{R}_+$ if $\Psi^{\tau}_{\mu}(u,j;v,k) \geq 0$ while $\varphi'_{v,k}(0) \in \mathbb{R}_-$ if $\Psi^{\tau}_{\mu}(u,j;v,k) \leq 0$. Hence, $\varphi(0) = \mathbf{1}$ and $\varphi'_{v,k}(0)\Psi^{\tau}_{\mu}(u,j;v,k) \geq 0$ for all

(v,k). Meanwhile, for the region $\mathcal{A}_1$ given in Item 3 of Claim 8.4.10, let $\psi : \mathcal{A}_1 \to \mathbb{D}(0,1)$ be a holomorphic function such that $\psi'(1) \in \mathbb{R}_+$. We will specify our choice of φ and ψ soon. Also, we point out here that our assumptions $\varphi'_{v,k}(0) \in \mathbb{R}_+/\mathbb{R}_-$ and $\psi'(1) \in \mathbb{R}_+$ would not cause strong restrictions; they can be easily satisfied by considering rotations $\varphi(e^{i\theta}z)$ and $e^{i\theta}\psi(z)$.

Given such φ and ψ, we define the holomorphic function $F : \mathbb{D}(0,1) \to \mathbb{D}(0,1)$ given by $F = \psi \circ f \circ \varphi$. Notice that $F(0) = \psi(1)$. The derivative $F'(0)$ at 0 is real and can be bounded by

$$\begin{aligned} F'(0) &= \psi'(1) \sum_{(v,k)\in\mathcal{P}'} \varphi'_{v,k}(0) \left.\frac{\partial f}{\partial \lambda_{v,k}}\right|_{\boldsymbol{\lambda}=\mathbf{1}} \\ &= \psi'(1) \sum_{(v,k)\in\mathcal{P}'} \varphi'_{v,k}(0) \Psi^{\tau}_{\mu}(u,j;v,k) \\ &\geq \psi'(1) \min_{(v,k)\in\mathcal{P}'} \left\{ \left|\varphi'_{v,k}(0)\right| \right\} S^{\tau}_{\mu}(u,j), \end{aligned} \tag{8.10}$$

where the second equality follows from Item 2 of Claim 8.4.10 and the inequality is due to our assumption that $\varphi'_{v,k}(0)\Psi^{\tau}_{\mu}(u,j;v,k) \geq 0$ for each (v,k). The Schwarz-Pick Theorem (Theorem 2.5.3) implies that $F'(0) \leq 1$, and hence we obtain

$$S^{\tau}_{\mu}(u,j) \leq \frac{1}{\psi'(1)} \left(\min_{(v,k)\in\mathcal{P}'} \left|\varphi'_{v,k}(0)\right| \right)^{-1}. \tag{8.11}$$

It remains to choose φ and ψ. Consider first the function φ. For each $(v,k) \in \mathcal{P}'$ we let

$$\delta_{v,k} = \operatorname{dist}(1, \partial\Gamma_{v,k}) \quad \text{and} \quad \chi_{v,k} = \operatorname{sgn}(\Psi^{\tau}_{\mu}(u,j;v,k)) = \begin{cases} +1, & \Psi^{\tau}_{\mu}(u,j;v,k) \geq 0; \\ -1, & \Psi^{\tau}_{\mu}(u,j;v,k) < 0. \end{cases}$$

We then define $\varphi_{v,k} : \mathbb{D}(0,1) \to \Gamma_{v,k}$ by

$$\varphi_{v,k}(z) = 1 + \chi_{v,k}\delta_{v,k}z. \tag{8.12}$$

Observe that φ is holomorphic, $\varphi(0) = \mathbf{1}$, and $\varphi'_{v,k}(z) = \chi_{v,k}\delta_{v,k}$ has the same sign as $\Psi^\tau_\mu(u,j;v,k)$ for each (v,k). Recall that $\delta = \min_{(v,k)\in\mathcal{P}_1} \operatorname{dist}(1, \partial\Gamma_{v,k})$, and thus

$$\min_{(v,k)\in\mathcal{P}'} \left|\varphi'_{v,k}(0)\right| = \min_{(v,k)\in\mathcal{P}'} \delta_{v,k} \geq \delta. \tag{8.13}$$

Next, we decide ψ. We will actually give two choices of ψ, denoted by ψ_1 and ψ_2 respectively, which correspond to the two bounds in Lemma 8.4.8.

We first consider the simpler choice ψ_1. Let $\delta_{u,j} = \operatorname{dist}(1, \partial\Gamma_{u,j}) \geq \delta$, and so $\mathbb{D}(1,\delta_{u,j}) \subseteq \Gamma_{u,j}$. Then, the region $\mathcal{A}$ from Item 3 of Claim 8.4.10 satisfies

$$\mathcal{A} = -\frac{1}{p^\tau_{u,j}}\left(\Gamma_{u,j} - 1\right)^{-1} \supseteq -\frac{1}{p^\tau_{u,j}}\left(\mathbb{D}(1,\delta_{u,j}) - 1\right)^{-1} = \frac{1}{p^\tau_{u,j}\delta_{u,j}}\overline{\mathbb{D}}(0,1)^{\mathsf{c}}.$$

It follows that

$$\mathcal{A}_1 \subseteq \overline{\mathcal{A}}^{\mathsf{c}} \subseteq \frac{1}{p^\tau_{u,j}\delta_{u,j}}\mathbb{D}(0,1).$$

We can define $\psi_1 : \mathcal{A}_1 \to \mathbb{D}(0,1)$ as

$$\psi_1(z) = p^\tau_{u,j}\delta_{u,j} z.$$

Then, ψ_1 is holomorphic, $\psi'_1(z) = p^\tau_{u,j}\delta_{u,j} \in \mathbb{R}_+$, and

$$\frac{1}{\psi'_1(1)} = \frac{1}{p^\tau_{u,j}\delta_{u,j}} \leq \frac{1}{b\delta}. \tag{8.14}$$

Combining Eqs. (8.11), (8.13) and (8.14), we obtain

$$S^\tau_\mu(u,j) \leq \frac{1}{b\delta^2}.$$

This shows the first bound in Lemma 8.4.8.

Next, we define ψ_2. Since $\emptyset \neq \mathcal{A}_1 \subsetneq \mathbb{C}$ is open and simply connected by Item 3 of

Claim 8.4.10, the Riemann Mapping Theorem (Theorem 2.5.4) implies that there exists a (unique) biholomorphic mapping $\psi_2 : \mathcal{A}_1 \to \mathbb{D}(0,1)$ such that $\psi_2(1) = 0$ and $\psi_2'(1) \in \mathbb{R}_+$. Write $h = \psi_2^{-1}$, which is a bijective holomorphic function from $\mathbb{D}(0,1)$ to $\mathcal{A}_1$ satisfying $h(0) = 1$. Then, Koebe's One-Quarter Theorem (Theorem 2.5.5) shows that

$$\frac{1}{4}|h'(0)| \le \operatorname{dist}(1, \partial\mathcal{A}_1) \le \operatorname{dist}(1, \mathcal{A}) = \operatorname{dist}\left(1, \mathcal{C}_{u,j}^{\tau}\right).$$

It follows that

$$\frac{1}{\psi_2'(1)} = h'(0) \le 4\operatorname{dist}\left(1, \mathcal{C}_{u,j}^{\tau}\right). \tag{8.15}$$

Combining Eqs. (8.11), (8.13) and (8.15), we get

$$S_{\mu}^{\tau}(u,j) \le \frac{4}{\delta}\operatorname{dist}\left(1, \mathcal{C}_{u,j}^{\tau}\right),$$

which is the second bound in Lemma 8.4.8.

Remark 8.4.11. The proof of Lemma 8.4.8 (and also Lemma 8.4.9 in Section 8.4.3) leaves the possibility of further improvements on the spectral independence bounds for specific problems. Here in the proof we are given regions $\mathcal{K}$ and $\mathcal{A}_1$ in abstract forms and the choices of φ and ψ may not be optimal for specific instances; in particular, the Riemann Mapping Theorem only shows the existence of a biholomorphic mapping and there is no guarantees that such a choice is the best possible. Hence, for specific problems and specific zero-free regions, one may be able to pick φ and ψ in a smarter way to achieve a better bound on spectral independence.

It remains to prove Claim 8.4.10. The following lemma is helpful to us.

Lemma 8.4.12. *Let $\mathcal{S} \subseteq \mathbb{C}$ be a non-empty open connected region such that $\mathcal{S}$ is unbounded and $\overline{\mathcal{S}} \neq \mathbb{C}$. If $\mathcal{S}_1$ is a connected component of $\overline{\mathcal{S}}^{\mathsf{c}}$, then $\mathcal{S}_1$ is open and simply connected.*

Proof. Clearly $\mathcal{S}_1$ is open and connected. If $\mathcal{S}_1$ is not simply connected, then there exists a

Jordan curve (simple closed curve) γ in $\mathcal{S}_1$ whose interior region contains a point $z_0 \notin \mathcal{S}_1$. Note that we can actually find a point z from the interior of γ such that $z \in \overline{\mathcal{S}}$; if not, then the whole interior of γ is contained in $\overline{\mathcal{S}}^{\mathsf{c}}$ and thus $z_0 \notin \mathcal{S}_1$ is connected to $\mathcal{S}_1$ in $\overline{\mathcal{S}}^{\mathsf{c}}$, contradicting to the assumption that $\mathcal{S}_1$ is a connected component of $\overline{\mathcal{S}}^{\mathsf{c}}$. Since the interior of γ is open, this further implies that the interior of γ contains a point $z \in \mathcal{S}$. Meanwhile, since $\mathcal{S}$ is unbounded the exterior of γ contains a point $w \in \mathcal{S}$. Now, as $\mathcal{S}$ is connected there exists a path p in $\mathcal{S}$ connecting z and w. Note that p must intersect with γ, because the interior and exterior of γ are disconnected. This yields a contradiction since $\gamma \subseteq \mathcal{S}_1 \subseteq \mathcal{S}^{\mathsf{c}}$ while $p \subseteq \mathcal{S}$. □

We complete the proof of Lemma 8.4.8 with the proof of Claim 8.4.10.

Proof of Claim 8.4.10. 1. Since $Z_w^\tau(\boldsymbol{\lambda} \cup \mathbf{1}_u) \neq 0$ whenever $\boldsymbol{\lambda} \in \mathcal{K}$ by our stability assumption, the function f is well-defined and holomorphic on $\mathcal{K}$. Also, by definition we have $f(\mathbf{1}) = 1$.

2. Let $(v,k) \in \mathcal{P}'$. Then one has

$$\frac{\partial f}{\partial \lambda_{v,k}} = \frac{1}{p_{u,j}^\tau}\left(\frac{1}{Z_w^\tau(\boldsymbol{\lambda} \cup \mathbf{1}_u)}\left(\frac{\partial}{\partial \lambda_{v,k}} Z_w^{\tau \cup (u,j)}(\boldsymbol{\lambda})\right) - \frac{Z_w^{\tau \cup (u,j)}(\boldsymbol{\lambda})}{Z_w^\tau(\boldsymbol{\lambda} \cup \mathbf{1}_u)^2}\left(\frac{\partial}{\partial \lambda_{v,k}} Z_w^\tau(\boldsymbol{\lambda} \cup \mathbf{1}_u)\right)\right).$$

Suppose τ is a pinning on $\Lambda \subseteq V$ and let $U = V^{\tau \cup (u,j)} = V \setminus \Lambda \setminus \{u\}$ be the set of unpinned vertices under the pinning $\tau \cup (u,j)$. We deduce that,

$$\begin{aligned}
\frac{\partial}{\partial \lambda_{v,k}} Z_w^{\tau \cup (u,j)}(\boldsymbol{\lambda}) &= \sum_{\sigma \in \Omega:\, \sigma_\Lambda = \tau, \sigma_u = j} w(\sigma) \cdot \frac{\partial}{\partial \lambda_{v,k}} \boldsymbol{\lambda}^{\sigma_U} \\
&= \sum_{\sigma \in \Omega:\, \sigma_\Lambda = \tau, \sigma_u = j, \sigma_v = k} w(\sigma) \boldsymbol{\lambda}^{\sigma_{U \setminus \{v\}}} \\
&= Z_w^{\tau \cup (u,j) \cup (v,k)}(\boldsymbol{\lambda}).
\end{aligned}$$

Similarly,

$$\frac{\partial}{\partial \lambda_{v,k}} Z_w^\tau(\boldsymbol{\lambda} \cup \mathbf{1}_u) = Z_w^{\tau \cup (v,k)}(\boldsymbol{\lambda} \cup \mathbf{1}_u).$$

We then get

$$\begin{aligned}
\left.\frac{\partial f}{\partial \lambda_{v,k}}\right|_{\boldsymbol{\lambda}=\mathbf{1}} &= \frac{1}{p_{u,j}^\tau} \left(\frac{Z_w^{\tau \cup (u,j) \cup (v,k)}(\mathbf{1})}{Z_w^\tau(\mathbf{1})} - \frac{Z_w^{\tau \cup (u,j)}(\mathbf{1}) \cdot Z_w^{\tau \cup (v,k)}(\mathbf{1})}{Z_w^\tau(\mathbf{1})^2} \right) \\
&= \frac{1}{\mu^\tau(\sigma_u = j)} \left(\mu^\tau(\sigma_u = j, \sigma_v = k) - \mu^\tau(\sigma_u = j)\mu^\tau(\sigma_v = k) \right) \\
&= \Psi_\mu^\tau(u, j; v, k),
\end{aligned}$$

as claimed.

