Spectral Independence

A New Tool to Analyze Markov Chains

Kuikui Liu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Shayan Oveis Gharan (Chair)

Anna Karlin

Cynthia Vinzant

Kevin Jamieson

Program Authorized to Offer Degree:

Paul G. Allen School of Computer Science & Engineering

University of Washington

Abstract

Spectral Independence: A New Tool to Analyze Markov Chains

Kuikui Liu

Chair of the Supervisory Committee:

Shayan Oveis Gharan

Paul G. Allen School of Computer Science & Engineering

We introduce a versatile technique called *spectral independence* for the analysis of Markov chain Monte Carlo algorithms in high-dimensional probability and statistics. We rigorously prove rapid mixing of practically usefully Markov chains for sampling from important classes of probability distributions arising in computer science, statistical physics, and pure mathematics, thus resolving several longstanding conjectures and open problems. In many cases, we obtain asymptotically *optimal* mixing time bounds. To achieve these results, we establish new *local-to-global phenomena* which translate spectral independence into mixing time bounds. Furthermore, we develop four distinct classes of techniques for establishing spectral independence by building new bridges with other fields.

To my family and friends.

Acknowledgements

There are way too many people to which I owe thanks. Nonetheless, I will do the utmost to include as many as possible here.

First and foremost, I would like to express my deepest gratitude to my doctoral advisor, Shayan Oveis Gharan, who introduced me to mathematical research as an undergraduate in the autumn of 2016. Shayan tirelessly guided me through all aspects of the research process, from formulating and solving a problem, to presenting ideas in a paper or talk. Somehow, Shayan always knew the right thing to say at the right time. I really appreciated his honesty, delivering critical feedback when I needed it. At numerous times over the course of my Ph.D., I felt "I wasn't going to make it". Many projects failed miserably, especially in the beginning. But Shayan stuck with me through it all. I am incredibly blessed to have had his mentorship.

I am also indebted to Anna Karlin, who I personally think of as my (informal) second advisor. She was someone I always felt comfortable going to for help and advice. She always gave incredibly insightful constructive feedback on my talks, writing, and teaching. I also had the great fortune of taking several advanced classes from her, serving as her teaching assistant, and participating in her research reading groups. Shayan and Anna were my guiding lights.

I would like to thank the wonderful people I had the privilege to collaborate with over the past few years. In particular, I would like to thank Dorna Abdolazimi, Nima Anari, Zongchen Chen, Eric Vigoda, Cynthia Vinzant, and Thuy-Duong (June) Vuong who I was lucky to coauthor papers with, and who I continue to work closely with. Equally, I also thank Aaron Archer, Farzam Ebrahimnejad, Ronen Eldan, Matthew Fahrbach, Alkis Gotovos, Fotis Iliopoulos, Jun-Ting (Tim) Hsieh, Sidhanth Mohanty, Swati Padmanabhan, Alireza Rezaei, Amin Saberi and Nikhil Srivastava for collaborations from which I learned so much, even if they didn't result in or haven't yet resulted in publication. I had so much fun working on various projects with each of these incredible researchers.

I was blessed to have Irit Dinur host me at the Institute for Advanced Study in winter of 2020, from who I learned a great deal about high-dimensional expanders. During that visit, I also had the privilege to speak with and learn from June Huh, Allan Sly, and Avi Wigderson, to whom I am also grateful. Thank you to Aaron Archer, Matthew Fahrbach, and Vahab Mirrokni, who gave me the opportunity to intern at Google Research in spring of 2022, and from who I learned what algorithms research in industry is like. I am particularly grateful for Matthew Fahrbach's mentorship at Google Research. His enthusiasm was infectious.

I also wish to thank Patrícia Alves-Oliveira, Paul Beame, Maya Cakmak, Karthik Desingh, Yuan (Charles) Gao, Kira Goldner, Kevin Jamieson, Sandy Kaplan, Yin-Tat Lee, James R. Lee, Anup Rao, Noah Smith, Thomas Rothvoss, Alistair Sinclair, Austin Stromme, Dan Suciu, Rekha Thomas, Alexander Tsun, and Avi Wigderson, for providing invaluable constructive feedback on my talks, research statements, and other writings. I thank Vedat Levi Alev, Pietro Caputo, Alan Frieze, Mark Jerrum, Tali Kaufman, Lap Chi Lau, Gary L. Miller, Justin Salez, Daniel Sleator, Yitong Yin, Pierre Youssef, and Chihao Zhang for insightful discussions, helpful comments on the papers I have coauthored, and for spotting bugs. I also thank the dedicated anonymous reviewers of the papers I have coauthored. Finally, many thanks to my committee members Kevin Jamieson, Anna Karlin, Shayan Oveis Gharan, and Cynthia Vinzant, for providing me extremely useful feedback on my research and on this thesis.

I am immensely grateful to the good friends I met along the way, from who I learned so much. I wish to thank Charlie Carlson, Anthony Della Pella, Siqi Liu, Xinyu Wu, and Elizabeth Yang, who I had the pleasure of hanging out with at the Simons Institute program on the Geometry of Polynomials in early 2018. I also thank Sami Davies, Sally Dong, Farzam Ebrahimnejad, Ashrujit Ghoshal, Kira Goldner, Siddharth Iyer, Haotian Jiang, Nathan Klein, Vincent Liew, Swati Pad-

manabhan, Siva Ramamoorthy, Victor Reis, Jaden Stock, and Robbie Weber from the UW CSE theory group, for introducing me to new ideas in other research areas, and also for their encouragement. More generally, I would also like to thank my officemates, the UW CSE theory group, and the broader UW CSE community for fostering an open and welcoming environment for me to explore and grow in.

Special thanks to Daniel Štefankovič, Prasad Tetali, and Eric Vigoda, for organizing a wonderful summer school in August 2022 on "New tools for optimal mixing of Markov chains: Spectral independence and entropy decay". The beautiful campus of the University of California, Santa Barbara was incredible. I am also grateful to the summer school participants, with whom I exchanged many simulating discussions.

I would like to thank our department's tennis group, and in particular, Elizabeth Clark, Ian Covert, Liang Luo and Chien-Yu Lin, for regularly hitting with me. Our evening tennis sessions were often the highlight of my week, especially when research was progressing slowly.

Last but not least, I wish to thank my wife Jiayao (Clara) Lu, my parents Junyu Liu and Wei (Joyce) Xu, my sister Lily Liu, my wife's parents Wenxin Lu and Yinghong Du, as well as my friends, for their helpful advice, warmth, and never-ending support. Without them, this thesis wouldn't be possible.

I wish to thank the COVID-19 pandemic for taking away over two years of meeting family, friends, and colleagues.

Contents

Chapter 1

Introduction

Probability distributions are everywhere. Their study forms the backbone of the entirety of statistics, and they are widely used in all areas of science and engineering, from quantum mechanics to epidemiology. Central to their ubiquity is their versatility in modeling important phenomena encountered throughout nature and society.

In the information age, and especially in the modern era of "big data", the probability distributions we encounter are often incredibly *complex* and *high-dimensional*. We are typically interested not only in precisely understanding a single hydrogen atom or a single infected person in isolation, but also the behavior of massive collections of interacting particles or populations of people. It is self-evident that the probability distributions used to accurately capture these complicated situations are necessarily complex. They are also high-dimensional, in the sense that any point drawn from the distribution describes the state of each member in a huge collection. As a result, these distributions have exponentially large or even infinitely large domains; at the very least, the number of possible outcomes is so large that even with the world's fastest supercomputer, the amount of time required to enumerate all possibilities will vastly exceed the age of the universe. *Finding ways to efficiently understand, process, manipulate, and generally work with such distributions is one of the central challenges of modern statistics and computing.*

So how do we work with such probability distributions? One of the most fundamental algorithmic primitives used is *sampling*. By sampling, I mean running some randomized process/algorithm/experiment whose random outcome is distributed according to the desired probabilities. In other words, the chance that the process/algorithm/experiment outputs any particular possibility is exactly (or approximately) its probability under the distribution you're trying to sample from. One can imagine tossing an unbiased coin as a way of sampling a random "heads" or "tails", each with probability one-half. As another example, one can imagine (repeatedly) shuffling a deck of cards as a way of sampling a uniformly random ordering of the cards.

Sampling is one of the most heavily used approaches to tractably tackling problems involving complex and high-dimensional probability distributions. At its most basic level, it gives us a way to probe what typical or likely states of the distribution look like. Algorithms designed for sampling can also be used to *simulate* the evolution of complex physical systems. The famous *Monte Carlo method* critically employs sampling to efficiently estimate useful and otherwise seemingly impossible-to-compute statistics. It truly is a universal tool for both practitioners and theorists alike. Below is a list of just a few applications of sampling from complex high-dimensional probability distributions to other scientific disciplines. It is not even close to being exhaustive.

- **Quantum & Statistical Mechanics** Physicists have a long history of using probability theory to model the behavior of large collections of interacting particles (e.g. the ferromagnetic Ising model of a magnet [Len20; Isi25]). In the quantum world, everything is inherently probabilistic. The probability distributions here are typically over configurations of states of atoms or molecules [Bov06; FV17]. Physicists are then interested in simulating the evolution of these systems so that we can observe physical phenomena (e.g. phase transitions, low-energy states, etc.). Chemists and materials scientists are also interested in estimating physical quantities associated to such systems in order to understand their properties as materials. All of these problems are often tackled using sampling techniques [Bro+11].

- **Learning & Inference in Statistics and Machine Learning** Following the physics tradition, *probabilistic (graphical) models (e.g. Markov random fields, Bayesian networks,*

etc.) are commonly used by statisticians and machine learning scientists to model observed data, dependencies between variables, and causality [WJ08; KF09]. Fundamental tasks such as *learning* these models from data, *inferring latent properties* of the underlying distribution or population given some (partially) observed variables, and *marginalizing* out irrelevant parameters, are all often solved using sampling algorithms (perhaps combined with the Monte Carlo method) [KF09].

- **Privacy, Fairness & Diversity in Algorithm Design** One of the most popular approaches to endowing algorithms with mathematically rigorous privacy guarantees [DN03; DN04; Blu+05; Dwo+06] is to add *tailored noise* to the algorithm [MT07]. Of course, noise just refers to random samples from some probability distribution, and so sampling algorithms again play a key role. Similarly, in machine learning, it is often beneficial for performance and fairness reasons to ensure there is *diversity* in both the inputs and outputs of a learned model. Concrete examples where diversity is crucial include feature selection, text summarization, and search result aggregation. Sampling from probability distributions which explicitly encourage diversity is a flexible and popular method to achieve this [Hou+06; LB12; KT12; MSJ18; RSJ19].

- **Epidemiology, Social Sciences, and More** Like in the preceding examples, scientists studying the behavior of populations of people also must work with massive datasets, incorporating huge quantities of information such as demographic features, social connections, community structures, etc. Here, probabilistic models and sampling again provide invaluable insights into how societies operate. Examples include modeling social networks [EK10; LKR12], quantitatively studying to what extent special interest groups and contractors can influence government policy [GW13; Gil14], epidemiology [HMR13] and specifically modeling the spread of the COVID-19 pandemic [DH21], etc. See [CH14] and references therein for more applications to the social sciences.

Even areas that seemingly have little to do with sampling or probability have benefited tremendously from employing sampling algorithms. Here are two examples of whole fields of study stemming from computer science, applied mathematics, engineering, and operations research, on which sampling has had a profound impact.

- **Optimization** Perhaps surprisingly, there are actually intimate connections between sampling and optimization. For instance, to solve the following optimization problem for some function $f : \Omega \to \mathbb{R}$ on some domain Ω,

$$\min_{x \in \Omega} f(x)$$

one can sample from the probability distribution

$$\mu_{f,\beta}(x) \propto \exp(-\beta \cdot f(x))$$

for a large $\beta > 0$, and output the sample as an (approximate) solution. The key is that the global minima of f will have vastly higher probability under $\mu_{f,\beta}$ than other points, and so a random sample from $\mu_{f,\beta}$ will very likely be (close to) a global minimizer. Based on this or similar insights, many practically useful algorithms for optimization were devised which directly use sampling such as *simulated annealing* [Pin70; KSV79; KSV81; LA87; KGV83].

However, while sampling and optimization are closely related, I would also argue that sampling is *strictly harder* than optimization in a certain (informal) sense. In optimization, since you have an objective function, in principle you can evaluate the quality of your algorithm's output, and even compare different algorithms based on how good of an objective value you get. In sampling, the algorithm just spits out some member of the underlying domain. It isn't clear at all if it was output with the correct probability, especially if the distribution is complicated and high-dimensional. Furthermore, at least at the time of this writing, the complexity theory of sampling problems is arguably less developed than the more traditional complexity theory of optimization problems.

- **Approximation Algorithms for NP-Hard Problems** In a similar vein to optimization, sampling has also played an important role in developing efficient approximation algorithms for classical NP-hard problem. An extremely useful recipe for designing algorithms for hard

combinatorial optimization problems (e.g. the traveling salesperson problem, MAXCUT, etc.) is to first solve an easier *continuous relaxation* of the problem (typically, a linear, semidefinite, or convex program), and then *round* the resulting *fractional solution* into a bonafide solution of the original problem. In many settings, this second step is done using sampling, taking into account the constraints of the problem as well as the fractional solution found in the first step. This is exactly what the famous *randomized rounding* technique [RT87] and its many variants do [GW95; AS04; ARV09; CVZ10; Asa+10; OSS11; Rot12; HO14]. See [KKO21] for a recent major advance in approximation algorithms employing such a scheme, and [Vaz03; WS11; SV14; Ove14] for further discussion.

Given the abundant applications, the practical question then becomes: How does one efficiently sample from such complex and high-dimensional probability distributions? Since the distributions are supported on such large domains, sampling efficiently without sacrificing fidelity is an incredibly challenging computational problem. On the theory side, we'd like to understand what exactly makes a high-dimensional problem easy or hard. *We seek simple, unifying principles which rigorously explain the computational complexity of all high-dimensional sampling problems, and perhaps more ambitiously, all high-dimensional statistical problems.* These are the driving questions behind this thesis and my research thus far. It also has been the subject of intense research for over a century.

Sampling via Markov chains One of the most useful classes of algorithms deployed in practice is the class of *Markov chain* sampling algorithms. Combined with the Monte Carlo method, we get the class of *Markov chain Monte Carlo (MCMC)* algorithms. Since their invention, MCMC algorithms have revolutionized scientific computing. They are ubiquitous in practice and simple to implement, but often require deep and highly sophisticated mathematics to analyze [Dia09; Bro+11].

At a high level, the main idea behind a Markov chain is to run a random process where simple, easy-to-compute, random updates are repeatedly applied to some initial starting point. Pictorially, one can imagine the Markov chain as a "particle" or "agent" randomly walking around the domain of the probability distribution; it will tend to move towards and stay within regions of the domain which have higher probability mass. By now, there are numerous methods for designing useful Markov chains for sampling. Here, generic recipes like the *Metropolis-Hastings method* or *Metropolis filter* [Met+53; Has70] readily yield Markov chains whose *stationary (or equilibrium) distribution* is precisely the distribution you want to sample from. Other examples of well-known classes of Markov chains used heavily in practice include Gibbs sampling/Glauber dynamics [Gla63; GG84], Langevin dynamics [RDF78; GM94; RT96; RR98; BZ20], Hamiltonian Monte Carlo (HMC) [Dua+87; Bet18; BZ20], etc.

Once you have a Markov chain you want to use, the central problem becomes the following.

Question 1. *How many steps should we run the Markov chain to ensure an accurate sample?*

The answer is captured by the *mixing time*, a fundamental parameter of the Markov chain that is notoriously challenging to understand, even empirically via simulation; entire textbooks have been devoted to analyzing mixing times [MT06; LPW17] and theoretical analyses more broadly [MT93; Bro+11]. In the absence of strong theoretical bounds on the mixing time, the conventional wisdom is to just run the Markov chain as long as possible and hope for the best. Mathematically rigorous bounds on the mixing time are also crucial to ensure accuracy when estimating statistics.

The following quote aptly summarizes the prevalence of MCMC, as well as the dire need for more theory.

> "I believe you can take any area of science, from hard to social, and find a burgeoning MCMC literature specifically tailored to that area. I note that *essentially none* of these applications is accompanied by any kind of practically useful running time analysis."
>
> – Persi Diaconis *"The Markov Chain Monte Carlo Revolution"* [Dia09]

The focus of this thesis is to tackle this problem for a wide variety of high-dimensional probability distributions.[1] In particular, this thesis describes an extensive new *theoretical toolkit* for analyzing

[1] Of course, we can reasonably debate about what "practically useful" means. For us, we will shoot for mixing times which are asymptotically optimal w.r.t. input size, perhaps up to logarithmic factors. In general, we will also be content with mixing times which scale polynomially in the input size. We will not concern ourselves with obtaining optimal constants.

Markov chain mixing times, which was built around a new and unifying concept called *spectral independence* we introduced in [ALO21]. This tool is based on the following high-level philosophical statement.

> *If the probability distribution of interest μ satisfies a "limited" or "structured" correlations property, then a simple Markov chain for sampling from μ mixes rapidly.*

Here, "limited" or "structured" should be interpreted in the broadest possible sense. Spectral independence is one particular way to instantiate this philosophy. For concreteness, we give an informal definition here. For simplicity, we restrict our attention to distributions over the discrete hypercube $\{0,1\}^n$.

Definition 1 (Spectral Independence (Informal); [ALO21])**.** *Let μ be a probability distribution over $\{0,1\}^n$, and define the **influence matrix** $\Psi_\mu \in \mathbb{R}^{n\times n}$ by*

$$\Psi_\mu(i,j) \stackrel{\text{def}}{=} \Pr_{\sigma\sim\mu}[\sigma_j = 1 \mid \sigma_i = 1] - \Pr_{\sigma\sim\mu}[\sigma_j = 1 \mid \sigma_i = 0].$$

*For $\eta \geq 0$, we say μ is η-**spectrally independent** if $\lambda_{\max}(\Psi_\mu) \leq 1+\eta$. We say μ is $(\eta,\dots,\eta)$-spectrally independent if every distribution obtained from μ by conditioning on the assignments for some subset of coordinates is η-spectrally independent.*

One should think of Ψ_μ as being an asymmetric version of a *correlation matrix*, or a normalization of a *covariance matrix*. However, we prefer to call the entries of Ψ_μ *influences* because each entry $\Psi_\mu(i,j)$ exactly quantifies how much knowing the assignment for coordinate i "affects" or "influences" the marginal probability of coordinate j (being assigned, say, 1). Having an upper bound on the maximum eigenvalue of Ψ_μ is thus a way of ensuring that the "total amount of pairwise correlation" in the distribution is bounded. Having a bound on just the pairwise correlations for μ itself is too weak, so we impose the same bound for all conditional distributions.

The word "independence" in "spectral independence" comes from the fact that product measures, where the assignments of coordinates are mutually independent, are $(0,\dots,0)$-spectrally independent under Definition 1. In an informal sense, how large the spectral independence parameter η is measures how close the distribution μ is to being a product distribution. The word "spectral" just comes from the fact that we are measuring the eigenvalues of the influence matrix. This explains the name "spectral independence".

We will show in this thesis how to go from spectral independence to rapid mixing of simple Markov chains. We will also refine such *local-to-global theorems* in several important settings to obtain *optimal mixing times*. We will then develop several classes of techniques for establishing spectral independence. We hope that the applications we will encounter along the way convincingly illustrate the versatility of this framework.

1.1 A Sample of Results

Our toolbox has been successfully applied to a wide variety of complex probability distributions on discrete combinatorial structures encountered throughout statistical and condensed matter physics, computer science, and pure mathematics. In this thesis, we discuss several fundamental, decades-old open problems and conjectures on the polynomial-time mixing of well-known Markov chains that we resolved using spectral independence. Along the way, we will see how spectral independence can further yield not only polynomial-time mixing guarantees, but optimal, *nearly-linear*[2] mixing times, well-within the regime of being practically useful. For the moment, we give a brief and high-level overview of the results. Throughout this introduction, all theorems are stated informally for simplicity and convenience.

1.1.1 Discrete Convexity and Discrete Log-Concave Measures

In the continuous world (e.g. continuous space $\mathbb{R}^n$), there is an incredibly rich theory of *convexity* which permits efficient optimization and sampling (see e.g. [BV04; LV07]). For continuous probability distributions, convexity naturally arises in a beautiful class of probability distributions

[2]The term *nearly-linear* (in some parameter n) means a function scaling as $n\log^{O(1)} n$, sometimes written $n\cdot\mathsf{polylog}(n)$ or $\tilde{O}(n)$.

frequently encountered in machine learning and statistics known as *log-concave* distributions. For instance, as special cases, this class includes the multivariate Gaussian and multivariate exponential distributions. These distributions have origins in high-dimensional *convex geometry*, and enjoy many useful *concentration of measure* properties. On the algorithmic side, dozens of different MCMC algorithms have been proposed and proved to run in polynomial-time for sampling from these distributions. Given its rich history and vast literature, it is natural to wonder if there are any parallels in the discrete world. However, such an analogous discrete theory proved elusive.

It turns out that for discrete probability distributions, the "correct" analog comes from log-concavity of the *generating polynomial* of the distribution [Gur09; Gur10; AOV21; BH20]. However, even with the correct definition in hand, the question of efficiently sampling from such *discrete log-concave distributions* was wide open. By connecting the analytic theory of multivariate polynomials with the theory of high-dimensional expanders and spectral independence, we managed to give the first efficient algorithm to sample from any discrete log-concave distribution [Ana+19], along with a subsequent improvement in running time [Ana+21c] (building on a follow-up work of [CGM21]).[3]

Theorem 1.1.1 (Rapid Mixing for Discrete Log-Concave Measures (Informal); [Ana+21c] building on [Ana+19; CGM21])**.** *Let $\mu : \binom{\mathscr{U}}{r} \to \mathbb{R}_{\geq 0}$ be a probability distribution on a size-r subsets of a fixed finite ground set $\mathscr{U}$. If the associated multivariate generating polynomial*

$$g_\mu(x_u : u \in \mathscr{U}) = \sum_{S \in \binom{\mathscr{U}}{r}} \mu(S) \prod_{u \in S} x_u$$

is log-concave on the positive orthant $\mathbb{R}^{\mathscr{U}}_{\geq 0}$, then the "natural" Markov chain with stationary distribution μ mixes in $O(r \log r)$-steps.

Matroids

As a special case of this result, we positively resolved an important 30-year-old conjecture due to Mihail–Vazirani [MV89] on the mixing time of a simple Markov chain on bases of *matroids*. To state this conjecture and its resolution, let us first define what a matroid is.

Definition 2 (Matroid; Independent Set Definition)**.** *A **matroid** $\mathcal{M}$ is a pair $(\mathscr{U}, \mathcal{X})$, where $\mathscr{U}$ is a finite ground set, and $\mathcal{X} \subseteq 2^{\mathscr{U}}$ is a family of subsets of $\mathscr{U}$ satisfying the following properties:*

- ***Downwards Closure:** If $T \in \mathcal{X}$ and $S \subseteq T$, then $S \in \mathcal{X}$ as well.*[4]
- ***Exchange Property:** If $S, T \in \mathcal{X}$ and $|T| > |S|$, then there exists $u \in T \setminus S$ such that $S \cup \{u\} \in \mathcal{X}$.*

*The sets in $\mathcal{X}$ are called **independent sets**, and the maximal independent sets are called **bases**. It is well-known that all bases have the same cardinality.*[5] *This common cardinality is called the **rank** of the matroid, which can be thought of as a kind of "dimension" parameter.*

Matroids were initially introduced in the 1930s [Whi35] as a combinatorial abstraction of the idea of *linear independence* in linear algebra (hence, the name "independent sets"). They possess a number of remarkable properties. For instance, in the discrete world, convexity manifests in the form of matroids [Mur03]. They have been intensely studied for decades in combinatorial optimization, polyhedral and topological combinatorics, discrete mathematics, mathematical economics, mathematical logic, and more [Oxl11]. Prototypical examples include spanning forests in a graph, and subsets of linearly independent vectors in a vector space.

It has been known since the early 1970s how to efficiently optimize over matroids [Edm71], but the question of sampling from them plagued researchers for over 30 years, with numerous works solving only very special cases (e.g. sampling random spanning trees). Practical motivations for sampling uniformly random bases of matroids include being able to estimate the *reliability* of networks and error-correcting codes, as well as the rigidity of shapes and structures. Additional applications can be found in algorithmic game theory [Kle21; Bea+22] and in credit network liquidity [Goe+15].

[3] Efficient algorithms were previously known only for a very special subclass known as *strongly Rayleigh* distributions [AOR16].

[4] In the language of algebraic topology, $\mathcal{X}$ is an *abstract simplicial complex*.

[5] In the language of algebraic topology, the simplicial complex $\mathcal{X}$ is *pure*

For a matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$, one can define a simple and natural Markov chain on the bases of $\mathcal{M}$ using the defining exchange property as follows. Starting from an arbitrary basis $B \in \mathcal{X}$, the random transition to another basis $B' \in \mathcal{X}$ under the Markov chain is described by the following two-step procedure:

1. Remove a uniformly random element $u \in B$.
2. Out of all $v \notin B \setminus \{u\}$ such that $B \setminus \{u\} \cup \{v\}$ is a base, pick one uniformly at random and set $B' = B \setminus \{u\} \cup \{v\}$. Note that one can choose $v = u$, in which case $B' = B$.

This Markov chain is known as the *bases exchange walk*. It is clear that each step of this Markov chain can be implemented efficiently. Using the *detailed balance* condition, it is also straightforward to verify that the uniform distribution over bases of M is stationary with respect to the bases exchange walk.

In 1989, Mihail and Vazirani posed the following tantalizing conjecture.[6]

Conjecture 1 (Informal; [MV89])**.** *For every rank-r matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$, the underlying graph of the bases exchange walk has "edge-expansion" at least* 1*. In particular, the mixing time of the bases exchange walk is polynomial in $r, |\mathscr{U}|$.*

Since it was known previously that the multivariate generating polynomial g_μ associated to the uniform distribution μ over the bases of a matroid (known as the *bases generating polynomial*) is log-concave [AOV21], our result Theorem 1.1.1 on discrete log-concave distributions allowed us to completely resolve the 30-year-old problem of sampling bases of matroids. Our proof also completely resolves Conjecture 1 in the affirmative.

Theorem 1.1.2 (Informal; [Ana+19])**.** *Conjecture 1 is true.*

As a nice side consequence, the fruitful connections we developed also led to a completely elementary and self-contained proof of log-concavity of the bases generating polynomial of a matroid, which completely avoids sophisticated machinery such combinatorial Hodge theory [AHK18; HSW21] used in previous proofs [AOV21]. Finally, building on a follow-up work of Cryan–Guo–Mousa [CGM21], we further reduced the mixing time of the bases exchange walk to the optimal $O(r \log r)$, where r is the rank of the matroid, independent of the size of the underlying ground set [Ana+21c].

Applications in Algebraic Combinatorics

Beyond rapidly mixing Markov chains and sampling algorithms, the theory we developed has also led to the resolution of long-standing conjectures in algebraic combinatorics. In [Ana+18a], we further develop the theory of log-concave polynomials, building on [Gur10; AOV21]. As our flagship application, we managed to resolve the strongest version of Mason's 50-year-old conjecture [Mas72] on the ultra-log-concavity of the independence numbers of matroids [Ana+18a].[7]

1.1.2 Spin Systems and Combinatorial Structures in Statistical Physics

As alluded to earlier, there is a long tradition of using probability distributions to model complex systems in physics. A notable class of examples, which we study extensively in this thesis, are *spin systems*. They abstractly represent large assemblies of interacting particles as follows. Imagine you have a large input graph $G = (V, E)$ (e.g. the integer lattice $\mathbb{Z}^d$ for $d = 2, 3$, perhaps truncated to a finite-length box), and each vertex has a *state* taking values in $[q]$ for some positive integer $q \geq 1$. For instance, if $q = 2$, one can think of state 1 (resp. 2) as representing the vertex being unoccupied (resp. occupied) by a particle, or representing the vertex as having a particle with spin value $+1$ (resp. -1). Interactions in the system are then captured by pairwise interactions between pairs of vertices connected by edges. More precisely, for a collection of symmetric *interaction matrices* $A = \{A_e \in \mathbb{R}^{q \times q}_{\geq 0} : e \in E\}$, and a collection of *vertex activities* (or *external fields*)

[6] Mihail and Vazirani actually posed a much broader conjecture which applies to all polytopes whose vertices only have coordinates in $\{0, 1\}$ [MV89]. This encompasses the case of matroids, as well as many other classes of polytopes arising from interesting combinatorial structures. However, at this level of generality, the conjecture remains open, and is well beyond the scope of this thesis. We refer interested readers to [Kai04] for other special cases which have been positively resolved, and [KLT22] for recent evidence against the full conjecture.

[7] This was also achieved independently by Brändén–Huh in the language of Lorentzian polynomials [BH18].

$h = \{h_u \in \mathbb{R}^q_{>0} : u \in V\}$, we define a probability distribution over all *configurations* $\sigma : V \to [q]$ of the system via

$$\mu_{G,A,h}(\sigma) \propto \prod_{e=\{u,v\}\in E} A_e(\sigma(u), \sigma(v)) \prod_{u\in V} h_u(\sigma(u)).$$

This distribution is often called the *Gibbs distribution* (or *Boltzmann distribution*) of the system. In most cases, we will be interested in the case where all interaction matrices A_e are the same. Notably, even though local particle-on-particle interactions are simple to describe, a huge diversity of intricate global structures and behaviors may arise by looking at the system in aggregate. By tuning the parameters A, h as well as the network topology G, one can produce a wide variety of intriguing behaviors which model interacting particle systems under various conditions.

For instance, by taking $q = 2$, all external fields h to be 1, and all interaction matrices to be

$$A_e = \begin{bmatrix} 1 & e^{-2\beta} \\ e^{-2\beta} & 1 \end{bmatrix}, \quad \forall e \in E$$

for some $\beta \geq 0$, one recovers the famous *ferromagnetic Ising model* of a magnet at inverse temperature β, which was originally invented by Lenz [Len20] and further studied by his student Ising [Isi25]. By ferromagnetic, we mean neighboring particles "prefer" to agree on their spin value; the distribution $\mu_{G,A,h}$ puts more probability mass on configurations where most particles have the same state. Just how much these particles prefer to agree depends on how large the parameter $\beta \geq 0$ is. On the other hand, by replacing $e^{-2\beta}$ with $e^{2\beta}$ (equivalently, making $\beta \leq 0$), we recover the *antiferromagnetic Ising model*, where neighboring particles favor different spin values. By taking all interaction matrices to be $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ instead, we wind up with another famous antiferromagnetic spin system called the *hardcore gas model* (see e.g. [BS94] and references therein), where there is a hard constraint enforcing that neighboring vertices cannot both simultaneously be in the second state, i.e. occupied by a particle.

These systems are not only of interest to physicists. Numerous fundamental *combinatorial structures* arising in mathematics and theoretical computer science can be found embedded in these distributions. For instance, underlying the ferromagnetic and antiferromagnetic Ising models is a distribution over *cuts* of the graph. The hardcore gas model may be alternatively stated as a weighted distribution over *independent sets* of the graph. By imposing additional structure on the graph G such as it being a "line graph", the hardcore gas model becomes equivalent to another famous statistical physics model called the *monomer-dimer model*, which at its core is a weighted distribution over *matchings* of the graph. As final example, a straightforward extension of the antiferromagnetic Ising model to a larger number of states q (and taking $\beta \to \infty$) yields the uniform distribution over *proper q-colorings* of the graph G.

Since these systems are so versatile in their ability to capture physical systems, physicists are extremely interested in simulating their behavior, and in particular, sampling from the Gibbs distribution $\mu_{G,A,h}$. For mathematicians, the Gibbs distribution $\mu_{G,A,h}$ also encodes detailed combinatorial information about combinatorial structures like cuts, independents, matchings, proper colorings, etc. For computer scientists, the study of Gibbs distributions forms the bedrock of the complexity theory of (approximate) counting and sampling. Mapping out the incredibly complicated complexity landscape for these sampling problems remains an extremely active research program.

The Hardcore Gas Model For simplicity, we focus the discussion here on our results for the hardcore gas model. Here, the Gibbs distribution may be more simply defined as follows: For an input graph $G = (V, E)$ and a "vertex activity" $\lambda \geq 0$ (sometimes called the *fugacity*), the Gibbs distribution of the hardcore model $\mu_{G,\lambda}$ is a distribution on *independent sets* I of the graph defined by

$$\mu_{G,\lambda}(I) \propto \lambda^{|I|},$$

where recall an independent set $I \subseteq V$ is just a subset of vertices such that no pair are connected by an edge. This model is a discretized version of the continuous *hard spheres gas model* [BNH80; HM13] in statistical physics and chemistry, which incidentally was the first motivation for the invention of the famous Metropolis-Hastings MCMC algorithm [Met+53; Has70].

Even for such a "simple" spin system, it was already known that the model exhibits a "physical" *phase transition* in the statistical properties of the model [Kel85]. More specifically, for the entire class of graphs with maximum degree at most some fixed parameter Δ, there is a precise *critical point* $\lambda_c(\Delta) \stackrel{\text{def}}{=} \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$ which delineates between the regime where "long-range correlations vanish", and the regime where "long-range correlations persist", in the limit as the number of vertices grows to infinity. This is also known as *uniqueness/nonuniqueness of the (infinite-volume) Gibbs measure*.

It was then believed that this physical phase transition *coincides* with a phase transition in the complexity of sampling from the Gibbs distribution of the hardcore model. The intuition here is that if λ is small (i.e. $\lambda < \lambda_c(\Delta)$), in the regime where long-range correlations vanish, then most of the probability mass is put on *small* independent sets like the empty set, which are "easy" to find and enumerate. On the other hand, when λ is large (i.e. $\lambda > \lambda_c(\Delta)$), in the regime where long-range correlations persist, significant probability mass is put on *large* independent sets, which are "hard" to even find [Hås99; BK99; Hås01], intuitively because they bump up against more independent set constraints. A massive research program began in the early 1990s trying to understand rigorously establish this complexity phase transition. This was finally resolved in the early 2010s, with [Wei06] furnishing a polynomial-time algorithm for sampling from $\mu_{G,\lambda}$ when $\lambda < \lambda_c(\Delta)$ (assuming $\Delta \leq O(1)$), and [Sly10] establishing that no efficient algorithm for sampling from $\mu_{G,\lambda}$ exists when $\lambda > \lambda_c(\Delta)$ unless $\mathsf{NP} = \mathsf{RP}$ (see also [SS14; Gal+14; GŠV15; GŠV16]).

While these two breakthroughs provided a theoretically complete picture for the complexity of sampling from $\mu_{G,\lambda}$, the algorithm of [Wei06], while mathematically beautiful, was unsatisfactory in the sense that its running time scales roughly as $n^{O((1/\delta)\log\Delta)}$ (which is not polynomial when Δ is unbounded), and it is rather complicated to implement. It remained open whether or not a much simpler and faster (Markov chain based) algorithm known as the *Glauber dynamics* (or *Gibbs sampler*) could efficiently sample from the hardcore distribution all the way up to the critical threshold $\lambda_c(\Delta)$. This problem had also been studied since the 1990s, but unfortunately, most prior works could only obtain rapid mixing when the fugacity λ of the model is far below from conjectured phase transition limit $\lambda_c(\Delta)$ [LV97; LV99; DG00; Vig01]. Using very sophisticated and complicated combinatorial techniques, several works were able to establish rapid mixing up to the critical threshold $\lambda_c(\Delta)$, but had to make additional and very restrictive structural assumptions on the input graph G (e.g. large girth, large maximum degree, subexponential growth, etc.) [Wei04; Wei06; Res+13; Eft+16]. Using spectral independence, we managed to break these old barriers and establish rapid mixing all the way up to the algorithmic phase transition threshold without making any additional assumptions on the graph [ALO21; CLV20]. For bounded-degree graphs, we subsequently proved optimal $O_\delta(n \log n)$ mixing [CLV21a]. Taken together, these results resolved several longstanding conjectures.

Before we state our result, we first formally define the *Glauber dynamics* for the hardcore model. This simple Markov chain is also sometimes called the *single-site dynamics*, or more commonly in the machine literature, the *Gibbs sampler*. It can be defined much more generally, but for now, we'll state it just for the hardcore model for clarity. Starting from an arbitrary independent set $I \subseteq V$, the random transition to another independent set $I' \subseteq V$ under the Markov chain is described by the following two-step procedure:

1. Select a uniformly random vertex $v \in V$.
2. If $v \in I$, then remove v from I (i.e. transition to $I' = I \setminus \{v\}$) with probability $\frac{1}{1+\lambda}$, and keep v (i.e. transition to $I' = I$) with probability $\frac{\lambda}{1+\lambda}$.

   If $v \notin I$, then set $I' = I$ with probability 1 if $I \cup \{v\}$ is not an independent set. If $I \cup \{v\}$ is an independent set, then add v to I (i.e. transition to $I' = I \cup \{v\}$) with probability $\frac{\lambda}{1+\lambda}$, and set $I' = I$ with probability $\frac{1}{1+\lambda}$.

It is relatively straightforward to verify, using the *detailed balanced condition*, that the Gibbs distribution $\mu_{G,\lambda}$ of the hardcore model is stationary with respect to this Markov chain. We proved the following in a sequence of works.

Theorem 1.1.3 (Rapid Mixing for Tree-Unique Hardcore Model (Informal); [ALO21; CLV20; CLV21a])**.** *Fix a graph $G = (V, E)$ with n vertices and maximum degree Δ, and some $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Then, the Glauber dynamics for the hardcore model $\mu_{G,\lambda}$ on G with fugacity λ mixes in $n^{O(1/\delta)}$-steps. If we additionally have $\Delta \leq O(1)$, then the Glauber dynamics mixes $O_{\delta,\Delta}(n \log n)$-steps, and in particular, yields an (approximate) sampling algorithm running in $O_{\delta,\Delta}(n \log n)$-time.*

Several follow-up works by other researchers then built on our techniques [Che+21b; Ana+21a], and eventually managed to establish the optimal $O_\delta(n \log n)$ mixing even without the bounded-degree (i.e. $\Delta \leq O(1)$) assumption [Ana+22c; Ana+21b; Che+22a; CE22].

Graph Colorings Another important model with a similar flavor to the hardcore model is the uniform distribution over proper colorings of a graph. More specifically, given a graph $G = (V, E)$ and a positive integer $q \geq 2$, let $\mu_{G,q}$ denote the uniform distribution over proper q-colorings of G, i.e. assignments $\chi : V \to [q]$ of "colors" (or states) in $[q] = \{1, \ldots, q\}$ to vertices of G such that no pair of vertices connected by an edge share the same color. In statistical physics jargon, this is also known as the "antiferromagnetic q-state Potts model at zero temperature". This model has attracted significant attention from the statistical physics and computer science community, in particular because it serves as a useful benchmark for new algorithms and techniques. Like the hardcore model, a phase transition occurs at $q = \Delta + 1$, where Δ is again the maximum degree of G. When $q \leq \Delta$, it is already known that no efficient sampling algorithm for $\mu_{G,q}$ exists unless $\mathsf{NP} = \mathsf{RP}$ [GŠV15]. Unlike the hardcore model, we currently only have efficient samplers when $q \geq \left(\frac{11}{6} - \epsilon\right) \Delta$ for some small constant $\epsilon \approx 10^{-5}$ [Vig00; Che+19] (see also [Liu21; Bla+22], and [Jer95] for a very simple and elegant argument recovering $q \geq 2\Delta$), even though we expect efficient sampling to be possible all the way down to $\Delta + 1$.

In the case of sampling uniformly random proper q-colorings, we can once again define the *Glauber dynamics*. Starting from an arbitrary proper q-coloring $\chi : V \to [q]$, the random transition to another proper q-coloring χ' under the Markov chain is described by the following two-step procedure:

1. Select a uniformly random vertex $v \in V$.
2. Out of all colors that are "available" to v, meaning all colors which are not currently assigned to a neighbor of v (including the current color $\chi(v)$ for v), select one such color c uniformly at random. We transition to $\chi' : V \to [q]$ given by $\chi'(u) = \chi(u)$ for all $u \neq v$, and $\chi'(v) = c$. In other words, we just update the color of v to be c.

Again, it is relatively straightforward to verify, using the *detailed balanced condition*, that the uniform distribution $\mu_{G,q}$ over proper q-colorings is stationary with respect to this Markov chain. We have the following conjecture, which remains wide open.

Conjecture 2 (Informal)**.** *The Glauber dynamics on proper q-colorings of a graph $G = (V, E)$ with maximum degree Δ mixes in a polynomial number of steps whenever $q \geq \Delta + 2$. Furthermore, there exists an efficient algorithm which (approximately) samples a uniformly random proper q-coloring of G whenever $q \geq \Delta + 1$.*

While we are a long ways from proving this conjecture, one can make progress on graphs which satisfy certain nice *local sparsity conditions*, namely there being no short cycle in the graph. The rough intuition here is that when the graph is locally sparse, the marginal distribution over colors assigned to vertices in a small ball around a fixed vertex v will have nice approximate "independence" properties. This allows one to use concentration arguments to show a strong lower bound on the number of available colors for v in a uniformly random proper q-coloring with high probability. These *local uniformity properties* are very useful, and have been used to significantly reduce the number of required colors q for locally sparse graphs within various parameter regimes [HV03; Dye+13] (see also [FV07] and references therein). More recently, two results bypassed these local uniformity methods by extending our techniques for the hardcore model, allowing them to establish rapid mixing of the Glauber dynamics when G is triangle-free and $q \leq \alpha^* \Delta$ where $\alpha^* \approx 1.763$ [Che+21d; Fen+21]. This was a new regime for sampling graph colorings with a less stringent girth requirement which also significantly improves upon the current $\left(\frac{11}{6} - \epsilon\right) \Delta$ threshold for general bounded-degree graphs.

However, how can one make progress on the general case? At the moment, we don't have an analog of Weitz's result [Wei06] that we had for the hardcore model. One way we can make further progress is to restrict our attention to *"locally dense"* graphs, with the hope that the techniques developed to attack such graphs can somehow eventually be combined with the techniques for locally sparse graphs to make progress on all graphs. Using spectral independence, we did exactly this for a well-known class of locally dense graphs known as *line graphs*.

A line graph is a graph that can be built out of the following process. Start with some "graph" $H = (U, F)$, and form the new graph $G = (V, E)$ by creating a vertex $v_f \in V$ for each edge $f \in F$,

and connecting two such vertices v_e, v_f if the two edges $e, f \in F$ share an endpoint in U. Any such graph G is called a line graph, and is locally dense in the sense that every vertex v_f of G (for some $f = \{x, y\} \in F$) participates in exactly two cliques: one clique of size $\deg_H(x)$ corresponding to v_e for all edges $e \in F$ incident to x, and one clique of size $\deg_H(y)$ corresponding to v_e for all edges $e \in F$ incident to y. In this setting, proper colorings of the vertices of G are in 1-to-1 correspondence with proper colorings on edges of H (i.e. assignments $\chi : F \to [q]$ such that no pair of edges sharing a vertex have the same color). For this class of graphs, we managed to make the following improvement for the Glauber dynamics for sampling proper colorings.

Theorem 1.1.4 (Rapid Mixing for Edge Colorings (Informal); [ALO22])**.** *Fix a line graph $G = (V, E)$ with n vertices and maximum degree Δ, an $\epsilon > 0$, and a positive integer q. If $\Delta \geq \Omega_\epsilon(1)$ and $q \geq \left(\frac{10}{6} + \epsilon\right)\Delta$, then the Glauber dynamics for sampling a uniformly random proper q-coloring on G mixes in $n^{O(1/\epsilon)}$-steps. If we additionally have $\Delta \leq O_\epsilon(1)$, then the Glauber dynamics mixes in $O_{\epsilon,\Delta}(n \log n)$-steps, and in particular, yields an (approximate) sampling algorithm running in $O_{\epsilon,\Delta}(n \log n)$-time.*

1.1.3 Fast Algorithms

While it is certainly useful from a complexity-theoretic perspective to show that polynomial-time sampling algorithms exist for a given problem, it is much more useful from a practical perspective to give sampling algorithms with running time *(nearly-)linear* in the problem's input size. For many distributions such as those mentioned previously mentioned, such algorithms would be (nearly-)optimal in running time, since the size of samples being output are at least linear in the input size of the problem. This is absolutely critical for large-scale and high-dimensional problems, e.g. statistical physics systems with billions of interacting particles.

Unfortunately, many of the landmark results in the approximate counting and sampling literature develop only polynomial-time algorithms where the running time has a large exponent. For instance, the breakthrough result of Jerrum-Sinclair-Vigoda [JSV04] on sampling perfect matchings in bipartite graphs (equivalently, computing the permanent of a $\{0, 1\}$-matrix) gave an algorithm running in $O(n^7 \log n)$-steps for a bipartite graph with n vertices and potentially up to $O(n^2)$ edges. Even though this was subsequently improved to $O(n^5 \log n)$-steps [Bez+08], this algorithm still remains impractical to use [NV20]. Unfortunately, this a common feature of many mixing results using the famous *canonical paths* technique, as well as many of the deterministic algorithms for approximate counting.

Using our notion of spectral independence, we were able to give *optimal, nearly-linear* mixing time bounds in almost all of the previously mentioned settings, including matroid bases and spin systems on bounded-degree graphs [Ana+21c; CLV21a]. In many cases, these algorithms also run in nearly-linear time, since each step of the Markov chain can be implemented very efficiently. Here is a concrete example of such a result for spanning trees; see the previously stated Theorems 1.1.1, 1.1.3 and 1.1.4 for other examples of fast sampling algorithms.

Theorem 1.1.5 (Fast Spanning Tree Sampling (Informal); [Ana+21c])**.** *Given a connected graph $G = (V, E)$ with n vertices and m edges, there is an algorithm running in $O(m \log^2 n)$-time which (approximately) samples a uniformly random spanning tree of G.*

The problem of sampling random spanning trees in graphs as mentioned in Theorem 1.1.5 has attracted significant attention due to connections with solving Laplacian linear systems [KM09; MST15; Dur+17a], electrical flows [Kir47], graph sparsification [GRV09; FH10; KS18; KKS21], and the traveling salesperson problem [Asa+10; OSS11; KKO21]. Surprisingly, we were not only able to give a faster algorithm but also provide a much simpler mathematical analysis. In particular, all previous approaches used highly intricate algorithms whose analyses are long and complicated, and rely crucially on properties specific to random spanning trees. This long line of work finally culminated in a staggering 215-page paper [Sch18] giving an *almost-linear time*[8] algorithm for this problem. On the other hand, our techniques led to the current fastest algorithm for (approximately) sampling random spanning trees via a simple and generic Markov chain, along with a simple and generic analysis requiring only a few pages [CGM21; Ana+21c].[9]

[8] The term *almost-linear* (in some parameter n) means a function scaling as $n^{1+o(1)}$.

[9] [Sch18] also gives an algorithm that *exactly* samples a uniformly random spanning tree, whereas MCMC samples from a distribution which is only close (in total variation) to the target distribution. Hence, our results do not completely subsume those of [Sch18]. However, our algorithm achieves any desired accuracy with only constant factor overhead, so it suffices for most purposes.

1.1.4 Additional Applications and Follow-Up Works

Spectral independence has had numerous additional applications beyond those already mentioned. It has led to optimal mixing analyses for far more complex Markov chains [Bla+22; Che+21b], as well as inspired the invention of new Markov chains [Ana+21b; Che+21b; CZ22] and new sampling algorithms [Bez+22; CMM22; Gal+22]. Outside the realm of sampling, it has significant ramifications for *concentration inequalities* in probability theory [KKS21], *strong data processing inequalities* in information theory, functional analytic inequalities, and approximation guarantees for discrete optimization algorithms [AV22]. It has also been generalized, extended, and strengthened in several ways [Ali+21; Ana+21a; Ana+21b; Ana+22c; QW22; CE22]. We defer a more detailed discussion of relevant follow-up works to Section 13.1.

1.2 A Unified Theory of Algorithmic Sampling via Spectral Independence

For discrete probability distributions, there are roughly three theoretically well-studied algorithmic paradigms in the theory of approximate counting and sampling, one being Markov chain Monte Carlo (MCMC), and the other two being *correlation decay* [Wei06] and *polynomial interpolation* [Bar16a]. Circa 2019, these methods were being developed independently, and the connection between them, was not well-understood. In fact, better understanding their relationship was one of the driving goals for the 2019 Simons Institute program on Geometry of Polynomials. What was particularly confounding about this state of affairs was that correlation decay and interpolation are inherently *deterministic methods*, yet they were outperforming randomized methods like MCMC on problems concerning probability distributions. There were many settings in which one of the other two methods succeeded, but Markov chains were only conjectured to work.

Besides the cognitive dissonance that begged to be resolved, there were also practical motivations for demonstrating the efficacy of Markov chains. As many of us in theoretical computer science learn early on, the price for ensuring that an algorithm is deterministic is often a compromise in speed. And as it turns out, these other two deterministic methods inherently have *galactic running times*, i.e. running times which may scale polynomially but have huge, impractically large exponents. With the promise of fast algorithms, there was a major push to study MCMC.

In a series of works, we showed that the key ingredients needed in the analysis of the correlation decay and interpolation approaches can be used to prove rapid mixing of the extensively studied Glauber dynamics [ALO21; CLV20; CLV21b]. *In other words, whenever one of these other two approaches works, local Markov chains also work.* This "unification" of algorithmic sampling techniques was achieved precisely through the lens of spectral independence. As a consequence of this confluence of ideas, we showed rapid mixing of the Glauber dynamics all the way up to the algorithmic phase transition threshold for several important statistical physics models such as the hardcore gas model (see e.g. Theorem 1.1.3), thus resolving several longstanding conjectures.

In the remainder of this section, we highlight some of the key recurring themes in this thesis.

Good Things Happen When Correlations are Controlled As we discuss earlier, one of the primary intuitions behind the results in this thesis is that if the correlations within the probability distribution of interest are "limited" in some sense, then it is possible to efficiently sample. In particular, correlation structure should govern the complexity of sampling. This intuition is not new. Indeed many prior works in statistical physics, and approximating counting and sampling make a statement of a similar flavor. For instance, rapid mixing of Markov chains had been previously established under negative correlation conditions [FM92; AOR16], certain decaying correlation conditions [Dye+04b; Wei04; Wei06], etc. However, previous work only managed to establish such a connection in very limited settings. Our notion of spectral independence turns out to be a broad and versatile way to capture this powerful intuition. The idea that controlling correlations enables algorithms is also pervasive in the theory of computing more broadly, e.g. k-wise independence in algorithms, cryptography and pseudorandomness (see e.g. [BR94; SSS95; LW06]), correlation inequalities and concentration inequalities in rounding schemes (see e.g. [Sri99]), etc.

Local-to-Global Phenomena When grappling with an intractably large object, a common and useful strategy is *decomposing* the object into smaller, more manageable pieces. One can

then try to recover global properties of the original object by studying these smaller pieces in isolation, perhaps via another round of decomposition, and then studying how they fit together. Prominent examples of this strategy in theoretical computer science include the development of probabilistically checkable proofs (PCPs) [Din07], the study of high-dimensional expanders [KM17; DK17; Lub19; KO20b], Markov chain decomposition techniques [MR02; Jer+04; Son04; MR06; MY09; AOR16; PS17; HS20], the recent breakthroughs on locally testable codes [Din+22; PK22], etc. Of course, for such a strategy to succeed, the *local properties* of these smaller pieces and how they fit together must still retain some amount of information about the global structure. This will fortunately be the case for us. Indeed, this entire thesis is based on this idea.

Algebraic Techniques in Probability and Combinatorics The recently emerging area of *geometry of polynomials* studies how geometric, analytic, and algebraic properties of multivariate polynomials can be leveraged to obtain useful probabilistic or combinatorial information. Examples of such properties include their behavior as functions (e.g. log-concavity) and the locations of their zeros in the complex plane. Examples of desirable probabilistic/combinatorial information include combinatorial log-concavity inequalities [Sta89] and correlation inequalities [Har60; Kle66; Gri67a; Gri67b; Gri67c; KS68; Gri69; GHS70; Gin70; Gal71; FKG71; Suz73; Hol74; AD78; She82; Fis84; RS93; AK96; Jan98; Roy14; LM17; DNS21; DSS22b] (see the books [Bov06; FV17] for further discussion). These ideas have led to significant advances in combinatorics, statistical mechanics, and quantum physics. The pioneering works of Lee-Yang [LY52] and Heilmann-Lieb [HL72] are prime examples of this connection; the Lee-Yang Theorem establishes phase transition properties of the Ising ferromagnet using intricate zero-freeness properties of its partition function, while the Heilmann-Lieb Theorem establishes lack of a phase transition for the monomer-dimer model using zero-freeness of the matching polynomial.

Theoretical computer scientists have since used these connections and techniques to great effect. They have had a profound impact on approximation algorithms [Asa+10; OSS11; Ove14; AO15; KKG20; KKO21], approximate counting [Bar16b; AOR16; Bar17a; Bar16a; PR17; AO17; SV17; Reg18; BCR20; Guo+21; AOV21; Ali+21; CLV21b], spectral graph theory [MSS14; MSS15a; MSS15c; Coh16; HPS18], algorithmic game theory [Ana+17; Ana+18b; Kle21; Bea+22], linear algebra [MSS15b; MSS22], and more. In several chapters, we will see fruitful applications of this powerful idea to establishing rapid mixing of discrete Markov chains on combinatorial structures.

The Bridge Between Mixing in Time and Space In the context of spin systems on graphs arising in statistical physics, there turns out to be a beautiful connection between *temporal mixing*, in the sense that correlations between the starting state and the current state in the Markov chain decay rapidly as time evolves, and *spatial mixing*, in the sense that correlations between sites decay rapidly as distance increases. This connection was made rigorous for physically relevant lattices like $\mathbb{Z}^d$, or more generally, graphs with a "subexponential growth" property [Wei04; Dye+04b; CP21a]. The seminal work of Weitz [Wei06] established a profound but weaker implication, namely that spatial mixing implies computational tractability using a deterministic algorithm which directly takes advantage of spatial mixing. This has been further explored in a long line of work [Bay+07; GK07; MS07; BG08; LLY12; GK12; LLY13; LY13; SSY13; Res+13; LLL14; LLZ14a; LWZ14; SST14; Sin+15; LL15b; LL15a; LYZ16; SYZ19; LSS20] (see also [Wei04; Sri14]). However, the full connection between temporal and spatial mixing proved elusive. In this thesis, we will resolve this longstanding challenge using spectral independence.

1.3 Organization of the Thesis

We conclude the bulk of this introduction with a high-level overview of how this thesis is organized. The remainder of this introduction, namely Section 1.4, is devoted to setting up some basic notation we will use throughout, as well as give the necessary background on Markov chains, approximate counting and sampling, etc.

In Chapter 2, we define spectral independence, the Markov chains we will analyze, and how these concepts are connected. We discuss these ideas at length, providing examples as well as two proofs of the *local-to-global theorem*, namely that spectral independence implies rapid mixing. This chapter forms the foundation of this thesis. After this, the thesis decomposes nicely into modular pieces; the remaining chapters may each be read essentially independently of each other.

We organize them into three parts. Part I develops tools to establish spectral independence. We provide four classes of techniques.

1. Chapters 3 and 4 build upon beautiful work of Oppenheim [Opp18] in the recently emerging area of high-dimensional expanders. The techniques here may roughly be put under the umbrella of *trickle-down methods*, which are another manifestation of local-to-global analysis. Chapter 3 specifically also has intimate connections with the study of log-concave polynomials [Gur05; AOV21; BH18; BH20], which is the focus of Chapter 5.

2. Chapters 5 and 6 develop methods originating in the study of the *geometry of polynomials*. We use algebraic and analytic properties of multivariate polynomials to deduce spectral independence. In Chapter 5, we leverage *log-concavity*, while in Chapter 6, we leverage *zero-freeness/stability*. The methods developed in Chapter 6 in particular will allow us to take advantage of powerful theorems in complex analysis and algebraic combinatorics. As previously mentioned, the ideas in Chapter 5 are intimately related to those present in Chapter 3.

3. Chapter 7 leverages classical statistical physics methods, and in particular, the *correlation decay property*. This is where we rigorously establish the connection between spatial and temporal mixing. Here, we will take the hardcore gas model as a case study, but we will also see how to tackle many other notable models in statistical physics. We will further apply this method in Appendices A to C.

4. In Chapter 8, we revisit a classical Markov chain mixing technique known as *coupling*, which is also fundamental to the theory of optimal transport. This chapter is more conceptual, and less technical. We clarify the relationship between the classically well-studied *Dobrushin influence* and the notion of influence used to define spectral independence. We also prove a blackbox comparison theorem between local Markov chains and the simplest single-site dynamics.

Part II develops methods for refined analyses of Markov chains which can yield *optimal mixing times*. We begin with a discussion of entropy decay and the functional analytic (modified) logarithmic Sobolev constants in Chapter 9. There, we also establish a local-to-global theorem for entropy contraction, which will form the foundation of optimal mixing analyses. This will be necessary for Chapters 10 and 11, where we prove optimal mixing theorems for graphical models and matroids, respectively. Finally, in Chapter 12, we use the optimal mixing results from Chapter 11 to build fast algorithms running in nearly-linear time. This will conclude the bulk of this thesis.

Part III is devoted to various additional results which we obtained using spectral independence. Appendices A to C are all essentially extensions or modifications of the ideas in Chapter 7. In Appendices A and B, we apply correlation decay to other (two-state) spin systems of interest beyond the hardcore gas model, including the Ising model and the monomer-dimer model. In Appendix D, we apply the techniques from Chapter 6 to obtain mixing results for weighted graph homomorphisms and tensor networks with sufficiently weak interactions. In Appendix E, we carry out spectral independence calculations for various statistical physics models on *infinite regular trees*, which are often the "worst case" instances. These calculations suggest concrete spectral independence lower bounds. Finally, Appendix F is a catch-all for various miscellaneous results which are related to spectral independence but aren't substantial enough to warrant a separate chapter.

Disclaimer: Of course, one motivation for writing this thesis was so I can graduate. But a much more significant motivation was so that some of the results in this area can be presented in a more unified and organized manner, with a more consistent notation. It is our hope that with the understanding we have now, the ideas, connections, and proofs can all be presented more clearly. Although we almost certainly have fallen short of this goal, we sincerely hope this can be useful to someone somewhere. While the majority of ideas in this thesis are already present in various papers I have (co)authored with others, all mistakes here are my own. The contents of this thesis are largely based on [Ana+18a; Ana+19; Ana+21c; ALO21; CLV20; CLV21a; CLV21b; Liu21; ALO22], although we have included a few additional results we obtained in our journey which, to the best of our knowledge, have not been written down anywhere yet.

1.4 Preliminaries and Notation

Throughout, we use the convention that $0 \log 0 = 0$ and $\frac{0}{0} = 0$. For a positive integer $n \geq 1$, we write $[n] = \{1, \ldots, n\}$. For a finite set $\mathscr{U}$ and an integer $0 \leq n \leq |\mathscr{U}|$, we write $\binom{\mathscr{U}}{n} = \{S \subseteq \mathscr{U} : |S| = n\}$, the collection of subsets of $\mathscr{U}$ with size n; note that if $n > |\mathscr{U}|$ or $n < 0$, then $\binom{\mathscr{U}}{n} = \emptyset$. We also write $2^{\mathscr{U}} = \{S \subseteq U\}$ for the family of all subsets of $\mathscr{U}$. If $S \subseteq \mathscr{U}$ and $u \in \mathscr{U}$, then we write $S - u$ for $S \setminus \{u\}$, and $S + u$ for $S \cup \{u\}$. If $S \subseteq \mathscr{U}$, we write $\mathbb{I}_S : \mathscr{U} \to \{0, 1\}$ for the $\{0, 1\}$-indicator function of S; for an element $j \in \mathscr{U}$, we write $\mathbb{I}_j$ as opposed to $\mathbb{I}_{\{j\}}$. If we wish to express $\mathbb{I}_S$ as a vector in $\mathbb{R}^{\mathscr{U}}$, we will typically write $\mathbf{1}_S$ instead. Note that the collection $\{\mathbf{1}_i : i \in \mathscr{U}\}$ form an orthonormal basis of $\mathbb{R}^{\mathscr{U}}$ called the *standard basis*. They are also called the *coordinate directions* in $\mathbb{R}^{\mathscr{U}}$; we reserve the common e_i notation for other objects in this thesis. We will use $\mathbb{I}[\cdot]$ to represent a generic indicator function which outputs 1 if the input "condition" is satisfied, and 0 otherwise.

A graph G is a pair (V, E), where V is a finite (or countable) set of *vertices* and $E \subseteq \binom{V}{2}$ is a set of *edges*. Throughout, our graphs will be undirected and simple, i.e. they will not have any parallel/repeated edges nor self-loops. We will write $\Delta = \Delta(G)$ for the maximum degree of G.

All probability distributions we encounter in this thesis will be over finite probability spaces, so there will be no issues of measurability, and we won't use any measure-theoretic notation. For a distribution μ on some finite state space Ω, we write $\mathrm{supp}(\mu) \stackrel{\mathsf{def}}{=} \{x \in \Omega : \mu(x) > 0\}$ for the support of μ. We call Ω the *state space* or *domain* of μ; we will tend to prefer the former, as the term "domain" already has a precise meaning in complex analysis. We write $\mu_{\min} \stackrel{\mathsf{def}}{=} \min_{x \in \mathrm{supp}(\mu)} \mu(x)$. If ν is another distribution on Ω with $\mathrm{supp}\,(\nu) \subseteq \mathrm{supp}\,(\mu)$, then write $\frac{d\nu}{d\mu} : \Omega \to \mathbb{R}$ for the *(relative) density of ν w.r.t. μ*, defined as $\frac{d\nu}{d\mu}(x) = \frac{\nu(x)}{\mu(x)}$ for all $x \in \Omega$.

For a set of points $S \subseteq \mathbb{R}^n$, we write

$$\mathrm{conv}(S) \stackrel{\mathsf{def}}{=} \left\{ \sum_{i=1}^{n} \alpha_i x_i : n \geq 1, x_1, \ldots, x_n \in S, \alpha_1, \ldots, \alpha_n \geq 0, \sum_{i=1}^{n} \alpha_i = 1 \right\} \subseteq \mathbb{R}^n$$

for the *convex hull of S*.

1.4.1 Linear Algebra

We use the notations $A \in \mathbb{R}^{n \times m}$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ interchangeably for matrices (or linear operators) mapping vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$. For a vector $v \in \mathbb{R}^n$, we write $\mathrm{diag}(v) \in \mathbb{R}^{n \times n}$ for the unique diagonal matrix with diagonal entries given by the entries of v. For two vectors $u \in \mathbb{R}^n, v \in \mathbb{R}^m$, we write $uv^\top = (u_i \cdot v_j)_{i \in [n], j \in [m]} \in \mathbb{R}^{n \times m}$ for the outer product of u and v. In the case when our vectors already have superscripts, perhaps because they depend on some other underlying parameters, we will prefer to write $u \otimes v$ instead of $uv^\top$. We generally use these two notations interchangeably.

If $u, v \in \mathbb{R}^{\mathscr{U}}$ are two vectors of the same dimension, then we write $u \odot v \in \mathbb{R}^{\mathscr{U}}$ for the *entrywise product* of u and v, i.e. $(u \odot v)_i = u_i v_i$ for all $i \in \mathscr{U}$. Similarly, if $A, B \in \mathbb{R}^{n \times m}$ are matrices of the same dimension, then we write $A \odot B \in \mathbb{R}^{n \times m}$ for the *entrywise product* of A and B, i.e. $(A \odot B)_{ij} = A_{ij} B_{ij}$ for all $i \in [n], j \in [m]$. For matrices, this is also sometimes called the *Hadamard product* or the *Schur product*.

As mentioned previously, we write $\mathbf{1}_{\mathscr{U}} \in \mathbb{R}^{\mathscr{U}}$ (resp. $\mathbf{1}_n \in \mathbb{R}^n$) for the all-ones vector, and $\mathbf{1}_S \in \mathbb{R}^{\mathscr{U}}$ (resp. $\mathbf{1}_S \in \mathbb{R}^n$) for the corresponding $\{0, 1\}$-indicator vector of $S \subseteq \mathscr{U}$ (resp. $S \subseteq [n]$). When the underlying ground set $\mathscr{U}$ or $[n]$ is clear from context, we drop the subscript.

Padding Let Ω be a finite set, and let $\Lambda \subseteq \Omega$ be some subset. If $v \in \mathbb{R}^\Lambda$ is a vector whose entries are indexed by the elements of Λ, then we may view v as a vector in $\mathbb{R}^\Omega$ by *padding* v with zero entries corresponding elements of $\Omega \setminus \Lambda$. We do the same for matrices, namely if $A \in \mathbb{R}^{\Lambda \times \Lambda}$ is a matrix whose rows and columns are indexed by elements of Λ, then we may view A as a matrix in $\mathbb{R}^{\Omega \times \Omega}$ by appropriately padding A with zero rows and columns corresponding to elements of $\Omega \setminus \Lambda$. Padded vectors and matrices will be convenient later on; since we will be decomposing various state spaces into smaller subsets, we will need a way to decompose vectors and matrices attached to such state spaces.

Matrix Norms We write $\|A\|_\infty = \max_{i=1,\dots,n} \sum_{j=1}^n |A(i,j)|$ for the maximum absolute row sum (also known as the *(induced) ℓ_∞-norm*), and $\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^n |A(i,j)|$ for the maximum absolute column sum (also known as the *(induced) ℓ_1-norm*). We write $\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$ for *operator norm of A* (also known as the *(induced) ℓ_2-norm*).

Eigenvalues and Inner Products For a matrix $A \in \mathbb{R}^{n \times n}$, we will often write $\lambda_1(A), \dots, \lambda_n(A)$ for the eigenvalues of A. If A has real eigenvalues, we typically order them as $\lambda_n(A) \leq \cdots \leq \lambda_1(A)$; in this case, we sometimes write $\lambda_{\max}(A)$ for the largest eigenvalue of A, and $\lambda_{\min}(A)$ for the smallest eigenvalue of A. We write $\rho(A) = \max_{i=1,\dots,n} \{|\lambda_i(A)|\}$ for the *spectral radius of A*, which is well-defined even if A has complex eigenvalues.

For an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ (not necessarily Euclidean), we say a matrix (or linear operator) A is *self-adjoint w.r.t.* $\langle \cdot, \cdot \rangle$ if $\langle u, Av \rangle = \langle Au, v \rangle$ for all $u, v \in \mathbb{R}^n$. In this thesis, we will often switch between different inner products. As such, we highlight the following variational characterization of eigenvalues for a matrix that is self-adjoint with respect to some inner product. In particular, the matrix need not be symmetric.

Theorem 1.4.1 (Courant-Fischer-Weyl Theorem; see e.g. [HJ13])**.** *Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator which is self-adjoint with respect to some (not necessarily Euclidean) inner product $\langle \cdot, \cdot \rangle$. Then for every integer $1 \leq k \leq n$, we have the following identity*

$$\lambda_k(A) = \inf_U \sup_{v \in U : v \neq 0} \frac{\langle v, Av \rangle}{\langle v, v \rangle},$$

where the infimum is taken over all $(n-k+1)$-dimensional subspaces $U \subseteq \mathbb{R}^n$.

We have the following useful and immediate consequence of Theorem 1.4.1.

Corollary 1.4.2 (Weyl's Inequality; see e.g. [HJ13])**.** *Let $A, B : \mathbb{R}^n \to \mathbb{R}^n$ be linear operators which are self-adjoint w.r.t. the same (not necessarily Euclidean) inner product $\langle \cdot, \cdot \rangle$. Then we have the inequality*

$$\lambda_1(A + B) \leq \lambda_1(A) + \lambda_1(B).$$

Note we do not assume that A, B are positive semidefinite.