3. We first show that $\mathrm{image}(f) \subseteq \mathcal{A}^{\mathsf{c}}$. Suppose for sake of contradiction that $f(\boldsymbol{\lambda}) \in \mathcal{A}$ for some $\boldsymbol{\lambda} \in \mathcal{K}$. Then there exists $1 \neq y \in \Gamma_{u,j}$ such that

$$-\frac{1}{p_{u,j}^\tau (y-1)} = f(\boldsymbol{\lambda}) = \frac{1}{p_{u,j}^\tau} \frac{Z_w^{\tau \cup (u,j)}(\boldsymbol{\lambda})}{Z_w^\tau(\boldsymbol{\lambda} \cup \mathbf{1}_u)}.$$

It follows that

$$Z_w^\tau(\boldsymbol{\lambda} \cup \boldsymbol{\lambda}_u) = y Z_w^{\tau \cup (u,j)}(\boldsymbol{\lambda}) + \sum_{j \neq j' \in \Omega_u^\tau} Z_w^{\tau \cup (u,j')}(\boldsymbol{\lambda}) = 0,$$

where $\boldsymbol{\lambda}_u$ is the vector of external fields at u defined by $\lambda_{u,j} = y$ and $\lambda_{u,j'} = 1$ for $j' \in \Omega_u^\tau \setminus \{0, j\}$. This contradicts to our stability assumption that $Z_w^\tau(\boldsymbol{\lambda} \cup \boldsymbol{\lambda}_u) \neq 0$. Therefore, we have shown that $\mathrm{image}(f) \subseteq \mathcal{A}^{\mathsf{c}}$.

Now, since $\mathcal{K}$ is open and connected and f is a non-constant holomorphic function, the Open Mapping Theorem (Theorem 2.5.6) implies that $\mathrm{image}(f)$ is open and connected. Thus, we have $\mathrm{image}(f) \subseteq (\mathcal{A}^{\mathsf{c}})^{\mathrm{o}} = \overline{\mathcal{A}}^{\mathsf{c}}$; note that in particular $1 \in \overline{\mathcal{A}}^{\mathsf{c}}$. Furthermore, since $\mathrm{image}(f)$ is connected one has $\mathrm{image}(f) \subseteq \mathcal{A}_1$, the connected component of $\overline{\mathcal{A}}^{\mathsf{c}}$ containing 1. The region $\mathcal{A}_1$ is open and connected by definition. It remains to show that $\mathcal{A}_1$ is simply connected, which follows immediately from Lemma 8.4.12 and the fact that

$\mathcal{A}$ is connected and unbounded. $\square$

Proof of Lemma 8.4.9

The proof of Lemma 8.4.9 is similar to that of Lemma 8.4.8. We will use the same notations and only emphasize a few key steps that differ.

Recall that

$$\mathcal{P}' = \{(v,k) \in \mathcal{P}_1^\tau : v \neq u\} \quad \text{and} \quad \mathcal{K} = \prod_{(v,k)\in\mathcal{P}'} \Gamma_{v,k}.$$

Define the multivariate complex function $g : \mathcal{K} \to \mathbb{C}$ as

$$g(\boldsymbol{\lambda}) = \frac{1}{p_{u,0}^\tau} \frac{Z_w^{\tau\cup(u,0)}(\boldsymbol{\lambda})}{Z_w^\tau(\boldsymbol{\lambda} \cup \mathbf{1}_u)}, \quad \text{for } \boldsymbol{\lambda} \in \mathcal{K} \tag{8.16}$$

where $\tau \cup (u,0) \in \mathcal{T}$ is the pinning combining τ and $\sigma_u = 0$, and $\boldsymbol{\lambda} \cup \mathbf{1}_u$ is the vector of external fields that combines $\boldsymbol{\lambda}$ and $\mathbf{1}_u$. The following claim is analogous to Claim 8.4.10 and summarizes key properties of the function g.

Claim 8.4.13. *Let $g : \mathcal{K} \to \mathbb{C}$ be the multivariate complex function defined by Eq.* (8.16).

1. *The function g is well-defined and holomorphic on $\mathcal{K}$, and $g(\mathbf{1}) = 1$.*

2. *For every $(v,k) \in \mathcal{P}'$,*
$$\left.\frac{\partial f}{\partial \lambda_{v,k}}\right|_{\boldsymbol{\lambda}=\mathbf{1}} = \Psi_\mu^\tau(u,0;v,k).$$

3. *Suppose that $g \not\equiv 1$. Let $\mathcal{B} \subseteq \mathbb{C}$ be an open region defined as*
$$\mathcal{B} = \frac{1}{p_{u,0}^\tau}\left((\Gamma_u - 1)^{-1} + 1\right).$$
Then $1 \notin \overline{\mathcal{B}}$. Let $\mathcal{B}_1$ be the connected component of $\overline{\mathcal{B}}^{\mathsf{c}}$ which contains 1. Then $\mathcal{B}_1$ is

open and simply connected, and

$$\mathrm{image}(g) \subseteq \mathcal{B}_1.$$

We may assume that $g \not\equiv 1$ since otherwise $S^\tau_\mu(u,0) = 0$ and the lemma is trivial. Again we choose holomorphic functions $\varphi : \mathbb{D}(0,1) \to \mathcal{K}$, $\psi : \mathcal{B}_1 \to \mathbb{D}(0,1)$ and consider the holomorphic function $G : \mathbb{D}(0,1) \to \mathbb{D}(0,1)$ defined as $G = \psi \circ g \circ \varphi$. Just as in the proof of Lemma 8.4.8, we let $\varphi : \mathbb{D}(0,1) \to \mathcal{K}$ be a holomorphic vector-valued function such that for every $(v,k) \in \mathcal{P}'$, the (v,k)-coordinate function $\varphi_{v,k} : \mathbb{D}(0,1) \to \Gamma_{v,k}$ is holomorphic and satisfies $\varphi_{v,k}(0) = 1$ and $\varphi'_{v,k}(0) \in \mathbb{R}_+$ if $\Psi^\tau_\mu(u,j;v,k) \geq 0$ while $\varphi'_{v,k}(0) \in \mathbb{R}_-$ if $\Psi^\tau_\mu(u,j;v,k) \leq 0$. Meanwhile, let $\psi : \mathcal{B}_1 \to \mathbb{D}(0,1)$ be a holomorphic function such that $\psi'(1) \in \mathbb{R}_+$. Hence, we have $G(0) = \psi(1)$, and by Claim 8.4.13 $G'(0) \in \mathbb{R}$ can be bounded by

$$G'(0) \geq \psi'(1) \min_{(v,k)\in\mathcal{P}'} \left\{ \left|\varphi'_{v,k}(0)\right| \right\} S^\tau_\mu(u,0),$$

which is analogous to Eq. (8.10). We then deduce the analog of Eq. (8.11) from the Schwarz-Pick Theorem (Theorem 2.5.3):

$$S^\tau_\mu(u,0) \leq \frac{1}{\psi'(1)} \left(\min_{(v,k)\in\mathcal{P}'} \left|\varphi'_{v,k}(0)\right| \right)^{-1}. \tag{8.17}$$

We specify next our choice of φ and ψ. The function φ is the same one as in the proof of Lemma 8.4.8, and is given by Eq. (8.12). In particular, Eq. (8.13) still holds. We also give two choices of the function ψ, denoted by ψ_3 and ψ_4 respectively, corresponding to the two bounds in Lemma 8.4.9.

Consider first ψ_3. Recall that $\Gamma_u \subseteq \mathbb{C}$ is defined to be the connected component of the intersection $\bigcap_{0\neq j\in\Omega^\tau_u} \Gamma_{u,j}$ that contains 1. Let $\delta_u = \mathrm{dist}\,(1, \partial\Gamma_u) \geq \delta$ and thus $\mathbb{D}(1,\delta_u) \subseteq$

Γ_u. Then we have

$$\mathcal{B} = \frac{1}{p^\tau_{u,0}} \left((\Gamma_u - 1)^{-1} + 1 \right) \supseteq \frac{1}{p^\tau_{u,0}} \left(\mathbb{D}(0, \delta_u)^{-1} + 1 \right) = \frac{1}{p^\tau_{u,0}\delta_u} \overline{\mathbb{D}}(\delta_u, 1)^{\mathsf{c}},$$

and hence

$$\mathcal{B}_1 \subseteq \overline{\mathcal{B}}^{\mathsf{c}} \subseteq \frac{1}{p^\tau_{u,0}\delta_u} \mathbb{D}\left(\delta_u, 1\right).$$

We define $\psi_3 : \mathcal{B}_1 \to \mathbb{D}(0,1)$ as

$$\psi_3(z) = p^\tau_{u,0}\delta_u z - \delta_u.$$

Observe that ψ_3 is holomorphic, $\psi_3'(z) = p^\tau_{u,0}\delta_u \in \mathbb{R}_+$, and

$$\frac{1}{\psi_3'(1)} = \frac{1}{p^\tau_{u,0}\delta_u} \leq \frac{1}{b\delta}. \tag{8.18}$$

Combining Eqs. (8.13), (8.17) and (8.18), we obtain

$$S^\tau_\mu(u, 0) \leq \frac{1}{b\delta^2}.$$

This shows the first bound in Lemma 8.4.9.

Finally, we define ψ_4. Since $\emptyset \neq \mathcal{B}_1 \subsetneq \mathbb{C}$ is open and simply connected by Claim 8.4.13, there exists a (unique) biholomorphic mapping $\psi_4 : \mathcal{B}_1 \to \mathbb{D}(0,1)$ such that $\psi_4(1) = 0$ and $\psi_4'(1) \in \mathbb{R}_+$ by the Riemann Mapping Theorem (Theorem 2.5.4). Let $h = \psi_4^{-1}$ be the holomorphic mapping from $\mathbb{D}(0,1)$ to $\mathcal{B}_1$ with $h(0) = 1$. We deduce from the Koebe's One-Quarter Theorem (Theorem 2.5.5) that

$$\frac{1}{4}|h'(0)| \leq \operatorname{dist}(1, \partial\mathcal{B}_1) \leq \operatorname{dist}(1, \mathcal{B}) = \operatorname{dist}\left(1, \mathcal{C}^\tau_{u,0}\right),$$

and hence

$$\frac{1}{\psi_4'(1)} = h'(0) \leq 4\mathrm{dist}\left(1, \mathcal{C}_{u,0}^{\tau}\right). \tag{8.19}$$

Combining Eqs. (8.13), (8.17) and (8.19), we get

$$S_{\mu}^{\tau}(u,0) \leq \frac{4}{\delta}\mathrm{dist}\left(1, \mathcal{C}_{u,0}^{\tau}\right),$$

which is the second bound in Lemma 8.4.9.

We end this section with the proof of Claim 8.4.13.

Proof of Claim 8.4.13. Item 1 follows from the stability of the partition function and Item 2 can be deduced by direct calculations. We omit the details here and refer to the proof of Claim 8.4.10.

For Item 3, again we first show that $\mathrm{image}(g) \subseteq \mathcal{B}^{\mathsf{c}}$. Suppose for sake of contradiction that $g(\boldsymbol{\lambda}) \in \mathcal{B}$ for some $\boldsymbol{\lambda} \in \mathcal{K}$. Then there exists $1 \neq y \in \Gamma_u \subseteq \bigcap_{0 \neq j \in \Omega_u^{\tau}} \Gamma_{u,j}$ such that

$$\frac{y}{p_{u,0}^{\tau}(y-1)} = g(\boldsymbol{\lambda}) = \frac{1}{p_{u,0}^{\tau}} \frac{Z_w^{\tau \cup (u,0)}(\boldsymbol{\lambda})}{Z_w^{\tau}(\boldsymbol{\lambda} \cup \mathbf{1}_u)}.$$

It follows that

$$Z_w^{\tau}(\boldsymbol{\lambda} \cup y\mathbf{1}_u) = Z_w^{\tau \cup (u,0)}(\boldsymbol{\lambda}) + \sum_{0 \neq j \in \Omega_u^{\tau}} y Z_w^{\tau \cup (u,j)}(\boldsymbol{\lambda}) = 0,$$

where $y\mathbf{1}_u$ represents the vector of external fields at u defined by $\lambda_{u,j} = y$ for all $0 \neq j \in \Omega_u^{\tau}$. This contradicts the stability assumption of the partition function. Therefore, we have $\mathrm{image}(g) \subseteq \mathcal{B}^{\mathsf{c}}$. The Open Mapping Theorem (Theorem 2.5.6) then implies that $\mathrm{image}(g) \subseteq \mathcal{B}_1$ which is the connected component of $\overline{\mathcal{B}}^{\mathsf{c}}$ containing 1. Meanwhile, notice that the region Γ_u is open and connected since it is a connected component of the open set $\bigcap_{0 \neq j \in \Omega_u^{\tau}} \Gamma_{u,j}$, and so $\mathcal{B}$ is open, connected, and unbounded. Hence, Lemma 8.4.12 shows that $\mathcal{B}_1$ is open and simply connected. This completes the proof of the claim. □

8.5 Optimal Mixing Results for Binary Symmetric Holant Problems

Let $G = (V, E)$ be a graph of maximum degree Δ. We consider the Holant problem in the binary symmetric case, which we now describe. Let $\{f_v\}_{v \in V} : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be a family of functions, one for each vertex $v \in V$ in the input graph. One should think of each f_v as representing a local constraint on the assignments to edges incident to v. Since we are restricting ourselves to the binary case, our configurations σ will map edges to $\{0, 1\}$. Furthermore, since we are restricting ourselves to the symmetric case, our local functions f_v will only depend on the number of edges incident to v which are mapped to 1. With these $\{f_v\}_{v \in V}$ in hand, we may write the multivariate partition function as

$$Z_G(\lambda) = \sum_{\sigma : E \to \{0,1\}} \prod_{v \in V} f_v(|\sigma_{E(v)}|) \prod_{e \in E} \lambda_e^{\mathbb{1}\{\sigma_e = 1\}}, \tag{8.20}$$

where $E(v)$ is the set of all edges adjacent to v, $\sigma_{E(v)}$ is the configuration restricted on $E(v)$, and $|\sigma_{E(v)}|$ is the number of edges in $E(v)$ with assignment 1.