We say a matrix $A \in \mathbb{R}^{n \times n}$ is *positive semidefinite* w.r.t. an inner product $\langle \cdot, \cdot \rangle$ if it is self-adjoint w.r.t. $\langle \cdot, \cdot \rangle$ and for every $v \in \mathbb{R}^n$, we have that $\langle v, Av \rangle \geq 0$. In this case, we write $A \succeq 0$. By Theorem 1.4.1, this is equivalent to A having nonnegative eigenvalues. We say A is negative semidefinite if $-A$ is positive semidefinite. In this case, we write $A \preceq 0$. Finally, when we have strict inequalities, we say that A is positive definite (resp. negative definite), and write $A \succ 0$ (resp. $A \prec 0$). When comparing two matrices A, B which are self-adjoint w.r.t. the same inner product, we write $A \preceq B$ if and only if $B - A \succeq 0$. This is the Loewner order w.r.t. to $\langle \cdot, \cdot \rangle$.

Throughout this thesis, we will use $\langle \cdot, \cdot \rangle$ to denote the usual Euclidean inner product, and $\preceq$ to denote the usual Loewner order for symmetric matrices which is naturally associated to the Euclidean inner product. When we switch to a new inner product, we will add appropriate subscripts to the inner product and to the Loewner order to make it clear which inner product we are working with. For instance, if $w \in \mathbb{R}^n_{>0}$ is a positive vector, then we write $\langle \cdot, \cdot \rangle_w$ for the inner product given by $\langle u, v \rangle_w \stackrel{\text{def}}{=} \langle u, \operatorname{diag}(w) v \rangle$ and $\preceq_w$ for the associated Loewner order. Similarly, if $A \in \mathbb{R}^{n \times n}$ is a symmetric positive semidefinite matrix, then we write $\langle \cdot, \cdot \rangle_A$ for the inner product given by $\langle u, v \rangle_A \stackrel{\text{def}}{=} \langle u, Av \rangle$ and $\preceq_A$ for the associated Loewner order.

Some Useful Linear Algebraic Results In the remainder of this subsection, we collect some additional useful linear algebraic facts.

Theorem 1.4.3 (Schur Product Theorem [HJ13, Thm 7.5.3])**.** *If $A, B \in \mathbb{R}^{n \times n}$ are symmetric positive semidefinite, then their entrywise product (or Hadamard product) $A \odot B$, whose entries are $(A \odot B)_{ij} \stackrel{\text{def}}{=} A_{ij} B_{ij}$, is symmetric positive semidefinite.*

Theorem 1.4.4 (Perron-Frobenius Theorem [HJ13, Ch. 8]). *Let $A \in \mathbb{R}^{n\times n}$ be matrix (not necessarily symmetric) with strictly positive entries. Then A has a real eigenvalue λ which is strictly positive. Furthermore, it has multiplicity one and its corresponding eigenvector v has strictly positive entries. The same is true if A only has nonnegative entries but is irreducible, i.e. the directed graph on vertex set $[n]$, which includes a directed edge $i \to j$ if and only if $A_{ij} > 0$, is strongly connected.*

Theorem 1.4.5 (Cauchy's Interlacing Theorem [HJ13, Corollary 4.3.9]). *For a symmetric matrix $A \in \mathbb{R}^{n\times n}$ and vector $v \in \mathbb{R}^n$, the eigenvalues of A interlace the eigenvalues of $A + vv^\top$. That is, for $B = A + vv^\top$,*

$$\lambda_n(A) \leqslant \lambda_n(B) \leqslant \lambda_{n-1}(A) \leqslant \cdots \leqslant \lambda_2(B) \leqslant \lambda_1(A) \leqslant \lambda_1(B).$$

The following is an immediate consequence:

Lemma 1.4.6. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and let $P \in \mathbb{R}^{m\times n}$. If A has at most one positive eigenvalue, then $PAP^\top$ has at most one positive eigenvalue.*

Proof. Since A has at most one positive eigenvalue, we can write $A = B + vv^\top$ for some vector $v \in \mathbb{R}^n$ and some negative semidefinite matrix B. Then $PAP^\top = PBP^\top + Pvv^\top P^\top$. First, observe that $PBP^\top \preceq 0$, since for $x \in \mathbb{R}^m$, $x^\top PBP^\top x = (P^\top x)^\top B(P^\top x) \leqslant 0$. Second, let $w = Pv \in \mathbb{R}^m$. Then $Pvv^\top P^\top = ww^\top$ and by Theorem 1.4.5, the eigenvalues of $PBP^\top$ interlace the eigenvalues of $PBP^\top + (Pv)(Pv)^\top$. Since all eigenvalues of $PBP^\top$ are nonpositive, $PAP^\top = PBP^\top + ww^\top$ has at most one positive eigenvalue. □

The following fact is well-known.

Fact 1.4.7. *Let $A \in \mathbb{R}^{n\times k}$ and $B \in \mathbb{R}^{k\times n}$ be arbitrary matrices. Then, non-zero eigenvalues of AB are equal to non-zero eigenvalues of BA with the same multiplicity.*

Lemma 1.4.8. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with at most one positive eigenvalue. Then, for any positive semidefinite matrix $B \in \mathbb{R}^{n\times n}$, BA has at most one positive eigenvalue.*

Proof. Since $B \succeq 0$, we can write $B = C^\top C$ for some $C \in \mathbb{R}^{n\times n}$. By Fact 1.4.7, $BA = C^\top CA$ has the same nonzero eigenvalues as the matrix $CAC^\top$. Since A has at most one positive eigenvalue, by Lemma 1.4.6, $CAC^\top$ has at most one positive eigenvalue and hence, so does BA. □

1.4.2 Multivariate Functions and Polynomials

Let $\mathbb{K}$ be one of the fields $\mathbb{R}$ or $\mathbb{C}$, and let $x = (x_i : i \in \mathscr{U})$ be a collection of variables taking values in $\mathbb{K}$. Let $f(x)$ be a multivariate function. For (partial) differentiation, we often write ∂_i instead of ∂_{x_i} for simplicity. Assuming appropriate smoothness or differentiability conditions (which will always be satisfied in this thesis), the *gradient* of f is given by the vector-valued function $\nabla f(x) \stackrel{\text{def}}{=} (\partial_i f(x))_{i\in\mathscr{U}}$ and the *Hessian* of f is given by the symmetric matrix-valued function $\nabla^2 f(x) \stackrel{\text{def}}{=} (\partial_i \partial_j f(x))_{i,j\in\mathscr{U}}$. For a vector $v \in \mathbb{R}^{\mathscr{U}}$, we write $\partial_v = \sum_{i\in\mathscr{U}} v_i \partial_i$ for the directional derivative w.r.t. v.

Let $\mathscr{U}$ be a nonempty finite set (e.g. $\mathscr{U} = [n]$ where $n \geq 1$). For a *multi-index* $\alpha \in \mathbb{Z}_{\geq 0}^{\mathscr{U}}$, we write $|\alpha| \stackrel{\text{def}}{=} \sum_{i\in\mathscr{U}} \alpha_i$ and $\alpha! \stackrel{\text{def}}{=} \prod_{i\in\mathscr{U}} \alpha_i!$. If $\beta \in \mathbb{Z}_{\geq 0}^{\mathscr{U}}$ is another multi-index, we write $\alpha \leq \beta$ to mean $\alpha_i \leq \beta_i$ for all $i \in \mathscr{U}$. If $\alpha \leq \beta$, we may then define $\binom{\beta}{\alpha} \stackrel{\text{def}}{=} \prod_{i\in\mathscr{U}} \binom{\beta_i}{\alpha_i} = \frac{\beta!}{\alpha!(\beta-\alpha)!}$

For a multi-index $\alpha \in \mathbb{Z}_{\geq 0}^{\mathscr{U}}$ (resp. $S \subseteq \mathscr{U}$), we often compactly write monomials as $x^\alpha \stackrel{\text{def}}{=} \prod_{i\in\mathscr{U}} x_i^{\alpha_i}$ (resp. $x^S \stackrel{\text{def}}{=} \prod_{i\in S} x_i$), and differential operators as $\partial^\alpha = \prod_{i\in\mathscr{U}} \partial_i^{\alpha_i}$ (resp. $\partial^S \stackrel{\text{def}}{=} \partial^{\mathbb{I}_S} = \prod_{i\in S} \partial_i$). Similarly, if $S \subseteq \mathscr{U}$, we sometimes write x_S as shorthand for the subset of variables $\{z_i : i \in S\}$.

A (multivariate) polynomial with variables $x = (x_i : i \in \mathscr{U})$ and *coefficients* in $\mathbb{K}$ is a function of the form

$$f(x) = \sum_{\alpha\in\mathcal{T}} c_\alpha x^\alpha,$$

where $\mathcal{T}$ is a finite subset of $\mathbb{Z}_{\geq 0}^{\mathscr{U}}$.

The *support of* f is defined as $\mathrm{supp}(f) \stackrel{\mathsf{def}}{=} \{\alpha \in \mathbb{Z}_{\geq 0}^{\mathscr{U}} : c_\alpha \neq 0\}$, and the *degree of* f is defined as $\max\{|\alpha| : \alpha \in \mathrm{supp}(f)\}$. We say f is *multiaffine* if $\alpha \leq \mathbf{1} \in \mathbb{Z}_{\geq 0}^{\mathscr{U}}$ for every $\alpha \in \mathrm{supp}(f)$, i.e. for each $i \in \mathscr{U}$, holding x_j fixed for every $j \neq i$ yields an *affine function* in x_i. We say f is *homogeneous (of degree-d)* if for every constant $c \in \mathbb{K}$, we have $f(cx) = c^d \cdot f(x)$; equivalently (for polynomials), f is homogeneous of degree-d if $|\alpha| = d$ for every $\alpha \in \mathrm{supp}(f)$. For brevity, we sometimes just say f is *d-homogeneous*. Note that if f is d-homogeneous and $\alpha \in \mathbb{Z}_{\geq 0}^{\mathscr{U}}$ is a multi-index, then $\partial^\alpha f$ is either identically zero or $(d - |\alpha|)$-homogeneous.

Fact 1.4.9 (Euler's Homogeneous Function Theorem)**.** *If $f(x)$ is a continuously differentiable multivariate function in variables $x = (x_i : i \in \mathscr{U})$ (e.g. a multivariate polynomial) which is homogeneous of degree-d, then we have the identity*

$$f(x) = \frac{1}{d} \sum_{i \in \mathscr{U}} x_i \partial_i f(x).$$

Corollary 1.4.10. *If $f(x)$ is a continuously differentiable multivariate function in variables $x = (x_i : i \in \mathscr{U})$ (e.g. a multivariate polynomial) which is homogeneous of degree-d, then we have the additional identities*

$$\nabla f(x) = \frac{1}{d-1} \sum_{i \in \mathscr{U}} x_i \cdot \nabla \partial_i f(x) = \frac{1}{d-1} \cdot (\nabla^2 f(x)) \cdot x$$
$$\nabla^2 f(x) = \frac{1}{d-2} \sum_{i \in \mathscr{U}} x_i \cdot \nabla^2 \partial_i f(x).$$

We write $\mathbb{K}[x_i : i \in \mathscr{U}]$ for the set of all multivariate polynomials f in the variables $x = (x_i : i \in \mathscr{U})$ with coefficients in $\mathbb{K}$. Similarly, we write $\mathbb{K}[x_i : i \in \mathscr{U}]_d$ for the subset of such polynomials which are homogeneous of degree-d.

We use $e_k(z_1, \ldots, z_n)$ to denote the degree-k elementary symmetric polynomial in $z_1, \ldots, z_n$. We sometimes abuse notation and write $e_k(u, z_S)$ to denote the degree-k elementary symmetric polynomial in variables $z_S \cup \{u\}$.

1.4.3 Entropy, Variance, and Distances Between Probability Measures

Let μ be some (reference) probability distribution on some finite state space Ω, and let $f : \Omega \to \mathbb{R}$ be some function. We define the *variance of f (w.r.t. μ)* by

$$\mathrm{Var}_\mu(f) \stackrel{\mathsf{def}}{=} \mathbb{E}_{x \sim \mu}\left[(f(x) - \mathbb{E}_\mu f)^2\right] = \mathbb{E}_\mu\left(f^2\right) - \mathbb{E}_\mu(f)^2.$$

Similarly, if $f : \Omega \to \mathbb{R}_{\geq 0}$ is a nonnegative function on Ω, we define the *entropy of f (w.r.t. μ)* by

$$\mathrm{Ent}_\mu(f) \stackrel{\mathsf{def}}{=} \mathbb{E}_\mu(f \log f) - \mathbb{E}_\mu(f) \log \mathbb{E}_\mu(f).$$

If ν is some other probability measure on Ω with $\mathrm{supp}(\nu) \subseteq \mathrm{supp}(\mu)$, then the *(relative) variance (or χ^2-divergence) of ν w.r.t. μ* is defined as

$$\mathscr{D}_{\chi^2}(\nu \| \mu) \stackrel{\mathsf{def}}{=} \mathrm{Var}_\mu\left(\frac{d\nu}{d\mu}\right).$$

Similarly, the *(relative) entropy (or* KL*-divergence) of ν w.r.t. μ* is defined as

$$\mathscr{D}_{\mathrm{KL}}(\nu \| \mu) \stackrel{\mathsf{def}}{=} \mathrm{Ent}_\mu\left(\frac{d\nu}{d\mu}\right).$$

One should view the (relative) variance and the (relative) entropy as ways of measuring how far two probability distributions are from each other, even though they are not metrics in a formal sense. We have the following useful alternative representation of (relative) entropy.

Proposition 1.4.11 (Donsker–Varadhan Variational Representation; [DV76; DV83])**.** *For two probability distributions μ, ν over a finite state space Ω, the* KL*-divergence between μ, ν admits the following variational formula:*

$$\mathscr{D}_{\mathrm{KL}}(\nu \| \mu) = \sup_{f : \Omega \to \mathbb{R}} \left\{\mathbb{E}_\nu(f) - \log \mathbb{E}_\mu\left(e^f\right)\right\}.$$

One can generalize these concepts further to form the class of *φ-entropies*, where we define $\mathrm{Ent}_\mu^\varphi(f) \stackrel{\mathsf{def}}{=} \mathbb{E}_\mu(\varphi \circ f) - \varphi(\mathbb{E}_\mu f)$ for some convex function φ with an appropriate domain; for instance, taking $\varphi(x) = x^2$ recovers variance and $\varphi(x) = x \log x$ recovers entropy. These then induce *φ-divergences* defined by $\mathscr{D}_\varphi(\nu\|\mu) \stackrel{\mathsf{def}}{=} \mathrm{Ent}_\mu^\varphi\left(\frac{d\nu}{d\mu}\right)$. These notions are fundamental to information theory, and it is interesting to study Markov chain mixing w.r.t. these "distances" between probability measures, but we won't need them in this thesis; we note that some of the results we prove can be extended in a straightforward manner to these objects.

We have the following useful fact.

Theorem 1.4.12 (Data Processing Inequalities)**.** *Let Ω, Λ be arbitrary finite state spaces, and let μ, ν be probability measures on Ω. Suppose $\mathsf{P} \in \mathbb{R}_{\geq 0}^{\Omega \times \Lambda}$ is a (row-)stochastic operator, i.e. the entries of P are nonnegative each row of P sums to 1 (so that each row of P is a probability distribution over Λ). Then we have the inequalities*

$$\mathscr{D}_{\chi^2}(\nu\mathsf{P}\|\mu\mathsf{P}) \leq \mathscr{D}_{\chi^2}(\nu\|\mu)$$
$$\mathscr{D}_{\mathrm{KL}}(\nu\mathsf{P}\|\mu\mathsf{P}) \leq \mathscr{D}_{\mathrm{KL}}(\nu\|\mu).$$

Remark 1. These inequalities can be generalized in a straightforward manner to all φ-divergences, and can be easily proved using Jensen's Inequality and convexity.

Let ν be some other probability distribution over Ω. We define the *total variation distance between μ and ν* by

$$\|\mu - \nu\|_{\mathsf{TV}} \stackrel{\mathsf{def}}{=} \frac{1}{2}\|\mu - \nu\|_1 = \frac{1}{2}\sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

A *coupling π of μ with ν* is a probability distribution over the product space $\Omega \times \Omega$ such that the marginals of π are given precisely by μ, ν, i.e.

$$\sum_{y \in \Omega} \pi(x, y) = \mu(x)$$
$$\sum_{x \in \Omega} \pi(x, y) = \nu(y)$$

for all $x, y \in \Omega$. We sometimes write $\Pi(\mu, \nu)$ for the space of all couplings of μ, ν.

Lemma 1.4.13 (Variational Representations of Total Variation Distance)**.** *1. We have the identities*

$$\|\mu - \nu\|_{\mathsf{TV}} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)| = \sup_{f:\Omega \to [0,1]} |\mathbb{E}_\mu(f) - \mathbb{E}_\nu(f)|.$$

2. Coupling Lemma (see e.g. [LPW17]): For every coupling π of μ, ν, we have the inequality

$$\Pr_{(x,y)\sim\pi}[x \neq y] \geq \|\mu - \nu\|_{\mathsf{TV}}.$$

Furthermore, the optimal coupling minimizing the left-hand side achieves equality, i.e.

$$\|\mu - \nu\|_{\mathsf{TV}} = \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \Pr_{(x,y)\sim\pi}[x \neq y] \right\}.$$

We will also need a few concepts from the theory of *optimal transport*. Assume our state space Ω is further endowed with a metric $d(\cdot, \cdot)$. Similar to the variational representations of total variation distance, we define the 1-*Wasserstein distance (or earth-mover distance) between μ and ν w.r.t. $d(\cdot, \cdot)$* to be

$$\mathscr{W}_1(\mu, \nu) \stackrel{\mathsf{def}}{=} \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \mathbb{E}_{(x,y)\sim\pi}[d(x, y)] \right\}. \tag{1.1}$$

This is an example of a *transportation distance*.

Theorem 1.4.14 (Kantorovich–Rubinstein Duality; see e.g. [Vil09]). *We may equivalently express the* 1*-Wasserstein distance between two probability measures* μ, ν *on a common metric space* (Ω, d) *as*

$$\mathscr{W}_1(\mu,\nu) = \sup_{f:\Omega\to\mathbb{R}} |\mathbb{E}_\mu(f) - \mathbb{E}_\nu(f)|,$$

where the supremum is taken over all functions $f : \Omega \to \mathbb{R}$ *which are* 1*-Lipschitz w.r.t.* d*, i.e. for all* $x, y \in \Omega$*, we have the inequality* $|f(x) - f(y)| \le d(x, y)$*.*

Remark 2. With this, one can actually view the total variation distance as the 1-Wasserstein distance w.r.t. the *discrete metric* on Ω, defined as $d(x,y) \stackrel{\text{def}}{=} \mathbb{I}[x \neq y]$. With this perspective, the Coupling Lemma from Lemma 1.4.13 is essentially a special case of duality, i.e. Theorem 1.4.14.

Finally, we will need to compare the relative entropy with the total variation distance.

Proposition 1.4.15 ([Pin64]; see also [Tsy09]). *For all probability measures* μ, ν *on a common finite set* Ω*, we have the inequality*

$$\|\mu - \nu\|_{\mathsf{TV}} \le \sqrt{\frac{1}{2}\mathscr{D}_{\mathrm{KL}}(\nu\|\mu)}.$$

1.4.4 Approximate Counting and Sampling

In this subsection, we briefly define what we mean by "efficient" approximate counting and approximate counting in a complexity-theoretic sense. We also mention the polynomial-time equivalence between approximate counting and sampling due to [JVV86].

Definition 3 (FPRAS (Slightly Informal); see e.g. [JVV86]). *Fix a nonempty finite alphabet* Σ *(e.g.* $\Sigma = \{0,1\}$*), and let* $f : \Sigma^* \to \mathbb{R}_{\ge 0}$ *be a nonnegative function on the set of strings of any length with letters in* Σ*. A* ***fully-polynomial randomized approximation scheme (***FPRAS*****)** *for estimating* f *is a randomized algorithm which, given an input/instance* $x \in \Sigma^n$*, an accuracy parameter* $\epsilon > 0$*, and a failure probability tolerance* $0 < \delta < 1$*, outputs a (random) number* Z *such that*

$$\Pr\left[(1-\epsilon)\cdot f(x) \le Z \le (1+\epsilon)\cdot f(x)\right] \ge 1 - \delta$$

in time $\mathsf{poly}\left(n, \frac{1}{\epsilon}, \log\frac{1}{\delta}\right)$*.*

For our purposes, one should think of each x as implicitly encoding some (typically exponentially large) state space Ω, and f as encoding a counting problem over that state space. For instance, a bit string $x \in \{0,1\}^{\binom{n}{2}}$ could encode an undirected graph $G = (V, E)$ on n vertices, and $f(x)$ could count the total number of independent sets in that graph. Here, the exponentially large state space Ω would simply be the collection of all independent sets.

Definition 4 (FPAUS (Slightly Informal); see e.g. [JVV86]). *Given a probability distribution* μ *over a finite state space* Ω*, encoded as, say, a binary string* $x \in \{0,1\}^n$*, a* ***fully-polynomial almost uniform sampler (***FPAUS*****)** *for* μ *is a randomized algorithm which, given* x *and an error tolerance* $0 < \delta < 1$*, outputs a random element* $\omega \in \Omega$ *drawn according to another distribution* ν *on* Ω *satisfying* $\|\mu - \nu\|_{\mathsf{TV}} \le \delta$ *in time* $\mathsf{poly}\left(n, \log\frac{1}{\delta}\right)$*.*

Again, one should think of x as encoding some probability distribution on an exponentially large domain.

The use of total variation distance for discrete probability distributions is very natural. For instance, it guarantees that for every bounded real-valued function f, the expectation $\mathbb{E}_\nu(f)$ is within an $O(\delta)$ additive error of the true statistic $\mathbb{E}_\mu(f)$; this is by Lemma 1.4.13. In particular, when combined with standard concentration inequalities, an FPAUS allows one to efficiently estimate the expectation of any bounded function f on Ω up to small additive error; one just needs to draw enough approximate samples $\omega_1, \ldots, \omega_T$ from ν via the FPAUS, and output the *empirical mean* $\frac{1}{T}\sum_{i=1}^T f(\omega_i)$ to estimate $\mathbb{E}_\nu(f)$, which is a close approximation of $\mathbb{E}_\mu(f)$.

One can intuitively think of the above intuition as hinting at a connection between approximate counting and approximate sampling. However, one significant issue is that the above only yields an additive approximation; obtaining a multiplicative approximation requires special care. Now, we informally state a remarkable result which shows the equivalence between the existence of an FPRAS for counting and the existence of an FPAUS.

Theorem 1.4.16 (Approximate Counting $\iff$ Approximate Sampling (Informal); see [JVV86])**.** *For "self-reducible" problems, there exists an* FPRAS *for estimating the partition function* $\mathcal{Z} = \sum_{\omega\in\Omega} c(\omega)$ *of a nonnegative weight function* $c : \Omega \to \mathbb{R}_{\geq 0}$ *on a finite state space* Ω *if and only if there exists an* FPAUS *for sampling the distribution* μ *on* Ω *defined by* $\mu(\omega) \propto c(\omega)$ *for all* $\omega \in \Omega$*.*

This is a beautiful landmark result in the theory of approximate counting and sampling. In this thesis, all of our problems are "self-reducible", since we will consider not only sampling from the distribution μ itself, but as well as all of its "conditional distributions"; we will formalize the type of conditioning we will use later on. The key point here is that Theorem 1.4.16 allows us to focus solely on the approximate sampling problem. Developing more efficient reductions between approximate counting and approximate sampling is also a very interesting problem, but we won't discuss this in the remainder of this thesis. We refer interested readers to [ŠVV09; Hub15; Kol18; HK21] and references therein for recent developments in this direction.

1.4.5 Basics of Markov Chains

In this thesis, we will obtain FPAUS (and hence, FPRAS) for many problems by running Markov chains. Here, we lay out the basic concepts from the classical theory of Markov chains that we will need. We refer the reader to [LPW17] for more background.

Let Ω be a finite state space. A *(discrete-time) Markov chain with state space* Ω is a sequence of random variables $\{X_t\}_{t=0}^{\infty} = X_0, X_1, \ldots, X_t, \ldots$ taking values in Ω satisfying the *Markov property*, namely that

$$\Pr[X_t = x_t \mid X_0 = x_0, \ldots, X_{t-1} = x_t] = \Pr[X_t = x_t \mid X_{t-1} = x_{t-1}] \tag{1.2}$$

holds under the joint distribution of $\{X_t\}_{t=0}^{\infty}$ for all $t \geq 0$ and all $x_0, \ldots, x_t \in \Omega$. In other words, the distribution of the next state X_t is independent of the history $X_0, \ldots, X_{t-2}$ given the immediately preceding state X_{t-1}. One should conceptually imagine the Markov chain $X_0, X_1, \ldots, X_t, \ldots$ as a *random walk* through Ω, where the *transition* from X_{t-1} to X_t constitutes one step of the Markov chain.

Throughout this thesis, our Markov chains will be *time-homogeneous*, in the sense that the transition probabilities $\Pr[X_t = y \mid X_{t-1} = x]$ are independent of the time step $t \geq 0$. In particular, the Markov chain will be generated by a *transition probability matrix* (or *Markov kernel*) $\mathsf{P} \in \mathbb{R}_{\geq 0}^{\Omega\times\Omega}$, along with an *initial distribution* $\mu_0 \in \mathbb{R}_{\geq 0}^{\Omega}$ over Ω. The entries $\mathsf{P}(x, y)$ of P (sometimes written $\mathsf{P}(x \to y)$) specify the probability of transitioning from state $X_{t-1} = x$ to state $X_t = y$ for every step $t \geq 0$, while μ_0 specifies the distribution of X_0. Thus, each row of P is a probability distribution over Ω, i.e. it is a *(row-)stochastic matrix*. The distribution μ_t of X_t over Ω may thus be described linear algebraically by the identity $\mu_t = \mu_0\mathsf{P}^t$ (if we view μ_0 and μ_t as row vectors). Most of the time, the most important part of a Markov chain is its transition probability matrix P. This is where most of our discussions and analyses will be centered on, and so we'll often abuse notation and refer to P as the Markov chain itself (even though technically, we also need to specify an initial distribution μ_0).

For a transition probability matrix P, we say a distribution μ on Ω is *stationary w.r.t.* P if $\mu\mathsf{P} = \mu$, i.e. $\mu(y) = \sum_{x\in\Omega} \mu(x) \cdot \mathsf{P}(x \to y)$ for all $y \in \Omega$. We sometimes also say μ is a stationary (or *equilibrium*) distribution of the Markov chain. In other words, the distribution μ is preserved under the action of P. A sufficient condition for μ to be stationary w.r.t. P is for μ and P to satisfy the *detailed balanced condition*

$$\mu(x) \cdot \mathsf{P}(x \to y) = \mu(y) \cdot \mathsf{P}(y \to x), \quad \forall x, y \in \Omega. \tag{1.3}$$

In this case, we say P is *reversible (w.r.t. μ)*. Throughout this thesis, all of our Markov chains will be reversible. This reversibility condition can alternatively be interpreted linear algebraically as saying the matrix P is self-adjoint w.r.t. the inner product $\langle\cdot,\cdot\rangle_\mu$ on $\mathbb{R}^\Omega$ induced by the distribution μ. In particular, P has all real eigenvalues.

Combinatorially, a Markov chain P which is reversible w.r.t. μ may also be viewed as a *simple random walk* on a weighted undirected graph $G_{\mathsf{P}} = (V = \Omega, E, c : E \to \mathbb{R}_{\geq 0})$, where the vertices correspond to elements of Ω, the set of edges is given by $E = \{\{x, y\} : \mathsf{P}(x \to y) > 0\}$, and the edge weights are given by $c(x, y) = \mu(x) \cdot \mathsf{P}(x \to y) = \mu(y) \cdot \mathsf{P}(y \to x)$ for every edge $\{x, y\} \in E$. This will sometimes be referred to as the *underlying (weighted) graph of* P. Conversely, if $G = (V, E, c : E \to \mathbb{R}_{\geq 0})$ is an undirected graph with nonnegative edge weights, then the *simple*

random walk P_G *on* G, which transitions from a vertex u to a neighbor v with probability $\propto c(u,v)$, is a Markov chain which is reversible w.r.t. the *weighted degree distribution* μ on V given by $\mu(u) \propto \sum_{e \in E : e \sim u} c(e) \stackrel{\text{def}}{=} c(u)$ for all $u \in V$. As a matrix, we may write $\mathsf{P}_G = D_G^{-1} A_G$, where $D_G = \text{diag}(c(u) : u \in V)$ is the *weighted degree matrix* of G, and A_G is the *weighted adjacency matrix* of G.

Definition 5 (Ergodicity)**.** *Fix a (reversible) Markov chain* P*. We say* P *is* ***irreducible*** *if for all* $x, y \in \Omega$*, there exists* $t \geq 0$ *such that* $\mathsf{P}^t(x \to y) > 0$*. In other words, there is a nonzero probability of walking from any starting state* x *to any other ending state* y*. The* ***period*** *of a state* $x \in \Omega$ *under* P *is defined as the greatest common divisor of* $\{t \geq 1 : \mathsf{P}^t(x,x) > 0\}$*. We say* P *is* ***aperiodic*** *if all states have period* 1*. We say* P *is* ***ergodic*** *if* P *is both irreducible and aperiodic.*

Equivalently, a reversible Markov chain P is irreducible (resp. aperiodic) if the underlying graph of P is connected (resp. not bipartite). Every periodic Markov chain P can be made into an "equivalent" aperiodic Markov chain by consider $\frac{\mathsf{Id}+\mathsf{P}}{2}$, where Id is the $\Omega \times \Omega$ identity matrix. This is sometimes called the *lazification* of P, since in each step, there is a $\frac{1}{2}$-probability of not moving to a state different from the current one.

Theorem 1.4.17 (Fundamental Theorem of Markov Chains; see e.g. [LPW17])**.** *Let* P *be an ergodic Markov chain on a state space* Ω*. Then* P *has a unique stationary distribution* μ *on* Ω*. Furthermore, for any initial distribution* μ_0*, the distribution* $\mu_t = \mu_0 \mathsf{P}^t$ *of* X_t *converges (pointwise) to* μ *as* $t \to \infty$*.*

This theorem highlights the relevance of Markov chains to sampling. Suppose you want to sample from some complicated probability distribution μ on some complicated state space Ω. If you could design an ergodic Markov chain P with μ as its equilibrium distribution such the transitions of P are efficiently implementable, then by selecting an arbitrary initial distribution μ_0 (or even selecting a single starting state $x_0 \in \Omega$ deterministically), you can efficiently sample from μ by simulating the dynamics described by P started at μ_0 (or x_0) for sufficiently long and outputting the final state X_T as your sample. Theorem 1.4.17 guarantees that if T is sufficiently large, then the distribution of X_T is "close" to μ. Exactly how large T needs to be is the subject of the next subsection.

Mixing Times Here, we define precisely what we mean by mixing time of a Markov chain. The goal is quantify precisely how quickly an ergodic Markov chain converges equilibrium. This is crucial since it directly controls the efficiency of our Markov chain sampling algorithms, as well as the "accuracy" of the samples we get out.

Definition 6 ((Total Variation) Mixing Time)**.** *Fix a reversible ergodic Markov chain* P *with stationary distribution* μ *on a state space* Ω*. We define the* ***(total variation) mixing time of*** P ***with error*** $\epsilon > 0$ ***(and initial distribution*** μ_0***)*** *by*

$$\mathrm{T}_{\mathsf{mix}}(\epsilon; \mu_0, \mathsf{P}) \stackrel{\text{def}}{=} \min\left\{t \geq 0 : \left\|\mu_0 \mathsf{P}^t - \mu\right\|_{\mathsf{TV}} \leq \epsilon\right\}$$
$$\mathrm{T}_{\mathsf{mix}}(\epsilon; \mathsf{P}) \stackrel{\text{def}}{=} \sup_{\mu_0} \mathrm{T}_{\mathsf{mix}}(\epsilon; \mu_0, \mathsf{P}).$$

When the Markov chain P *is clear from context, we drop the* P*, and write* $\mathrm{T}_{\mathsf{mix}}(\epsilon; \mu_0)$ *and* $\mathrm{T}_{\mathsf{mix}}(\epsilon)$*, respectively. We define the total variation mixing time of* P *to be* $\mathrm{T}_{\mathsf{mix}} \stackrel{\text{def}}{=} \mathrm{T}_{\mathsf{mix}}(1/4)$*.*

The constant $1/4$ is arbitrary, and can be chosen to be any constant less than $1/2$. We define it this way because $\mathrm{T}_{\mathsf{mix}}(\epsilon) \leq \mathrm{T}_{\mathsf{mix}} \cdot \log(1/\epsilon)$; see e.g. [LPW17]. Note that one can define a version of mixing time with respect to any metric on probability measures. Total variation distance is the most standard in the literature, since convergence in total variation implies convergence w.r.t. every bounded statistic.

A Markov chain whose mixing time is bounded by a polynomial in the size of the problem input is said to be *rapidly mixing* (or *fast mixing*). Otherwise, it is *torpidly mixing* (or *slow mixing*). In the context of sampling from a distribution μ, the goal is then to construct a rapidly mixing Markov chain P whose transitions can be implemented efficiently.

1.4.6 Markov Chain Mixing via Functional Analytic Techniques

Our goal here is to state particularly useful tools in the analysis of mixing times, which appeal to ideas from *functional analysis*. For instance, through the connection between spectral gaps and mixing times, we will see one example of the beautiful interplay between probability theory and linear algebra.

Mixing via Eigenvalues It is well-known that the largest eigenvalue in magnitude of any stochastic matrix is 1, with corresponding right eigenvector being all-ones vector $\mathbf{1}$. Write $-1 \leq \lambda_{|\Omega|}(\mathsf{P}) \leq \cdots \leq \lambda_2(\mathsf{P}) \leq \lambda_1(\mathsf{P}) = 1$ for the eigenvalues of P, which are real if P is reversible. We write $\gamma(\mathsf{P}) \stackrel{\text{def}}{=} 1-\lambda_2(\mathsf{P})$ for the *spectral gap* of P. Let $\lambda^*(\mathsf{P}) \stackrel{\text{def}}{=} \max_{i>1}\{\lambda_i(\mathsf{P})\} = \max\{\lambda_2(\mathsf{P}), |\lambda_n(\mathsf{P})|\}$ be the second largest eigenvalue of P in absolute value, and write $\gamma^*(\mathsf{P}) \stackrel{\text{def}}{=} 1-\lambda^*(\mathsf{P})$ for the *absolute spectral gap* of P. Most of the Markov chains P we analyze in this thesis will have the convenient property that P is positive semidefinite, i.e. all of its eigenvalues are nonnegative. In this case, $\lambda^*(\mathsf{P}) = \lambda_2(\mathsf{P})$ and $\gamma^*(\mathsf{P}) = \gamma(\mathsf{P})$.

With these notions in hand, we have the following well-known fact.

Fact 1.4.18. *A reversible Markov chain P is irreducible/connected if and only if $\lambda_2(\mathsf{P}) < 1$, or equivalently, $\gamma(\mathsf{P}) > 0$.*

Later on, we will need a "quantitative" version of this fact, which says that a reversible Markov chain P is "very well-connected" if and only if $\gamma(\mathsf{P})$ is "large". For now, we focus on the relevance of $\gamma(\mathsf{P})$ to rapid mixing. The following result is folklore.

Theorem 1.4.19 (Spectral Gap Implies Rapid Mixing; see e.g. [LPW17])**.** *Let P be a reversible ergodic Markov chain with stationary distribution μ on a state space Ω. Then for every $\epsilon > 0$, the following inequality holds*

$$\mathrm{T}_{\mathsf{mix}}(\epsilon) \leq \frac{1}{\gamma^*(\mathsf{P})}\left(\frac{1}{2}\log\frac{1}{\mu_{\min}} + \log\frac{1}{2\epsilon}\right),$$

where recall that $\mu_{\min} = \min_{x\in\Omega:\mu(x)>0}\mu(x)$.

This is one of the basic tools in the theory of Markov chain mixing times which we will appeal to repeatedly throughout this thesis. We will also occasionally appeal to the following convenient inequality when performing spectral analyses of Markov chains.

Lemma 1.4.20. *Let P be an reversible Markov chain on a finite state space Ω with stationary distribution μ. Then we have the following spectral inequality*

$$\mathsf{P} - \mathbf{1}\otimes\mu \preceq_\mu \lambda_2(\mathsf{P})\cdot(\mathsf{Id} - \mathbf{1}\otimes\mu).$$

or equivalently,

$$\mathsf{Id} - \mathsf{P} \succeq_\mu \gamma(\mathsf{P})\cdot(\mathsf{Id} - \mathbf{1}\otimes\mu).$$

Proof. One can rewrite the first inequality as $\mathsf{P} - (1-\lambda_2(\mathsf{P}))\cdot\mathbf{1}\otimes\mu \preceq_\mu \lambda_2(\mathsf{P})\cdot\mathsf{Id}$. This inequality is very intuitive. It says that after shifting the top eigenvalue 1 down to $\lambda_2(\mathsf{P})$, the largest eigenvalue of the resulting matrix is $\lambda_2(\mathsf{P})$. This shifting is achieved just by subtracting a multiple of the rank-1 matrix $\mathbf{1}\otimes\mu$ representing the top eigenspace of P.

Let us now give a more formal argument. First, note that the inequality is vacuously true if P is not irreducible since $\lambda_2(\mathsf{P}) = 1$. Hence, we may assume that P is irreducible in the rest of the proof. Consider the symmetric matrix $\mathrm{diag}(\mu)^{1/2}\mathsf{P}\,\mathrm{diag}(\mu)^{-1/2}$[10], which is similar to P and hence has the same eigenvalues. By the Spectral Theorem, $\mathrm{diag}(\mu)^{1/2}\mathsf{P}\,\mathrm{diag}(\mu)^{-1/2}$ admits a rank-1 decomposition $\sum_{i=1}^{|\Omega|}\lambda_i(\mathsf{P})v_iv_i^\top$ where $\{v_i\}_{i=1}^{|\Omega|}$ is an orthonormal basis of eigenvectors. Note that $\mathrm{diag}(\mu)^{1/2}\mathsf{P}\,\mathrm{diag}(\mu)^{-1/2}\sqrt{\mu} = \mathrm{diag}(\mu)^{1/2}\mathsf{P}\mathbf{1} = \mathrm{diag}(\mu)^{1/2}\mathbf{1} = \sqrt{\mu}$ shows that $\sqrt{\mu}$ is an eigenvector with eigenvalue 1. Since P is irreducible, the eigenvector with eigenvalue 1 is unique up to scaling, and so $v_1 = \sqrt{\mu}$.

[10] If P is the simple random walk on a graph, then this is also called the *normalized adjacency matrix* in the spectral graph theory literature.

Now, converting to symmetric matrices, the desired inequality is equivalent to

$$\operatorname{diag}(\mu)\mathsf{P} - (1-\lambda_2(\mathsf{P}))\cdot\mu\mu^\top \preceq \lambda_2(\mathsf{P})\cdot\operatorname{diag}(\mu),$$

which can be rewritten as

$$\operatorname{diag}(\mu)^{1/2}\cdot\left(\operatorname{diag}(\mu)^{1/2}\mathsf{P}\operatorname{diag}(\mu)^{-1/2} - (1-\lambda_2(\mathsf{P}))\sqrt{\mu}\sqrt{\mu}^\top\right)\cdot\operatorname{diag}(\mu)^{1/2}$$
$$\preceq \lambda_2(\mathsf{P})\cdot\operatorname{diag}(\mu)^{1/2}\cdot\mathsf{Id}\cdot\operatorname{diag}(\mu)^{1/2}.$$

By viewing $\operatorname{diag}(\mu)^{1/2}$ as a change of coordinates, this is equivalent to

$$\operatorname{diag}(\mu)^{1/2}\mathsf{P}\operatorname{diag}(\mu)^{-1/2} - (1-\lambda_2(\mathsf{P}))\sqrt{\mu}\sqrt{\mu}^\top \preceq \lambda_2(\mathsf{P})\cdot\mathsf{Id}.$$

The left-hand side is nothing but $\lambda_2(\mathsf{P})v_1v_1^\top + \sum_{i=2}^{|\Omega|}\lambda_i(\mathsf{P})v_iv_i^\top$ using the orthonormal eigendecomposition of $\operatorname{diag}(\mu)^{1/2}\mathsf{P}\operatorname{diag}(\mu)^{-1/2}$. Since $\lambda_i(\mathsf{P}) \le \lambda_2(\mathsf{P})$ for all $2 \le i \le |\Omega|$ and $\{v_i\}_{i=1}^{|\Omega|}$ is an orthonormal basis (so that $\mathsf{Id} = \sum_{i=1}^{|\Omega|} v_iv_i^\top$), the inequality holds and we are done. □

The Dirichlet Form A useful functional analytic way to view the spectral gap is as follows. For two real-valued functions $f, g : \Omega \to \mathbb{R}$, define the *Dirichlet form* as the following bilinear form:

$$\mathcal{E}_{\mathsf{P}}(f,g) \stackrel{\mathsf{def}}{=} \langle f, (I-\mathsf{P})g\rangle_\mu = \frac{1}{2}\sum_{x,y\in\Omega}\mu(x)\mathsf{P}(x\to y)\cdot(f(x)-f(y))(g(x)-g(y)). \tag{1.4}$$

The *Dirichlet energy* of a function $f : \Omega \to \mathbb{R}$ is defined as $\mathcal{E}_{\mathsf{P}}(f,f)$. We have the following variational characterization of the spectral gap, which is essentially a special case of Theorem 1.4.1.

Fact 1.4.21. *Let P be a Markov chain which is reversible w.r.t. a distribution μ on a finite state space Ω. Then we have the identity*

$$\gamma(\mathsf{P}) = \inf_{f:\Omega\to\mathbb{R}}\left\{\frac{\mathcal{E}_{\mathsf{P}}(f,f)}{\operatorname{Var}_\mu(f)} : \operatorname{Var}_\mu(f) \neq 0\right\}.$$

This identity is essentially a consequence of Theorem 1.4.1. Since inequalities of the form $\mathcal{E}_{\mathsf{P}}(f,f) \ge \gamma\operatorname{Var}_\mu(f)$ are often called *Poincaré Inequalities*, the spectral gap $\gamma(\mathsf{P})$ of P is also called the *Poincaré constant of* P. We will use these terms interchangeably.

1.4.7 Combinatorial Structures

In this subsection, we briefly define the main combinatorial structures we will be working with in this thesis. These structures have all be studied in multiple communities for decades.

Matroids

In several chapters, we will study certain abstract combinatorial structures called *matroids*. We will use a few of the many *cryptomorphic definitions* of a matroid; we already saw one such definition earlier in Definition 2. We first state the standard definition in terms of basis exchange. For the equivalence with other prominent definitions of a matroid, and more generally references to facts stated here see [Oxl11].

Definition 7 (Matroid; Basis Exchange Definition)**.** *Fix a finite set/universe $\mathscr{U}$. We say that a family $\mathscr{B} \subseteq \binom{\mathscr{U}}{r}$ is the family of **bases** of a matroid $\mathcal{M}$ if it satisfies the following **basis exchange property**: For every $A, B \in \mathscr{B}$ and every $a \in A \setminus B$, there exists $b \in B \setminus A$ such that $(A \setminus \{a\}) \cup \{b\} \in \mathscr{B}$.*

One useful geometric definition is in terms of the polytope of its bases.

Theorem 1.4.22 (Matroid; Geometric Definition; [Gel+87])**.** *Fix a finite set/universe $\mathscr{U}$. A family $\mathscr{B} \subseteq \binom{\mathscr{U}}{r}$ is the family of **bases** of a matroid $\mathcal{M}$ if and only if all edges of the polytope $\operatorname{conv}\{\mathbf{1}_B : B \in \mathscr{B}\}$ are parallel to a vector of the form $\mathbf{1}_i - \mathbf{1}_j$ for some $i \neq j$ in $\mathscr{U}$. In other words, all edges of the polytope have the minimum possible length, namely $\sqrt{2}$. This polytope is called the **(matroid) basis polytope** of the matroid with bases $\mathscr{B}$.*

For a collection of bases $\mathscr{B} \subseteq \binom{\mathscr{U}}{r}$, we call r the *rank* of the matroid, and $\mathscr{U}$ the *ground set* of the matroid. The family of *independent sets* of the matroid is then

$$\mathcal{X} \stackrel{\text{def}}{=} \{S \subseteq \mathscr{U} : \exists B \in \mathscr{B} \text{ s.t. } S \subseteq B\}$$

i.e. $\mathcal{X}$ is the downwards closure of $\mathscr{B}$. We often write $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ for the matroid. The subsets of $\mathscr{U}$ not in $\mathcal{X}$ are called *dependent*. We say an element $i \in \mathscr{U}$ is a *loop* if $\{i\}$ is dependent. We say two distinct elements $i, j \in \mathscr{U}$ are *parallel* (to each other), if $\{i, j\}$ is dependent.

Associated to each matroid is a *rank function* $\text{rank}_{\mathcal{M}} : 2^{\mathscr{U}} \to \mathbb{N}$ defined as

$$\text{rank}_{\mathcal{M}}(S) = \max\{|T| : T \subseteq S, T \in \mathcal{X}\}.$$

We sometimes simply write $r_{\mathcal{M}}(\cdot)$ for the rank function. When the matroid $\mathcal{M}$ is clear from context, we often drop the subscript.

A well-known fact about matroids, that can be easily derived from Theorem 1.4.22, is that the *dual* of a matroid, defined below, is another matroid.

Proposition 1.4.23 (Matroid Duality)**.** *If $\mathscr{B} \subseteq \binom{\mathscr{U}}{r}$ is the family of bases of a matroid, then the following is also the family of bases of another matroid, called the* ***dual matroid****:*

$$\mathscr{B}^* \stackrel{\text{def}}{=} \{\mathscr{U} \setminus B : B \in \mathscr{B}\} \subseteq \binom{\mathscr{U}}{|\mathscr{U}| - r}.$$

We write $\mathcal{M}^*$ for the dual matroid with bases $\mathscr{B}^*$. The rank function $\text{rank}^*_{\mathcal{M}} = \text{rank}_{\mathcal{M}^*} : 2^{\mathscr{U}} \to \mathbb{N}$ can be written in terms of $\text{rank}_{\mathcal{M}}$ as

$$\text{rank}^*_{\mathcal{M}}(S) = \text{rank}_{\mathcal{M}}(\mathscr{U} \setminus S) + |S| - \text{rank}_{\mathcal{M}}(\mathscr{U}). \tag{1.5}$$

One can also find within a matroid $\mathcal{M}$ smaller "submatroids". These are induced by "deletion" and "contraction" operations. For a matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ and a subset of elements $S \subseteq \mathscr{U}$, we define the *deletion* of S from $\mathcal{M}$ as the matroid

$$\mathcal{M} \setminus S \stackrel{\text{def}}{=} (\mathscr{U} \setminus S, \{T \in \mathcal{X} : T \subseteq \mathscr{U} \setminus S\}).$$

This is also sometimes called the *restriction* of $\mathcal{M}$ to $\mathscr{U} \setminus S$. If $S \in \mathcal{X}$, i.e. S is an independent set, then we define the *contraction* of $\mathcal{M}$ w.r.t. S as the matroid

$$\mathcal{M}/S \stackrel{\text{def}}{=} (\mathscr{U} \setminus S, \{T \setminus S : T \in \mathcal{X}, T \supseteq S\}).$$

If $S \subseteq \mathscr{U}$ is an arbitrary set which isn't necessarily independent, we define the contraction $\mathcal{M}/S$ of $\mathcal{M}$ w.r.t. S to be the contraction $\mathcal{M}/I$ for any maximal independent set $I \subseteq S$. Finally, for a matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ and $0 \leq k \leq \text{rank}(\mathcal{M})$, we define the *rank-$k$ truncation* of $\mathcal{M}$ as the matroid $\mathcal{M}_k \stackrel{\text{def}}{=} (\mathscr{U}, \{S \in \mathcal{X} : |S| \leq k\})$.

Having defined these useful matroid operations, we now turn to classical examples of matroids, some of which will be important for us. The first class we will consider comes from graphs, and will be the primary example for us. These are called *graphic matroids*.

Proposition 1.4.24 (Graphic Matroid)**.** *Let $G = (V, E)$ be a graph. Then the following is the family of bases $\mathscr{B}$ of a matroid $\mathcal{M}$ with ground set E called the* ***graphic matroid of*** *G:*

$$\mathscr{B} \stackrel{\text{def}}{=} \{T \subseteq E : T \textit{ forms a spanning forest}\}.$$

Note that the rank of the graphic matroid is $\leq |V| - 1$ and the ground set is E. If G is connected, then the bases are spanning trees of G, and the rank is exactly equal to $|V| - 1$.

Graphic matroids are a special case of a much more general class of matroids called *linear matroids*.

Proposition 1.4.25 (Linear Matroid)**.** *Fix a vector space V over an arbitrary field $\mathbb{F}$, and let $E \subseteq V$ be a finite subset of vectors. Then the following is the family of bases $\mathscr{B}$ of a matroid $\mathcal{M}$ with ground set E called the* ***linear matroid of*** *E:*

$$\mathscr{B} \stackrel{\text{def}}{=} \{T \subseteq E : T \textit{ linearly independent and maximal}\}.$$

Matroids which can be realized as the linear matroid of some finite set of vectors in a vector space over a field $\mathbb{F}$ are called *representable over* $\mathbb{F}$ (or *$\mathbb{F}$-representable*).

Spin Systems

Here, we define spin systems on vertices of graphs, which encompass numerous classically studied combinatorial structures in computer science. In Chapter 6 and Appendix D, we will also consider other combinatorial structures defined on edges of graphs, as opposed to vertices. For now, we content ourselves with vertex spin systems.

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

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

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

$$\mathcal{Z}_G(A, h) = \sum_{\sigma : V \to [q]} \prod_{\{u,v\} \in E} A(\sigma(u), \sigma(v)) \prod_{v \in V} h(\sigma(v)).$$

One can of course further generalize these models by having a separate edge interaction matrix A_e for each edge $e = \{u, v\} \in E$ and a separate positive vector h_v of external fields for every vertex $v \in V$. However, we won't need this level of generality for this discussion. Typically, we consider the parameters A, h as fixed in the definition of the sampling problem, while the graph G is viewed as the input to the algorithm.

Many classical models in statistical physics as well as distributions over combinatorial objects on graphs may be found as special cases of spin systems:

1. *Ising Model of Magnetism (i.e. Cuts)*: $A = \begin{bmatrix} e^{\beta} & 1 \\ 1 & e^{\beta} \end{bmatrix}$ and $h > 0$ is an external magnetic field
2. *Hardcore Gas Model (i.e. Independent Sets)*: $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ and $h = \lambda \mathbf{1}$ where $\lambda > 0$
3. *Monomer-Dimer Model (i.e. Matchings)*: $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ and $h = \lambda \mathbf{1}$ where $\lambda > 0$ (the same parameters as the hardcore model), with the restriction that G is a line graph
4. *Zero-Temperature Antiferromagnetic Potts Model (i.e. Proper q-Colorings)*: $A = J_q - I_q$ and $h = \mathbf{1}$, where J_q is the $q \times q$ all-1s matrix, and I_q is the $q \times q$ identity matrix

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

Chapter 2

Spectral Independence and Rapid Mixing of High-Dimensional Walks

In this chapter, we introduce our notion of *spectral independence*, and show how it implies rapid mixing of a class of simple and natural Markov chains walking around a high-dimensional discrete space. Our notion of spectral independence is a manifestation of the following philosophy stated in Chapter 1:

> *If the probability distribution of interest μ satisfies a "limited" or "structured" correlations property, then a simple Markov chain for sampling from μ mixes rapidly.*

We will see later how the connection between spectral independence and rapid mixing can be tightened, as well as techniques for establishing spectral independence.

2.1 High-Dimensional Discrete Distributions and Walks

Before we talk about spectral independence, let us first be clear on what we mean by a high-dimensional discrete probability distribution. We'll then define what Markov chains we'll use for sampling, and establish some basic but useful properties. Along the way, we'll see famous examples of discrete distributions and associated Markov chain samplers which are captured by our framework.

General Set-Up Let $\mathscr{U}$ be a finite *ground set* (or *universe*), and let $0 \leq n \leq |\mathscr{U}|$ be a "dimension" parameter, which we think of as growing to infinity (hence, "high-dimensional"). Given this, we will then take our probability distribution μ to be over the discrete collection $\binom{\mathscr{U}}{n}$ of size-n subsets of $\mathscr{U}$. We sometimes say our distributions are supported on *homogeneous set systems*. For convenience, we will say a subset $\tau \subseteq \mathscr{U}$ is *feasible* if $\tau \subseteq \sigma$ for some $\sigma \in \mathrm{supp}(\mu)$.

While this might seem strange at first sight, and indeed it isn't entirely standard in some communities, it will be convenient to formulate things this way since it will be sufficiently general to capture all of the discrete probability distributions we will be interested in e.g. distributions on matroids and Gibbs distributions of graphical models as discussed in the introduction. Here are some classically studied combinatorial structures on graphs which can be easily put in this framework.

Example 1 (Spanning Trees in Graphs). Fix an undirected connected graph $G = (V, E)$. Then the uniform distribution μ over spanning trees of G can be viewed as a distribution over $\binom{\mathscr{U}}{n}$ where $\mathscr{U} = E$ and $n = |V| - 1$. In this case, the feasible sets are the spanning forests of the graph, since any spanning forest can be completed into a spanning tree. This generalizes to any distribution over the bases of any matroid $\mathcal{M}$, where the feasible sets are exactly the independent sets in $\mathcal{M}$, and the maximal feasible sets are exactly the bases of $\mathcal{M}$.

Example 2 (Proper Colorings of Graphs). Fix an undirected graph $G = (V, E)$ and a positive integer $q \geq 1$. Then the uniform distribution μ over proper q-colorings of the vertices of G can be viewed as a distribution over $\binom{\mathscr{U}}{n}$ where $\mathscr{U} = \{(v, c) : v \in V, c \in [q]\}$ is the set of vertex-color (or vertex-assignment) pairs, and $n = |V|$. In this case, the feasible sets can be thought of as partial colorings on subsets of vertices which can be completed into a full valid coloring. This generalizes

to any distribution over a discrete product space $\Sigma_1 \times \cdots \times \Sigma_n$, where $\Sigma_1, \ldots, \Sigma_n$ are finite sets (e.g. the discrete hypercube $\{0,1\}^n$, or equivalently, $\{-1,+1\}^n$ or $2^{[n]}$).

Given this general set-up, we can now state the Markov chains that we will work with in this thesis. These Markov chains are very natural, and are examples of the Metropolis–Hastings algorithm (with a specific proposal distribution and acceptance probability). They will turn out to be highly local as well; each step will be easily implementable in linear or even sublinear time. Furthermore, they generalize several classes of famous Markov chains previously studied in the literature, including the bases exchange walk on matroid bases, the Glauber dynamics on configurations in graphical models, and the Kawasaki dynamics. They are sometimes called the *high-order*, *high-dimensional*, or *down-up* walks on the support of a high-dimensional discrete probability distribution [KM17; DK17; KO20b; AL20]. Throughout, we use the name *down-up walk*.

These down-up walks are so named because they are formed by first taking a *down step* (where elements are randomly removed and the cardinality of the set goes down), followed by an *up step* (where elements are added and the cardinality of the set goes up). Hence, we first define the down and up operators describing these steps. For the following definitions, we first fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$.

Definition 8 (Down Operators)**.** *For each $0 \le k \le n$, define the **down operator** $\mathcal{D}_\mu^{n \searrow k}$ to be the Markov operator acting on $\mathrm{supp}(\mu)$ by taking a given $\sigma \in \mathrm{supp}(\mu)$ and outputting a uniformly random subset $\sigma_k \subseteq \sigma$ with $|\sigma_k| = k$. This down operator induces the **(level-k) marginal distribution** μ_k on feasible size-k subsets, defined by $\mu_k \stackrel{\mathsf{def}}{=} \mu\mathcal{D}_\mu^{n \searrow k}$; note that $\mu_n = \mu$. We more generally define down operators $\mathcal{D}_\mu^{\ell \searrow k}$ for each $0 \le k \le \ell \le n$ via the same action, namely take a given $\sigma_\ell \in \mathrm{supp}(\mu_\ell)$ and output a uniformly random subset $\sigma_k \subseteq \sigma_\ell$ with $|\sigma_k| = k$. The entries of $\mathcal{D}_\mu^{\ell \searrow k} \in \mathbb{R}^{\mathrm{supp}(\mu_\ell) \times \mathrm{supp}(\mu_k)}$ are given by*

$$\mathcal{D}_\mu^{\ell \searrow k}(\sigma_\ell, \sigma_k) = \begin{cases} \frac{1}{\binom{\ell}{k}}, & \text{if } \sigma_k \subseteq \sigma_\ell \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

for every $\sigma_\ell \in \mathrm{supp}(\mu_\ell), \sigma_k \in \mathrm{supp}(\mu_k)$.

Clearly, $\mathcal{D}_\mu^{n \searrow k} = \mathcal{D}_\mu^{n \searrow \ell} \cdot \mathcal{D}_\mu^{\ell \searrow k}$ for every $0 \le k \le \ell \le n$. We emphasize that these down operators *do not* depend on the distribution μ one is analyzing, so one does not truly need the subscript μ. We will keep the subscript regardless just to make clear which distribution μ we are working with. The corresponding up operators we will define later will depend on the distribution μ.

We have the following useful and intuitive expressions for the entries of the distributions μ_k, which also justify why we call them "marginal" distributions.

Lemma 2.1.1. *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. Then for every $0 \le k \le n$, the entries of the distribution μ_k are given by*

$$\mu_k(\tau) = \frac{1}{\binom{n}{k}} \Pr_{\sigma \sim \mu}[\sigma \supseteq \tau], \quad \forall \tau \subseteq \mathscr{U}, |\tau| = k.$$

More generally, for every $0 \le k \le \ell \le n$, we have the identity

$$\mu_k(\sigma_k) = \frac{1}{\binom{\ell}{k}} \Pr_{\sigma_\ell \sim \mu_\ell}[\sigma_\ell \supseteq \sigma_k], \quad \forall \sigma_k \in \mathrm{supp}(\mu_k). \tag{2.2}$$

Proof. By definition, $\mu_k = \mu\mathcal{D}_\mu^{n \searrow k}$, and so $\mu_k(\tau)$ is precisely the probability of obtaining τ by first sampling a set σ according to μ and then independently selecting a uniformly random subset of size-k. This is possible if and only if the set $\sigma \sim \mu$ contains τ, in which case there is exactly $\frac{1}{\binom{n}{k}}$ probability of picking $\tau \subseteq \sigma$. The exact same line of reasoning yields the second, more general identity. □

Having defined the marginal distributions, we can now define the up operators, which admit a nice and succinct expression using Lemma 2.1.1.

Definition 9 (Up Operators)**.** *For each $0 \le k \le n$, define the **up operator** $\mathcal{U}_\mu^{k\nearrow n}$ to be the Markov operator acting on $\mathrm{supp}(\mu_k)$ by taking a given $\sigma_k \in \mathrm{supp}(\mu_k)$ and outputting a random superset $\sigma_n \supseteq \sigma_k$ in $\mathrm{supp}(\mu)$ with probability proportional to $\mu(\sigma_n)$. We more generally define up operators $\mathcal{U}_\mu^{k\nearrow\ell}$ for each $0 \le k \le \ell \le n$ via the same action, namely take a given $\sigma_k \in \mathrm{supp}(\mu_k)$ and output a random superset $\sigma_\ell \supseteq \sigma_k$ in $\mathrm{supp}(\mu_\ell)$ with probability proportional to $\mu_\ell(\sigma_\ell)$. The entries of $\mathcal{U}_\mu^{k\nearrow\ell} \in \mathbb{R}^{\mathrm{supp}(\mu_k)\times\mathrm{supp}(\mu_\ell)}$ are given by*

$$\mathcal{U}_\mu^{k\nearrow\ell}(\sigma_k,\sigma_\ell) = \begin{cases} \frac{\mu_\ell(\sigma_\ell)}{\binom{\ell}{k}\cdot\mu_k(\sigma_k)}, & \text{if } \sigma_\ell \supseteq \sigma_k \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

for every $\sigma_\ell \in \mathrm{supp}(\mu_\ell), \sigma_k \in \mathrm{supp}(\mu_k)$.

Again, it is clear that $\mathcal{U}_\mu^{k\nearrow n} = \mathcal{U}_\mu^{k\nearrow\ell}\cdot\mathcal{U}_\mu^{\ell\nearrow n}$ for every $0 \le k \le \ell \le n$. With the down and up operators in hand, we may now finally define the down-up (and up-down) random walks.

Definition 10 (Down-Up and Up-Down Walks)**.** *Given the down and up Markov operators, for every $0 \le k \le \ell \le n$, we may now define the $\ell \leftrightarrow k$ **down-up walk** as the Markov chain on $\mathrm{supp}(\mu_\ell)$ whose transition probability matrix is described by $\mathcal{D}_\mu^{\ell\searrow k}\mathcal{U}_\mu^{k\nearrow\ell}$. Similarly, we define the $k \leftrightarrow \ell$ **up-down walk** as the Markov chain on $\mathrm{supp}(\mu_k)$ whose transition probability matrix is described by $\mathcal{U}_\mu^{k\nearrow\ell}\mathcal{D}_\mu^{\ell\searrow k}$. When used without qualification, we typically refer to $\mathcal{P}_\mu \overset{\mathsf{def}}{=} \mathcal{D}_\mu^{n\searrow n-1}\mathcal{U}_\mu^{n-1\nearrow n}$ as "the" down-up walk, which has entries*

$$\mathcal{P}_\mu(\sigma\to\sigma') = \begin{cases} \frac{1}{n}\cdot\frac{\mu(\sigma')}{n\cdot\mu_{n-1}(\sigma\cap\sigma')}, & \text{if } |\sigma\cap\sigma'| = n-1 \\ \frac{1}{n}\sum_{u\in\sigma}\frac{\mu(\sigma)}{n\cdot\mu_{n-1}(\sigma-u)}, & \text{if } \sigma=\sigma' \\ 0, & \text{otherwise} \end{cases} \tag{2.4}$$

for every $\sigma,\sigma' \in \mathrm{supp}(\mu)$.

The down-up walk can alternatively be described as follows. Starting from a state $\sigma \in \mathrm{supp}(\mu)$, we randomly transition to another state σ' under $\mathcal{P}_\mu$ via the following simple two-step procedure:

1. Down-Step: Remove a uniformly random element $u \in \sigma$.
2. Up-Step: Out of all $v \in \mathscr{U}$ such that $\sigma - u + v \in \mathrm{supp}(\mu)$, pick one with probability proportional to $\mu(\sigma - u + v)$. Set $\sigma' = S - u + v$. Note that one can choose $v = u$, in which case $\sigma' = \sigma$.

The main goal of this thesis is to build tools to establish fast mixing of the down-up walk $\mathcal{P}_\mu$. Our tools will naturally extend to more general down-up walks (e.g. $\mathcal{D}_\mu^{n\searrow k}\mathcal{U}_\mu^{k\nearrow n}$ for some k), but our main interest will be in $\mathcal{P}_\mu$, as it is the simplest to describe and the easiest to implement in practice. Let us now see a few concrete manifestations of these walks.

Example 3 (The Exchange Walk on Spanning Trees of Graphs). Going back to Example 1, let $G = (V,E)$ be a connected undirected simple graph, and let μ be the uniform distribution over all spanning trees of G, which is a subset of $\binom{E}{n-1}$ where $n = |V|$. Then the down and up steps can be described as follows.

1. Down-Step: Starting from a spanning tree $T \subseteq E$, remove a uniformly random edge $e \in T$. The result $T \setminus e$ is a union of two subtrees which are not connected to each other.
2. Up-Step: Out of all edges $f \in E$ such that $T \setminus \{e\} \cup \{f\}$ becomes a spanning tree, pick one such edge uniformly at random and add it to $T \setminus \{e\}$.

Thus, after one down-step followed by one up-step, we have "exchanged" a random edge e for another random edge f. In particular, the resulting down-up walk recovers the well-known bases exchange walk on spanning trees of the graph. This generalizes to all matroids.

Example 4 (The Glauber Dynamics on Proper Colorings of Graphs). Returning to Example 2, let $G = (V,E)$ be an undirected graph and let μ be the uniform distribution over proper q-colorings of the vertices of G with $q \ge 1$ colors. Recall that we may view μ as a distribution over $\binom{\mathscr{U}}{n}$ where $n = |V|$ and $\mathscr{U} = \{(v,c) : v \in V, c \in [q]\}$ is the set of vertex-color pairs. Then the down and up steps can be described as follows.

1. Down-Step: Starting from a full coloring $\chi : V \to [q]$, remove a uniformly random vertex-color pair $(v, \chi(v))$. In other words, *uncolor* a uniformly random vertex $v \in V$.

2. Up-Step: Out of all vertex-color pairs which can complete $\chi \setminus \{(v, \chi(v))\}$ into a full coloring, select one uniformly at random and add it. Since v is the only uncolored vertex, this is equivalent to resampling a uniformly random available color for v.

Phrased in a more streamlined fashion, in each step, the down-up walk simply picks a uniformly random vertex and resamples the assignment to that vertex conditioned on the current assignments for all other vertices. This exactly recovers the Glauber dynamics/Gibbs sampler commonly used in statistical physics and machine learning. This generalizes to any distribution over a discrete product space.

The remainder of this section is devoted to establishing basic but useful properties of these down and up operators. The goal is to also build some intuition on what these constructions look like. For instance, the following lemma shows that the down-up and up-down walks have the "correct" stationary distributions, and that the down and up operators are intimately related.

Lemma 2.1.2 (Basic Properties of the Down and Up Operators)**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. Then for every $0 \le k \le \ell \le n$, the following properties hold.*

1. *For every $\tau \in \operatorname{supp}(\mu_k), \sigma \in \operatorname{supp}(\mu_\ell)$, we have the identity*

$$\mu_k(\tau) \cdot \mathcal{U}_\mu^{k \nearrow \ell}(\tau \to \sigma) = \mu_\ell(\sigma) \cdot \mathcal{D}_\mu^{\ell \searrow k}(\sigma \to \tau).$$

   In particular, $\mu_k = \mu_\ell \mathcal{D}_\mu^{\ell \searrow k}$ and $\mu_\ell = \mu_k \mathcal{U}_\mu^{k \nearrow \ell}$. Furthermore, for every pair of functions $f : \operatorname{supp}(\mu_k) \to \mathbb{R}, g : \operatorname{supp}(\mu_\ell) \to \mathbb{R}$, we have the following adjointness identity

$$\left\langle f, \mathcal{U}_\mu^{k \nearrow \ell} g \right\rangle_{\mu_k} = \left\langle \mathcal{D}_\mu^{\ell \searrow k} f, g \right\rangle_{\mu_\ell}.$$

   In other words, $\mathcal{D}_\mu^{\ell \searrow k}, \mathcal{U}_\mu^{k \nearrow \ell}$ are time reversals of each other w.r.t. the distributions μ_k, μ_ℓ.