This class of problems is already incredibly rich, and encompasses many classical objects studied in combinatorics and statistical physics including the following:

- *Matchings/Monomer-Dimer Model:* Assume all f_v are the same and given by the "at-most-one" function:

$$f_v(k) = \begin{cases} 1, & \text{if } k = 0, 1; \\ 0, & \text{if } k \geq 2. \end{cases}$$

  Then $Z_G(\mathbf{1})$ yields the number of matchings (of any size) in G.

- *Weighted Edge Covers:* Assume all f_v are the same and given by the weighted "at-

least-one" function:

$$f_v(k) = \begin{cases} \rho, & \text{if } k = 0; \\ 1, & \text{if } k \geq 1. \end{cases}$$

In the case $\rho = 0$, then $Z_G(\mathbf{1})$ yields the number of edge covers of G, that is, subsets of edges such that every vertex is incident to at least one selected edge.

- *Weighted Even Subgraphs:* In this case, all f_v are the same and given by the weighted "parity" function. More specifically, for a fixed positive parameter $\rho > 0$, we have

$$f_v(k) = \begin{cases} 1, & \text{if } k \text{ is even}; \\ \rho, & \text{if } k \text{ is odd}. \end{cases}$$

In the case $\rho = 0$, then $Z_G(\mathbf{1})$ counts the number of even subgraphs, that is, subsets of edges such that all vertices have even degrees in the resulting subgraph. (Note that when $\rho = 0$, the Glauber dynamics is not ergodic.)

- *Ising Model on Line Graphs:* In this case, each f_v depends on the degree of v. If $\beta > 0$ is some fixed parameter (independent of v), and $d = \deg(v)$, then we have

$$f_v(k) = \begin{cases} \beta^{\binom{k}{2}}\beta^{\binom{d-k}{2}}, & \text{if } 0 \leq k \leq d; \\ 0, & \text{o/w.} \end{cases}$$

In all of the above examples, prior works managed to show that the Glauber dynamics admits an inverse polynomial spectral gap ([76] for matchings, [73] for edge covers, [77] for weighted even subgraphs, and [53] for the Ising model in the antiferromagnetic $\beta < 1$ regime). Furthermore, all of these results were obtained via the canonical paths method [76], and its winding extension [102]. However, one down-side behind these results is that the spectral gap bounds are suboptimal, and do not yield optimal mixing times nor sub-Gaussian concentration estimates. In contrast, by combining our framework with known

zero-free regions for these models and the local-to-global mixing result of [45], we obtain optimal mixing times and sub-Gaussian concentration results for these problems in the bounded-degree regime.

Note that though Theorem 1.2.1 is stated only for spin systems, it holds for Holant problems and tensor network contractions (see Sections 8.5 and 8.6 for definitions) as well since one can view these as spin systems defined on hypergraphs (also known as Markov random fields) and the proof approach of [45, 23] still works when the underlying graph has bounded maximum degree.

One of the convenient aspects of our approach is that establishing the required root-region for the complicated multivariate partition function can be boiled down to establishing stability for a bounded-degree univariate polynomial with coefficients coming from the local functions f_v. This was one of the main insights of [126, 66, 21]. More specifically, if Δ is the maximum degree of the input graph $G = (V, E)$, and $f_v : [d] \to \mathbb{R}_{\geq 0}$ is the local function for some vertex $v \in V$, where $d = \deg(v) \leq \Delta$, then define the corresponding local polynomial at v by

$$P_v(z) = \sum_{k=0}^{d} \binom{d}{k} f_v(k) z^k. \tag{8.21}$$

A circular region on the complex plane is the interior or exterior of a disk, or an open half-plane. [66] showed using Asano-Ruelle contractions [7, 114] that in the case all f_v are the same, and all P_v are Φ-stable for an open half-plane $\Phi \subseteq \mathbb{C}$, the multivariate partition function is Γ-stable where $\Gamma = [-(\Phi^{\mathsf{c}})^2]^{\mathsf{c}}$. This result actually holds for any circular region $\Phi \subseteq \mathbb{C}$ assuming that either Φ is convex or every local polynomial P_v has degree $\deg(v)$; under these assumptions one can apply the famous Grace-Walsh-Szegö Coincidence Theorem to the local polynomials, see [66, 26]. A straightforward generalization of their techniques yields the following.

Theorem 8.5.1 ([66])**.** *Let $G = (V, E)$ be a graph. Let $\{f_v\}_{v \in V} : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be a family of local functions, and let $\{\Phi_v\}_{v \in V}$ be a family of circular regions containing 0 such that*

for every $v \in V$, either Φ_v is convex or $f_v(\deg(v)) > 0$. If for every $v \in V$, the local polynomial P_v is Φ_v-stable, then the multivariate partition function $Z_G(\lambda)$ is $\prod_{e \in E} \Gamma_e$ stable, where for each edge $e = \{u, v\}$, $\Gamma_e = (-\Phi_u^{\mathsf{c}} \cdot \Phi_v^{\mathsf{c}})^{\mathsf{c}} \subseteq \mathbb{C}$.

Using Theorem 8.5.1, [66] established zero-free regions for a large class of Holant problems satisfying generalized second-order recurrences, including matchings, weighted edge covers, and weighted even subgraphs. Our main theorems Theorems 8.1.1 and 8.1.3 and Theorem 1.1.6 build upon these zero-free results as well as Theorem 8.2.1 and Theorem 1.2.1 (note that we can obtain spectral independence for matchings from Theorems 8.5.1 and 8.2.1, which was already known in [45] with a better bound by correlation decay proofs). Zero-free regions were also established for weighted edge covers and the antiferromagnetic Ising model on line graphs in [21], using techniques from [126].

Before proving the main theorems, we will need the following simple lemma concerning the case where the regions Φ_u are half-planes. Recall that $\mathbb{H}_\varepsilon = \{x + iy : x < -\varepsilon\}$ and $\overline{\mathbb{H}}_\varepsilon = \{x + iy : x \leq -\varepsilon\}$ for $\varepsilon \in \mathbb{R}_+$.

Lemma 8.5.2 (Lemma 5 in [66])**.** *For $\varepsilon > 0$, let $\Gamma = (-\overline{\mathbb{H}}_\varepsilon^2)^{\mathsf{c}}$ be a region. Then Γ contains $\mathbb{R}_+$, and for every $\lambda \in \mathbb{R}_+$ we have $\operatorname{dist}(\lambda, \partial\Gamma) = \lambda + \varepsilon^2$ if $\lambda \in (0, \varepsilon^2)$, and $\operatorname{dist}(\lambda, \partial\Gamma) = 2\varepsilon\sqrt{\lambda}$ if $\lambda \in [\varepsilon^2, \infty)$.*

For completeness, we provide a proof in Section 8.5.2. With these tools in hand, we deduce strong zero-free regions for the above examples. We use these to prove our main mixing results Theorems 8.1.1 and 8.1.3 and Theorem 1.1.6. Note that by Lemma 8.4.1 and Theorem 8.4.2, one can in fact establish rapid mixing results for these models with non-uniform external fields, though we only state the uniform case for simplicity.

Proof of Theorem 8.1.1. By Theorem 1.2.1, it suffices to prove η-spectral independence for $\eta = O_{\Delta,\lambda,\rho}(1)$. By Theorem 8.2.1, it suffices to prove that the multivariate partition function Eq. (8.20) is Γ-stable, where $\Gamma \subseteq \mathbb{C}$ is an open connected region containing $\mathbb{R}_+$ and $\delta = \frac{1}{\lambda}\operatorname{dist}(\lambda, \partial\Gamma) = \Omega_{\Delta,\lambda,\rho}(1)$.

It is more convenient for us to work with the model on complements of weighted edge covers, whose partition function is the inversion of that for weighted edge covers. For this, the local polynomial is given by

$$P_v(z) = (1+z)^{\deg(v)} - (1-\rho)z^{\deg(v)},$$

which is $\overline{\mathbb{H}}^{\mathsf{c}}_{1/2}$-stable. Then by Theorem 8.5.1, the inversion of the weighted edge cover partition function $Z_G(\lambda)$ is $\left(-\overline{\mathbb{H}}^2_{1/2}\right)^{\mathsf{c}}$-stable, and therefore $Z_G(\lambda)$ is Γ-stable for

$$\Gamma = \left[\left(-\overline{\mathbb{H}}^2_{1/2}\right)^{\mathsf{c}}\right]^{-1} = [-\overline{\mathbb{D}}(-1,1)^2]^{\mathsf{c}}.$$

This region Γ is also derived in [21]. We remark that the region $-\overline{\mathbb{D}}(-1,1)^2$ is cardioid-shaped, and its complement Γ is an open connected region containing $\mathbb{R}_+$; see Lemma 3.9 and Figure 1 in [21]. Hence, we have $\mathbb{R}_+ \subseteq \Gamma$ and $\delta = \Omega_{\Delta,\lambda,\rho}(1)$ as wanted. □

Proof of Theorem 1.1.6. We may assume $\rho \in (0,1)$ since if $\rho = 1$ then we get a trivial product distribution. Once again, by Theorem 1.2.1, it suffices to prove η-spectral independence for $\eta = O_{\Delta,\lambda,\rho}(1)$, and by Theorem 8.2.1, it suffices to prove that the multivariate partition function Eq. (8.20) is Γ-stable, where $\Gamma \subseteq \mathbb{C}$ is an open connected region containing $\mathbb{R}_+$ and $\delta = \frac{1}{\lambda}\mathrm{dist}(\lambda, \partial\Gamma) = \Omega_{\Delta,\lambda,\rho}(1)$.

For this, observe that the local polynomial is given by

$$\begin{aligned} P_v(z) &= \sum_{k=0}^{\deg(v)} \binom{\deg(v)}{k} \left(\frac{1+\rho}{2} + \frac{1-\rho}{2}(-1)^k\right) z^k \\ &= \frac{1+\rho}{2}(1+z)^{\deg(v)} + \frac{1-\rho}{2}(1-z)^{\deg(v)}. \end{aligned}$$

Since $0 < \rho < 1$, the roots of P_v are given by $\frac{\omega - t_v}{\omega + t_v}$ where $\omega \in \mathbb{C}$ satisfies $\omega^{\deg(v)} = -1$, and $t_v \in \mathbb{R}_+$ is given by

$$t_v = \left(\frac{1+\rho}{1-\rho}\right)^{1/\deg(v)} > 1.$$

It follows that P_v is $\left[\overline{\mathbb{D}}\left(-\frac{t_v^2+1}{t_v^2-1}, \frac{2t_v}{t_v^2-1}\right)\right]^{\mathsf{c}}$-stable. Then by Theorem 8.5.1, the partition function $Z_G(\lambda)$ is $\prod_{e\in E}\Gamma_e$-stable, where for each edge $e = uv \in E$,

$$\Gamma_e = \left[-\overline{\mathbb{D}}\left(-\frac{t_u^2+1}{t_u^2-1}, \frac{2t_u}{t_u^2-1}\right) \cdot \overline{\mathbb{D}}\left(-\frac{t_v^2+1}{t_v^2-1}, \frac{2t_v}{t_v^2-1}\right)\right]^{\mathsf{c}}.$$

In particular, $Z_G(\lambda)$ is Γ-stable for

$$\Gamma = \left[-\overline{\mathbb{D}}\left(-\frac{t^2+1}{t^2-1}, \frac{2t}{t^2-1}\right)^2\right]^{\mathsf{c}} \subseteq \Gamma_e, \quad \forall e \in E, \quad \text{where } t = \left(\frac{1+\rho}{1-\rho}\right)^{1/\Delta} > 1.$$

The region Γ is open and connected. Observe that we have $\Gamma \supseteq \left(-\overline{\mathbb{H}}^2_{\frac{t-1}{t+1}}\right)^{\mathsf{c}}$. Hence, by Lemma 8.5.2 we have $\mathbb{R}_+ \subseteq \Gamma$ and $\delta = \Omega_{\Delta,\lambda,\rho}(1)$ as wanted. □

Proof of Theorem 8.1.3. By Theorem 1.2.1 it suffices to prove η-spectral independence for $\eta = O_{\Delta,\beta,\gamma,\lambda}(1)$. By Theorem 8.2.1 it suffices to prove that the multivariate partition function Eq. (8.20) is Γ-stable, where $\Gamma \subseteq \mathbb{C}$ is an open connected region containing $\mathbb{R}_+$ and $\delta = \frac{1}{\lambda}\mathrm{dist}(\lambda, \partial\Gamma) = \Omega_{\Delta,\beta,\gamma,\lambda}(1)$.

For this, observe that the local polynomial is given by

$$P_v(z) = \sum_{k=0}^{\deg(v)} \binom{\deg(v)}{k} \beta^{\binom{k}{2}} \gamma^{\binom{\deg(v)-k}{2}} z^k.$$

By Proposition 8.5.3 below (see Section 8.5.1 for the proof), all roots of this polynomials are strictly negative reals, i.e. they are contained in $(-\infty, -\varepsilon_{\deg(v)}]$ for some constant $\varepsilon_{\deg(v)} = \varepsilon_{\deg(v)}(\beta,\gamma) > 0$ depending only on $\deg(v), \beta, \gamma$. Then by Theorem 8.5.1, $Z_G(\lambda)$ is $\prod_{e\in E}\Gamma_e$-stable, where for each edge $e = uv \in E$,

$$\Gamma_e = \left(-\overline{\mathbb{H}}_{\varepsilon_{\deg(u)}} \cdot \overline{\mathbb{H}}_{\varepsilon_{\deg(v)}}\right)^{\mathsf{c}}.$$

In particular, $Z_G(\lambda)$ is Γ-stable for $\Gamma = (-\overline{\mathbb{H}}^2_\varepsilon)^{\mathsf{c}}$ where $\varepsilon = \min_{1\le d\le\Delta}\varepsilon_d$ depends only on Δ, β, γ. The region Γ is open and connected, and by Lemma 8.5.2 it contains $\mathbb{R}_+$ and we

have $\delta = \Omega_{\Delta,\beta,\gamma,\lambda}(1)$ as wanted. $\square$

8.5.1 Stability for Antiferromagnetic Two-Spin Edge Models

In this subsection, we analyze the roots of the local polynomial for antiferromagnetic two-spin edge models, which is needed in the proof of Theorem 8.1.3 above. We generalize a result due to [21] which proves that the local polynomial for the antiferromagnetic edge Ising model has strictly negative real roots. We achieve this by generalizing their arguments to all antiferromagnetic two-spin edge models.

Proposition 8.5.3 (Generalization of Lemma 4.3 in [21])**.** *For every $\beta \geq 0$, $\gamma > 0$ with $\beta\gamma < 1$ and every positive integer $d \geq 1$, the univariate polynomial*

$$P_d(z) = \sum_{k=0}^{d} \binom{d}{k} \beta^{\binom{k}{2}} \gamma^{\binom{d-k}{2}} z^k$$

has strictly negative real roots.

We prove this via an inductive approach, relying on the following decomposition of P_d.

Lemma 8.5.4. *For every $\beta \geq 0$, $\gamma > 0$ and every positive integer $d \geq 1$, we have that*

$$P_{d+1}(z) = \gamma^d P_d(z/\gamma) + zP_d(\beta z).$$

Proof. We have

$$\begin{aligned} P_{d+1}(z) &= \sum_{k=0}^{d+1} \underbrace{\binom{d+1}{k}}_{=\binom{d}{k}+\binom{d}{k-1}} \beta^{\binom{k}{2}} \gamma^{\binom{d+1-k}{2}} z^k \\ &= \sum_{k=0}^{d} \binom{d}{k} \beta^{\binom{k}{2}} \gamma^{\binom{d+1-k}{2}} z^k + \sum_{k=0}^{d} \binom{d}{k} \beta^{\binom{k+1}{2}} \gamma^{\binom{d-k}{2}} z^{k+1} \\ &= \sum_{k=0}^{d} \binom{d}{k} \beta^{\binom{k}{2}} \gamma^{\binom{d-k}{2}} \gamma^{d-k} z^k + z \sum_{k=0}^{d} \binom{d}{k} \beta^{\binom{k}{2}} \gamma^{\binom{d-k}{2}} \beta^k z^k \end{aligned}$$

$$= \gamma^d P_d(z/\gamma) + zP_d(\beta z). \qquad \square$$

Proof of Proposition 8.5.3. If $\beta = 0$ then P_d is liner and the proposition is immediate. We may assume $\beta > 0$. We prove via induction the following stronger claim: The roots $r_1 > \cdots > r_d$ of P_d are distinct, real, and strictly negative, and further satisfy $r_i/r_{i+1} < \beta\gamma$. The cases $d = 0, 1$ are vacuous. When $d = 2$, the polynomial $P_2(z) = \beta z^2 + 2z + \gamma$ has roots $(-1 \pm \sqrt{1-\beta\gamma})/\beta$, which are distinct, real, and strictly negative since $\beta\gamma < 1$. One can also check that $r_1/r_2 < \beta\gamma$ via a straightforward calculation. This establishes the base case.

Assume the stronger conclusion holds for some $d \geq 2$. By Lemma 8.5.4, we may write $P_{d+1}(z) = \gamma^d P_d(z/\gamma) + zP_d(\beta z)$. If $r_1 > \cdots > r_d$ are the roots of P_d, then $\gamma r_1 > \cdots > \gamma r_d$ are the roots of $\gamma^d P_d(z/\gamma)$, and $0 = r_0/\beta > r_1/\beta > \cdots > r_d/\beta$ are the roots of $zP_d(\beta z)$, where for convenience we define $r_0 = 0$. First, we claim that the roots of $\gamma^d P_d(z/\gamma)$ interlace the roots of $zP_d(\beta z)$, i.e.,

$$0 = r_0/\beta > \gamma r_1 > r_1/\beta > \gamma r_2 > \cdots > r_{d-1}/\beta > \gamma r_d > r_d/\beta.$$

To see this, observe that $\gamma r_i > r_i/\beta$ since $\beta\gamma < 1$, and $r_{i-1}/\beta > \gamma r_i$ since $r_{i-1}/r_i < \beta\gamma$ by the induction hypothesis for P_d.

Now, we claim that for each $i = 2, \ldots, d$, the evaluations

$$P_{d+1}(\gamma r_i) = \gamma r_i P_d(\beta\gamma r_i) \quad \text{and} \quad P_{d+1}(r_{i-1}/\beta) = \gamma^d P_d(r_{i-1}/\beta\gamma)$$

are nonzero and have different signs. Observe that $\beta\gamma r_i, r_{i-1}/\beta\gamma \in (r_i, r_{i-1})$; hence, the evaluations $P_d(\beta\gamma r_i)$ and $P_d(r_{i-1}/\beta\gamma)$ are nonzero and have the same sign, and we deduce the claim by $r_i < 0$. It then follows from the Intermediate Value Theorem that P_{d+1} has a root $s_i \in (\gamma r_i, r_{i-1}/\beta)$ for each $i = 2, \ldots, d$.

Moreover, P_{d+1} also has a root $s_1 \in (\gamma r_1, 0)$ and a root $s_{d+1} \in (-\infty, r_d/\beta)$. Observe

that the evaluations $P_{d+1}(\gamma r_1) = \gamma r_1 P_d(\beta\gamma r_1)$ and $P_{d+1}(0) = \gamma^d P_d(0)$ are nonzero and have different signs since $0 > \beta\gamma r_1 > r_1$, and the Intermediate Value Theorem implies a root $s_1 \in (\gamma r_1, 0)$. Meanwhile, $P_{d+1}(r_d/\beta) = \gamma^d P_d(r_d/\beta\gamma)$ and $P_d(-\infty)$ are nonzero and have the same sign since $-\infty < r_d/\beta\gamma < r_d$. Also, $P_d(-\infty)$ and $P_{d+1}(-\infty)$ have different signs since the two polynomials differ in degree by 1. This shows that $P_{d+1}(r_d/\beta)$ and $P_{d+1}(-\infty)$ are nonzero and have different signs, and the Intermediate Value Theorem shows the existence of a root $s_{d+1} \in (-\infty, r_d/\beta)$.

To summarize, we prove that P_{d+1} has roots $s_1 > \cdots > s_{d+1}$ which are distinct strictly negative real numbers and (taking $r_0 = 0$ and $r_{d+1} = -\infty$ for convenience) satisfy $s_i \in (\gamma r_i, r_{i-1}/\beta)$ for any $i = 1, \ldots, d+1$. To finish the induction, we need to show that $s_i/s_{i+1} < \beta\gamma$ for all $i = 1, \ldots, d$, which follows by $s_i/s_{i+1} < (\gamma r_i)/(r_i/\beta) = \beta\gamma$. □

8.5.2 Proofs of Technical Lemmas

Proof of Lemma 8.5.2. It was shown in [66] that

$$\Gamma = \left(-\overline{\mathbb{H}}_\varepsilon^2\right)^{\mathsf{c}} = \left\{\rho e^{i\theta} : \rho < \frac{2\varepsilon^2}{1-\cos\theta},\, 0 < \theta < 2\pi\right\}.$$

To make this more interpretable, we rewrite the set in Cartesian coordinates. If $z = \rho e^{i\theta}$, then by Euler's formula we may write $z = x + iy$ where $x = \rho\cos\theta$ and $y = \rho\sin\theta$. We then obtain

$$\begin{aligned}
& \rho < \frac{2\varepsilon^2}{1-\cos\theta} \\
\iff \quad & \rho(1-\cos\theta) < 2\varepsilon^2 \\
\iff \quad & \rho < x + 2\varepsilon^2 \\
\iff \quad & x^2 + y^2 < (x + 2\varepsilon^2)^2 \\
\iff \quad & y^2 < 4\varepsilon^2(x + \varepsilon^2).
\end{aligned}$$

Therefore, we see that

$$\Gamma = \left\{x + iy : y^2 < 4\varepsilon^2(x + \varepsilon^2)\right\},$$

which clearly contains $\mathbb{R}_+$.

Furthermore, for $\lambda \in \mathbb{R}_+$ we have

$$\begin{aligned}
\operatorname{dist}(\lambda, \partial\Gamma) &= \inf_{z \in \partial\Gamma} |z - \lambda| \\
&= \inf_{(x,y) \in \mathbb{R}^2 :\, y^2 = 4\varepsilon^2(x+\varepsilon^2)} \sqrt{(x-\lambda)^2 + y^2} \\
&= \inf_{x \in [-\varepsilon^2, \infty)} \sqrt{(x-\lambda)^2 + 4\varepsilon^2(x+\varepsilon^2)} \\
&= \begin{cases} \lambda + \varepsilon^2, & \lambda \in (0, \varepsilon^2); \\ 2\varepsilon\sqrt{\lambda}, & \lambda \in [\varepsilon^2, \infty). \end{cases}
\end{aligned}$$

This establishes the lemma. □

8.6 Further Optimal Mixing Results

In this section, we study spectral independence for general tensor network contractions and weighted graph homomorphisms. Unlike binary symmetric Holant problems, where rapid mixing of the Glauber dynamics was already known for our main examples such as matchings [76], Ising model on line graphs [53], edge covers [73], and weighted even subgraphs [77], in the setting we consider here, rapid mixing for any local Markov chain was not known before. Prior works [16, 17, 112, 110] had studied these problems but only from the perspective of deterministic approximation algorithms using Barvinok's polynomial interpolation method [10]. While these algorithms run in polynomial time for bounded-degree graphs, the exponent typically depends on the maximum degree, and are more difficult to implement.

Here, we show that the Glauber dynamics mixes in $O(n \log n)$ steps for these problems on bounded-degree graphs, yielding significantly faster and simpler algorithms for

computing the partition function. We again reduce rapid mixing to spectral independence via Theorem 1.2.1, and then reduce spectral independence to the existence of a sufficiently large zero-free region for the multivariate partition function via Theorem 8.4.2. Fortunately, such zero-free regions were already obtained in prior works, as they are the entire basis for Barvinok's polynomial interpolation method. We leverage them here in a completely black-box manner.

8.6.1 Weighted Graph Homomorphisms

Here, we study weighted graph homomorphisms, which may also be viewed as spin systems on vertices. In the bounded-degree setting, we show that the Glauber dynamics on vertex configurations for these models mixes in $O(n \log n)$ steps, provided the weights are sufficiently close to 1. This is analogous to classical mixing results stating the Glauber dynamics mixes rapidly in the "high-temperature" regime.

Theorem 8.6.1 (Spectral Independence for Weighted Graph Homomorphisms)**.** *Fix a positive integer $q \geq 2$, let $G = (V, E)$ be a graph with maximum degree $\leq \Delta$, and for each edge $uv \in E$, let $A^{uv} \in \mathbb{R}_{\geq 0}^{q \times q}$ be a (not necessarily symmetric) nonnegative matrix. There exists a universal constant $\gamma \approx 0.56$ independent of $q, G, \{A^{uv}\}_{uv \in E}$ such that if $|A^{uv}(j,k) - 1| \leq \frac{\gamma}{\Delta + \gamma} - \varepsilon$ for some $\varepsilon > 0$, all $uv \in E$ and all $j, k \in [q]$, then the associated graph homomorphism distribution μ on vertex configurations $\sigma : V \to [q]$ given by*

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

is η-spectrally independent for some constant $\eta = \eta(\Delta, \varepsilon)$. In particular, if $\Delta, \varepsilon = \Theta(1)$, then the Glauber dynamics for sampling from μ mixes in $O(n \log n)$ steps.

Remark 8.6.2. A straightforward application of the classical Dobrushin uniqueness condition yields rapid mixing when $|A^{uv}(j,k) - 1| < \frac{1}{2\Delta}$ for all $uv \in E$ and $j, k \in [q]$.

The zero-free region for the graph homomorphism partition function was studied in [17]. We state here a slightly more general theorem, the proof of which is included in Section 8.6.3 for completeness.

Theorem 8.6.3 (Zeros for Weighted Graph Homomorphisms; [17])**.** *Fix a positive integer $q \geq 2$, let $G = (V, E)$ be a graph with maximum degree $\leq \Delta$, and for each edge $e = uv \in E$, let $A^{uv} \in \mathbb{C}^{q \times q}$ be a (not necessarily symmetric or Hermitian) complex matrix. There exists a universal constant $\gamma \approx 0.56$ independent of $q, G, \{A^{uv}\}_{uv \in E}$ such that if $|A^{uv}(j,k) - 1| < \frac{\gamma}{\Delta+\gamma}$ for all $uv \in E$ and all $j, k \in [q]$, then for every $S \subseteq V$ and every $\phi : S \to [q]$, the graph homomorphism partition function*

$$\sum_{\substack{\sigma: V \to [q] \\ \sigma|_S = \phi}} \prod_{uv \in E} A^{uv}(\sigma(u), \sigma(v))$$

with pinning ϕ is nonzero.

We give below the proof of Theorem 8.6.1.

Proof of Theorem 8.6.1. By Theorem 8.4.2, it suffices to prove that the multivariate partition function

$$\sum_{\substack{\sigma: V \to [q] \\ \sigma|_S = \phi}} \prod_{uv \in E} A^{uv}(\sigma(u), \sigma(v)) \prod_{v \in V} \lambda_{v, \sigma(v)} \tag{8.22}$$

is nonzero in the polydisk $\mathcal{D} = \left\{\lambda \in \mathbb{C}^{V \times [q]} : |\lambda_{v,k} - 1| < c, \forall v \in V, \forall k \in [q]\right\}$ for all pinnings ϕ, where $c = c(\Delta, \varepsilon) > 0$ is some constant depending only on Δ, ε but not G. Define a new set of matrices $\{\tilde{A}^{uv}\}_{uv \in E}$ by

$$\tilde{A}^{uv}(j,k) = A^{uv}(j,k) \cdot \lambda_{u,j}^{1/\deg(u)} \cdot \lambda_{v,k}^{1/\deg(v)}, \quad \forall uv \in E, \forall j, k \in [q].$$

Note that the partition function for $G, \{\tilde{A}^{uv}\}_{uv \in E}$ is precisely given in Eq. (8.22).