2. *The down-up walk $\mathcal{D}_\mu^{\ell \searrow k} \mathcal{U}_\mu^{k \nearrow \ell}$ on $\operatorname{supp}(\mu_\ell)$ is reversible w.r.t. μ_ℓ and is positive semidefinite w.r.t. the inner product $\langle \cdot, \cdot \rangle_{\mu_\ell}$. Similarly, the up-down walk $\mathcal{U}_\mu^{k \nearrow \ell} \mathcal{D}_\mu^{\ell \searrow k}$ on $\operatorname{supp}(\mu_k)$ is reversible w.r.t. μ_k and is positive semidefinite w.r.t. the inner product $\langle \cdot, \cdot \rangle_{\mu_k}$. Furthermore, these two Markov chains have the exact same eigenvalues (with multiplicity), up to the multiplicity of the zero eigenvalue.*

Proof. 1. The first identity follows immediately by observing that both sides of the equation equal $\frac{1}{\binom{\ell}{k}} \cdot \mu_\ell(\tau)$. The adjointness identity follows from the first identity by noting that

$$\begin{aligned}
\left\langle f, \mathcal{U}_\mu^{k \nearrow \ell} g \right\rangle_{\mu_k} &= \sum_\tau \sum_\sigma f(\tau) \cdot g(\sigma) \cdot \mu_k(\tau) \mathcal{U}_\mu^{k \nearrow \ell}(\tau \to \sigma) \\
&= \sum_\tau \sum_\sigma f(\tau) \cdot g(\sigma) \cdot \mu_\ell(\sigma) \mathcal{D}_\mu^{\ell \searrow k}(\sigma \to \tau) \\
&= \left\langle \mathcal{D}_\mu^{\ell \searrow k} f, g \right\rangle_{\mu_\ell}.
\end{aligned}$$

2. Reversibility of $\mathcal{D}_\mu^{\ell \searrow k} \mathcal{U}_\mu^{k \nearrow \ell}$ follows immediately from the adjointness identity in Item 1. To show that this matrix is positive semidefinite w.r.t. $\langle \cdot, \cdot \rangle_{\mu_\ell}$, we again use adjointness. For every $f : \operatorname{supp}(\mu_\ell) \to \mathbb{R}$, we have that

$$\left\langle f, \mathcal{D}_\mu^{\ell \searrow k} \mathcal{U}_\mu^{k \nearrow \ell} f \right\rangle_{\mu_\ell} = \left\langle \mathcal{U}_\mu^{k \nearrow \ell} f, \mathcal{U}_\mu^{k \nearrow \ell} f \right\rangle_{\mu_k} \ge 0.$$

   An identical argument yields reversibility and positive semidefiniteness for $\mathcal{U}_\mu^{k \nearrow \ell} \mathcal{D}_\mu^{\ell \searrow k}$ w.r.t. μ_k. That the $\ell \leftrightarrow k$ down-up and $k \leftrightarrow \ell$ up-down Markov chains have the same eigenvalues (up to multiplicity of the zero eigenvalue) follows from Fact 1.4.7, a general linear algebraic fact.

□

Conditional Distributions One of the key intuitions behind the connection between spectral independence and rapid mixing is the idea of "decomposing" our target distribution μ into "smaller" distributions supported on smaller portions of the domain $\mathrm{supp}(\mu)$. More specifically, we will decompose μ into its "conditional" distributions, where we condition on the event that a sample from μ contains some subset $\tau \subseteq \mathscr{U}$ of elements. This conditioning turns out to be very useful, and so we discuss the notation here.

For a feasible $\tau \subseteq \mathscr{U}$, we write μ^τ for the *conditional distribution* induced by conditioning μ on the event that $\sigma \sim \mu$ contains τ. More formally, μ^τ is supported on $\binom{\mathscr{U}\setminus\tau}{n-|\tau|}$, where $\mu^\tau(\tau') \propto \mu(\tau \cup \tau')$ for each $\tau' \subseteq \mathscr{U} \setminus \tau$. We will also equivalently think of μ^τ as a distribution still on $\binom{\mathscr{U}}{n}$, but where $\mu^\tau(\sigma) \propto \mu(\sigma)$ if $\sigma \supseteq \tau$, and $\mu^\tau(\sigma) = 0$ otherwise. If $\tau = \{u\}$ is a singleton, we write μ^u instead of $\mu^{\{u\}}$.

For example, if μ is the uniform distribution over bases of a matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ and $\tau \in \mathcal{X}$ is an independent set, then μ^τ is the uniform distribution over the bases of the *contraction* $\mathcal{M}/\tau = (\mathscr{U} \setminus \tau, \{\sigma \setminus \tau : \sigma \in \mathcal{X}, \sigma \supseteq \tau\})$, a smaller (sub-)matroid contained in $\mathcal{M}$. If μ is the Gibbs distribution of a q-state graphical model on a graph $G = (V, E)$ and $\tau_\Lambda : \Lambda \to [q]$ is a partial assignment on a subset of vertices $\Lambda \subseteq V$ (viewed as a collection of vertex-assignment pairs), then μ^{τ_Λ} is the conditional Gibbs distribution where we *pin* the assignment of each $v \in \Lambda$ to $\tau_\Lambda(v)$. In statistical physics jargon, τ_Λ is a *pinning*.

Each conditional distribution μ^τ itself comes equipped with its own down and up operators $\mathcal{D}^{\ell \searrow k}_{\mu^\tau}, \mathcal{U}^{k \nearrow \ell}_{\mu^\tau}$, as well as the corresponding $\ell \leftrightarrow k$ down-up and $k \leftrightarrow \ell$ up-down walks. We write $\mathcal{P}_{\mu^\tau}$ for "the" down-up walk on $\mathrm{supp}(\mu^\tau)$. The corresponding level-k marginal distribution μ^τ_k is supported on $\binom{\mathscr{U}\setminus\tau}{k}$, for each $0 \le k \le n - |\tau|$. Their entries are given by

$$\mu^\tau_k(\tau') = \frac{1}{\binom{n-|\tau|}{k}} \Pr_{\sigma\sim\mu}[\sigma \supseteq \tau' \mid \sigma \supseteq \tau], \quad \forall \tau' \in \mathrm{supp}(\mu^\tau_k) \subseteq \binom{\mathscr{U}\setminus\tau}{k} \tag{2.5}$$

When the distribution μ is clear from context, we drop μ from the subscript. Again, if $\tau = \{u\}$ is a singleton, we write u instead of $\{u\}$ in all subscripts and superscripts.

The Support of these Distributions To reduce notational clunkiness, we will sometimes write $\mathcal{X}_k$ for $\mathrm{supp}(\mu_k)$ for each $0 \le k \le n$. Thus, $\mathcal{X}_n$ is synonymous with the support of the distribution μ itself. Similarly, for each feasible $\sigma \subseteq \mathscr{U}$ and each $0 \le k \le n - |\sigma|$, we write $\mathcal{X}^\sigma_k$ for $\mathrm{supp}(\mu^\sigma_k)$. Of course, we can assume without loss of generality that $\mathcal{X}_1 = \mathscr{U}$. We will sometimes write $\mathcal{X}^\sigma_{\le k} \stackrel{\mathsf{def}}{=} \mathcal{X}^\sigma_0 \sqcup \cdots \sqcup \mathcal{X}^\sigma_k$ for the *downwards closure* of $\mathcal{X}^\sigma_k$, for each feasible $\sigma \subseteq \mathscr{U}$ and $0 \le k \le n - |\sigma|$. We write $\mathcal{X} = \mathcal{X}_{\le n}$ for the collection of feasible sets; similarly, we write $\mathcal{X}^\sigma = \mathcal{X}^\sigma_{\le n-|\sigma|}$ for each feasible $\sigma \subseteq \mathscr{U}$. This notational choice is consistent with our notation for matroids, where we write $\mathcal{X}$ for the collection of independent sets of a matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$.

In the language of algebraic topology, $\mathcal{X}$ is known as a *abstract simplicial complex*. Since $\mathcal{X}$ is generated by taking downwards closure of $\mathcal{X}_n$, the simplicial complex $\mathcal{X}$ is called *pure*. The elements of $\mathcal{X}$, which are subsets of $\mathscr{U}$, are called *faces*, while the elements of $\mathcal{X}_n$ are called *facets*. $\mathcal{X}_{\le k}$ is known as the $(k-1)$-*skeleton* of $\mathcal{X}$, and each $\mathcal{X}^\sigma$ is known as the *link* of $\mathcal{X}$ w.r.t. σ. However, we will eschew the use of this terminology to cut down on jargon.

2.2 Spectral Independence: Limited Correlations in High-Dimensional Discrete Distributions

We now define spectral independence. For the following definitions, μ is again some probability distribution over $\binom{\mathscr{U}}{n}$ for a finite ground set $\mathscr{U}$ and $0 \le n \le |\mathscr{U}|$.

Definition 11 ((One-Sided) Influence; [ALO21])**.** *For every pair of elements $i, j \in \mathscr{U}$, define the **(one-sided) influence of** i **on** j by*

$$\mathcal{I}_\mu(i \to j) \stackrel{\mathsf{def}}{=} \Pr_{\sigma\sim\mu}[j \in \sigma \mid i \in \sigma] - \Pr_{\sigma\sim\mu}[j \in \sigma],$$

*with the convention that $\Pr_{\sigma\sim\mu}[i \in \sigma \mid i \in \sigma] = 1$ for all $i \in \mathscr{U}$. We also define the **(one-sided) influence matrix** $\mathcal{I}_\mu \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ to be the matrix with entries $\mathcal{I}_\mu(i,j) = \mathcal{I}_\mu(i \to j)$ for every $i, j \in \mathscr{U}$.*

Remark 3. If μ is supported on a more structured space such as the discrete hypercube $\{0,1\}^n$, or more generally, a discrete product space $\Sigma_1 \times \cdots \times \Sigma_n$ for some finite sets $\Sigma_1, \ldots, \Sigma_n$, then there are slightly "cleaner" but essentially equivalent formulations of influence. We use these other formulations in later chapters, but refrain from introducing them here for simplicity and clarity; see Section 2.5 for more details and discussion.

We also note that [Ali+21; Ana+22c] tend to refer to the one-sided influence matrix in Definition 11 above as the *correlation matrix* (and refer to the alternative definitions in Section 2.5 below as influence matrices). We will not use this terminology to avoid confusion, since the term correlation matrix is already a well-defined concept in statistics, and is typically defined in a way which ensures it is symmetric (the correlation of a random variable X with another random variable Y is the same as the correlation of Y with X). This of course comes at the expense of overloading the term "influence" or lengthening the term with appropriate qualifications (e.g. "one-sided influence"). This is also in line with [KKS21].

Readers familiar with the concept of *Dobrushin influence* or *path coupling* may wonder what the connection with this notion of influence is. We discuss this briefly later in this chapter. For a more detailed discussion of the connections, see Chapter 8.

Given these influence matrices, we can now define spectral independence.

Definition 12 (Spectral Independence; [ALO21])**.** *For a real number $0 \leq \eta \leq n-1$, we say μ is η-**spectrally independent** if $\lambda_{\max}(\mathcal{I}_\mu) \leq 1+\eta$*[1]*. Analogously, for real numbers $\eta_0, \ldots, \eta_{n-2}$ satisfying $0 \leq \eta_k \leq n-k-1$ for every $0 \leq k \leq n-2$, we say μ is $(\eta_0, \ldots, \eta_{n-2})$-**spectrally independent** if for $0 \leq k \leq n-2$ and every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k$, we have that the conditional distribution μ^σ is η_k-spectrally independent, i.e. $\lambda_{\max}(\mathcal{I}_{\mu^\sigma}) \leq 1+\eta_k$.*

Remark 4. In many cases, we will establish something stronger, namely $\|\mathcal{I}_\mu\|$ for some matrix norm $\|\cdot\|$. For spin systems, we will often work with the norm $\|\cdot\|_\infty$. In Appendix C, we work with $\|\cdot\|_1$. However, none of the mixing/algorithmic results in this thesis depend on matrix norm bounds in an essential way. We refer interested readers to [KKS21] for an example of where $\|\cdot\|_\infty$-bounds on influence matrices are essential.

Again, when the distribution μ is clear from context, we drop the subscript, and write $\mathcal{I}(i \to j)$ instead for simplicity. When conditioning is present, we will replace the subscript μ^σ with a superscript σ (e.g. $\mathcal{I}^\sigma(i \to j)$).

A priori, since $\mathcal{I}_\mu$ is asymmetric and in general admits both positive and negative entries, it isn't clear that $\mathcal{I}_\mu$ even has real eigenvalues. However, one can actually view influence as an appropriate normalization of *covariance*. More specifically, if we view $\sigma \sim \mu$ as a random indicator vector $X \in \{0,1\}^{\mathscr{U}}$, then we have that

$$\begin{aligned}
\mathrm{Cov}_\mu(X_i, X_j) &\stackrel{\mathsf{def}}{=} \mathbb{E}_\mu\left[(X_i - \mathbb{E}_\mu X_i)\cdot(X_j - \mathbb{E}_\mu X_j)\right] \\
&= \mathbb{E}_\mu\left[X_i X_j\right] - \mathbb{E}_\mu\left[X_i\right]\mathbb{E}_\mu\left[X_j\right] \\
&= \Pr_\mu\left[X_i = X_j = 1\right] - \Pr_\mu\left[X_i = 1\right]\cdot\Pr_\mu\left[X_j = 1\right] \\
&= \Pr_\mu[i,j \in \sigma] - \Pr_\mu[i \in \sigma]\cdot\Pr_\mu[j \in \sigma] \\
&= \frac{1}{\Pr_\mu[i \in \sigma]}\cdot\mathcal{I}_\mu(i \to j).
\end{aligned}$$

Hence, if $\Sigma_\mu = [\mathrm{Cov}_\mu(X_i, X_j)]_{i,j \in \mathscr{U}} \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ is the covariance matrix, then we have that

$$\mathcal{I}_\mu = \mathrm{diag}\left(\Pr_\mu[i \in \sigma] : i \in \mathscr{U}\right)^{-1}\cdot\Sigma_\mu. \tag{2.6}$$

Since $\Sigma_\mu = \mathbb{E}_\mu\left[(X - \mathbb{E}_\mu[X])\cdot(X - \mathbb{E}_\mu[X])^\top\right]$ is symmetric positive semidefinite, and we are hitting it with a diagonal matrix with positive diagonal entries, we immediately see that $\mathcal{I}_\mu$ has nonnegative real eigenvalues. In particular, $\lambda_{\max}(\mathcal{I}_\mu)$ is well-defined. Using this relationship, one can further show that the bounds $0 \leq \lambda_{\max}(\mathcal{I}_\mu) \leq n$ always hold. For this reason, one can also view influence a *correlation*, which is why we call spectral independence a "limited correlations" property of the distribution μ. We'll see later how $\mathcal{I}_\mu$ is also intimately related to the transition

[1] The additional 1+ in the upper bound on $\lambda_{\max}(\mathcal{I}_\mu)$ is there for technical reasons, namely to "cancel" out the $\Pr_{\sigma\sim\mu}[i \in \sigma \mid i \in \sigma] = 1$ component on the diagonal of $\mathcal{I}_\mu$.

probability matrix of a certain reversible Markov chain, which hints at a more concrete connection between spectral independence and Markov chain mixing.

For now, let us first see a few examples.

Example 5. Consider the uniform distribution μ over the independent sets of a star graph with center vertex r and two leaves u, v. There are five such independent sets, namely

$$\emptyset, \qquad \{r\}, \qquad \{u\}, \qquad \{v\}, \qquad \{u, v\},$$

each occurring with probability $1/5$.

Instead of viewing μ as a distribution over $\{0,1\}^3$, we may view it as a distribution over $\binom{\mathscr{U}}{3}$ where $\mathscr{U} = \{(r,\text{“in”}), (r,\text{“out”}), (u,\text{“in”}), (u,\text{“out”}), (v,\text{“in”}), (v,\text{“out”})\}$ is the collection of vertex-assignment pairs, following Example 2. In this case, $\mathcal{I}_\mu$ is a 6×6 matrix. An example of an entry in $\mathcal{I}_\mu$ is

$$\mathcal{I}_\mu((u,\text{“out”}) \to (r,\text{“in”})) = \Pr_\mu[r \text{ “in”} \mid u \text{ “out”}] - \Pr_\mu[r \text{ “in”}] = \frac{1}{3} - \frac{1}{5} = \frac{2}{15}.$$

The full matrix is given by

$$\mathcal{I}_\mu = \begin{bmatrix} 3/5 & -3/5 & -1/5 & 1/5 & 1/10 & -1/10 \\ -2/5 & 2/5 & 2/15 & -2/15 & -1/15 & 1/15 \\ -2/5 & 2/5 & 4/5 & -4/5 & -2/5 & 2/5 \\ 1/10 & -1/10 & -1/5 & 1/5 & 1/10 & -1/10 \\ 1/10 & -1/10 & -1/5 & 1/5 & 3/5 & -3/5 \\ -1/15 & 1/15 & 2/15 & -2/15 & -2/5 & 2/5 \end{bmatrix},$$

which has eigenvalues $\left\{0, 0, 0, \frac{1}{2}, \frac{5}{6}, \frac{5}{3}\right\}$. It follows that this distribution is $\frac{2}{3}$-spectrally independent.

In Section 2.5, we will provide an explanation for why $\mathcal{I}_\mu$ has eigenvalue 0 with multiplicity 3. There, we will also introduce a slightly simpler alternative notion of an influence matrix Ψ_μ, which for this example is only 3×3, and has entries

$$\Psi_\mu = \begin{bmatrix} 1 & -1/2 & -1/2 \\ -1/3 & 1 & 1/6 \\ -1/3 & 1/6 & 1 \end{bmatrix}$$

and eigenvalues $\left\{\frac{1}{2}, \frac{5}{6}, \frac{5}{3}\right\}$. As one can see through this example, Ψ_μ will capture the “nontrivial” spectrum of $\mathcal{I}_\mu$.

Example 6 (Negatively Correlated Distributions). Suppose μ is negatively correlated in the sense that

$$\Pr_{\sigma\sim\mu}[j \in \sigma \mid i \in \sigma] \le \Pr_{\sigma\sim\mu}[j \in \sigma], \quad \forall \text{ distinct } i, j \in \mathscr{U}. \tag{2.7}$$

Negative dependence conditions of this flavor, as well as stronger variants of this inequality, have been extensively studied in probability theory [Pem00; BBL09; PP14], combinatorics [FM92; AOR16], approximation algorithms [Sri01; Gan+06; Asa+10; CVZ10], and more. A classic example of such a distribution is the uniform distribution over spanning trees of a connected graph.

For negatively correlated distributions, we see that the off-diagonal entries of $\mathcal{I}_\mu$ are nonpositive. Furthermore, if we fix an arbitrary $i \in \mathscr{U}$, then

$$\begin{aligned} \sum_{j\in\mathscr{U}} |\mathcal{I}_\mu(i,j)| &= 1 - \Pr_{\sigma\sim\mu}[i \in \sigma] + \sum_{j\neq i} \left| \Pr_{\sigma\sim\mu}[j \in \sigma \mid i \in \sigma] - \Pr_{\sigma\sim\mu}[j \in \sigma] \right| \\ &= 1 - \Pr_{\sigma\sim\mu}[i \in \sigma] + \Bigg| \underbrace{\sum_{j\neq i} \Pr_{\sigma\sim\mu}[j \in \sigma \mid i \in \sigma]}_{=n-1} - \underbrace{\sum_{j\neq i} \Pr_{\sigma\sim\mu}[j \in \sigma]}_{=n-\Pr_{\sigma\sim\mu}[i\in\sigma]} \Bigg| && \text{(Eq. (2.7))} \\ &= 2 \cdot \left(1 - \Pr_{\sigma\sim\mu}[i \in \sigma]\right). && \text{(Homogeneity)} \end{aligned}$$

It follows that $\lambda_{\max}(\mathcal{I}_\mu) \le \|\mathcal{I}_\mu\|_\infty \le 2$ so that μ is 1-spectrally independent. Note that the final step really does follow just by homogeneity, since

$$\begin{aligned}\sum_{i\in\mathscr{U}} \Pr_{\sigma\sim\mu}[i\in\sigma] &= \sum_{i\in\mathscr{U}} \mathbb{E}_{\sigma\sim\mu}[\mathbb{I}[i\in\sigma]] \\ &= \mathbb{E}_{\sigma\sim\mu}[|\sigma|] && \text{(Linearity of Expectation)} \\ &= n && (|\sigma| = n \text{ for all } \sigma\in\operatorname{supp}(\mu), \text{ i.e. } n\text{-homogeneity})\end{aligned}$$

as desired.

Example 7 (Distribution with Disconnected Support). Let $\mathscr{U} = [2n]$, and let μ be the distribution which places probability mass $\frac{1}{2}$ on each of the two sets $A = \{1,\dots,n\}$ and $B = \{n+1,\dots,2n\}$. In this case, $\mathcal{I}_\mu = \mathbb{R}^{2n\times 2n}$ has entries given by

$$\mathcal{I}_\mu(i\to j) = \begin{cases}\frac{1}{2}, & \text{if } i,j\in A \text{ or } i,j\in B\\ -\frac{1}{2}, & \text{if } i\in A, j\in B \text{ or } i\in B, j\in A\end{cases}$$

In particular, $\mathcal{I}_\mu$ is given by the block matrix $\frac{1}{2}\begin{bmatrix} J_n & -J_n\\ -J_n & J_n\end{bmatrix} = \frac{1}{2}\cdot[\mathbf{1}_n, -\mathbf{1}_n]\otimes[\mathbf{1}_n,-\mathbf{1}_n]$, where $\mathbf{1}_n\in\mathbb{R}^n$ is the all-ones vector, and $J_n = \mathbf{1}_n\otimes\mathbf{1}_n$ is the $n\times n$ all-ones matrix. In particular, $\lambda_{\max}(\mathcal{I}_\mu) = n$, so μ is $(n-1)$-spectrally independent. Notice how this distribution has "disconnected" support, in the sense that one needs to change many elements in order to make a move from A to B or vice versa. Hence, one intuitively sees that η-spectral independence for large η is generally undesirable, since in that case, local Markov chains may not even be irreducible, let alone rapidly mixing.

Let us see a few basic properties of the influence matrix and spectral independence, which will help build intuition and will be useful later on.

Lemma 2.2.1 (Block Structure of Influence Matrices for Product Distributions)**.** *Let μ,ν be probability distributions on $\binom{\mathscr{U}}{n}$ and $\binom{\mathcal{V}}{m}$, respectively, where $\mathscr{U},\mathcal{V}$ are disjoint finite sets and $0\le n\le|\mathscr{U}|, 0\le m\le|\mathcal{V}|$ are nonnegative integers. Let $\mu\otimes\nu$ be the **product distribution** over $\binom{\mathscr{U}\sqcup\mathcal{V}}{n+m}$ defined by*

$$(\mu\otimes\nu)(\tau\sqcup\sigma) = \mu(\tau)\cdot\nu(\sigma),\quad \forall\tau\in\binom{\mathscr{U}}{n}, \sigma\in\binom{\mathcal{V}}{m}.$$

Then we have the identity

$$\mathcal{I}_{\mu\otimes\nu} = \begin{bmatrix}\mathcal{I}_\mu & 0\\ 0 & \mathcal{I}_\nu\end{bmatrix}.$$

In particular, if μ is η^μ-spectrally independent and ν is η^ν-spectrally independent, then $\mu\otimes\nu$ is $\max\{\eta^\mu,\eta^\nu\}$-spectrally independent.

Proof. We check the matrix equality entry-by-entry. Clearly, the diagonal entries are equal, since the marginals of $\mu\otimes\nu$ on $\mathscr{U},\mathcal{V}$ are μ,ν, respectively. For the remaining entries, the key is to use the fact that $\tau\sqcup\sigma\sim\mu\otimes\nu$ is given by independent samples $\tau\sim\mu$ and $\sigma\sim\nu$.

Let $i,j\in\mathscr{U}\sqcup\mathcal{V}$ be distinct. If $i\in\mathscr{U}$ and $j\in\mathcal{V}$, then

$$\Pr_{\tau\sqcup\sigma\sim\mu\otimes\nu}[j\in\tau\sqcup\sigma \mid i\in\tau\sqcup\sigma] = \Pr_{\tau\sim\mu,\sigma\sim\nu \text{ independent}}[j\in\sigma\mid i\in\tau] = \Pr_{\sigma\sim\nu}[j\in\sigma].$$

Via a similar argument, we also see that $\Pr_{\tau\sqcup\sigma\sim\mu\otimes\nu}[j\in\tau\sqcup\sigma] = \Pr_{\sigma\sim\nu}[j\in\sigma]$. It follows that $\mathcal{I}_\mu(i\to j) = 0$. An identical calculation shows $\mathcal{I}_{\mu\otimes\nu}(i\to j) = 0$ if $i\in\mathcal{V}$ and $j\in\mathscr{U}$ instead.

Now, assume $i,j\in\mathscr{U}$ are distinct. Then

$$\Pr_{\tau\sqcup\sigma\sim\mu\otimes\nu}[j\in\tau\sqcup\sigma \mid i\in\tau\sqcup\sigma] = \Pr_{\tau\sim\mu,\sigma\sim\nu \text{ independent}}[j\in\tau\mid i\in\tau] = \Pr_{\tau\sim\mu}[j\in\tau\mid i\in\tau].$$

Via a similar argument, we also see that $\Pr_{\tau\sqcup\sigma\sim\mu\otimes\nu}[j\in\tau\sqcup\sigma] = \Pr_{\tau\sim\mu}[j\in\tau]$. Hence, $\mathcal{I}_{\mu\otimes\nu}(i\to j) = \mathcal{I}_\mu(i\to j)$. A nearly identical calculation shows that $\mathcal{I}_{\mu\otimes\nu}(i\to j) = \mathcal{I}_\nu(i\to j)$ if $i,j\in\mathcal{V}$ instead. □

Our Probabilistic Terminology	Terminology of High-Dimensional Expanders
Distribution μ	(Weighted) Simplicial Complex
Support $\mathrm{supp}(\mu)$	Facets/Maximal Faces
Cardinality of τ	Dimension $(+1)$ of τ
Homogeneity	Purity
Conditioning	Link
Feasible Subset of $\mathscr{U}$	Face
Total Connectivity	Connectivity of All Links
Marginal Probabilities	Localized Weight Function
Influence Matrix	Local Random Walk
Spectral Independence	Local Spectral Expansion

Table 2.1: A rough correspondence between terminology used by the different communities.

Relation to Alternative Notions of Influence There are alternative notions of "influence" which have been extensively studied previously in the literature. For instance, the *Dobrushin influence* and the associated *Dobrushin uniqueness condition* [Dob70] were studied extensively in statistical physics and used to establish rapid mixing of Markov chains such as the Glauber dynamics (see also the *Dobrushin-Shlosman condition* [DS85a; DS85b; DS87b], the ℓ_2-Dobrushin condition [Hay06], and other *"Dobrushin-type" conditions* [DGJ09]). Dobrushin influences are intimately related with the beautiful *path coupling technique* in the analysis of Markov chains [BD97a]; we discuss formal connections with our notion of influence in Chapter 8.

There is also a well-known notion of influence stemming from the analysis of Boolean functions and social choice theory which for clarity, we will refer to as *coordinate influence* (or *voter influence*) [ODo14]. In that context, our notion of influence is sometimes referred to as the *conditional influence* (see e.g. [GG06]), although we were not aware of this when we introduced our notion of influence; we discuss some loose connections in Appendix D.2.

Connections with High-Dimensional Expanders Our notion of spectral independence was originally derived from a notion of *spectral expansion* for high-dimensional *simplicial complexes* [KM17; DK17; Opp18; KO20b]. These *high-dimensional expanders (HDXs)* have recently gained significant attention due to emerging applications and connections to classical and quantum error-correcting codes [Ale+20; EKZ21; KT21; Din+21b; Din+22; PK22], property testing [DK17; DD19; KM20; KO20b], analysis of Boolean functions [Dik+18; Baf+22a; Baf+22b; GLL22], the theory of constraint-satisfaction problems (CSPs) [AGT19a; Din+21a], high-dimensional combinatorics [Fox+12], etc. Besides these applications, a key motivation for developing such a theory is to extend the incredibly fruitful theory of *expander graphs* to "higher dimensional" objects such as hypergraphs, simplicial complexes, and posets.

Sophisticated algebraic constructions of such high-dimensional expanders called *Ramanujan complexes* have been known since the works of [Bal00; ISŻ03; Li04; Sar04; LSV05a; LSV05b],[2] building on the seminal work of Lubotzky–Phillips–Sarnak [LPS88] on *Ramanujan graphs*. Elementary constructions now also exist [KO18; KO20a; LMY20; Gol21; OP22]. The entirety of our set-up, as well as all of our results, can be stated in their language, although we avoid doing so for simplicity and clarity. For instance, our probability distributions can be said to be supported over the *facets of a pure (abstract) simplicial complex*. Table 2.1 gives correspondences in terminology.

2.3 Spectral Independence Implies Rapid Mixing

Our goal now is to see how spectral independence, a static property of the (stationary) distribution μ, is intimately connected to the mixing time of the down-up walk $\mathcal{P}_\mu$, a dynamical property of a Markov chain sampling from μ.

Theorem 2.3.1 (Spectral Independence $\implies$ Rapid Mixing of $\mathcal{P}_\mu$; [ALO21] building on [AL20])**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution*

[2] The fascinating mixing properties of such complexes have also been studied, see e.g. [LLP20].

over $\binom{\mathscr{U}}{n}$. *If* μ *is* $(\eta_0, \dots, \eta_{n-2})$*-spectrally independent, then the spectral gap of the down-up walk satisfies the following lower bound*

$$\gamma\left(\mathcal{P}_\mu\right) \geq \frac{1}{n} \prod_{k=0}^{n-2} \left(1 - \frac{\eta_k}{n-k-1}\right). \tag{2.8}$$

In particular, if there exist constants $\eta, C > 0$ *s.t.* $\eta_k \leq \min\{\eta, C \cdot (n-k-1)\}$ *for all* $0 \leq k \leq n-2$*, then* $\gamma\left(\mathcal{P}_\mu\right) \geq \Omega\left(n^{-(1+\eta)}\right)$ *and* $\mathcal{P}_\mu$ *mixes in* $O\left(n^{1+\eta} \log \frac{1}{\mu_{\min}}\right)$*-steps.*

Remark 5. One can also show that spectral independence implies rapid mixing of more general down-up walks $\mathcal{D}_\mu^{n \searrow k} \mathcal{U}_\mu^{k \nearrow n}$. We will need this later, but won't discuss this point further in this chapter.

Remark 6. It turns out the first proof we provide actually yields a slightly stronger bound of

$$\gamma\left(\mathcal{P}_\mu\right) \geq \frac{1}{n} \min_{\emptyset = \sigma_0 \subsetneq \sigma_1 \subsetneq \cdots \subsetneq \sigma_n \in \mathrm{supp}(\mu)} \prod_{k=0}^{n-2} \left(1 - \frac{\lambda_{\max}\left(\mathcal{I}_{\mu^{\sigma_k}}\right)}{n-k-1}\right). \tag{2.9}$$

We call this the "flag" version of Theorem 2.3.1, since a strictly increasing chain of subsets $\emptyset = \sigma_0 \subsetneq \sigma_1 \subsetneq \cdots \subsetneq \sigma_n$ is typically called a *flag*. Note that it must be that $|\sigma_k| = k$ for every k.

This was further improved in [Ali+21], where they essentially showed that the minimum over all flags can be replaced by a minimum over $\sigma_n \in \mathrm{supp}(\mu)$ and an *average* over the remaining choices $\emptyset = \sigma_0 \subsetneq \sigma_1 \subsetneq \cdots \subsetneq \sigma_{n-1}$ contained in σ_n (equivalently, an average over an ordering of the elements of σ_n). While Eq. (2.9) hasn't found any applications that couldn't be achieved using the simpler Eq. (2.8), the improved version contained in [Ali+21] has; see [ALV22].

Given this theorem, establishing rapid mixing of these down-up walks reduces to establishing $(\eta_0, \dots, \eta_{n-2})$-spectral independence with $\eta_0, \dots, \eta_{n-2} \leq O(1)$ for the associated distribution μ. We'll see numerous examples of this later in this thesis. In Section 2.6, we give an illustrative example of a distribution μ where $\eta_0 \geq \Omega(n)$ and $\mathcal{P}_\mu$ is torpidly mixing. At this juncture, readers primarily interested in techniques for establishing spectral independence can safely proceed to later chapters.

The remainder of this chapter will be devoted to setting up and proving this theorem. We will also discuss extensions of this result, as well as further surrounding context. We will provide two proofs of Theorem 2.3.1. Both proof methods are inductive in nature, and proceed by iteratively decomposing the target measure μ into its conditional distributions. The first proof is based directly examining spectral gaps. It will be simpler and more intuitive, while appealing more directly to linear algebra. The second proof is based on the Law of Total Variance and contraction of variance under a Markov operator. It will be notationally more complex, but ultimately will enable important extensions and improvements. At the foundation of both proofs is the equivalence between spectral independence and very strong "local" mixing conditions, such as *"nearly-instant"* mixing of certain "local" Markov chains. For this reason, theorems of this form are often called *local-to-global theorems*. They are one form of local-to-global analysis.

2.3.1 Local Mixing Conditions

We now introduce the "local" Markov chains. We'll see how spectral independence is equivalent to very fast mixing of these local Markov chains.