Since $|A^{uv}(j,k) - 1| \leq \frac{\gamma}{\Delta+\gamma} - \varepsilon$, there exists our desired $c(\Delta, \varepsilon) > 0$ such that $|\lambda_{u,j} - 1|, |\lambda_{v,k} - 1| < c(\Delta, \varepsilon)$ implies $|\tilde{A}^{uv}(j,k) - 1| < \frac{\gamma}{\Delta+\gamma}$, for all $uv \in E$ and all $j, k \in [q]$. It follows from Theorem 8.6.3 that the multivariate partition function Eq. (8.22) is nonzero. As this holds for all $\lambda \in \mathcal{D}$, we are done. □

8.6.2 Tensor Network Contractions

Here, we study general tensor network contractions, which is a partition function of a distribution over configurations on edges of a graph. Tensor networks are heavily studied in quantum computing [96, 6, 109] and are also used to model Holant problems [31, 32, 33]. In the bounded-degree setting, we also show that the Glauber dynamics on edge configurations for these models mixes in $O(n \log n)$ steps, provided the weights are sufficiently close to 1. Again, this is analogous to classical mixing results stating the Glauber dynamics mixes rapidly in the "high-temperature" regime.

To state our main result, let us first define tensor network contraction. Given a graph $G = (V, E)$ and a collection of local functions $\{f_v : [q]^{E(v)} \to \mathbb{R}_{\geq 0}\}_{v \in V}$ on configurations on edges, we define the associated tensor network distribution μ over edge configurations $\sigma : E \to [q]$ to be given by

$$\mu(\sigma) \propto \prod_{v \in V} f_v(\sigma \mid_{E(v)}). \tag{8.23}$$

The associated partition function, known as a tensor network contraction, is given by

$$\sum_{\sigma : E \to [q]} \prod_{v \in V} f_v(\sigma \mid_{E(v)}).$$

The name "tensor network" comes from the fact that each f_v may be viewed as a tensor with axes corresponding to edges in $E(v)$ and indexed by $[q]$. This is a vast generalization of the Holant problems considered in Section 8.5 (see, for instance, Eq. (8.20)), where $q = 2$ and each local function f_v is symmetric. Zeros for tensor network contractions were

analyzed in [112] in the symmetric case.

Theorem 8.6.4 (Spectral Independence for Tensor Network Distribution)**.** *Fix a positive integer $q \geq 2$, let $G = (V, E)$ be a graph with maximum degree $\leq \Delta$, and for each vertex $v \in V$, let $f_v : [q]^{E(v)} \to \mathbb{R}_{\geq 0}$ be a nonnegative function on configurations of edges incident to v. There exists a universal constant $\gamma \approx 0.56$ independent of $q, G, \{f_v\}_{v \in V}$ such that if $|f_v(\alpha) - 1| \leq \frac{\gamma}{\Delta+1+\gamma} - \varepsilon$ for some $\varepsilon > 0$, all $v \in V$ and all $\alpha : E(v) \to [q]$, then the tensor network distribution μ on edge configurations $\sigma : E \to [q]$ given by Eq.* (8.23) *is η-spectrally independent for some constant $\eta = \eta(\Delta, \varepsilon)$. In particular, if $\Delta, \varepsilon = \Theta(1)$, then the Glauber dynamics for sampling from μ mixes in $O(n \log n)$ steps.*

To establish this spectral independence, we need a sufficiently large zero-free region. This was proved by [112] in the symmetric case, where each local function f_v depends only on the number of incident edges that are mapped to each color in $[q]$. It turns out using nearly identical arguments, one can obtain the following more general theorem. We provide a proof in Section 8.6.3 for completeness.

Theorem 8.6.5 (Zeros of Tensor Network Contractions; [112])**.** *Fix a positive integer $q \geq 2$, let $G = (V, E)$ be a graph with maximum degree $\leq \Delta$, and for each vertex $v \in V$, let $f_v : [q]^{E(v)} \to \mathbb{C}$ be a complex function on configurations of edges incident to v. There exists a universal constant $\gamma \approx 0.56$ independent of $q, G, \{f_v\}_{v \in V}$ such that if $|f_v(\alpha) - 1| < \frac{\gamma}{\Delta+1+\gamma}$ for all $v \in V$ and all $\alpha : E(v) \to [q]$, then for every $F \subseteq E$ and every $\phi : F \to [q]$, the tensor network contraction*

$$\sum_{\substack{\sigma : E \to [q] \\ \sigma|_F = \phi}} \prod_{v \in V} f_v(\sigma \mid_{E(v)})$$

with pinning ϕ is nonzero.

We give below the proof of Theorem 8.6.4.

Proof of Theorem 8.6.4. By Theorem 8.4.2, it suffices to prove that the multivariate partition function

$$\sum_{\substack{\sigma:E\to[q]\\ \sigma|_F=\phi}} \prod_{v\in V} f_v(\sigma\mid_{E(v)}) \prod_{e\in E} \lambda_{e,\sigma(e)} \tag{8.24}$$

is nonzero whenever λ lies in the polydisk $\mathcal{D} = \{\lambda \in \mathbb{C}^{E\times[q]} : |\lambda_{e,k} - 1| < c, \forall e \in E, \forall k \in [q]\}$ for all pinnings ϕ, where $c = c(\Delta, \varepsilon) > 0$ is some constant depending only on Δ, ε but not G. Define a new set of local constraint functions $\{\tilde{f}_v\}_{v\in V}$ by

$$\tilde{f}_v(\alpha) = f_v(\alpha) \cdot \prod_{e\in E(v)} \lambda_{e,\alpha(e)}^{1/2}, \quad \forall v \in V, \forall \alpha : E(v) \to [q].$$

Note that the partition function for $G, \{\tilde{f}_v\}_{v\in V}$ is precisely given in Eq. (8.24).

Since $|f_v(\alpha)-1| \le \frac{\gamma}{\Delta+1+\gamma} - \varepsilon$, there exists our desired $c(\Delta, \varepsilon) > 0$ such that $|\lambda_{e,k}-1| < c(\Delta, \varepsilon)$ for all $e \in E(v)$ implies $|\tilde{f}_v(\alpha)-1| < \frac{\gamma}{\Delta+1+\gamma}$, for all $v \in V$ and all $\alpha : E(v) \to [q]$. It follows from Theorem 8.6.5 that the multivariate partition function Eq. (8.24) is nonzero. As this holds for all $\lambda \in \mathcal{D}$, we are done. □

8.6.3 Proofs of Zero-Free Results

In this section, we supply proofs of the main zero-free statements used in Section 8.6. As noted earlier, for technical reasons, we need straightforward generalizations of prior results which do not make symmetry assumptions. We manage to adapt previous arguments without much additional effort, which we provide here for completeness.

The main idea in these zero-free proofs is to do induction by conditioning on the assignment of fewer and fewer vertices (respectively, edges) for weighted homomorphisms (respectively, tensor networks). However, one needs to strengthen the inductive hypothesis beyond simple zero-freeness. To the best of our knowledge, this type of argument was first pioneered by Barvinok, and has had a wide range of applications; see [9, 12, 15, 14] for

applications besides those discussed in this chapter.

The crucial tool is the following geometric lemma, which provides a kind of "reverse triangle inequality". The version below is due to Boris Bukh; a weaker version, with $\cos(\theta/2)$ replaced by $\sqrt{\cos\theta}$, was known due to [11]. See [10] for a proof.

Lemma 8.6.6 (Angle Lemma)**.** *Let $x_1, \ldots, x_n \in \mathbb{C}$ be nonzero complex numbers viewed as vectors in $\mathbb{R}^2$. Suppose there is an angle $0 \leq \theta < 2\pi/3$ such that for all i, j, the angle between x_i, x_j is at most θ. Then we have the lower bound $|\sum_{i=1}^n x_i| \geq \cos(\theta/2) \sum_{i=1}^n |x_i|$.*

Proofs for Weighted Graph Homomorphisms

Our goal in this subsection is to prove Theorem 8.6.3, i.e. that the weighted graph homomorphism partition function

$$Z_\phi^S(A) = \sum_{\substack{\sigma: V \to [q] \\ \sigma|_S = \phi}} \prod_{uv \in E} A^{uv}(\sigma(u), \sigma(v))$$

is nonzero in a large polydisk around 1, where $S \subseteq V, \phi : S \to [q]$, and we view $Z_\phi^S(A)$ as a polynomial with variables $\{A^{uv}(j,k)\}_{uv \in E, j,k \in [q]}$. For convenience, for a $\delta > 0$, define

$$\mathcal{U}(\delta) = \{A = \{A^{uv}\}_{uv \in E} : |A^{uv}(j,k) - 1| < \delta, \forall uv \in E, \forall j, k \in [q]\}.$$

Additionally, for a partial configuration $\phi : S \to [q]$, a vertex $u \in V \setminus S$ and a spin $j \in [q]$, we write $\phi_{u,j} : S \cup \{u\} \to [q]$ for the unique extension of ϕ with $\phi_{u,j}(u) = j$.

We will need the following lemmas to implement an inductive approach.

Lemma 8.6.7 (Lemma 3.3 from [17])**.** *Let $\tau, \delta > 0$, and suppose $A \in \mathcal{U}(\delta)$. Let $S \subseteq V$, $\phi : S \to [q]$, $u \in V \setminus S$ be arbitrary. Assume the following hold:*

(1) $Z_{\phi_{u,j}}^{S \cup \{u\}}(A) \neq 0$ for every $u \in V \setminus S$ and every $j \in [q]$;

(2) For every $u \in V \setminus S$ and every $j \in [q]$, we have

$$|Z^{S\cup\{u\}}_{\phi_{u,j}}(A)| \geq \frac{\tau}{\Delta} \sum_{v\sim u} \sum_{k\in[q]} |A^{uv}(j,k)| \cdot \left| \frac{\partial}{\partial A^{uv}(j,k)} Z^{S\cup\{u\}}_{\phi_{u,j}}(A) \right|.$$

Then for every $u \in V \setminus S$ and every $j, k \in [q]$, the angle between $Z^{S\cup\{u\}}_{\phi_{u,j}}(A)$ and $Z^{S\cup\{u\}}_{\phi_{u,k}}(A)$ in $\mathbb{C}$ is at most $\frac{2\delta\Delta}{\tau(1-\delta)}$.

Proof. By assumption (1), the relevant partition functions are nonzero, and so the logarithm is well-defined when applied to these partition functions and we may bound the angle between $Z^{S\cup\{u\}}_{\phi_{u,j}}(A)$ and $Z^{S\cup\{u\}}_{\phi_{u,k}}(A)$ by

$$\left| \log Z^{S\cup\{u\}}_{\phi_{u,j}}(A) - \log Z^{S\cup\{u\}}_{\phi_{u,k}}(A) \right|. \tag{8.25}$$

The strategy is to write $Z^{S\cup\{u\}}_{\phi_{u,k}}(A)$ as $Z^{S\cup\{u\}}_{\phi_{u,j}}(\tilde{A})$ for some $\tilde{A} \in \mathcal{U}(\delta)$ which differs from A by a small number of coordinates, and then apply the Fundamental Theorem of Calculus and assumption (2). For every $v \sim u$, we set $\tilde{A}^{uv}(j,c) = A^{uv}(k,c)$ for every $c \in [q]$, and $\tilde{A}^{uv}(\ell,c) = A^{uv}(\ell,c)$ for all $\ell \neq j$. For all other edges $vw \in E$, we set $\tilde{A}^{vw} = A^{vw}$.

It is clear that $Z^{S\cup\{u\}}_{\phi_{u,k}}(A) = Z^{S\cup\{u\}}_{\phi_{u,j}}(\tilde{A})$. By the Fundamental Theorem of Calculus, we may upper bound Eq. (8.25) by

$$\begin{aligned}
&\max_{B\in\mathcal{U}(\delta)} \sum_{v\sim u} \sum_{c\in[q]} \left| \frac{\partial}{\partial A^{uv}(j,c)} \log Z^{S\cup\{u\}}_{\phi_{u,j}}(B) \right| \cdot \underbrace{|A^{uv}(j,c) - \tilde{A}^{uv}(j,c)|}_{\leq 2\delta \text{ since } A,\tilde{A}\in\mathcal{U}(\delta)} \\
&\leq \frac{2\delta}{1-\delta} \max_{B\in\mathcal{U}(\delta)} \underbrace{\sum_{v\sim u} \sum_{c\in[q]} |A^{uv}(j,c)| \cdot \frac{1}{|Z^{S\cup\{u\}}_{\phi_{u,j}}(B)|} \cdot \left| \frac{\partial}{\partial A^{uv}(j,c)} Z^{S\cup\{u\}}_{\phi_{u,j}}(B) \right|}_{\leq \Delta/\tau \text{ by assumption (2)}} \\
&\leq \frac{2\delta\Delta}{\tau(1-\delta)}.
\end{aligned}$$

□

Lemma 8.6.8 (Lemma 3.4 from [17])**.** *Let $0 \leq \theta < 2\pi/3$, $\delta > 0$, and suppose $A \in \mathcal{U}(\delta)$. Let $S \subseteq V$, $\phi : S \to [q]$ be arbitrary. Assume the following hold:*

(1) $Z^{S\cup\{u\}}_{\phi_{u,j}}(A) \neq 0$ *for every* $u \in V \setminus S$ *and every* $j \in [q]$*;*

(2) The angle between $Z^{S\cup\{u\}}_{\phi_{u,j}}(A)$ *and* $Z^{S\cup\{u\}}_{\phi_{u,k}}(A)$ *in* $\mathbb{C}$ *is at most* θ*, for every* $u \in V \setminus S$ *and every* $j, k \in [q]$*.*

Then for every $u \in S$*, we have the lower bound*

$$|Z^S_\phi(A)| \geq \frac{\cos(\theta/2)}{\Delta} \sum_{v\sim u} \sum_{k\in[q]} |A^{uv}(\phi(u),k)| \cdot \left| \frac{\partial}{\partial A^{uv}(\phi(u),k)} Z^S_\phi(A) \right|.$$

Proof. If $v \in S$ as well, then there is a unique $k \in [q]$ for which $\frac{\partial}{\partial A^{uv}(\phi(u),k)} Z^S_\phi(A) \neq 0$, namely $k = \phi(v)$. In this case, $A^{uv}(\phi(u),k) \cdot \frac{\partial}{\partial A^{uv}(\phi(u),k)} Z^S_\phi(A) = Z^S_\phi(A)$. Otherwise, $v \notin S$ and $\frac{\partial}{\partial A^{uv}(\phi(u),k)} Z^S_\phi(A) = \frac{1}{A^{uv}(\phi(u),k)} \cdot Z^{S\cup\{v\}}_{\phi_{v,k}}(A)$, where $\phi_{v,k}$ is the unique extension of ϕ mapping v to k.