Definition 13 (Local Walk; [KM17; DK17; KO20b])**.** *We define the* ***(nonlazy) local walk*** *on* $\mathscr{U}$ *as*

$$\mathcal{Q}_\mu \stackrel{\text{def}}{=} 2 \cdot \left(\mathcal{U}_\mu^{1 \nearrow 2} \mathcal{D}_\mu^{2 \searrow 1} - \frac{1}{2} \cdot \mathsf{Id}\right).$$

Fact 2.3.2 (Properties of the Local Walk)**.** *1. The entries of* $\mathcal{Q}_\mu \in \mathbb{R}^{\mathscr{U} \times \mathscr{U}}$ *are given by*

$$\mathcal{Q}_\mu(i \to j) = \frac{1}{n-1} \Pr_{\sigma \sim \mu}[j \in \sigma \mid i \in \sigma]$$

for $i \neq j$*, and* $\mathcal{Q}_\mu(i \to i) = 0$ *for all* $i \in \mathscr{U}$*. In particular, row* $i \in \mathscr{U}$ *of* $\mathcal{Q}_\mu$ *is precisely the conditional marginal distribution* μ_1^i *(except padded with an additional zero at entry* i*).*

2. *$\mathcal{Q}_\mu$ is reversible w.r.t. the marginal distribution μ_1 on $\mathscr{U}$, which we recall is given by*

$$\mu_1(i) = \frac{1}{n} \Pr_{\sigma\sim\mu}[i \in \sigma], \quad \forall i \in \mathscr{U}.$$

3. *We have the additional identity*

$$\mathcal{Q}_\mu = \frac{n}{n-1} \cdot \left(\mathcal{U}_\mu^{1\nearrow n} \mathcal{D}_\mu^{n\searrow 1} - \frac{1}{n} \cdot \mathsf{Id} \right).$$

4. *For every $f_1 : \operatorname{supp}(\mu_1) \to \mathbb{R}$, we have that*

$$\langle f_1, \mathcal{Q}_\mu f_1 \rangle_{\mu_1} = \mathbb{E}_{\{i,j\}\sim\mu_2}[f_1(i) \cdot f_1(j)]$$

Proof. 1. Using Lemma 2.1.1, we calculate that for $i \neq j$,

$$\left(\mathcal{U}_\mu^{1\nearrow 2} \mathcal{D}_\mu^{2\searrow 1}\right)(i \to j) = \frac{1}{2} \cdot \frac{\mu_2(i,j)}{2 \cdot \mu_1(i)} = \frac{1}{2} \cdot \mu_1^i(j).$$

In particular, $\sum_{j\neq i} \left(\mathcal{U}_\mu^{1\nearrow 2} \mathcal{D}_\mu^{2\searrow 1}\right)(i \to j) = \frac{1}{2}$, so the remaining transition probability for each row of $\mathcal{U}_\mu^{1\nearrow 2} \mathcal{D}_\mu^{2\searrow 1}$ is precisely $\left(\mathcal{U}_\mu^{1\nearrow 2} \mathcal{D}_\mu^{2\searrow 1}\right)(i \to i) = \frac{1}{2}$ for all $i \in \mathscr{U}$. Removing this diagonal and rescaling finally gives

$$\mathcal{Q}_\mu(i \to j) = \mu_1^i(j) = \frac{1}{n-1} \Pr_{\sigma\sim\mu}[j \in \sigma \mid i \in \sigma]$$

as desired.

2. Using Item 1, we see that

$$\mu_1(i) \cdot \mathcal{Q}_\mu(i \to j) = \frac{1}{n(n-1)} \cdot \Pr_{\sigma\sim\mu}[j \in \sigma \mid i \in \sigma] \cdot \Pr_{\sigma\sim\mu}[i \in \sigma] = \frac{1}{n(n-1)} \Pr_{\sigma\sim\mu}[i,j \in \sigma]$$

is symmetric in i, j. Hence, the detailed balance condition holds, establishing reversibility.

3. Following a similar calculation to Item 1, using Lemma 2.1.1, we calculate that for $i \neq j$,

$$\left(\mathcal{U}_\mu^{1\nearrow n} \mathcal{D}_\mu^{n\searrow 1}\right)(i \to j) = \frac{1}{n} \sum_{\sigma\in\operatorname{supp}(\mu):\sigma\supseteq\{i,j\}} \frac{\mu_n(\sigma)}{n \cdot \mu_1(i)} = \frac{n-1}{n} \cdot \mu_1^i(j).$$

Removing the remaining $\frac{1}{n}$ on the diagonal and rescaling gives the desired equality.

4. This follows immediately Items 1 and 2.

□

Historically, these local walks were first defined in the study of expansion phenomena in high-dimensional simplicial complexes [KM17; DK17; KO20b]. There, mathematicians were interested in establishing strong spectral gap lower bounds for these local walks $\mathcal{Q}_\mu$, a condition called *(one-sided) local spectral expansion.*

Definition 14 ((One-Sided) Local Spectral Expansion; [KO20b])**.** *We say μ satisfies* ***(one-sided)*** $(\lambda_0, \ldots, \lambda_{n-2})$***-local spectral expansion*** *if for every $0 \leq k \leq n-2$ and every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k$, we have $\lambda_2(\mathcal{Q}_{\mu^\sigma}) \leq \lambda_k$, or equivalently, $\gamma(\mathcal{Q}_{\mu^\sigma}) \geq 1 - \lambda_k$. We sometimes call μ a* ***(one-sided)*** $(\lambda_0, \ldots, \lambda_{n-2})$***-local spectral expander.***

Remark 7. Local spectral expanders are also sometimes simply called *local expanders*, *link expanders* [Dik+18], or *skeleton expanders* [KM17]. Although we will not need this, we note that *two-sided* local spectral expansion, where we additional require a lower bound on the smallest eigenvalue of $\mathcal{Q}_\mu$, has also been studied extensively in the literature. This was first introduced in [DK17], and subsequently developed in [Dik+18; DD19; GLL22; Baf+22b], with applications to solving constraint satisfaction problems (CSPs) [AGT19b; Baf+22a].

We discovered such local spectral expansion conditions are actually equivalent to spectral independence.

Lemma 2.3.3 (Spectral Independence $\iff$ Local Spectral Expansion; [ALO21])**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. Then we have the following identity:*

$$\mathcal{I}_\mu = \mathsf{Id} + (n-1) \cdot \left(\mathcal{Q}_\mu - \frac{n}{n-1} \cdot \mathbf{1} \otimes \mu_1 \right).$$

In particular, $\mathcal{I}_\mu$ has real eigenvalues, and for every $\eta_0, \dots, \eta_{n-2} \in \mathbb{R}$, the following are equivalent.

1. *μ is $(\eta_0, \dots, \eta_{n-2})$-spectrally independent.*
2. *μ is a $\left(\frac{\eta_0}{n-1}, \frac{\eta_1}{n-2}, \dots, \eta_{n-2}\right)$-local spectral expander.*

Proof. The identity can be established by applying Fact 2.3.2 entrywise. Note that the identity shows that the eigenvalues $\mathcal{I}_\mu$ are precisely given by applying $\lambda \mapsto 1 + (n-1)\lambda$ to the eigenvalues of $\mathcal{Q}_\mu$, except for one copy of the top eigenvalue 1 for $\mathcal{Q}_\mu$, which becomes an additional zero eigenvalue for $\mathcal{I}_\mu$. Conversely, the eigenvalues of $\mathcal{Q}_\mu$ are precisely given by applying $\lambda \mapsto \frac{\lambda - 1}{n-1}$ to the eigenvalues of $\mathcal{I}_\mu$, except for one copy of the zero eigenvalue, which becomes an additional copy of eigenvalue 1 for $\mathcal{Q}_\mu$. The remaining claims immediately follow. □

Remark 8. With this, one can restate Theorem 2.3.1 as

$$\gamma(\mathcal{P}_\mu) \ge \frac{1}{n} \min_{\emptyset = \sigma_0 \subsetneq \sigma_1 \subsetneq \cdots \subsetneq \sigma_n \in \mathrm{supp}(\mu)} \prod_{k=0}^{n-2} \gamma\left(\mathcal{Q}_{\mu^{\sigma_k}}\right) \ge \frac{1}{n} \prod_{k=0}^{n-2} (1 - \lambda_k)$$

under the assumption that μ is a $(\lambda_0, \dots, \lambda_{n-2})$-local spectral expander. In other words, informally, the spectral gap of the global down-up walk is at least the product of the local spectral gaps (with an extra $1/n$ factor).

Throughout this thesis, we will assume these local walks are *connected*, i.e. $\lambda_2(\mathcal{Q}_\mu) < 1$; from the perspective of spectral independence, this just says that the spectral independence of μ is strictly less than $n-1$. In fact, we will need this for all conditional distributions as well, which we formalize as follows.

Definition 15 (Total Connectivity)**.** *We say a distribution μ on $\binom{\mathscr{U}}{n}$ for a finite ground set $\mathscr{U}$ and positive integer $n \ge 1$ is **totally connected** if for every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| \le n-2$, the local walk $\mathcal{Q}_{\mu^\sigma}$ for the conditional distribution μ^σ is connected, i.e. $\lambda_2(\mathcal{Q}_{\mu^\sigma}) < 1$ or equivalently $\gamma(\mathcal{Q}_{\mu^\sigma}) > 0$. Note that by Lemma 2.3.3, this is also equivalent to $\lambda_{\max}(\mathcal{I}_{\mu^\sigma}) < n - |\sigma| - 1$ for every feasible $\sigma \subseteq \mathscr{U}$.*

We emphasize that this is a very mild technical condition which will be convenient, and will be satisfied by all of the distributions we encounter in this thesis.

2.3.2 The First Proof of Theorem 2.3.1: Decomposition via Conditioning and Spectral Analysis

We now give the first proof of Theorem 2.3.1. We note that previously, it was known how to deduce a spectral gap lower bound on the global down-up walk $\mathcal{P}_\mu$ given spectral gap lower bounds on the local walks $\mathcal{Q}_\mu$, in an analogous way to Theorem 2.3.1, although using a slightly more complicated and less intuitive induction [AL20]. Weaker versions of this implication were known even prior to that [KM17; DK17; KO20b], building on a beautiful work of Garland [Gar73].

For convenience, write $\mathcal{P}$ for $\mathcal{P}_\mu$, and recall that for every (feasible) element $u \in \mathscr{U}$, we have the associated down-up walks $\mathcal{P}^u = \mathcal{P}_{\mu^u}$ for each conditional distribution μ^u. The key is to "decompose" $\mathcal{P}$ into all of the $\mathcal{P}^u$, and then relate their spectral gaps. We have the following crucial proposition, which is strongly reminiscent of previous Markov chain decomposition results based on partitioning the domain. Here, the analog of the "restriction chains" are the $\mathcal{P}^u$, while the analog of the "projection chain" is the local walk $\mathcal{Q} = \mathcal{Q}_\mu$.

Proposition 2.3.4. *We have the following lower bound.*

$$\gamma(\mathcal{P}) \ge \frac{n-1}{n} \cdot \gamma(\mathcal{Q}) \cdot \min_{u \in \mathscr{U}} \{\gamma(\mathcal{P}^u)\}$$

In particular, by combining with Lemma 2.3.3, we see that if μ is η-spectrally independent, then we have the lower bound

$$\gamma(\mathcal{P}) \geq \frac{n-1}{n} \cdot \left(1 - \frac{\eta}{n-1}\right) \cdot \min_{u \in \mathscr{U}} \{\gamma(\mathcal{P}^u)\}.$$

By inductively applying this inequality to each $\mathcal{P}^u$, and each $\mathcal{P}^\sigma$ more generally, Theorem 2.3.1 immediately follows. Hence, all that remains is to establish Proposition 2.3.4, which is the heart of the proof.

Proof of Proposition 2.3.4. Since our proof is linear algebraic, and relies on decomposing $\mathcal{P}$ into the $\mathcal{P}^u$, every matrix or vector with subscript u (indicating it is living in $\mathbb{R}^{\mathrm{supp}\left(\mu^u_{n-1}\right) \times \mathrm{supp}\left(\mu^u_{n-1}\right)}$ or $\mathbb{R}^{\mathrm{supp}\left(\mu^u_{n-1}\right)}$, respectively), will be sufficiently padded with zeros so that they live in $\mathbb{R}^{\mathrm{supp}(\mu) \times \mathrm{supp}(\mu)}$ or $\mathbb{R}^{\mathrm{supp}(\mu)}$, respectively (see the discussion on padding in Section 1.4.1). Note that with this padding, the matrices $\mathcal{P}^u$ become self-adjoint w.r.t. the inner product $\langle \cdot, \cdot \rangle_\mu$, since for each $u \in \mathscr{U}$, one can rescale μ^u_{n-1} by a constant so that it agrees with μ on $\mathrm{supp}(\mu^u_{n-1})$. While this padding increases the multiplicity of the zero eigenvalue, this won't affect anything in our analysis of the spectral gaps.

With this in mind, we have the following identities

$$\mathcal{P} = \frac{1}{n} \sum_{u \in \mathscr{U}} \mathcal{P}^u$$
$$\mathsf{Id} = \frac{1}{n} \sum_{u \in \mathscr{U}} \mathsf{Id}^u,$$

where Id^u is the identity matrix on $\mathbb{R}^{\mathrm{supp}\left(\mu^u_{n-1}\right) \times \mathrm{supp}\left(\mu^u_{n-1}\right)}$ (appropriately padded), and we use the fact that each $\sigma \in \mathrm{supp}(\mu)$ contains exactly n distinct elements in $\mathscr{U}$. Viewing the stationary distribution $\mu^u = \mu^u_{n-1}$ of $\mathcal{P}^u$ as a row vector (appropriately padded), it follows that

$$\begin{aligned}
\mathsf{Id} - \mathcal{P} &= \frac{1}{n} \sum_{u \in \mathscr{U}} (\mathsf{Id}^u - \mathcal{P}^u) \\
&\succeq_\mu \frac{1}{n} \sum_{u \in \mathscr{U}} \gamma(\mathcal{P}^u) \cdot (\mathsf{Id}^u - \mathbf{1}^u \otimes \mu^u) && \text{(Lemma 1.4.20)} \\
&\succeq_\mu \min_{u \in \mathscr{U}} \{\gamma(\mathcal{P}^u)\} \cdot \Bigg(\mathsf{Id} - \underbrace{\frac{1}{n} \sum_{u \in \mathscr{U}} \mathbf{1}^u \otimes \mu^u}_{(*)} \Bigg) \\
&= \min_{u \in \mathscr{U}} \{\gamma(\mathcal{P}^u)\} \cdot \left(\mathsf{Id} - \mathcal{D}^{n \searrow 1} \mathcal{U}^{1 \nearrow n}\right) && (\mathcal{D}^{n \searrow 1} \mathcal{U}^{1 \nearrow n} = (*) \text{ by inspection.})
\end{aligned}$$

Comparing the eigenvalues of these matrices, we obtain the desired lower bound, since

$$\begin{aligned}
\gamma(\mathcal{P}) &\geq \gamma\left(\mathcal{D}^{n \searrow 1} \mathcal{U}^{1 \nearrow n}\right) \cdot \min_{u \in \mathscr{U}} \{\gamma(\mathcal{P}^u)\} \\
&= \gamma\left(\mathcal{U}^{1 \nearrow n} \mathcal{D}^{n \searrow 1}\right) \cdot \min_{u \in \mathscr{U}} \{\gamma(\mathcal{P}^u)\} && \text{(Fact 1.4.7)} \\
&= \frac{n-1}{n} \cdot \gamma(\mathcal{Q}) \cdot \min_{u \in \mathscr{U}} \{\gamma(\mathcal{P}^u)\}. && \text{(Fact 2.3.2)}
\end{aligned}$$

□

Remark 9. One can generalize the above analysis significantly. For instance, one can show that for every k, if we write $\mathcal{P}^\sigma = \mathcal{P}_{\mu^\sigma}$ for the down-up walk with stationary distribution μ^σ, then we have the decomposition

$$\mathsf{Id} - \mathcal{P} = \frac{1}{\binom{n}{k}} \sum_{\sigma \in \mathrm{supp}(\mu_k)} (\mathsf{Id}^\sigma - \mathcal{P}^\sigma)$$

which readily yields the spectral inequality

$$\begin{aligned}\mathsf{Id} - \mathcal{P} &\succeq_\mu \min_{\sigma \in \mathrm{supp}(\mu_k)} \{\gamma\left(\mathcal{P}^\sigma\right)\} \cdot \left(\mathsf{Id} - \frac{1}{\binom{n}{k}} \sum_{\sigma \in \mathrm{supp}(\mu_k)} \mathbf{1}^\sigma \otimes \mu^\sigma\right) \\ &= \min_{\sigma \in \mathrm{supp}(\mu_k)} \{\gamma\left(\mathcal{P}^\sigma\right)\} \cdot \left(\mathsf{Id} - \mathcal{D}^{n \searrow k} \mathcal{U}^{k \nearrow n}\right).\end{aligned}$$

This can be thought of as a way to *compare* the mixing behavior of $\mathcal{D}^{n \searrow n-1} \mathcal{U}^{n-1 \nearrow n}$ with the mixing behavior of $\mathcal{D}^{n \searrow k} \mathcal{U}^{k \nearrow n}$ for general k. This is useful because one intuitively expects $\mathcal{D}^{n \searrow k} \mathcal{U}^{k \nearrow n}$ to mix *faster* as k decreases. Indeed, the smaller k is, the less local $\mathcal{D}^{n \searrow k} \mathcal{U}^{k \nearrow n}$ is, and the more computationally difficult it is to implement each step. For instance, in the extremal case where $k = 0$, the Markov chain $\mathcal{D}^{n \searrow 0} \mathcal{U}^{k \nearrow 0} = \mathbf{1} \otimes \mu$ is the trivial one which mixes in one step, but requires one to already be able to perfectly sample from μ efficiently, defeating the whole point of using Markov chains for sampling from μ.

One unfortunate feature of this comparison inequality, however, is the fact that we have to look at the *worst* spectral gap $\gamma\left(\mathcal{P}^\sigma\right)$ over all $\sigma \in \mathrm{supp}\left(\mu_k\right)$. Despite this, we will still leverage this intuition in later chapters (e.g. Chapter 10), although under a different formulation which permits a more delicate analysis that goes beyond the worst case.

Remark 10. The original proof of Theorem 2.3.1 given in [AL20], which closely follows the strategy in [KO18], is to analyze the global random walks "layer by layer". More specifically, one can study the spectral gap of $\mathcal{D}_\mu^{k+1 \searrow k} \mathcal{U}_\mu^{k \nearrow k+1}$ by doing the following:

1. First, bound the spectral gap of $\mathcal{D}_\mu^{k \searrow k-1} \mathcal{U}_\mu^{k-1 \nearrow k}$. This can be done inductively. The base case is $\mathcal{D}_\mu^{2 \searrow 1} \mathcal{U}_\mu^{1 \nearrow 2}$, which shares the same eigenvalues as $\mathcal{U}_\mu^{1 \nearrow 2} \mathcal{D}_\mu^{2 \searrow 1} = \frac{1}{2}\left(\mathcal{Q}_\mu + \mathsf{Id}\right)$.

2. Roughly speaking, the difference between $\mathcal{D}_\mu^{k \searrow k-1} \mathcal{U}_\mu^{k-1 \nearrow k}$ and $\mathcal{U}_\mu^{k \nearrow k+1} \mathcal{D}_\mu^{k+1 \searrow k}$ decomposes as a sum over local walks of conditional distributions μ^ξ, where $\xi \subseteq \mathscr{U}$ is feasible and $|\xi| = k - 1$. Hence, one can control the eigenvalues of $\mathcal{U}_\mu^{k \nearrow k+1} \mathcal{D}_\mu^{k+1 \searrow k}$ in terms of the eigenvalues of $\mathcal{D}_\mu^{k \searrow k-1} \mathcal{U}_\mu^{k-1 \nearrow k}$ and the spectral independence of the μ^ξ. This can also be viewed as a Markov chain comparison statement.

3. Use the fact that $\mathcal{U}_\mu^{k \nearrow k+1} \mathcal{D}_\mu^{k+1 \searrow k}$ and $\mathcal{D}_\mu^{k+1 \searrow k} \mathcal{U}_\mu^{k \nearrow k+1}$ share the same eigenvalues to bound the spectral gap of the latter.

This approach has the advantage that one can get some control on *all* eigenvalues of the global random walks, not just the spectral gap. We chose not to present this "layer-by-layer" spectral analysis in full detail as it is not ours. We discovered the above proof after the release of [AL20], which we believe can be useful pedagogically. We refer interested readers to [AL20] for the details of this "layer-by-layer approach".

2.4 Switching Perspectives: Variance Contraction

The first proof of Theorem 2.3.1 was nice and intuitive since it only used elementary linear algebra, and directly decomposed the down-up walk $\mathcal{P}_\mu$. However, this analysis is much more difficult to extend beyond the study of spectral gaps. It is also challenging to extend to distributions where we do not necessarily have strong spectral independence guarantees for all conditional distributions.

Our goal in this section is to provide a second proof of Theorem 2.3.1 using tools with a much more probabilistic flavor. We will view mixing as contraction of (relative) variance under the Markov chain, and show how certain "local" variance contraction conditions imply "global" variance contraction under $\mathcal{P}_\mu$. These "local" variance contraction conditions are intimately related to fast mixing conditions for the local walk $\mathcal{Q}_\mu$ and spectral independence discussed previously. We will then see later on how this analysis is easily extended to contraction of (relative) entropy, which yields bounds on the (modified) log-Sobolev constants. This opens up the potential for tight analyses of the mixing time.

First, we set up some more notation, which will allow from a more granular analysis.

Global Functions and Local Functions We typically refer to a function $f_n : \mathrm{supp}(\mu) \to \mathbb{R}$ as a *global function*, since it is supported on the entire domain. One should think of f_n as being the *density* $\frac{d\nu}{d\mu}$ of some other probability measure ν on $\binom{\mathscr{U}}{n}$ w.r.t. μ, such as the initial distribution of the Markov chain. By applying the up operators, we can *project* a global function $f_n : \mathrm{supp}(\mu) \to \mathbb{R}$ down onto lower levels, producing *projections* $f_k \stackrel{\mathsf{def}}{=} \mathcal{U}_\mu^{k\nearrow n} f_n : \mathrm{supp}(\mu_k) \to \mathbb{R}$ for each $0 \le k \le n$.

We can further *localize* a global function $f_n : \mathrm{supp}(\mu) \to \mathbb{R}$ by looking at its induced *local functions* $f_k^\sigma : \mathrm{supp}(\mu_k^\sigma) \to \mathbb{R}$ for each feasible $\sigma \subseteq \mathscr{U}$ and $0 \le k \le n - |\sigma|$. These are defined by considering the projection f_ℓ for $\ell = k + |\sigma|$ and *restricting* to $\mathrm{supp}(\mu_k^\sigma)$, more specifically, setting $f_k^\sigma(\sigma') \stackrel{\mathsf{def}}{=} f_\ell(\sigma \cup \sigma')$ for every $\sigma' \in \mathrm{supp}(\mu_k^\sigma)$. One can imagine each f_k^σ as providing a "local point of view" of the function f_n from the perspective of σ.

The idea that properties of these local functions f_k^σ can be used to deduce useful properties about the global function f_n, especially in regards to how they interact with the down and up operators, was pioneered by Garland [Gar73] in studying the cohomology of discrete groups and simplicial complexes. This is sometimes known as *Garland's method* in the literature, and has since been used to great effect in many other contexts as well; see e.g. [BŚ97; GW12; Opp18].

We have the following useful facts.

Fact 2.4.1. *Let $f_n : \mathrm{supp}(\mu) \to \mathbb{R}$ be a global function. Then the following hold.*

1. *For every $0 \le k \le \ell \le n$, we have the identity*

$$f_k(\tau) = \mathbb{E}_{\sigma\sim\mu_\ell}\left[f_\ell(\sigma) \mid \sigma \supseteq \tau\right], \quad \forall \tau \in \mathrm{supp}(\mu_k).$$

Similarly, for every feasible $\tau \subseteq \mathscr{U}$, every $0 \le k \le n - |\tau|$, and every $k + |\tau| \le \ell \le n$, the entries of the local function f_k^τ satisfy

$$f_k^\tau(\tau') = \mathbb{E}_{\sigma\sim\mu_\ell}\left[f_\ell(\sigma) \mid \sigma \supseteq \tau \cup \tau'\right], \quad \forall \tau' \in \mathrm{supp}(\mu_k^\tau).$$

2. *For every $0 \le k \le \ell \le n$, the projections f_k, f_ℓ have the same expectation, i.e.*

$$\mathbb{E}_{\sigma_k\sim\mu_k}\left[f_k(\sigma_k)\right] = \mathbb{E}_{\sigma_\ell\sim\mu_\ell}\left[f_\ell(\sigma_\ell)\right].$$

3. *If f_n is the density $\frac{d\nu}{d\mu}$ of some other probability measure ν on $\binom{\mathscr{U}}{n}$ w.r.t. μ, then for every feasible $\tau \subseteq \mathscr{U}$ and every $0 \le k \le n - |\tau|$, we have the following alternative expression for the entries of the local function f_k^τ:*

$$f_k^\tau(\tau') = \frac{\Pr_{\sigma\sim\nu}\left[\sigma \supseteq \tau \cup \tau'\right]}{\Pr_{\sigma\sim\mu}\left[\sigma \supseteq \tau \cup \tau'\right]}.$$

4. *Fix $0 \le k \le n$. If f_n is the density of some other probability measure ν on $\mathrm{supp}(\mu)$ w.r.t. μ, then $\mathcal{U}_\mu^{k\nearrow n} f_n : \mathrm{supp}(\mu_k) \to \mathbb{R}_{\ge 0}$ is the density of $\nu\mathcal{D}_\mu^{n\searrow k}$ w.r.t. μ_k. Similarly, if f_k is the density of some other probability measure ν_k on $\mathrm{supp}(\mu_k)$ w.r.t. μ_k, then $\mathcal{D}_\mu^{n\searrow k} f_k : \mathrm{supp}(\mu) \to \mathbb{R}_{\ge 0}$ is the density of $\nu_k\mathcal{U}_\mu^{k\nearrow n}$ w.r.t. μ. The same claim holds mutatis mutandis for the conditional and marginal distributions of μ as well.*

Proof. 1. By definition, $f_k = \mathcal{U}_\mu^{k\nearrow n} f_n = \mathcal{U}_\mu^{k\nearrow \ell}\mathcal{U}_\mu^{\ell\nearrow n} f_n = \mathcal{U}_\mu^{k\nearrow \ell} f_\ell$. The row of $\mathcal{U}_\mu^{k\nearrow \ell}$ indexed by S is given precisely by the conditional distribution $\sigma \mapsto \frac{\mu_\ell(\sigma)}{\binom{\ell}{k}\mu_k(\tau)}$ whenever $\sigma \supseteq \tau$, from which the claim follows immediately. The second identity follows from the first since $f_k^\tau(\tau') = f_{k+|\tau|}(\tau \cup \tau')$ by definition.

2. We may express the expectation as an inner product with the constant all-ones function, on which we can apply the adjointness property from Lemma 2.1.2. In particular, we see that

$$\mathbb{E}_{\sigma_k\sim\mu_k}\left[f_k(\sigma_k)\right] = \left\langle \mathcal{U}_\mu^{k\nearrow n} f_n, \mathbf{1}\right\rangle_{\mu_k} = \left\langle f_n, \mathcal{D}_\mu^{n\searrow k}\mathbf{1}\right\rangle_{\mu_n} = \left\langle f_n, \mathbf{1}\right\rangle_{\mu_n} = \mathbb{E}_{\sigma_n\sim\mu_n}\left[f_n(\sigma_n)\right]$$

is independent of k.

3. Using Item 1, we have that

$$\begin{aligned} f_k^\tau(\tau') &= \mathbb{E}_{\sigma\sim\mu}\left[\frac{\nu(\sigma)}{\mu(\sigma)} \mid \sigma \supseteq \tau\cup\tau'\right] \\ &= \sum_{\sigma\in\mathrm{supp}(\mu):\sigma\supseteq\tau\cup\tau'} \frac{\mu(\sigma)}{\Pr_{\sigma'\sim\mu}\left[\sigma'\supseteq\tau\cup\tau'\right]}\cdot\frac{\nu(\sigma)}{\mu(\sigma)} \\ &= \frac{\Pr_{\sigma\sim\nu}\left[\sigma\supseteq\tau\cup\tau'\right]}{\Pr_{\sigma\sim\mu}\left[\sigma\supseteq\tau\cup\tau'\right]}. \end{aligned}$$

4. For the first claim, we use Eq. (2.3). For every $\sigma\in\mathrm{supp}(\mu_k)$,

$$\left(\mathcal{U}_\mu^{k\nearrow n}\frac{d\nu}{d\mu}\right)(\sigma) = \frac{1}{\binom{n}{k}}\sum_{\tau\in\mathrm{supp}(\mu):\tau\supseteq\sigma}\frac{\mu(\tau)}{\mu_k(\sigma)}\cdot\frac{\nu(\tau)}{\mu(\tau)} = \frac{\nu_k(\sigma)}{\mu_k(\sigma)}.$$

Similarly, using Eqs. (2.1) and (2.3), we have that for every $\tau\in\mathrm{supp}(\mu)$,

$$\begin{aligned} \frac{d\left(\nu_k\mathcal{U}_\mu^{k\nearrow n}\right)}{d\mu}(\tau) &= \frac{1}{\mu(\tau)}\sum_{\sigma\in\mathrm{supp}(\mu_k):\sigma\subseteq\tau}\nu_k(\sigma)\cdot\frac{\mu(\tau)}{\binom{n}{k}\cdot\mu_k(\sigma)} \\ &= \frac{1}{\binom{n}{k}}\sum_{\sigma\in\mathrm{supp}(\mu_k):\sigma\subseteq\tau}\frac{\nu_k(\sigma)}{\mu_k(\sigma)} = \left(\mathcal{D}_\mu^{k\searrow n}\frac{d\nu_k}{d\mu_k}\right)(\tau). \end{aligned}$$

□

(Relative) Variance For convenience and clarity, especially when the distribution μ on $\binom{\mathscr{U}}{n}$ is clear from context, we will write $\mathrm{Var}_k(\cdot)$ instead of Var_{μ_k}. Similarly, if we are conditioning on a feasible $\sigma\subseteq\mathscr{U}$, we will write $\mathrm{Var}_k^\sigma(\cdot)$ instead of $\mathrm{Var}_{\mu_k^\sigma}$.

The following variance decomposition result will be incredibly useful. It is essentially the Law of Total Variance.

Lemma 2.4.2 (Variance Decomposition)**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0\le n\le|\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. Then for every $0\le k\le\ell\le n$ and every function $f_\ell:\mathrm{supp}(\mu_\ell)\to\mathbb{R}$, we have that*

$$\mathrm{Var}_\ell(f_\ell) = \mathrm{Var}_k(f_k) + \mathbb{E}_{\sigma\sim\mu_k}\left[\mathrm{Var}_{\ell-k}^\sigma\left(f_{\ell-k}^\sigma\right)\right]$$

where f_k and $f_{\ell-k}^\sigma$ are local functions induced by f_ℓ.

Proof. Using the Law of Total Expectation and the fact that $f_\ell(\tau) = f_{\ell-k}^\sigma(\tau\setminus\sigma)$ whenever $\tau\supseteq\sigma$, we have that

$$\begin{aligned} \mathbb{E}_{\tau\sim\mu_\ell}\left[f_\ell(\tau)^2\right] &= \mathbb{E}_{\sigma\sim\mu_k}\left[\mathbb{E}_{\tau\sim\mu_\ell}\left[f_\ell(\tau)^2 \mid \tau\supseteq\sigma\right]\right] \\ &= \mathbb{E}_{\sigma\sim\mu_k}\left[\mathbb{E}_{\sigma'\sim\mu_{\ell-k}^\sigma}\left[f_{\ell-k}^\sigma(\sigma')\right]^2\right] + \mathbb{E}_{\sigma\sim\mu_k}\left[\mathrm{Var}_{\ell-k}^\sigma\left(f_{\ell-k}^\sigma\right)\right] \\ &= \mathbb{E}_{\sigma\sim\mu_k}\left[f_k(\sigma)^2\right] + \mathbb{E}_{\sigma\sim\mu_k}\left[\mathrm{Var}_{\ell-k}^\sigma\left(f_{\ell-k}^\sigma\right)\right]. \end{aligned}$$

Since $\mathbb{E}_{\mu_\ell}[f_\ell] = \mathbb{E}_{\mu_k}[f_k]$, subtracting (the square of) both sides gives the desired result. □

2.4.1 Equivalent Global Mixing Conditions: Spectral Gap and Variance Contraction

Before giving our second proof of Theorem 2.3.1, we first discuss different ways of thinking about rapid mixing of the global down-up walk $\mathcal{P}_\mu$. Having these different perspectives will be invaluable later on, especially when we wish to establish stronger local-to-global theorems.

Proposition 2.4.3 (Variance and Spectral Gap)**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0\le n\le|\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. For every $0\le\beta\le1$, the following are equivalent.*

1. ***(Global) Variance Contraction:*** *For every global function* $f_n : \mathrm{supp}(\mu) \to \mathbb{R}$ *with level-*$(n-1)$ *projection* $f_{n-1} = \mathcal{U}_\mu^{n-1\nearrow n} f_n$*, we have the inequality*

$$\mathrm{Var}_{n-1}(f_{n-1}) \le \beta \cdot \mathrm{Var}_n(f_n). \tag{2.10}$$

2. ***Spectral Gap:*** *We have that* $\gamma(\mathcal{P}_\mu) \ge 1-\beta$*.*

Furthermore, these conditions imply the following.

1. ***(Relative) Variance Decay:*** *For every probability distribution* ν *on* $\mathrm{supp}(\mu)$*, we have that*

$$\mathscr{D}_{\chi^2}(\nu\mathcal{P}_\mu \| \mu) \le \beta^2 \cdot \mathscr{D}_{\chi^2}(\nu\|\mu).$$

2. ***Mixing:*** *The down-up walk* $\mathcal{P}_\mu$ *mixes in* $O\left(\frac{1}{1-\beta}\log\frac{1}{\mu_{\min}}\right)$*-steps.*

Proof. First, we establish the equivalence between Items 1 and 2. Observe that

$$\begin{aligned}
\mathrm{Var}_{n-1}(f_{n-1}) &= \left\langle \mathcal{U}^{n-1\nearrow n} f_n, \mathcal{U}^{n-1\nearrow n} f_n \right\rangle_{\mu_{n-1}} - \left\langle \mathcal{U}^{n-1\nearrow n} f_n, \mathbf{1} \right\rangle_{\mu_{n-1}}^2 \\
&= \langle \mathcal{P} f_n, f_n \rangle_\mu - \langle f_n, \mathbf{1}\rangle_\mu^2 && \text{(Lemma 2.1.2)} \\
\mathrm{Var}_n(f_n) &= \langle f_n, f_n\rangle_\mu - \langle f_n, \mathbf{1}\rangle_\mu^2
\end{aligned}$$

It follows (by rearranging) that Item 1 is equivalent to

$$(1-\beta)\cdot \mathrm{Var}_n(f_n) \le \langle (\mathsf{Id} - \mathcal{P}) f_n, f_n\rangle_\mu = \mathcal{E}_{\mathcal{P}}(f_n, f_n)$$

for every global function $f_n : \mathrm{supp}(\mu) \to \mathbb{R}$. This is exactly a Poincaré Inequality with constant $1-\beta$ for $\mathcal{P}$, which is equivalent to $\gamma(\mathcal{P}) \ge 1-\beta$ by Fact 1.4.21.

To establish the rate of decay in (relative) variance, we observe that by Fact 2.4.1, if we write $f = \frac{d\nu}{d\mu} : \Omega \to \mathbb{R}_{\ge 0}$, then

$$\begin{aligned}
\mathscr{D}_{\chi^2}(\nu\mathcal{P}_\mu\|\mu) &= \mathrm{Var}_\mu\left(\frac{d(\nu\mathcal{P}_\mu)}{d\mu}\right) = \mathrm{Var}_\mu(\mathcal{P}_\mu f) \\
&= \langle \mathcal{P}_\mu f, \mathcal{P}_\mu f\rangle_\mu - \langle \mathcal{P}_\mu f, \mathbf{1}\rangle_\mu^2 = \langle f, \mathcal{P}_\mu^2 f\rangle - \langle f, \mathbf{1}\rangle_\mu^2 \\
\mathscr{D}_{\chi^2}(\nu\|\mu) &= \mathrm{Var}_\mu(f) = \langle f, f\rangle_\mu - \langle f, \mathbf{1}\rangle_\mu^2
\end{aligned}$$

Hence, after rearranging, we see that decay of (relative) variance with rate β^2 in Item 1 is equivalent to $(1-\beta^2)\cdot \mathrm{Var}_\mu(f) \le \mathcal{E}_{\mathcal{P}_\mu^2}(f,f)$, which is implied by the spectral gap lower bound

$$\gamma(\mathcal{P}_\mu^2) = 1 - (1-\gamma(\mathcal{P}_\mu))^2 \ge 1-\beta^2.$$

That Items 1 and 2 imply mixing follows immediately from Theorem 1.4.19. □

In information theory, variance contraction Eq. (2.10) is more traditionally referred to as a *(strong) data processing inequality* with constant β for the Markov operator (or *channel*) $\mathcal{D}_\mu^{n\searrow n-1}$ w.r.t. χ^2-divergence; see e.g. [Rag16] and references therein.

2.4.2 The Second Proof of Theorem 2.3.1: Local-to-Global Variance Contraction

Having set up the necessary notation, we now provide our second proof of Theorem 2.3.1. As previously mentioned, this second proof based more on direct decomposition of the (relative) variance. We will show that a local version of variance contraction implies the global version of variance contraction, which recall is equivalent to a spectral gap lower bound by Proposition 2.4.3. The local version of variance contraction will turn out to be implied by, and almost equivalent to, spectral independence. We define it now.

Definition 16 (Local Variance Contraction; [CLV21a]). *For $0 \le \alpha \le 1$, we say μ satisfies α-**local variance contraction** if for every global function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}$, we have that the induced projections f_1, f_2 satisfy*

$$\operatorname{Var}_1(f_1) \le \alpha \cdot \operatorname{Var}_2(f_2).$$

*Similarly, for $0 \le \alpha_0, \dots, \alpha_{n-2} \le 1$, we say μ satisfies $(\alpha_0, \dots, \alpha_{n-2})$-**local variance contraction** if for every global function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}$, every $0 \le k \le n-2$, and every $\sigma \in \operatorname{supp}(\mu_k)$, the induced local functions f_1^σ, f_2^σ satisfy*

$$\operatorname{Var}_1^\sigma(f_1^\sigma) \le \alpha_k \cdot \operatorname{Var}_2^\sigma(f_2^\sigma).$$

Lemma 2.4.4 (Spectral Independence $\implies$ Local Variance Contraction). *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a distribution over $\binom{\mathscr{U}}{n}$. If μ is η-spectrally independent for some $\eta \in \mathbb{R}$, then μ satisfies α-local variance contraction with $\alpha = \frac{1}{2}\left(1 + \frac{\eta}{n-1}\right)$. Similarly, if μ is $(\eta_0, \dots, \eta_{n-2})$-spectrally independent, then μ satisfies $(\alpha_0, \dots, \alpha_{n-2})$-local variance contraction with $\alpha_k = \frac{1}{2}\left(1 + \frac{\eta_k}{n-k-1}\right)$ for every $0 \le k \le n-2$.*

Remark 11. The original definition of local variance contraction stated in [CLV21a] requires

$$\operatorname{Var}_1\left(\mathcal{U}_\mu^{1 \nearrow 2} f\right) \le \alpha \cdot \operatorname{Var}_2(f)$$

to hold for every function $f : \operatorname{supp}(\mu_2) \to \mathbb{R}$, rather than restricting attention to local functions f_2 induced by some global function $f_n : \operatorname{supp}(\mu_n) \to \mathbb{R}$. This stronger version of local variance contraction is equivalent to spectral independence, using a proof highly analogous to the proof of Proposition 2.4.3. We also note that the version of local variance contraction stated above was also independently studied in [KM20].

We have the following result, which shows that local variance contraction in the sense of Definition 16 implies global variance contraction. By combining this with Lemma 2.4.4, which says that spectral independence implies local variance contraction, and Proposition 2.4.3, which says that spectral gap for $\mathcal{P}_\mu$ is equivalent to global variance contraction, we obtain an alternative proof of Theorem 2.3.1, albeit with a slightly worse dependence on the spectral independence parameters.[3]

Theorem 2.4.5 (Local-to-Global Variance Contraction; [CLV21a]). *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. If μ satisfies $(\alpha_0, \dots, \alpha_{n-2})$-local variance contraction, then for every $0 \le k \le \ell \le n$ and every global function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}$, the induced projections $f_k = \mathcal{U}^{k \nearrow n} f_n : \operatorname{supp}(\mu_k) \to \mathbb{R}$ and $f_\ell = \mathcal{U}^{\ell \nearrow n} f_n : \operatorname{supp}(\mu_\ell) \to \mathbb{R}$ satisfy the inequality*

$$\frac{\operatorname{Var}_k(f_k)}{\sum_{j=0}^{k-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)} \le \frac{\operatorname{Var}_\ell(f_\ell)}{\sum_{j=0}^{\ell-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}. \tag{2.11}$$

In particular, for every $0 \le k \le n$, the distribution μ satisfies global variance contraction with constant β where

$$\beta \le \frac{\sum_{j=0}^{n-2} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}{\sum_{j=0}^{n-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}.$$

Proof. It suffices to show that for every $0 \le k < n$, we have the inequality

$$\frac{\operatorname{Var}_k(f_k)}{\sum_{j=0}^{k-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)} \le \frac{\operatorname{Var}_{k+1}(f_{k+1})}{\sum_{j=0}^{k} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)} \tag{2.12}$$

since the general case Eq. (2.11) follows by chaining these together. We prove Eq. (2.12) by induction. The case where $k = 0$ is trivial since the left-hand side is 0, and the first nontrivial base

[3] One can imagine that the dependence on the spectral independence parameter η is slightly worse because we technically didn't use the full power of an eigenvalue bound. Indeed, in this second proof, we only used variance contraction for local functions induced by some global function, rather than variance contraction for all possible local functions. For a more detailed comparison of dependencies on η, see [GM20; CLV21a].

case $k = 1$ follows immediately by the definition of local variance contraction. For the inductive step, we have that

$$\begin{aligned}
\mathrm{Var}_{k+1}(f_{k+1}) - \mathrm{Var}_{k-1}(f_{k-1}) &= \mathbb{E}_{\sigma\sim\mu_{k-1}}\left[\mathrm{Var}_2^{\sigma}(f_2^{\sigma})\right] && \text{(Lemma 2.4.2)}\\
&\geq \frac{1}{\alpha_{k-1}} \cdot \mathbb{E}_{\sigma\sim\mu_{k-1}}\left[\mathrm{Var}_1^{\sigma}(f_1^{\sigma})\right] && \text{(Definition 16)}\\
&= \frac{1}{\alpha_{k-1}} \cdot \left(\mathrm{Var}_k(f_k) - \mathrm{Var}_{k-1}(f_{k-1})\right) && \text{(Lemma 2.4.2)}
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{Var}_{k+1}(f_{k+1}) &\geq \frac{1}{\alpha_{k-1}} \cdot \mathrm{Var}_k(f_k) - \left(\frac{1}{\alpha_{k-1}} - 1\right) \cdot \mathrm{Var}_{k-1}(f_{k-1})\\
&\geq \left(\frac{1}{\alpha_{k-1}} - \left(\frac{1}{\alpha_{k-1}} - 1\right) \cdot \frac{\sum_{j=0}^{k-2}\prod_{i=0}^{j-1}\left(\frac{1}{\alpha_i} - 1\right)}{\sum_{j=0}^{k-1}\prod_{i=0}^{j-1}\left(\frac{1}{\alpha_i} - 1\right)}\right) \cdot \mathrm{Var}_k(f_k)\\
&\qquad\qquad \text{(Inductive Hypothesis)}\\
&= \frac{\sum_{j=0}^{k}\prod_{i=0}^{j-1}\left(\frac{1}{\alpha_i} - 1\right)}{\sum_{j=0}^{k-1}\prod_{i=0}^{j-1}\left(\frac{1}{\alpha_i} - 1\right)} \cdot \mathrm{Var}_k(f_k)
\end{aligned}$$

as desired. □

All that remains is to prove Lemma 2.4.4. We do this now.

Proof of Lemma 2.4.4. We follow the proof of Proposition 2.4.3. Observe that

$$\begin{aligned}
\mathrm{Var}_1(f_1) &= \left\langle \mathcal{U}^{1\nearrow 2} f_2, \mathcal{U}^{1\nearrow 2} f_2\right\rangle_{\mu_1} - \left\langle \mathcal{U}^{1\nearrow 2} f, \mathbf{1}\right\rangle_{\mu_1}^2\\
&= \left\langle \mathcal{D}^{2\searrow 1}\mathcal{U}^{1\nearrow 2} f_2, f_2\right\rangle_{\mu_2} - \left\langle f_2, \mathbf{1}\right\rangle_{\mu_2}^2 && \text{(Lemma 2.1.2)}\\
\mathrm{Var}_2(f_2) &= \left\langle f_2, f_2\right\rangle_{\mu_2} - \left\langle f_2, \mathbf{1}\right\rangle_{\mu_2}^2
\end{aligned}$$

It follows (by rearranging) that local variance contraction is equivalent to

$$(1-\alpha)\cdot \mathrm{Var}_2(f_2) \leq \mathcal{E}_{\mathcal{D}^{2\searrow 1}\mathcal{U}^{1\nearrow 2}}(f_2, f_2)$$

for all f_2 induced by a global function f_n, which is implied by a spectral gap lower bound $\gamma\left(\mathcal{D}^{2\searrow 1}\mathcal{U}^{1\nearrow 2}\right) \geq 1-\alpha$. Hence, it suffices to show that η-spectral independence is equivalent to $\gamma\left(\mathcal{D}^{2\searrow 1}\mathcal{U}^{1\nearrow 2}\right) \geq \frac{1}{2}\left(1 - \frac{\eta}{n-1}\right)$. Using Lemma 2.3.3, which recall establishes the equivalence between spectral independence and a spectral gap lower bound on the corresponding local walk $\mathcal{Q} = 2\cdot\left(\mathcal{U}^{1\nearrow 2}\mathcal{D}^{2\searrow 1} - \frac{1}{2}\cdot \mathsf{Id}\right)$, we see that η-spectral independence is equivalent to

$$1 - \frac{\eta}{n-1} \leq \gamma(\mathcal{Q}_\mu) = 2\cdot\gamma\left(\mathcal{U}^{1\nearrow 2}\mathcal{D}^{2\searrow 1}\right) = 2\cdot\gamma\left(\mathcal{D}^{2\searrow 1}\mathcal{U}^{1\nearrow 2}\right).$$

□

2.5 Alternative Formulations of Spectral Independence in Structured Spaces

In this section, we consider alternative formulations of spectral independence on spaces which are more structured than the more general setting we considered previously. Specifically, when our distribution is supported over a discrete product space, such as the discrete hypercube $\{0,1\}^n$, there are simpler and cleaner formulations of influence which turn out to be equivalent to the original formulation in Definition 11. At a high level, this is because the product structure induces additional nonrelevant or *trivial* eigenvalues which we can annihilate[4]. We will use these other notions of influence in later chapters.

[4] In the language of algebraic topology, the simplicial complexes which naturally arise in these settings are n-*partite*.

2.5.1 The Discrete Hypercube

We begin with the simplest setting, namely distributions μ over the discrete hypercube $\{0,1\}^n$, which can also be thought of as distributions over all subsets of $[n]$. Recall that one can fit this within our original formalism by taking the ground set $\mathscr{U}$ to be the set of coordinate-assignment pairs (i,s), where $i \in [n]$ and $s \in \{0,1\}$, and then viewing a point in $\{0,1\}^n$ as a size-n collection of such coordinate-assignment pairs. Within this formalism, the one-sided influence matrix defined in Definition 11 is $2n \times 2n$. However, in this setting, there is a much more intuitive formulation which turns out to be equivalent spectrally.

Definition 17 ((Two-Sided) Influence; [ALO21])**.** *Let μ be a probability distribution over $\{0,1\}^n$. For every pair of coordinates $i,j \in [n]$, define the **(two-sided) influence of i on j** by*

$$\Psi_\mu(i \to j) \stackrel{\text{def}}{=} \Pr_{\sigma\sim\mu}[\sigma_j = 1 \mid \sigma_i = 1] - \Pr_{\sigma\sim\mu}[\sigma_j = 1 \mid \sigma_i = 0],$$

*with the convention that $\Pr_{\sigma\sim\mu}[\sigma_i = 1 \mid \sigma_i = 1] = 1$ and $\Pr_{\sigma\sim\mu}[\sigma_i = 1 \mid \sigma_i = 0] = 0$ for all $i \in [n]$. We also define the **(two-sided) influence matrix** $\Psi_\mu \in \mathbb{R}^{n\times n}$ to be the matrix with entries $\Psi_\mu(i,j) = \Psi_\mu(i \to j)$ for every $i,j \in [n]$.*

With this alternative notion of influence comes an alternative notion of spectral independence.

Definition 18 (Spectral Independence in the Hypercube; [ALO21])**.** *For $\eta \in \mathbb{R}$, we say μ is **η-spectrally independent** if $\lambda_{\max}(\Psi_\mu) \le 1+\eta$. Analogously, for $\eta_0,\dots,\eta_{n-2} \in \mathbb{R}$, we say μ is $(\eta_0,\dots,\eta_{n-2})$-spectrally independent if for $0 \le k \le n-2$ and every feasible partial assignment $\tau : S \to \{0,1\}$ on a subset of coordinates $S \subseteq [n]$ with $|S| = k$, we have that the conditional distribution μ^τ is η_k-spectrally independent, i.e. $\lambda_{\max}(\Psi_{\mu^\tau}) \le 1+\eta_k$.*

One may wonder how $\mathcal{I}$ and Ψ are related, and indeed, they are intimately related in a spectral sense. In particular, they essentially share the same eigenvalues, so these notions of spectral independence are completely equivalent. We prove this now.

Proposition 2.5.1 ([ALO21])**.** *Let μ be a probability distribution over $\{0,1\}^n$. Then $\mathcal{I}_\mu - \mathsf{Id}$ and $\Psi_\mu - \mathsf{Id}$ share the exact same eigenvalues with multiplicity, up to multiplicity of the eigenvalue -1. In other words, Definitions 11 and 17 are equivalent, and μ is η-spectrally independent in the sense of Definition 12 if and only if μ is η-spectrally independent in the sense of Definition 18.*

Note that this also immediately implies Ψ_μ has real eigenvalues, since $\mathcal{Q}_\mu$ (and hence, $\mathcal{I}_\mu$) has real eigenvalues. The rest of this subsection is devoted to proving Proposition 2.5.1.

We first give a high level overview. The key observation is that for distributions μ over the discrete hypercube $\{0,1\}^n$, the local random walk $\mathcal{Q}_\mu$ not only has right eigenvector $\mathbf{1}$ with eigenvalue 1, but also $n-1$ copies of the eigenvalue $-\frac{1}{n-1}$ with eigenvectors which are *independent* of μ. These eigenvectors and eigenvalues are considered *"trivial"* since they only depend on the fact that μ is supported over $\{0,1\}^n$, and have nothing to do with the structure of μ nor the values it takes. These trivial eigenvalues are analogs of the fact that the random walk matrix corresponding to the simple random walk on a bipartite graph has eigenvalue 1 and eigenvalue -1, the latter simply due to bipartiteness. Our setting, the local random walk $\mathcal{Q}_\mu$ will be n-*partite*, which yields the $n-1$ additional trivial eigenvalues.

The fact that $\mathcal{Q}_\mu$ has eigenvalue 1 with multiplicity 1 and eigenvalue $-\frac{1}{n-1}$ with multiplicity $n-1$ translates into the (one-sided) influence matrix $\mathcal{I}_\mu - \mathsf{Id}$ having eigenvalue -1 with multiplicity n. What we will show is that by "shifting" $\mathcal{I}_\mu$ so that these -1 eigenvalues become 0 (while maintaining the remaining n eigenvalues), the resulting $2n \times 2n$ matrix has a nice block structure from which we can recover $\Psi_\mu - \mathsf{Id}$ via a simple projection operation; this projection operation simply annihilates these trivial eigenspaces.

Let us now implement this strategy. Recall that for the Boolean hypercube $\{0,1\}^n$, we may view μ instead as a distribution over $\binom{\mathscr{U}}{n}$ where $\mathscr{U} = \{(i,s) : i \in [n], s \in \{0,1\}\}$ is the set of coordinate-assignment pairs. For each coordinate $i \in [n]$, let $\mathbf{1}_i \in \mathbb{R}^{\mathscr{U}}$ be the indicator vector that the coordinate-assignment pair (j,s) satisfies $j = i$. Similarly, let $\pi_i \in \mathbb{R}^{\mathscr{U}}$ denote the marginal distribution of coordinate i, sufficiently padded with zeros, i.e. $\pi_i((i,s)) = \Pr_{\sigma\sim\mu}[\sigma(i) = s]$ and $\pi_i((j,s)) = 0$ whenever $j \ne i$; also let $\pi = n \cdot \mu_1 = \sum_{i=1}^n \pi_i \in \mathbb{R}^{\mathscr{U}}$. Given this, define the matrix

$$\widetilde{\mathcal{I}}_\mu \stackrel{\text{def}}{=} \mathcal{I}_\mu - \mathsf{Id} + \sum_{i=1}^n \mathbf{1}_i \otimes \pi_i \tag{2.13}$$

which will be the appropriate "shift" of the original one-sided influence matrix $\mathcal{I}_\mu$. We prove the following two claims.

Claim 2.5.2 (Relating $\mathcal{I}_\mu - \mathsf{Id}$ to $\widetilde{\mathcal{I}}_\mu$)**.** *For every $i \in [n]$, $\mathcal{I}_\mu - \mathsf{Id}$ has right eigenvector $\mathbf{1}_i \in \mathbb{R}^{\mathscr{U}}$ and left eigenvector $\pi_i \in \mathbb{R}^{\mathscr{U}}$, both with corresponding "trivial" eigenvalue -1. Furthermore, the spectrum of $\widetilde{\mathcal{I}}_\mu$ (as a multiset) is precisely the spectrum of $\mathcal{I}_\mu - \mathsf{Id}$ (as a multiset), except with the n copies of the trivial eigenvalue -1 replaced by n copies of 0.*

Claim 2.5.3 (Relating $\widetilde{\mathcal{I}}_\mu$ to $\Psi_\mu - \mathsf{Id}$)**.** *The spectrum of $\widetilde{\mathcal{I}}_\mu$ (as a multiset) is precisely the union of the spectrum of $\Psi_\mu - \mathsf{Id}$ (as a multiset) with n additional copies of 0.*

Proposition 2.5.1 follows immediately as a consequence of these two claims, which we now prove.

Proof of Claim 2.5.2. Observe that for every $i, j \in [n]$ and $s \in \{0,1\}$,

$$\mathcal{I}_\mu((i,s) \to (j,0)) = -\mathcal{I}_\mu((i,s) \to (j,1)),$$

simply because $\Pr_{\sigma\sim\mu}[\sigma(j) = 0] = 1 - \Pr_{\sigma\sim\mu}[\sigma(j) = 1]$ (and the same holds even in the presence of conditioning on $\sigma(i) = s$). It follows that $\mathcal{I}_\mu \mathbf{1}_j = 0$, whence $(\mathcal{I}_\mu - \mathsf{Id})\mathbf{1}_j = -\mathbf{1}_j$, for all $j \in [n]$. A nearly identical calculation using the fact that $\Pr_{\sigma\sim\mu}[\sigma(j) = 0, \sigma(i) = s] = \Pr_{\sigma\sim\mu}[\sigma(i) = s] - \Pr_{\sigma\sim\mu}[\sigma(j) = 1, \sigma(i) = s]$ demonstrates that $\pi_i^\top \mathcal{I}_\mu = 0$, whence $\pi_i^\top (\mathcal{I}_\mu - \mathsf{Id}) = \pi_i^\top$, for all $i \in [n]$. This establishes the first claim.

Since the $\mathbf{1}_i$ are mutually orthogonal (eigen)vectors w.r.t. the inner product $\langle \cdot, \cdot \rangle_\pi$, $\mathbb{R}^{2n} = \mathbb{R}^{\mathscr{U}}$ admits a basis $\{u_k\}_{k=1}^{2n}$ of eigenvectors of $\mathcal{I}_\mu - \mathsf{Id}$ which contains the $\mathbf{1}_i$ and which are mutually orthogonal w.r.t. $\langle \cdot, \cdot \rangle_\pi$. For any such u_k, we have that

$$\begin{aligned}
\widetilde{\mathcal{I}}_\mu u_k &= (\mathcal{I}_\mu - \mathsf{Id}) \cdot u_k + \sum_{i=1}^{n} \langle \pi_i, u_k \rangle \cdot \mathbf{1}_i \\
&= \lambda_k (\mathcal{I}_\mu - \mathsf{Id}) \cdot u_k + \sum_{i=1}^{n} \langle \mathbf{1}_i, u_k \rangle_\pi \cdot \mathbf{1}_i \\
&= \begin{cases} 0, & \text{if } u_k = \mathbf{1}_i \text{ for some } i \\ \lambda_k(\mathcal{I}_\mu - \mathsf{Id}) \cdot u_k, & \text{otherwise.} \end{cases}
\end{aligned}$$

The claim follows. □

Proof of Claim 2.5.3. First, we observe that $\widetilde{\mathcal{I}}_\mu$ has the following convenient block structure:

$$\widetilde{\mathcal{I}}_\mu = \begin{bmatrix} A & -A \\ B & -B \end{bmatrix}$$

where $A, B \in \mathbb{R}^{n \times n}$ are matrices with entries

$$\begin{aligned}
A(i,j) &= \left(\Pr_{\sigma\sim\mu}[\sigma(j) = 1 \mid \sigma(i) = 1] - \Pr_{\sigma\sim\mu}[\sigma(j) = 1] \right) \cdot \mathbb{I}[i \neq j] \\
B(i,j) &= \left(\Pr_{\sigma\sim\mu}[\sigma(j) = 1 \mid \sigma(i) = 0] - \Pr_{\sigma\sim\mu}[\sigma(j) = 1] \right) \cdot \mathbb{I}[i \neq j].
\end{aligned}$$

Furthermore, it is straightforward to see that $\Psi_\mu - \mathsf{Id} = A - B$. If we assume the truth of this observation, then the claim follows, since we have that the characteristic polynomial of $\widetilde{\mathcal{I}}_\mu$ is given by

$$\begin{aligned}
\det(xI - \widetilde{\mathcal{I}}_\mu) &= \det \begin{bmatrix} xI - A & A \\ -B & xI + B \end{bmatrix} \\
&= \det((xI - A)(xI + B) + AB) && \text{(using e.g. [Sil00, Theorem 3])} \\
&= \det(x^2 I - xA + xB) \\
&= x^n \det(xI - (A - B)) \\
&= x^n \det(xI - (\Psi_\mu - \mathsf{Id})).
\end{aligned}$$

It remains to justify the observation that $\widetilde{\mathcal{I}}_\mu$ may be expressed as such as block matrix. This can be done by comparing entry-by-entry. The case when $i \neq j$ is immediate by inspection. The case $i = j$ follows simply because adding $\sum_{i=1}^n \mathbf{1}_i \otimes \pi_i$ to $\mathcal{I}_\mu - \mathsf{Id}$ in Eq. (2.13) zeros out the 2×2 blocks corresponding to pairs $(i,s),(i,s') \in \mathscr{U}$ where $s, s' \in \{0,1\}$. □

Remark 12. Similar to Eq. (2.6), if we write Σ_μ for the $n \times n$ covariance matrix of the random $\{0,1\}$-vector $\sigma \sim \mu$ in $\{0,1\}^n$, then

$$\Psi_\mu = \operatorname{diag}\left(\Pr_{\sigma\sim\mu}[\sigma_i = 0] \cdot \Pr_{\sigma\sim\mu}[\sigma_i = 1]\right)_{i=1,\dots,n}^{-1} \cdot \Sigma_\mu.$$

Remark 13. One can exhibit simple and small examples of probability distributions μ which are 0-spectrally independent, but for which the conditional distributions clearly are not 0-spectrally independent. The following example is often attributed to S. Bernstein, which we found in the textbook [HMC19]. Consider the uniform distribution μ over $\{(0,0,0),(1,1,0),(1,0,1),(0,1,1)\} \subseteq \{0,1\}^3$. A straightforward computation reveals that this distribution is *pairwise independent*, i.e. the off-diagonal entries of Ψ_μ are zero. In particular, $\Psi_\mu - \mathsf{Id}$ is the zero matrix, so μ is 0-spectrally independent. On the other hand, conditioning on the value of any coordinate results in a distribution over $\{0,1\}^2$ which is 1-spectrally independent. Indeed, all pairs of distinct points in the support of μ have Hamming distance 2 from each other, so the Glauber dynamics is not even connected.

2.5.2 General Discrete Product Spaces

Like the discrete hypercube $\{0,1\}^n$, for distributions over more general discrete product spaces (e.g. $[q]^n$ for some $q > 2$), the local random walks $\mathcal{Q}_\mu$ are still n-partite and have $n-1$ copies of the trivial eigenvalue $-\frac{1}{n-1}$. Hence, $\mathcal{I}_\mu - \mathsf{Id}$ also inherits these trivial eigenvalues, and one can try to use this fact to find a slightly "simpler" version of the influence matrix $\mathcal{I}_\mu$.

Let μ be a probability distribution over $\Sigma_1 \times \cdots \times \Sigma_n$ for some finite sets $\Sigma_1, \dots, \Sigma_n$. Again, recall that one can fit this within our original formalism by taking the ground set $\mathscr{U}$ to be the set of coordinate-assignment pairs (i,s) where $i \in [n]$ and $s \in \Sigma_i$, and then viewing a point in this product space as a size-n collection of such coordinate-assignment pairs. For each $i \in [n]$, we let $\mathbf{1}_i \in \mathbb{R}^{\mathscr{U}}$ be the indicator vector of $\{(i,s) : s \in \Sigma_i\}$. Similarly, we let $\pi_i \in \mathbb{R}^{\mathscr{U}}$ be the marginal vector of coordinate i (appropriately padded with zeros elsewhere), where $\pi_i((i,s)) = \Pr_{\sigma\sim\mu}[\sigma(i) = s] = n \cdot \mu_1((i,s))$ for all $s \in \Sigma_i$; also let $\pi = n \cdot \mu_1 = \sum_{i=1}^n \pi_i \in \mathbb{R}^{\mathscr{U}}$.

We show the following, which is an analog of Claim 2.5.2.

Claim 2.5.4 ([Opp18; ALO21; Che+21d])**.** *Let μ be a probability distribution over a discrete product space $\Sigma_1 \times \cdots \times \Sigma_n$, and let $\mathscr{U} = \{(i,s) : i \in [n], s \in \Sigma_i\}$. Then for every $i = 1, \dots, n$, $\mathcal{I}_\mu$ has right eigenvector $\mathbf{1}_i \in \mathbb{R}^{\mathscr{U}}$ and left eigenvector $\pi_i \in \mathbb{R}^{\mathscr{U}}$, both with corresponding "trivial" eigenvalue -1. Furthermore, if we define $\Psi_\mu \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ by*

$$\Psi_\mu((i,s_i) \to (j,s_j)) = \begin{cases} \Pr_{\sigma\sim\mu}[\sigma(j) = s_j \mid \sigma(i) = s_i] - \Pr_{\sigma\sim\mu}[\sigma(j) = s_j], & \text{if } i \neq j \\ 1, & \text{if } i = j, s_i = s_j, \\ 0, & \text{otherwise} \end{cases}$$

then the spectrum of $\Psi_\mu - \mathsf{Id}$ (as a multiset) is precisely the spectrum of $\mathcal{I}_\mu - \mathsf{Id}$ (as a multiset), except with n copies of the trivial eigenvalue -1 replaced by n copies of 0. In particular, μ is η-spectrally independent if and only if $\lambda_{\max}(\Psi_\mu) \leq 1 + \eta$.

Remark 14. Stated in the language of simplicial complexes, this result says that for any n-dimensional n-*partite* complex $(\mathscr{U}, \mathcal{X})$ with parts $\mathscr{U} = \mathscr{U}_1 \sqcup \cdots \sqcup \mathscr{U}_n$, the indicator vectors $\mathbf{1}^{\mathscr{U}_1}, \dots, \mathbf{1}^{\mathscr{U}_n}$ are eigenvectors of $\mathcal{Q}_\mu - \frac{n}{n-1}\mathbf{1}\mu_1^\top$ with eigenvalue $-\frac{1}{n-1}$. This was previously observed in [Opp18], although we were not aware of this. In [Che+21d], the influence matrix was defined as $\Psi_\mu - \mathsf{Id}$. These differences are immaterial.

Proof. The proof is the same as the proof of Claim 2.5.2. To show that $(\mathcal{I}_\mu - \mathsf{Id}) \cdot \mathbf{1}_i = -\mathbf{1}_i$ (or equivalently $\mathcal{I}_\mu \mathbf{1}_i = 0$), one just uses the fact that for every $i, j \in [n]$ and every $s_i \in \Sigma_i$,

$$\sum_{s_j \in \Sigma_j} \mathcal{I}_\mu((i,s_i) \to (j,s_j)) = \underbrace{\sum_{s_j \in \Sigma_j} \Pr_{\sigma\sim\mu}[\sigma(j) = s_j \mid \sigma(i) = s_i]}_{=1} - \underbrace{\sum_{s_j \in \Sigma_j} \Pr_{\sigma\sim\mu}[\sigma(j) = s_j]}_{=1} = 0.$$

Similarly, to show that $\pi_i^\top(\mathcal{I}_\mu - \mathsf{Id}) = -\pi_i^\top$ (or equivalently $\pi_i^\top \mathcal{I}_\mu = 0$), one uses that for every $i, j \in [n]$ and every $s_j \in \Sigma_j$,

$$\begin{aligned}
&\sum_{s_i \in \Sigma_i} \left(\Pr_{\sigma\sim\mu}[\sigma(j) = s_j \mid \sigma(i) = s_i] \cdot \Pr_{\sigma\sim\mu}[\sigma(i) = s_i] - \Pr_{\sigma\sim\mu}[\sigma(i) = s_i] \cdot \Pr_{\sigma\sim\mu}[\sigma(j) = s_j] \right) \\
&= \underbrace{\sum_{s_i \in \Sigma_i} \Pr_{\sigma\sim\mu}[\sigma(j) = s_j, \sigma(i) = s_i]}_{=\Pr_{\sigma\sim\mu}[\sigma(j)=s_j]} - \Pr_{\sigma\sim\mu}[\sigma(j) = s_j] \cdot \underbrace{\sum_{s_i \in \Sigma_i} \Pr_{\sigma\sim\mu}[\sigma(i) = s_i]}_{=1} \\
&= 0.
\end{aligned}$$

The matrix Ψ_μ may then be expressed as $\Psi_\mu = \mathcal{I}_\mu + \sum_{i=1}^n \mathbf{1}_i \otimes \pi_i$. Noting that $\Psi_\mu - \mathsf{Id}$ is simply $\widetilde{\mathcal{I}}_\mu$ from Eq. (2.13), the rest of the argument follows in the same manner as the proof of Claim 2.5.2. □

Remark 15. Note that Ψ_μ is a $\mathscr{U} \times \mathscr{U}$ matrix just like $\mathcal{I}_\mu$, whereas in the case of the discrete hypercube $\{0,1\}^n$, we could "compress" Ψ_μ down to be $n \times n$ rather than $2n \times 2n$. One can ask if there is such a "compact" $n \times n$ influence matrix for general discrete product spaces. This indeed is the case, and was discovered in [Fen+21]. In particular, one can quantify the total influence of a coordinate i on another coordinate j by using the total variation distance in a similar way to the definition of the *Dobrushin influence matrix* (see Definition 39).