Combining these two observations, we obtain

$$\begin{aligned}
&\sum_{v\sim u} \sum_{k\in[q]} |A^{uv}(\phi(u),k)| \cdot \left| \frac{\partial}{\partial A^{uv}(\phi(u),k)} Z^S_\phi(A) \right| \\
&= |N(u)\cap S| \cdot |Z^S_\phi(A)| + \sum_{v\sim u: v\notin S} \sum_{k\in[q]} |Z^{S\cup\{v\}}_{\phi_{v,k}}(A)| \\
&\leq |N(u)\cap S| \cdot |Z^S_\phi(A)| + \frac{1}{\cos(\theta/2)} \underbrace{\left| \sum_{v\sim u: v\notin S} \sum_{k\in[q]} Z^{S\cup\{v\}}_{\phi_{v,k}}(A) \right|}_{=|N(u)\setminus S|\cdot|Z^S_\phi(A)|} && \text{(Lemma 8.6.6)} \\
&\leq \frac{\Delta}{\cos(\theta/2)} \cdot |Z^S_\phi(A)|.
\end{aligned}$$

Rearranging yields the desired result. □

With these lemmas in hand, we can now prove the main zero-free result.

Proof of Theorem 8.6.3. Let $0 < \theta < 2\pi/3$ be a parameter to be determined later, set $\tau = \cos(\theta/2)$, and let $\delta > 0$ satisfy $\theta = \frac{2\delta\Delta}{\tau(1-\delta)}$; in particular, $\delta = \frac{\frac{1}{2\Delta}\theta\cos(\theta/2)}{1+\frac{1}{2\Delta}\theta\cos(\theta/2)}$. We show by descending induction on $|S|$ that the following three statements are all true:

(i) For every $S \subseteq V$, $\phi : S \to [q]$ and $A \in \mathcal{U}(\delta)$, we have $Z_\phi^S(A) \neq 0$.

(ii) For every $S \subseteq V$, $u \in V \setminus S$, $\phi : S \to [q]$, $A \in \mathcal{U}(\delta)$ and $j, k \in [q]$, the angle between $Z_{\phi_{u,j}}^{S \cup \{u\}}(A)$ and $Z_{\phi_{u,k}}^{S \cup \{u\}}(A)$ in $\mathbb{C}$ is at most θ.

(iii) For every $S \subseteq V$, $u \in S$, $A \in \mathcal{U}(\delta)$, we have the inequality

$$|Z_\phi^S(A)| \geq \frac{\cos(\theta/2)}{\Delta} \sum_{v \sim u} \sum_{k \in [q]} |A^{uv}(\phi(u), k)| \cdot \left| \frac{\partial}{\partial A^{uv}(\phi(u), k)} Z_\phi^S(A) \right|.$$

The base case $S = V$ is easily verified since $Z_\phi^S(A) = \prod_{uv \in E} A^{uv}(\phi(u), \phi(v))$, a product of nonzero complex numbers.

Now, let $S \subseteq V$ with $|S| < |V|$.

(i) Let $u \in V \setminus S$, which exists since $|S| < |V|$. It follows that (i) holds for $S \cup \{u\}$ by the inductive hypothesis. Since $Z_\phi^S(A) = \sum_{k \in [q]} Z_{\phi_{u,k}}^{S \cup \{u\}}(A)$, Lemma 8.6.6 applied to $Z_\phi^S(A)$ yields (i) assuming that (ii) holds. We prove (ii) below.

(ii) Let $u \in V \setminus S$, which exists since $|S| < |V|$. Then (i) and (iii) hold for $S \cup \{u\}$ by the inductive hypothesis. (ii) then follows by Lemma 8.6.7.

(iii) Let $u \in S$. Then (i) holds for $S \cup \{u\}$ by the inductive hypothesis. Since (ii) holds for S (as proved earlier), we may then apply Lemma 8.6.8, yielding (iii) for S.

Now, we choose $0 < \theta < 2\pi/3$. As we wish to maximize the size of our zero-free region, i.e. δ, we need to maximize $\theta \cos(\theta/2)$. As shown in [112], the maximum is attained when $2/\theta = \tan(\theta/2)$, which has solution $\theta^* \approx 1.72067$ and has objective value $x^* = \theta^* \cos(\theta^*/2) \approx 1.12219$. This yields $\delta = \frac{\frac{x^*}{2}}{\Delta + \frac{x^*}{2}}$ as claimed. □

Proofs for Tensor Network Contractions

Our goal in this subsection is to prove Theorem 8.6.5, i.e. that the tensor network partition function

$$Z_\phi^F(h) = \sum_{\substack{\sigma: E \to [q] \\ \sigma|_F = \phi}} \prod_{v \in V} h_v(\sigma\,|_{E(v)})$$

is nonzero in a large polydisk around 1, where $F \subseteq E, \phi : F \to [q]$, and we view $Z_\phi^F(\cdot)$ as a polynomial with variables $\{h_v(\alpha)\}_{v,\alpha}$. We prove the following stronger result.

Theorem 8.6.9 (Generalization of Theorem 6 from [112])**.** *Let $G = (V, E)$ be a graph of maximum degree $\leq \Delta$. Then for every $F \subseteq E$, $\phi : F \to [q]$, $\eta > 0$, and $0 \leq \theta < 2\pi/3$, the function $Z_\phi^F(h)$ is nonzero whenever $h \in \prod_{v \in V} S_v(\delta, \eta)$, where*

$$S_v(\delta, \eta) = \left\{ h_v : [q]^{E(v)} \to \mathbb{C} : \begin{smallmatrix} |h_v(\alpha) - h_v(\beta)| < \delta, \forall \alpha, \beta : E(v) \to [q], \\ |h_v(\alpha)| \geq \eta, \forall \alpha : E(v) \to [q] \end{smallmatrix} \right\}$$

and $\delta = \eta \cdot \min\left\{1, \frac{\theta \cos(\theta/2)}{\Delta+1}\right\}$.

Before we prove this result, let us see how this gives Theorem 8.6.5.

Proof of Theorem 8.6.5. Observe that $S_v(\delta, \eta)$ contains a disk around 1 of radius given by $\min\{\delta/2, 1 - \eta\}$. Using Theorem 8.6.9 and given that $\delta = \eta \cdot \min\left\{1, \frac{\theta \cos(\theta/2)}{\Delta+1}\right\}$, where $0 < \theta < 2\pi/3$, our goal is to maximize $\theta \cos(\theta/2)$ over $0 < \theta < 2\pi/3$ to obtain the largest zero-free disk. As shown in [112], this maximum is attained when $2/\theta = \tan(\theta/2)$, which has solution $\theta^* \approx 1.72067$ and has objective value $x^* = \theta^* \cos(\theta^*/2) \approx 1.12219$. Given this, to obtain the largest possible radius disk, we equalize $1 - \eta$ and $\delta/2 = \eta \cdot \frac{x^*}{2(\Delta+1)}$. Solving, we obtain $\eta = \frac{1}{1 + \frac{x^*}{2(\Delta+1)}}$, yielding radius $\frac{\frac{x^*}{2(\Delta+1)}}{1 + \frac{x^*}{2(\Delta+1)}}$ as desired. □

It remains to prove Theorem 8.6.9. We will need the following lemmas to implement an inductive approach.

Lemma 8.6.10 (Lemma 8 from [112]). *Let $\tau > 0$, $F \subseteq E$, $\phi : F \to [q]$ and $u \in V$ be arbitrary. Suppose for all $h \in \prod_{v \in V} S_v(\delta, \eta)$ and all $\psi : F \cup E(u) \to [q]$ extending ϕ, the following hold:*

(1) $Z_\psi^{F \cup E(u)}(h) \neq 0$;

(2) For all $v \in N(u) \cup \{u\}$, we have

$$|Z_\psi^{F \cup E(u)}(h)| \geq \tau \sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \psi}} |h_v(\alpha)| \cdot \left| \frac{\partial}{\partial h_v(\alpha)} Z_\psi^{F \cup E(u)}(h) \right|.$$

Then for all extensions $\psi, \tilde{\psi} : F \cup E(u) \to \mathbb{C}$ of ϕ, the angle between $Z_\psi^{F \cup E(u)}$ and $Z_{\tilde{\psi}}^{F \cup E(u)}(h)$ in $\mathbb{C}$ is at most $\frac{\delta(\Delta+1)}{\tau\eta}$.

Proof. By assumption (1), the relevant partition functions are nonzero, and so the logarithm is well-defined when applied to these partition functions and we may bound the angle between $Z_\psi^{F \cup E(u)}(h)$ and $Z_{\tilde{\psi}}^{F \cup E(u)}(h)$ by

$$\left| \log Z_\psi^{F \cup E(u)}(h) - \log Z_{\tilde{\psi}}^{F \cup E(u)}(h) \right|. \tag{8.26}$$

The strategy is to write $Z_{\tilde{\psi}}^{F \cup E(u)}(h)$ as $Z_\psi^{F \cup E(u)}(\tilde{h})$ for some $\tilde{h} \in \prod_{v \in V} S_v(\delta, \eta)$ which differs from h by a small number of coordinates, and then apply the Fundamental Theorem of Calculus and assumption (2). Let $v \in V$. We consider three cases.

- $\mathbf{v \notin N(u) \cup \{u\}}$: In this case, $\psi, \tilde{\psi}$ agree on $E(v)$ and so we may simply take $h_v = \tilde{h}_v$.

- $\mathbf{v \in N(u)}$: In this case, $\psi, \tilde{\psi}$ differ only on the single edge uv. If $\alpha : E(v) \to [q]$ agrees with ψ on uv, then let $\alpha' : E(v) \to [q]$ be given by replacing $\alpha(uv) = \psi(uv)$ with $\tilde{\psi}(uv)$, and take $\tilde{h}_v(\alpha) = h_v(\alpha')$. Otherwise, just set $\tilde{h}_v(\alpha) = h_v(\alpha)$. (Note that it does not really matter what we set $\tilde{h}_v(\alpha)$ to since $Z_\psi^{F \cup E(u)}(h)$ only has the term

$h_v(\alpha)$ when α agrees with ψ on uv. However, we wish to minimize the number of coordinates in which $h, \tilde{h}$ differ.)

- $\mathbf{v = u}$: In this case, just set $\tilde{h}_v(\psi \mid_{E(v)}) = h_v(\tilde{\psi} \mid_{E(v)})$ and $\tilde{h}_v(\alpha) = h_v(\alpha)$ for all $\alpha \neq \psi \mid_{E(v)}$.

It is clear that $Z_{\tilde{\psi}}^{F \cup E(u)}(h) = Z_{\psi}^{F \cup E(u)}(\tilde{h})$. By the Fundamental Theorem of Calculus , we may upper bound Eq. (8.26) by

$$\begin{aligned}
&\max_{x \in \prod_{v \in V} S_v(\delta,\eta)} \sum_{v \in N(u) \cup \{u\}} \sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \psi}} \left| \frac{\partial}{\partial h_v(\alpha)} \log Z_{\psi}^{F \cup E(u)}(x) \right| \cdot \left| h_v(\alpha) - \tilde{h}_v(\alpha) \right| \\
&\leq \frac{\delta}{\eta} \max_{x \in \prod_{v \in V} S_v(\delta,\eta)} \underbrace{\sum_{v \in N(u) \cup \{u\}}}_{\leq \Delta + 1} \underbrace{\sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \psi}} |h_v(\alpha)| \cdot \frac{1}{|Z_{\psi}^{F \cup E(u)}(x)|} \cdot \left| \frac{\partial}{\partial h_v(\alpha)} Z_{\psi}^{F \cup E(u)}(x) \right|}_{\leq 1/\tau \text{ by assumption (2)}} \\
&\qquad \text{(Definition of } S_v(\delta,\eta)) \\
&\leq \frac{\delta(\Delta+1)}{\tau\eta}. \qquad \square
\end{aligned}$$

Lemma 8.6.11 (Lemma 9 from [112]). *Let $0 \leq \theta < 2\pi/3$, $u \in V$, $F \subseteq E$ satisfying $F \supseteq E(u)$, and $\phi : F \to [q]$. Suppose for all $v \in N(u) \cup \{u\}$, all $h \in \prod_{v \in V} S_v(\delta,\eta)$, and all extensions $\psi, \tilde{\psi} : F \cup E(v) \to [q]$ of ϕ, the following hold:*

(1) $Z_{\psi}^{F \cup E(v)}(h) \neq 0$;

(2) The angle between $Z_{\psi}^{F \cup E(v)}(h)$ and $Z_{\tilde{\psi}}^{F \cup E(v)}(h)$ in $\mathbb{C}$ is at most θ.

Then for all $v \in N(u) \cup \{u\}$ and all $h \in \prod_{v \in V} S_v(\delta,\eta)$, we have

$$|Z_{\phi}^{F}(h)| \geq \cos(\theta/2) \sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \phi}} |h_v(\alpha)| \cdot \left| \frac{\partial}{\partial h_v(\alpha)} Z_{\phi}^{F}(h) \right|.$$

Proof. The conclusion is trivially true if $v = u$, since by the assumption $E(u) \subseteq F$, there is only one $\alpha : E(v) \to [q]$ compatible with ϕ, namely $\phi \mid_{E(u)}$ itself. In this case, $h_v(\alpha)$

divides $Z_\phi^F(h)$ and we can replace $\cos(\theta/2)$ by 1.

Suppose $v \in N(u)$. Since $Z_\phi^F(h) = \sum_{\substack{\psi:F\cup E(v)\to[q]\\ \psi|_F=\phi}} Z_\psi^{F\cup E(v)}(h)$, assumptions (1) and (2) make Lemma 8.6.6 applicable, yielding

$$\begin{aligned}|Z_\phi^F(h)| &\geq \cos(\theta/2) \sum_{\substack{\psi:F\cup E(v)\to[q]\\ \psi|_F=\phi}} |Z_\psi^{F\cup E(v)}(h)| \\ &= \cos(\theta/2) \sum_{\substack{\alpha:E(v)\to[q]\\ \text{compatible with } \psi}} |h_v(\alpha)| \cdot \left|\frac{\partial}{\partial h_v(\alpha)} Z_\phi^F(h)\right|\end{aligned}$$

as desired. □

With these lemmas in hand, we may now proceed with the proof of Theorem 8.6.9.

Proof of Theorem 8.6.9. Let $\eta > 0$ and $0 \leq \theta < 2\pi/3$ be arbitrary, and take $\tau = \cos(\theta/2)$, $\delta = \eta \cdot \min\left\{1, \frac{\theta\tau}{\Delta+1}\right\}$. We show by descending induction on $|F|$ that the following three statements are all true:

(i) For every $F \subseteq E$, $\phi : F \to [q]$ and $h \in \prod_{v\in V} S_v(\delta, \eta)$, we have $Z_\phi^F(h) \neq 0$.

(ii) For every $F \subseteq E$, $u \in V$, $\phi : F \to [q]$, $h \in \prod_{v\in V} S_v(\delta, \eta)$ and $\psi, \tilde{\psi} : F\cup E(u) \to [q]$ extending ϕ, the angle between $Z_\psi^{F\cup E(u)}(h)$ and $Z_{\tilde{\psi}}^{F\cup E(u)}(h)$ in $\mathbb{C}$ is at most θ.

(iii) For every $F \subseteq E$, $u \in V$ satisfying $E(u) \subseteq F$, $\phi : F \to [q]$, $h \in \prod_{v\in V} S_v(\delta, \eta)$ and $v \in N(u) \cup \{u\}$, we have the inequality

$$|Z_\phi^F(h)| \geq \cos(\theta/2) \sum_{\substack{\alpha:E(v)\to[q]\\ \text{compatible with } \phi}} |h_v(\alpha)| \cdot \left|\frac{\partial}{\partial h_v(\alpha)} Z_\phi^F(h)\right|.$$

The base case $F = E$ is easily verified since $Z_\phi^F(h) = \prod_{v\in V} h_v(\phi\ |_{E(v)})$, a product of nonzero complex numbers.

Now, let $F \subseteq E$ with $|F| < |E|$.

(i) Let $v \in V$ with $E(v) \not\subseteq F$. Since $|F \cup E(v)| > |F|$, (i) holds for $F \cup E(v)$ by the inductive hypothesis. Since $Z_\phi^F(h) = \sum_{\substack{\psi: F \cup E(v) \to [q] \\ \psi|_F = \phi}} Z_\psi^{F \cup E(v)}(h)$, Lemma 8.6.6 applied to $Z_\psi^{F \cup E(v)}(h)$ yields (i) assuming that (ii) holds. We prove (ii) below.

(ii) Let $u \in V$ and $\phi : F \to [q]$. If $E(u) \subseteq F$, then the claim is trivially true since $\psi = \tilde{\psi} = \phi$. Otherwise, assume $E(u) \not\subseteq F$ and let $\psi, \tilde{\psi} : F \cup E(u) \to [q]$ extend ϕ. Since $|F \cup E(u)| > |F|$, (i) and (iii) hold for $F \cup E(u)$ by the inductive hypothesis. Applying Lemma 8.6.10 to $F \cup E(u)$ then yields (ii).

(iii) Let $u \in V$ with $E(u) \subseteq F$. Without loss of generality, we may assume such an u exists since otherwise, there is nothing to prove. Let $v \in N(u) \cup \{u\}$. If $E(v) \subseteq F$, then (iii) trivially holds with $\cos(\theta/2)$ replaced by 1, since there is only one term in the summation, namely $\alpha = \phi \mid_{E(v)}$. Hence, assume $E(v) \not\subseteq F$. In this case, $|F \cup E(v)| > |F|$ and so (i) holds for $F \cup E(v)$ by the inductive hypothesis. Since (ii) for F holds (as proved earlier), we may then apply Lemma 8.6.11, yielding (iii) for F. $\square$

CHAPTER 9
CONCLUSION

In this thesis we establish optimal mixing time of the Glauber dynamics for many classes of spin systems, including Ising model, hardcore model, colorings, monomer-dimer model, and weighted even subgraphs. The key concept for proving our results is the notion of spectral independence. We show that for spin systems, if the Gibbs distribution is spectrally independent and the model is defined on bounded-degree graphs with constant marginal probabilities, then the Glauber dynamics mixes in $Cn \log n$ steps where n is the number of vertices of the underlying graph and C is a constant depending on the maximum degree and parameters of the model. We further show how to establish spectral independence using current algorithmic approach of approximate counting and sampling, including coupling arguments for MCMC, correlation decay methods, and polynomial interpolation methods.

A natural open direction is to improve the constant C with better dependencies on all parameters. In particular, our current approach does not work when the underlying graph has unbounded maximum degree, e.g., the complete graph. It is an interesting open problem to establish optimal mixing of Glauber dynamics for spin systems defined on dense graphs.

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, pp. 1198–1211.

[2] Alen Alexanderian. "On continuous dependence of roots of polynomials on coefficients". Technical notes. 2013.

[3] 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 Symposium on Theory of Computing (STOC)*. 2021.

[4] 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 Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2020, pp. 1319–1330.

[5] 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 (STOC)*. 2019, pp. 1–12.

[6] Itai Arad and Zeph Landau. "Quantum computation and the evaluation of tensor networks". In: *SIAM Journal on Computing* 39.7 (2010), pp. 3089–3121.

[7] Taro Asano. "Theorems on the partition functions of the Heisenberg ferromagnets". In: *Journal of the Physical Society of Japan* 29.2 (1970), pp. 350–359.

[8] Franck Barthe, Dario Cordero-Erausquin, Michel Ledoux, and Bernard Maurey. "Correlation and Brascamp-Lieb inequalities for Markov semigroups". In: *International Mathematics Research Notices* 10 (2011), pp. 2177–2216.

[9] Alexander Barvinok. "Computing the partition function for cliques in a graph". In: *Theory of Computing* 11.13 (2015), pp. 339–355.

[10] Alexander Barvinok. *Combinatorics and Complexity of Partition Functions*. Vol. 30. Springer Algorithms and Combinatorics, 2016.

[11] Alexander Barvinok. "Computing the permanent of (some) complex matrices". In: *Foundations of Computational Mathematics* 16.2 (2016), pp. 329–342.

[12] Alexander Barvinok. “Approximating permanents and hafnians”. In: *Discrete Analysis* 2 (2017).

[13] Alexander Barvinok. “Computing the partition function of a polynomial on the Boolean cube”. In: *A Journey Through Discrete Mathematics* (2017), pp. 135–164.

[14] Alexander Barvinok and Nicholas Barvinok. “More on zeros and approximation of the Ising partition function”. In: *ArXiv preprint, arXiv:2005.11232* (2021).

[15] Alexander Barvinok and Anthony Della Pella. “Testing for dense subsets in a graph via the partition function”. In: *SIAM Journal on Discrete Mathematics* 34 (1 2020), pp. 308–327.

[16] Alexander Barvinok and Pablo Soberón. “Computing the partition function for graph homomorphisms with multiplicities”. In: *Journal of Combinatorial Theory, Series A* 137 (2016), pp. 1–26.

[17] Alexander Barvinok and Pablo Soberón. “Computing the partition function for graph homomorphisms”. In: *Combinatorica* 37 (2017), pp. 633–650.

[18] 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.

[19] Mohsen Bayati, David Gamarnik, Dimitriy Katz, Chandra Nair, and Prasad Tetali. “Simple deterministic approximation algorithms for counting matchings”. In: *Proceedings of the 39th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*. 2007, pp. 122–127.

[20] Ferenc Bencs. “On trees with real-rooted independence polynomial”. In: *Discrete Mathematics* 341.12 (2018), pp. 3321–3330.

[21] Ferenc Bencs, Péter Csikvári, and Guus Regts. “Some applications of Wagner’s weighted subgraph counting polynomial”. In: *ArXiv preprint, arXiv:2012.00806* (2020).

[22] Ivona Bezáková and William A. Rummler. “Sampling edge covers in 3-regular graphs”. In: *Proceedings of the 34th International Symposium on Mathematical Foundations of Computer Science (MFCS)*. 2009, pp. 137–148.

[23] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. “On mixing of Markov chains: coupling, spectral independence, and entropy factorization”. In: *ArXiv preprint, arXiv:2103.07459* (2021).

[24] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. "Entropy decay in the Swendsen-Wang dynamics". In: *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*. 2021.

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

[26] Julius Borcea and Petter Brändén. "The Lee-Yang and Pólya-Schur programs. I. linear operators preserving stability". In: *Inventiones mathematicae* 177.3 (2009), p. 541.

[27] Magnus Bordewich, Catherine Greenhill, and Viresh Patel. "Mixing of the Glauber dynamics for the ferromagnetic Potts model". In: *Random Structures & Algorithms* 48.1 (2016), pp. 21–52.

[28] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. "Left and right convergence of graphs with bounded degree". In: *Random Structures & Algorithms* 42.1 (2013), pp. 1–28.

[29] Guy Bresler and Dheeraj Nagaraj. "Stein's method for stationary distributions of Markov chains and application to Ising models". In: *The Annals of Applied Probability* 29.5 (2019), pp. 3230–3265.

[30] Russ Bubley and Martin E. Dyer. "Path coupling: a technique for proving rapid mixing in Markov chains". In: *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 1997, pp. 223–231.

[31] Jin-Yi Cai, Sangxia Huang, and Pinyan Lu. "From Holant to #CSP and back: dichotomy for Holant^c problems". In: *Algorithmica* 64.3 (2012), pp. 511–533.

[32] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. "Holant problems and counting CSP". In: *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*. 2009, pp. 715–724.

[33] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. "Computational complexity of Holant problems". In: *SIAM Journal on Computing* 40 (4 2011), pp. 1101–1132.

[34] Pietro Caputo. "Spectral gap inequalities in product spaces with conservation laws". In: *Stochastic analysis on large scale interacting systems* 39 (2004), pp. 53–88.

[35] Pietro Caputo and Fabio Martinelli. "Relaxation time of anisotropic simple exclusion processes and quantum Heisenberg models". In: *The Annals of Applied Probability* 13.2 (2003), pp. 691–721.

[36] Pietro Caputo, Georg Menz, and Prasad Tetali. “Approximate tensorization of entropy at high temperature”. In: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 24. 4. 2015, pp. 691–716.

[37] Pietro Caputo and Daniel Parisi. “Block factorization of the relative entropy via spatial mixing”. In: *ArXiv preprint, arXiv:2004.10574* (2020).

[38] Eric A. Carlen, Maria C. Carvalho, and Michael Loss. “Determination of the spectral gap for Kac’s master equation and related stochastic evolution”. In: *Acta Mathematica* 191 (2003), pp. 1–54.

[39] Eric A. Carlen and Dario Cordero-Erausquin. “Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities”. In: *Geometric and Functional Analysis* 19 (2009), pp. 373–405.

[40] Eric A. Carlen, Elliott H. Lieb, and Michael Loss. “A sharp analog of Young’s inequality on S^N and related entropy inequalities”. In: *The Journal of Geometric Analysis* 14 (2004), pp. 487–520.

[41] Filippo Cesi. “Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields”. In: *Probability Theory and Related Fields* 120.4 (2001), pp. 569–584.

[42] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. “Improved bounds for randomly sampling colorings via linear programming”. In: *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2019, pp. 2216–2234.

[43] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. “Rapid mixing for colorings via spectral independence”. In: *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2021, pp. 1548–1557.

[44] Zongchen Chen, Kuikui Liu, and Eric Vigoda. “Rapid mixing of Glauber dynamics up to uniqueness via contraction”. In: *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2020, pp. 1307–1318.

[45] 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 Symposium on Theory of Computing (STOC)*. 2021.

[46] Zongchen Chen, Kuikui Liu, and Eric Vigoda. “Spectral independence via stability and applications to Holant-type problems”. In: *ArXiv preprint, arXiv:2106.03366* (2021).

[47] Mary Cryan, Heng Guo, and Giorgos Mousa. “Modified log-Sobolev inequalities for strongly log-concave distributions”. In: *The Annals of Probability* 49.1 (2021), pp. 506–525.

[48] Persi Diaconis and Laurent Saloff-Coste. “Logarithmic Sobolev inequalities for finite Markov chains”. In: *The Annals of Applied Probability* 6.3 (1996), pp. 695–750.

[49] Irit Dinur and Tali Kaufman. “High dimensional expanders imply agreement expanders”. In: *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2017, pp. 974–985.

[50] John Duchi. “Information theory and statistics”. In: *Lecture Notes for Statistics 311 and Electrical Engineering 377* (2019), p. 304.

[51] Martin Dyer, Alan Frieze, Thomas P. Hayes, and Eric Vigoda. “Randomly coloring constant degree graphs”. In: *Random Structures & Algorithms* 43.2 (2013), pp. 181–200.

[52] Martin Dyer, Leslie A. Goldberg, and Mark Jerrum. “Matrix norms and rapid mixing for spin systems”. In: *The Annals of Applied Probability* 19.1 (2009), pp. 71–107.

[53] Martin Dyer, Marc Heinrich, Mark Jerrum, and Haiko Müller. “Polynomial-time approximation algorithms for the antiferromagnetic Ising model on line graphs”. In: *Combinatorics, Probability and Computing* (2021), pp. 1–17.

[54] Ronen Eldan, James R. Lee, and Joseph Lehec. “Transport-entropy inequalities and curvature in discrete-space Markov chains”. In: *A Journey Through Discrete Mathematics*. 2017, pp. 391–406.

[55] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. “Rapid mixing from spectral independence beyond the Boolean domain”. In: *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2021, pp. 1558–1577.

[56] Cornelius Marius Fortuin and Piet W. Kasteleyn. “On the random-cluster model I. introduction and relation to other models”. In: *Physica* 57.4 (1972), pp. 536–564.

[57] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. “Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region”. In: *Journal of the ACM* 62.6 (2015), pp. 1–60.

[58] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models". In: *Combinatorics, Probability and Computing* 25.4 (2016), pp. 500–559.

[59] David Gamarnik, Dmitriy Katz, and Sidhant Misra. "Strong spatial mixing of list coloring of graphs". In: *Random Structures & Algorithms* 46.4 (2015), pp. 599–613.

[60] Christopher David Godsil. *Algebraic Combinatorics*. 1st ed. Chapman & Hall, Inc., 1993.

[61] Leslie A. Goldberg, Russell Martin, and Mike Paterson. "Strong spatial mixing with fewer colors for lattice graphs". In: *SIAM Journal on Computing* 35.2 (2005), pp. 486–517.

[62] Friedrich Götze, Holger Sambale, and Arthur Sinulis. "Higher order concentration for functions of weakly dependent random variables". In: *Electronic Journal of Probability* 24 (2019).

[63] Geoffrey Grimmett. *The Random-Cluster Model*. Vol. 333. Springer-Verlag, Berlin, 2006.

[64] Geoffrey Grimmett and Svante Janson. "Random even graphs". In: *The Electronic Journal of Combinatorics* 16.1 (2009), R46.

[65] Alice Guionnet and Bogusław Zegarlinksi. "Lectures on logarithmic sobolev inequalities". In: *Séminaire de Probabilités, XXXVI, volume 1801 of Lecture Notes in Math.* Springer, Berlin, 2003, pp. 1–134.

[66] Heng Guo, Chao Liao, Pinyan Lu, and Chihao Zhang. "Zeros of Holant problems: locations and algorithms". In: *ACM Transactions on Algorithms* 17.1 (2021), pp. 1–25.

[67] Heng Guo and Pinyan Lu. "Uniqueness, spatial mixing, and approximation for ferromagnetic 2-spin systems". In: *ACM Transactions on Computation Theory* 10.4 (2018), pp. 1–25.

[68] Heng Guo and Giorgos Mousa. "Local-to-global contraction in simplicial complexes". In: *ArXiv preprint, arXiv:2012.14317* (2020).

[69] Olle Häggström. "The random-cluster model on a homogeneous tree". In: *Probability Theory and Related Fields* 104.2 (1996), pp. 231–253.

[70] Thomas P. Hayes. "Local uniformity properties for Glauber dynamics on graph colorings". In: *Random Structures & Algorithms* 43.2 (2013), pp. 139–180.

[71] Thomas P. Hayes and Alistair Sinclair. “A general lower bound for mixing of single-site dynamics on graphs”. In: *The Annals of Applied Probability* 17.3 (2007), pp. 931–952.

[72] Jonathan Hermon and Justin Salez. “Modified log-Sobolev inequalities for strong-Rayleigh measures”. In: *ArXiv preprint, arXiv:1902.02775* (2019).

[73] Lingxiao Huang, Pinyan Lu, and Chihao Zhang. “Canonical paths for MCMC: from art to science”. In: *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2016, pp. 514–527.

[74] Mark Jerrum. “A very simple algorithm for estimating the number of k-colorings of a low-degree graph”. In: *Random Structures & Algorithms* 7.2 (1995), pp. 157–165.

[75] Mark Jerrum. *Counting, Sampling and Integrating: Algorithms and Complexity*. Lectures in Mathematics, ETH Zürich. Birkhäuser Basel, 2003.

[76] Mark Jerrum and Alistair Sinclair. “Approximating the permanent”. In: *SIAM Journal on Computing* 18.6 (1989), pp. 1149–1178.

[77] Mark Jerrum and Alistair Sinclair. “Polynomial-time approximation algorithms for the Ising model”. In: *SIAM Journal on Computing* 22.5 (1993), pp. 1087–1116.

[78] Mark Jerrum, Jung-Bae Son, Prasad Tetali, and Eric Vigoda. “Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains”. In: *The Annals of Applied Probability* 14.4 (2004), pp. 1741–1765.

[79] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. “Random generation of combinatorial structures from a uniform distribution”. In: *Theoretical Computer Science* 43 (1986), pp. 169–188.

[80] Tali Kaufman and David Mass. “High dimensional random walks and colorful expansion”. In: *Proceedings of the 8th Innovations in Theoretical Computer Science Conference (ITCS)*. 2017, 4:1–4:27.

[81] Tali Kaufman and David Mass. “Local-to-global agreement expansion via the variance method”. In: *Proceedings of the 11th Innovations in Theoretical Computer Science Conference (ITCS)*. 2020, 74:1–74:14.

[82] Tali Kaufman and Izhar Oppenheim. “High order random walks: beyond spectral gap”. In: *Combinatorica* 40.2 (2020), pp. 245–281.

[83] Vladimir Kolmogorov. “A faster approximation algorithm for the Gibbs partition function”. In: *Proceedings of the 31st Annual Conference on Learning Theory (COLT)*. Vol. 75. 2018, pp. 228–249.

[84] Jacob Korevaar and Jan Wiegerinck. *Several complex variables*. Korteweg-de Vries Institute for Mathematics, 2017.

[85] Michel Ledoux. “Concentration of measure and logarithmic Sobolev inequalities”. In: *Séminaire de Probabilités XXXIII* (1999), pp. 120–216.

[86] Tsung-Dao Lee and Chen-Ning Yang. “Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model”. In: *Physical Review* 87.3 (1952), pp. 410–419.

[87] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, 2017.

[88] Liang Li, Pinyan Lu, and Yitong Yin. “Approximate counting via correlation decay in spin systems”. In: *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2012, pp. 922–940.

[89] Liang Li, Pinyan Lu, and Yitong Yin. “Correlation decay up to uniqueness in spin systems”. In: *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2013, pp. 67–84. The latest version is at arXiv:1111.7064.

[90] Chengyu Lin, Jingcheng Liu, and Pinyan Lu. “A simple FPTAS for counting edge covers”. In: *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2014, pp. 341–348.

[91] Jingcheng Liu, Pinyan Lu, and Chihao Zhang. “FPTAS for counting weighted edge covers”. In: *Proceedings of the 22nd Annual European Symposium on Algorithms (ESA)*. 2014, pp. 654–665.

[92] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. “A deterministic algorithm for counting colorings with 2Δ colors”. In: *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2019, pp. 1380–1404.

[93] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. “Fisher zeros and correlation decay in the Ising model”. In: *Journal of Mathematical Physics* 60.10 (2019), p. 103304.

[94] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. “The Ising partition function: zeros and deterministic approximation”. In: *Journal of Statistical Physics* 174.2 (2019), pp. 287–315.

[95] Neal Madras and Dana Randall. “Markov chain decomposition for convergence rate analysis”. In: *The Annals of Applied Probability* 12.2 (2002), pp. 581–606.

[96] Igor L. Markov and Yaoyun Shi. “Simulating quantum computation by contracting tensor networks”. In: *SIAM Journal on Computing* 38 (3 2008), pp. 963–981.

[97] Russell A. Martin and Dana Randall. “Sampling adsorbing staircase walks using a new markov chain decomposition method”. In: *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2000, pp. 492–502.

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

[99] Fabio Martinelli and Enzo Olivieri. “Approach to equilibrium of Glauber dynamics in the one phase region. II. the general case”. In: *Communications in Mathematical Physics* 161.3 (1994), pp. 487–514.

[100] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. “The Ising model on trees: boundary conditions and mixing time”. In: *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2003, pp. 628–639.

[101] Katalin Marton. “Logarithmic Sobolev inequalities in discrete product spaces”. In: *Combinatorics, Probability & Computing* 28.6 (2019), pp. 919–935.

[102] Colin McQuillan. “Approximating Holant problems by winding”. In: *ArXiv preprint, arXiv:1301.2880* (2013).

[103] Carl D. Meyer. *Matrix analysis and applied linear algebra*. Vol. 71. SIAM, 2000.

[104] Laurent Miclo. “Remarques sur l’hypercontractivité et l’évolution de l’entropie pour des chaînes de Markov finies”. In: *Séminaire de Probabilités XXXI*. Springer Berlin Heidelberg, 1997, pp. 136–167.

[105] Ravi Montenegro and Prasad Tetali. “Mathematical aspects of mixing times in Markov chains”. In: *Foundations and Trends in Theoretical Computer Science* 1.3 (2006), pp. 236–236.

[106] 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.

[107] Yann Ollivier. “Ricci curvature of Markov chains on metric spaces”. In: *Journal of Functional Analysis* 256.3 (2009), pp. 810–864.

[108] Izhar Oppenheim. “Local spectral expansion approach to high dimensional expanders part I: descent of spectral gaps”. In: *Discrete and Computational Geometry* 59.2 (2018), pp. 293–330.

[109] Román Orús. “A practical introduction to tensor networks: matrix product states and projected entangled pair states”. In: *Annals of Physics* 349 (2014), pp. 117–158.

[110] Viresh Patel and Guus Regts. “Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials”. In: *SIAM Journal on Computing* 46.6 (2017), pp. 1893–1919.

[111] Yuval Peres and Peter Winkler. “Can extra updates delay mixing?” In: *Communications in Mathematical Physics* 323 (2013), pp. 1007–1016.

[112] Guus Regts. “Zero-free regions of partition functions with applications to algorithms and graph limits”. In: *Combinatorica* 38 (2018), pp. 987–1015.

[113] Gesine Reinert and Nathan Ross. “Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs”. In: *The Annals of Applied Probability* 29.5 (2019), pp. 3201–3229.

[114] David Ruelle. “Extension of the Lee-Yang circle theorem”. In: *Physical Review Letters* 26.6 (1971), pp. 303–304.

[115] Laurent Saloff-Coste. “Lectures on finite Markov chains”. In: *Lectures on Probability Theory and Statistics, volume 1665 of Lecture Notes in Math.* Springer, Berlin, 1997, pp. 301–413.

[116] Holger Sambale and Arthur Sinulis. “Logarithmic Sobolev inequalities for finite spin systems and applications”. In: *Bernoulli* 26.3 (2020), pp. 1863–1890.

[117] Alexander D. Scott and Alan D. Sokal. “The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma”. In: *Journal of Statistical Physics* 118.5 (2005), pp. 1151–1261.

[118] Shuai Shao and Yuxin Sun. “Contraction: a unified perspective of correlation decay and zero-freeness of 2-spin systems”. In: *Proceedings of the 47th International Colloquium on Automata, Languages, and Programming (ICALP)*. Vol. 168. 2020, 96:1–15.

[119] 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.

[120] Allan Sly. “Computational transition at the uniqueness threshold”. In: *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2010, pp. 287–296.

[121] Allan Sly and Nike Sun. “The computational hardness of counting in two-spin models on d-regular graphs”. In: *The Annals of Probability* 42.6 (2014), pp. 2383–2416.

[122] Daniel Štefankovič, Santosh Vempala, and Eric Vigoda. “Adaptive simulated annealing: a near-optimal connection between sampling and counting”. In: *Journal of the ACM* 56.3 (2009), pp. 1–36.

[123] Daniel W. Stroock and Boguslaw Zegarlinski. “The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition”. In: *Communications in Mathematical Physics* 144.2 (1992), pp. 303–323.

[124] Mario Ullrich. “Rapid mixing of Swendsen-Wang dynamics in two dimensions”. In: *Dissertationes Mathematicae* 502 (2014), pp. 1–65.

[125] Eric Vigoda. “Improved bounds for sampling colorings”. In: *Journal of Mathematical Physics* 41.3 (2000), pp. 1555–1569.

[126] David G. Wagner. “Weighted enumeration of spanning subgraphs with degree constraints”. In: *Journal of Combinatorial Theory, Series B* 99.2 (2009), pp. 347–357.

[127] Dror Weitz. “Counting independent sets up to the tree threshold”. In: *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*. 2006, pp. 140–149.

VITA

Zongchen Chen received the Ph.D. degree in Algorithms, Combinatorics, and Optimization from the School of Computer Science at Georgia Institute of Technology in 2021. He was fortunate to be advised by Eric Vigoda. Before that, he received the Bachelor's degree in Math and Applied Math from Zhiyuan College at Shanghai Jiao Tong University in 2016. He has broad interests in randomized algorithms, discrete probability, and machine learning. Currently, he works on Markov chain Monte Carlo (MCMC) methods for sampling from Gibbs distributions, and machine learning problems related to graphical models.