For every pair of coordinates $i, j \in [n]$, define the **(absolute coordinate) influence of i on j** by

$$\Upsilon_\mu(i \to j) \stackrel{\text{def}}{=} \max_{s,s' \in \Sigma_i} \left\| \mu_j^{i \leftarrow s} - \mu_j^{i \leftarrow s'} \right\|_{\mathsf{TV}},$$

where $\mu_j^{i\leftarrow s}$ is the marginal distribution of coordinate j over Σ_j conditioned on the event that coordinate i is assigned $s \in \Sigma_i$. Note that $\Upsilon_\mu(i \to i) = 1$ for all $i \in [n]$. We also define the **(absolute coordinate) influence matrix** $\Upsilon_\mu \in \mathbb{R}^{n\times n}$ by $\Upsilon_\mu(i,j) = \Upsilon_\mu(i \to j)$ for every $i, j \in [n]$.

In the case of the discrete hypercube $\{0,1\}^n$, $\Upsilon_\mu \in \mathbb{R}^{n\times n}$ is simply the entrywise absolute value of $\Psi_\mu \in \mathbb{R}^{n\times n}$. Unfortunately, Υ_μ does not necessarily have real eigenvalues. However, we can still define spectral independence using the spectral radius $\rho(\Upsilon_\mu)$ of Υ_μ instead. In particular, the entries of Υ_μ are nonnegative, the Perron-Frobenius Theorem (see Theorem 1.4.4) applies and $\rho(\Upsilon_\mu)$ is a nonnegative real eigenvalue of Υ_μ with an entrywise nonnegative eigenvector. It was shown in [Fen+21] using a coupling argument that $\frac{\rho(\Upsilon_\mu)-1}{n-1}$ is an upper bound on the second largest eigenvalue $\lambda_2(\mathcal{Q}_\mu)$ of the local walk, similar to how $\frac{\lambda_{\max}(\Psi_\mu)-1}{n-1}$ equals $\lambda_2(\mathcal{Q}_\mu)$.

This notion of influence matrix was recently generalized to products of arbitrary measure spaces which are not necessarily discrete [QW22]. There, the authors prove an analog of Theorem 2.3.1 for the natural extension of the Glauber dynamics/Gibbs sampler to products of arbitrary measure spaces.

2.6 An Illustrative Example: Independent Sets in $K_{n,n}$

In this section, we analyze the spectral independence of a simple but instructive example distribution. For this distribution, we will see that the spectral independence is $\Omega(n)$, and that the Glauber dynamics/down-up walk mixes only after $\exp(\Omega(n))$-steps. This example also rules out a simple and natural "average-case" formulation of Theorem 2.3.1, where one tries to require "good" spectral independence only for "most" conditional distributions.

For a positive integer $n \geq 1$ and a positive real parameter $\lambda > 0$, let $\mu = \mu_{n,\lambda}$ be the probability distribution over independent sets of the complete bipartite graph $K_{n,n}$, with n vertices on each side, given by

$$\mu(I) \propto \lambda^{|I|}.$$

This is the Gibbs distribution of the hardcore model with parameter λ on $K_{n,n}$.

Let $L \sqcup R$ denote the two sides of the bipartition, which we colloquially refer to as the "left" and "right" sides of the graph, respectively. Since there is an edge between every pair of vertices $u \in L, v \in R$, it is clear from the independent set constraint that the independent sets of $K_{n,n}$ consist simply of:

- The empty set $\emptyset$.
- All nonempty subsets of L. Denote this collection by Ω_L.
- All nonempty subsets of R. Denote this collection by Ω_R.

Hence, the normalizing constant (or partition function) for μ is simply $2(1+\lambda)^n - 1$.

2.6.1 Torpid Mixing of the Glauber Dynamics

Now, since the Glauber dynamics is a local Markov chain which only updates the status of a single vertex at a time, intuitively a bottleneck is formed at the $\emptyset$. If $\lambda \geq \Omega(1)$, the probability of $\emptyset$ under the stationary distribution is exponentially small. Hence, it should take exponential time to move from the maximal independent set L to the maximal independent set R, since one has to go through $\emptyset$.

One can formally prove that the Glauber dynamics has exponentially small spectral gap by simply considering the *conductance* of the collection Ω_L. Indeed, the mass of Ω_L and Ω_R are

$$\mu(\Omega_L) = \mu(\Omega_R) = \frac{1}{2}(1 - \mu(\emptyset)) = \frac{1}{2}\left(1 - \frac{1}{2(1+\lambda)^n - 1}\right).$$

On the other hand, the total flow across the cut $(\Omega_L, \Omega_R \sqcup \{\emptyset\})$ is

$$\mu(\emptyset) \cdot \sum_{u \in L} \mathcal{P}_\mu(\emptyset \to \{u\}) \leq \mu(\emptyset) = \frac{1}{2(1+\lambda)^n - 1}.$$

It follows that the conductance of this cut (and hence, of the Glauber dynamics), is exponentially small. This shows the Glauber dynamics requires $\exp(\Omega_\lambda(n))$-steps to mix.

2.6.2 Spectral Independence Calculations Using Ψ_μ

Now, let us compute the spectral independence of this distribution. It will be convenient to use $\Psi_\mu \in \mathbb{R}^{2n \times 2n}$ from Definition 17 instead of $\mathcal{I}_\mu \in \mathbb{R}^{4n \times 4n}$ from Definition 11. Since we assumed $|L| = |R| = n$ for convenience, by symmetry, we have that for every pair of vertices u, v

$$\Psi_\mu(u \to v) = \begin{cases} -\frac{\lambda(1+\lambda)^{n-1}}{(1+\lambda)^n + (1+\lambda)^{n-1} - 1} \stackrel{\text{def}}{=} -\beta, & \text{if } u \in L, v \in R \text{ or } u \in R, v \in L \\ \frac{\lambda}{1+\lambda} - \frac{\lambda(1+\lambda)^{n-2}}{(1+\lambda)^n + (1+\lambda)^{n-1} - 1} \stackrel{\text{def}}{=} \alpha, & \text{if } u, v \in L \text{ or } u, v \in R \text{ with } u \neq v \\ 1, & \text{if } u = v \end{cases}$$

Now, note that $\Psi_\mu - (1-\alpha) \cdot I$ can be written as the tensor product $\begin{bmatrix} \alpha & -\beta \\ -\beta & \alpha \end{bmatrix} \otimes \mathbf{1}\mathbf{1}^\top$. Hence, the spectral independence of the distribution is given by

$$\begin{aligned} \eta = \boldsymbol{\lambda}_{\max}(\Psi_\mu) - 1 &= n \cdot \boldsymbol{\lambda}_{\max} \begin{bmatrix} \alpha & -\beta \\ -\beta & \alpha \end{bmatrix} - \alpha \\ &= n \cdot (\alpha + \beta) - \alpha \\ &\geq (n-1) \cdot (\alpha + \beta) \\ &\geq \frac{\lambda}{1+\lambda} \cdot \frac{1+3\lambda}{1+2\lambda} \cdot (n-1), \end{aligned}$$

where we use that $\alpha, \beta \geq 0$ and

$$\alpha + \beta = \frac{\lambda}{1+\lambda} + \frac{\lambda^2(1+\lambda)^{n-2}}{(1+\lambda)^n + (1+\lambda)^{n-1} - 1} \geq \frac{\lambda}{1+\lambda} + \frac{\lambda^2}{(1+\lambda)^2 + (1+\lambda)} = \frac{\lambda}{1+\lambda} \cdot \frac{1+3\lambda}{1+2\lambda}.$$

If $\lambda \geq \Omega(1)$ (e.g. $\lambda = 1$), then $\eta \geq \Omega(n)$. The calculations performed here also demonstrate the utility of using influence matrices rather than local random walk matrices. Explicit calculations, especially of the eigenvalues, are often simpler to carry out.

One can ask about the conditional distributions of μ. Observe that pinning a vertex to being in the independent set forces all vertices on the other side of the bipartition to be pinned to be

out of the independent set. The remaining unpinned vertices are thus all on the same side of the bipartition, and hence have no edges among them. It follows that all such conditional distributions are *product measures* and hence, 0-spectrally independent.

To find "bad" pinnings, one should then just pin some subset of vertices to be "out". Intuitively, if one wishes to maximize the total strength of correlations in the distribution, one should aim to keep as many edges as possible, since interactions are only across edges. Hence, one should try to keep the number of vertices pinned to be balanced across the two sides of the bipartition. If one pins k vertices to be "out", where k is even, the worst resulting conditional distribution is simply the Gibbs distribution of the hardcore model on $K_{n-\frac{k}{2},n-\frac{k}{2}}$. The above analysis again shows this conditional distribution is $\Omega_\lambda(n-k)$-spectrally independent. All of this can be justified formally just by calculation.

This example actually rules out a natural but naïve average-case formulation of Theorem 2.3.1. Indeed, it would be wonderful if something like the following could be true.

> "Let μ be a probability distribution over $\binom{\mathcal{U}}{n}$ for a finite set $\mathcal{U}$ and positive integer n. If for every $0 \le k \le n-2$, $\lambda_{\max}\left(\Psi_{\mu^\xi}\right) \le O(1)$ with very high probability (e.g. $1-\exp(\Theta(k))$) over the choice of ξ drawn from the induced level-k distribution μ_k, then the down-up walk has spectral gap $\Omega(1/\mathsf{poly}(n))$."

Indeed, for each fixed $0 \le k \le n$ and each $S \subseteq V$ with $|S| = k$, the marginal probability that no vertex of S is assigned 1 is exactly

$$\frac{(1+\lambda)^{n-|S\cap L|} + (1+\lambda)^{n-|S\cap R|} - 1}{2(1+\lambda)^n - 1}.$$

Hence, the total probability that such a "bad" pinning occurs in a sample from μ_k is

$$\sum_{\ell=0}^{k} \frac{\binom{n}{\ell}\binom{n}{k-\ell}}{\binom{2n}{k}} \cdot \frac{(1+\lambda)^{n-\ell} + (1+\lambda)^{n-k+\ell} - 1}{2(1+\lambda)^n - 1}.$$

In the $\lambda \to +\infty$ limit, this probability becomes

$$\frac{2\cdot\binom{n}{k}}{\binom{2n}{k}} = \prod_{j=1}^{n}\left(1 - \frac{k}{n+j}\right) = \exp(-\Theta(k)).$$

The conditional distribution for all other pinnings is 0-spectrally independent, the best possible. At some level, the fact that such an average-case local-to-global theorem fails for this example makes sense, since the whole reason the Markov chain is slow mixing is because there is a bottleneck state which has extremely low probability under the stationary distribution.

Part I

Techniques for Establishing Spectral Independence

Chapter 3

Matroids and Oppenheim's Trickle-Down Method

In this chapter, we explore a beautiful technique for establishing spectral independence known as the Trickle-Down Theorem. This was first discovered by Oppenheim [Opp18] in the study of expansion phenomena in simplicial complexes, and has since been used repeatedly in recent constructions of bounded-degree high-dimensional expanders [KO18; KO20a; OP22]. Trickling-down phenomena has since been studied in posets as well [KT22]. In the context of sampling, the Trickle-Down Theorem was used to give efficient sampling algorithms for bases of matroids [Ana+19], and subsequently, for fixed-size independent sets in graphs and common independent sets of two partition matroids [AL20]. This chapter is devoted to the former result. We will show how this intriguing technique can be used to establish optimal spectral independence bounds for the uniform distribution over the bases of any matroid. By Theorem 2.3.1, as well as the results of Chapters 9 and 11 later on, this implies fast mixing for the bases exchange walk. Along the way, we will resolve the *Mihail–Vazirani Conjecture* [MV89] for matroid basis polytopes. This chapter is based primarily on [Ana+19] and [Opp18].

3.1 Counting and Sampling Bases of Matroids

We begin by stating the first main result of this chapter.

Theorem 3.1.1. *Let μ be the uniform distribution over the bases of an n-element rank-r matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$, and let $\mathcal{P}_\mu$ be the transition matrix of the corresponding bases exchange walk. Then $\gamma(\mathcal{P}_\mu) \geq \frac{1}{r}$ and $\mathcal{P}_\mu$ mixes in $O(r^2 \log n)$ steps.*

In particular, combined with standard reductions from approximate counting to approximate sampling (see Section 1.4.4), there is an FPRAS for approximately counting bases of matroids. As an immediate corollary, for any $0 \leq k \leq r$, we can approximately count the number of independent sets of $\mathcal{M}$ of size-k. This is because if we truncate $\mathcal{M}$ to independent sets of size at most k it remains a matroid. As a consequence, we can generate approximately uniformly random forests in a given graph, and compute the *reliability polynomial*

$$C_{\mathcal{M}}(p) \stackrel{\text{def}}{=} \sum_{\tau \subseteq \mathscr{U} : \operatorname{rank}(\tau) = r} (1-p)^{|\tau|} p^{n-|\tau|}$$

for any matroid and $0 \leq p \leq 1$, all in polynomial-time. Note this latter fact follows from the ability to approximately count the number of independent sets of a fixed size, as the complements of rank-r subsets $\tau \subseteq \mathscr{U}$ are precisely the independent sets of the *dual* of $\mathcal{M}$. Prior to this result, we could only compute the reliability polynomial for graphic matroids due to [GJ19].

Estimating the reliability polynomial of a matroid has a number of important applications. When the matroid is the graphic matroid corresponding to some underlying graph $G = (V, E)$ (e.g. a road network or a communications network), the reliability polynomial gives the probability that the network remains connected under independent edge deletions (e.g. network links fail independently with probability p). When the matroid is representable over $\mathbb{F}_2$, then viewing the underlying $\mathbb{F}_2$-vectors as columns of the *parity check matrix* of an error-correcting code, the

reliability polynomial yields the probability of being able to recover from independent *erasures* of bits when the codewords are sent across a noisy channel.

We prove Theorem 3.1.1 by showing that the uniform distribution over the bases of a matroid satisfies strong spectral independence bounds; in fact, they turn out to be 0-spectrally independent, which is optimal. This fact combined with Theorem 2.3.1 immediately implies Theorem 3.1.1. Hence, all that remains is to prove the following.

Theorem 3.1.2. *Let μ be the uniform distribution over the bases of an n-element rank-r matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$. Then μ is $(0, \dots, 0)$-spectrally independent.*

Our approach has a close connection to the original plan of Feder–Mihail [FM92], who used the negative correlation property of balanced matroids to show that the bases exchange walk mixes rapidly. Again, many interesting matroids unfortunately do not satisfy negative correlation; some interesting examples are collected in [HSW21] and in references therein. However, Theorem 3.1.2 says that the uniform distribution μ over the bases of any matroid is 0-spectrally independent, which can be thought of as a *spectral negative dependence* property. Indeed, negative correlation says that all off-diagonal entries of $\mathcal{I}_\mu - \mathsf{Id}$ are nonpositive, while 0-spectral independence says that all eigenvalues of $\mathcal{I}_\mu - \mathsf{Id}$ are nonpositive. Spectrally negative correlations turns out to be precisely what one needs to bound the mixing time of the bases exchange walk. This spectral negative dependence property was first observed in [AHK18; HW17; AOV21], but through the lens and language of multivariate polynomials and log-concavity. We will discuss this further in Chapter 5.

Prior Work on Counting Bases of Matroids There is a long line of work on designing approximation algorithms to count the bases of a matroid. Most of these works focus on expansion properties of bases exchange graph. [FM92] showed that for a special class of matroids known as *balanced matroids* [MS91; FM92], the bases exchange graph has *edge expansion* at least 1. A matroid $\mathcal{M}$ is balanced if for any *minor* of $\mathcal{M}$ (including $\mathcal{M}$ itself), the uniform distribution over its bases satisfies the *pairwise negative correlation property* (see e.g. Example 6). Many of the extensive results in this area [Gam99; JS02; Jer+04; Jer06; Clo10; CTY15; AOR16] only study approximation algorithms for this limited class of matroids, and not much is known beyond the class of balanced matroids. Unfortunately, many interesting matroids are not balanced. An important example is the matroid of all acyclic subsets of edges of a graph $G = (V, E)$ with size at most k (for some $k < |V| - 1$) [FM92].

There have been other approaches for approximately counting bases. [GJ21] used the popping method to approximately count bases of *bicircular matroids*. [BS07] designed a *randomized* algorithm that gives, roughly, a $\log(n)^r$-multiplicative approximation to the number of bases of a given n-element rank-r matroid. [AOV21] gave a deterministic e^r approximation to the number of bases using *log-concavity of the bases generating polynomial*. For deterministic algorithms, this is essentially the best possible due an unconditional lower bound of [ABF94], which says that no deterministic algorithm given only access to an independence oracle can approximate the number of bases of an n-element matroid within a multiplicative factor of $2^{O(n/\log^2 n)}$. We discuss log-concave polynomials in Chapter 5.

3.2 Spectral Expansion Trickles Down

To prove Theorem 3.1.2, we use an incredible useful result due to Oppenheim [Opp18] now known as the Trickle-Down Theorem. At a high level, in the language of spectral independence, it says that strong enough bounds on the spectral independence for conditional distributions of μ imply nontrivial bounds on the spectral independence of μ itself. With such a tool in hand, the strategy is then to show that all "1-dimensional" conditional distributions satisfy strong spectral independence bounds. These "1-dimensional" distributions are much easier to analyze precisely because they are low-dimensional. Furthermore, the weights which arise in these "1-dimensional" distributions are much more controlled, since they are given directly by μ.

This result was original discovered in the study of high-dimensional expansion phenomena in simplicial complexes, and has been used repeatedly to certify that certain (often algebraically-constructed) families of simplicial complexes are indeed high-dimensional expanders. It turns out, the cleanest formulation of this theorem uses the local walks $\mathcal{Q}_\mu$ defined in Definition 13. Hence, throughout this chapter, we primarily work with the local walks $\mathcal{Q}_\mu$ as opposed to influence matrices $\mathcal{I}_\mu$.

Theorem 3.2.1 (Oppenheim's Vanilla Trickle-Down Theorem; [Opp18])**.** *Let μ be a probability distribution on $\binom{\mathscr{U}}{n}$, where $\mathscr{U}$ is a finite ground set and $n \geq 1$ is a positive integer. Assume the following conditions:*

- ***Connectivity:*** *The local walk $\mathcal{Q}_\mu$ is irreducible (i.e. connected).*
- ***Expansion for Conditional Local Walks:*** *For some $0 \leq \lambda \leq 1/2$, we have the bound $\lambda_2\left(\mathcal{Q}_{\mu^i}\right) \leq \lambda$ for every $i \in \mathscr{U}$.*

Then the local walk $\mathcal{Q}_\mu$ satisfies the bound $\lambda_2\left(\mathcal{Q}_\mu\right) \leq \frac{\lambda}{1-\lambda}$.

Remark 16. The original statement in [Opp18] does not have the assumption $\lambda \leq 1/2$, but the two are completely equivalent, since if $\lambda > 1/2$, then $\frac{\lambda}{1-\lambda} > 1$, making the statement vacuously true.

Corollary 3.2.2 (Iterated Vanilla Trickle-Down Theorem; [Opp18])**.** *Let μ be a totally connected probability distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and positive integer $n \geq 1$. Suppose there exists $0 \leq \lambda < \frac{1}{n-1}$ such that for every $\tau \in \mathcal{X}(n-2)$, we have the bound $\lambda_2\left(\mathcal{Q}_{\mu^\tau}\right) \leq \lambda$. Then for every feasible $\tau \subseteq \mathscr{U}$ with $|\tau| = k \leq n-2$, we have the bound $\lambda_2\left(\mathcal{Q}_{\mu^\tau}\right) \leq \frac{\lambda}{1-(n-k-2)\lambda}$.*

The utility of this result lies in the fact that it allows us to restrict our attention to the conditional distributions μ^σ where $|\sigma| = n-2$. Such distributions can be thought of as "1-dimensional", since they are supported on size-2 subsets of $\mathscr{U}$,[1] and hence, are much simpler to analyze. For instance, the entries of $\mathcal{Q}_{\mu^\sigma}$ come directly from μ itself (appropriately renormalized in a nice way) when $|\sigma| = n-2$, whereas the entries of $\mathcal{Q}_\mu$ are in general hard to compute. In general, the distribution μ^σ is much more "local" when $|\sigma| = n-2$, so in a way, Corollary 3.2.2 is another kind of *local-to-global analysis*.[2]

To give you a feeling for the parameters, if all the local walks $\mathcal{Q}_{\mu^\sigma}$ are connected for every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| \leq n-2$, then we have the following:

- If the "1-dimensional" distributions μ^σ satisfy $\lambda_2\left(\mathcal{Q}_{\mu^\sigma}\right) \leq 0$ for all feasible $\sigma \subseteq \mathscr{U}$ satisfying $|\sigma| = n-2$, i.e. they are 0-spectrally independent, then *all* conditional distributions of μ are 0-spectrally independent. This implies the down-up walk has spectral gap $\geq 1/n$.
- If the "1-dimensional" distributions μ^σ satisfy $\lambda_2\left(\mathcal{Q}_{\mu^\sigma}\right) \leq \frac{1}{2n}$ for all feasible $\sigma \subseteq \mathscr{U}$ satisfying $|\sigma| = n-2$, i.e. they are $\frac{1}{2n}$-spectrally independent, then $\lambda_2\left(\mathcal{Q}_{\mu^\sigma}\right) \leq \frac{1}{n+|\sigma|+2} \leq \frac{1}{n}$ for all feasible $\sigma \subseteq \mathscr{U}$. In particular, *all* conditional distributions of μ are 1-spectrally independent or better, and the down-up walk has spectral gap $\geq \Omega(1/n)$.
- If the "1-dimensional" distributions μ^σ satisfy $\lambda_2\left(\mathcal{Q}_{\mu^\sigma}\right) \leq \frac{1}{n}$ for all feasible $\sigma \subseteq \mathscr{U}$ satisfying $|\sigma| = n-2$, i.e. they are $\frac{1}{n}$-spectrally independent, then $\lambda_2\left(\mathcal{Q}_{\mu^\sigma}\right) \leq \frac{1}{|\sigma|+2}$ for all feasible $\sigma \subseteq \mathscr{U}$. Furthermore, the down-up walk has spectral gap $\geq \Omega(1/n^2)$.

We will later on see settings in which this theorem cannot be usefully applied. Nevertheless, this theorem is essentially the only known elementary method for obtaining spectral independence for the uniform distribution over bases of a matroid.[3]

3.2.1 0-Spectral Independence for Matroids

Before we prove this theorem, let us first see how this can be used to prove Theorem 3.1.2. We will need the following crucial property of matroids. This property is actually a *cryptomorphism* [Oxl11]; it can be used to give another definition of a matroid which is completely equivalent to the standard one Definition 7. It is a special case of the definition of a matroid via *flats* (or *hyperplanes*).

[1] Indeed, in algebraic topology, the dimension of a face σ in an abstract simplicial complex $\mathcal{X}$ is defined as $|\sigma| - 1$ rather than $|\sigma|$. This makes sense geometrically, since for instance, the standard n-vertex simplex in $\mathbb{R}^n$ is contained in an $(n-1)$-dimensional affine subspace.

[2] This strategy also bears some resemblance to the idea of *localization* in high-dimensional convex geometry [KLS95; Eld13], where one also tries to decompose a high-dimensional probability distribution into "1-dimensional" distributions, sometimes called *needles*, which are easier to analyze. However, one should be careful not confuse this notion of "localization" with the notion of "localization" in the study of high-dimensional expanders.

[3] There is a significantly more sophisticated and involved proof [AOV21] using the recently developed *combinatorial Hodge theory* for matroids [AHK18]. This is well-beyond the scope of this thesis.

Fact 3.2.3 (Matroid Partition Property)**.** *Let $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ be a matroid. Without loss of generality, we may assume $\mathcal{M}$ has no loops. Then there is a partition $S_1 \sqcup \cdots \sqcup S_\ell$ of $\mathscr{U}$ such that $\{u, v\} \notin \mathcal{X}$ if and only if $u, v \in \mathscr{U}$ belong to the same block of the partition.*

Proof. Define a relation $\sim$ on $\mathscr{U}$ via $u \sim v \iff \{u, v\} \notin \mathcal{X}$; in other words, $u, v \in \mathscr{U}$ are related if they are dependent (or *parallel* to each other). If we can show that $\sim$ is an *equivalence relation*, then we can simply partition $\mathscr{U}$ into the *equivalence classes* under $\sim$. Let $u, v, w \in \mathscr{U}$ be distinct. If $u \not\sim w$, then applying the exchange axiom to the two independent sets $\{v\}, \{u, w\}$, we see that at least one of $\{u, v\}$ or $\{v, w\}$ must also be independent; in particular, $u \not\sim v$ or $v \not\sim w$. Taking the contrapositive proves that $\sim$ is an equivalence relation as desired. □

Proof of Theorem 3.1.2. By inductively applying Theorem 3.2.1, it suffices to show the following:

- Connectivity: The local walk $\mathcal{Q}_{\mu^\sigma}$ for the conditional distribution μ^σ is connected for every independent set $\sigma \in \mathcal{X}$ with $|\sigma| \leq r - 2$.
- 0-Spectral Independence for 1-Dimensional Conditional Distributions: For every independent set $\sigma \in \mathcal{X}$ with $|\sigma| = r - 2$, the conditional distribution μ^σ is 0-spectrally independent.

Connectivity of the local walks follows immediately from the exchange property for matroids. Hence, it remains to show that μ^σ is 0-spectrally independent, or equivalently $\boldsymbol{\lambda}_2(\mathcal{Q}_{\mu^\sigma}) \leq 0$ (by Lemma 2.3.3), for all rank-$(r-2)$ independent sets $\sigma \in \mathcal{X}$. Since μ^σ is simply the uniform distribution over the bases of the contraction $\mathcal{M}/\sigma$, a rank-2 matroid, it suffices to prove the claim for all rank-2 matroids.

Let $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ be a rank-2 matroid and let μ be the uniform distribution over the bases of $\mathcal{M}$. Observe that $\mathcal{Q}_\mu = D_{\mathcal{M}}^{-1} A_{\mathcal{M}}$, where $A_{\mathcal{M}} \in \mathbb{R}^{\mathscr{U} \times \mathscr{U}}$ is the adjacency matrix of an unweighted undirected simple graph with entries

$$A_{\mathcal{M}}(u, v) = \begin{cases} 1, & \text{if } \{u, v\} \in \mathcal{X} \\ 0, & \text{otherwise} \end{cases}$$

and $D_{\mathcal{M}} \in \mathbb{R}^{\mathscr{U} \times \mathscr{U}}$ is the associated diagonal degree matrix. By Lemma 1.4.8, to show $\boldsymbol{\lambda}_2(\mathcal{Q}_\mu) \leq 0$, it suffices to show that $\boldsymbol{\lambda}_2(A_{\mathcal{M}}) \leq 0$. By Fact 3.2.3, there is a partition $S_1 \sqcup \cdots \sqcup S_\ell$ of $\mathscr{U}$ such that $\{u, v\} \in \mathcal{X}$ if and only if $u, v \in \mathscr{U}$ belong to different blocks of the partition. In particular, $A_{\mathcal{M}}$ is the adjacency matrix of a *complete multipartite graph*, and can be written as $A_{\mathcal{M}} = \mathbf{1}\mathbf{1}^\top - \sum_{i=1}^{\ell} \mathbf{1}_{S_i} \mathbf{1}_{S_i}^\top \preceq \mathbf{1}\mathbf{1}^\top$. That $\boldsymbol{\lambda}_2(A_{\mathcal{M}}) \leq 0$ (and hence, $\boldsymbol{\lambda}_2(\mathcal{Q}_\mu) \leq 0$) follows immediately. □

We close this subsection with an observation concerning the structure of the local random walks for 0-spectrally independent distributions. One can use Fact 3.2.3 to say that in fact, if μ is the uniform distribution over bases of an arbitrary matroid $\mathcal{M}$ of any rank, the unweighted undirected graph supporting the underlying graph of $\mathcal{Q}_\mu$ is a complete multipartite graph. There is a (partial) converse to this in the sense that if the adjacency matrix of a weighted graph has at most one positive eigenvalue, i.e. its second largest eigenvalue is upper bounded by 0, then its support *must* be a complete multipartite graph. We formalize this as follows.

Lemma 3.2.4. *Let $G = (V, E, c : E \to \mathbb{R}_{>0})$ be a weighted undirected loopless graph without isolated vertices. Let $A \in \mathbb{R}_{\geq 0}^{V \times V}$ be its weighted adjacency matrix, and assume A has at most one positive eigenvalue. Then G must be supported on a complete multipartite graph, in the sense that there exists a partition $V = V_1 \sqcup \cdots \sqcup V_k$ of the vertices such that $A(u, v) > 0$ if and only if $u \in V_i, v \in V_j$ for some $i \neq j$.*

Proof. It suffices to show that the following transitivity property holds: for distinct vertices x, y, z, if $A(x, y) = A(y, z) = 0$, then $A(x, z) = 0$. Such a property immediately implies that the relation $x \sim y$ if and only if $A(x, y) = 0$ is an *equivalence relation*, and the partition is immediately given by the *equivalence classes* of this relation.

To establish this transitivity property, we construct a suitable test vector to extract the entries $A(x, y), A(y, z), A(x, z)$. Consider the vector ϕ with entries

$$\phi(u) = \begin{cases} 1, & \text{if } u = x, z \\ -2, & \text{if } u = y \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$\phi^\top A\phi = \sum_{u,v} A(u,v)\phi(u)\phi(v) = 2(A(x,z) - 2A(x,y) - 2A(y,z)) = 2 \cdot A(x,z),$$

where in the last step, we used the assumption that $A(x,y) = A(y,z) = 0$. By assumption $A(x,z) \geq 0$ so it suffices to prove $A(x,z) \leq 0$. For this, it suffices to show $\phi^\top A\phi \leq 0$.

Here is where we use that A has exactly one positive eigenvalue. Since A has nonnegative entries and isn't identically zero, $\mathbf{1}^\top A\mathbf{1} > 0$. Since $\langle \phi, \mathbf{1}\rangle = 0$, if $\phi^\top A\phi > 0$, then A would be positive definite on $\mathrm{span}\{\mathbf{1}, \phi\}$, a subspace of dimension-2, whence A would have at least two strictly positive eigenvalues. Hence, $\phi^\top A\phi \leq 0$ as desired and we are done. □

Lemma 3.2.4 is also intimately related to the structural properties of the support of 0-spectrally independent distributions; see Section 5.6 for further discussion.

3.2.2 Decomposing the Local Walks via Conditioning

Finally, we prove the Trickle-Down Theorem. We will make use of the following lemma, which is essentially just an application of the Law of Total Probability.

Lemma 3.2.5. *Let μ be a probability distribution on $\binom{\mathscr{U}}{n}$, where $\mathscr{U}$ is a finite ground set and $n \geq 1$ is a positive integer. For each $u \in \mathscr{U}$, view the conditional local walk $\mathcal{Q}_{\mu^u}$ as a matrix in $\mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ and the conditional marginal distribution μ_1^u as a vector in $\mathbb{R}^{\mathscr{U}}$ by appropriately padding them with zeros. Then for every $\phi, \psi \in \mathbb{R}^{\mathscr{U}}$, the following identities hold:*

$$\langle \phi, \psi\rangle_{\mu_1} = \mathbb{E}_{u\sim\mu_1}\langle \phi, \psi\rangle_{\mu_1^u} \tag{3.1}$$

$$\langle \phi, \mathcal{Q}_\mu\psi\rangle_{\mu_1} = \mathbb{E}_{u\sim\mu_1}\langle \phi, \mathcal{Q}_{\mu^u}\psi\rangle_{\mu_1^u} \tag{3.2}$$

$$\langle \mathcal{Q}_\mu\phi, \mathcal{Q}_\mu\psi\rangle_{\mu_1} = \mathbb{E}_{u\sim\mu_1}\left[\langle \mathbf{1}, \phi\rangle_{\mu_1^u} \cdot \langle \mathbf{1}, \psi\rangle_{\mu_1^u}\right] \tag{3.3}$$

Proof. For the first identity, we apply the Law of Total Probability to obtain

$$\mathbb{E}_{u\sim\mu_1}\langle \phi, \psi\rangle_{\mu_1^u} = \sum_{v\in\mathscr{U}} \phi(v)\psi(v) \cdot \mathbb{E}_{u\sim\mu_1}\left[\mu_1^u(v)\right] = \sum_{v\in\mathscr{U}} \phi(v)\psi(v) \cdot \mu_1(v) = \langle \phi, \psi\rangle_{\mu_1}.$$

Similarly, additionally using the fact that the rows of $\mathcal{Q}_\mu$ are given precisely by each μ_1^u for $u \in \mathscr{U}$, we have that

$$\begin{aligned}
\mathbb{E}_{u\sim\mu_1}\langle \phi, \mathcal{Q}_{\mu^u}\psi\rangle_{\mu_1^u} &= \mathbb{E}_{u\sim\mu_1}\mathbb{E}_{v\sim\mu_1^u}\left[\phi(v) \cdot \left(\mathcal{Q}_{\mu^u}\psi\right)(v)\right] \\
&= \mathbb{E}_{u\sim\mu_1}\mathbb{E}_{v\sim\mu_1^u}\mathbb{E}_{w\sim\mu_1^{uv}}\left[\phi(v) \cdot \psi(w)\right] \\
&= \mathbb{E}_{v\sim\mu_1}\mathbb{E}_{w\sim\mu_1^v}\left[\phi(v) \cdot \psi(w)\right] \\
&= \mathbb{E}_{v\sim\mu_1}\left[\phi(v) \cdot \left(\mathcal{Q}_\mu\psi\right)(v)\right] \\
&= \langle \phi, \mathcal{Q}_\mu\psi\rangle_{\mu_1}.
\end{aligned}$$

Finally, using a nearly identical calculation, we see that

$$\begin{aligned}
\langle \mathcal{Q}_\mu\phi, \mathcal{Q}_\mu\psi\rangle_{\mu_1} &= \mathbb{E}_{u\sim\mu_1}\left[\left(\mathcal{Q}_\mu\phi\right)(u) \cdot \left(\mathcal{Q}_\mu\psi\right)(u)\right] \\
&= \mathbb{E}_{u\sim\mu_1}\left[\mathbb{E}_{\mu_1^u}[\phi] \cdot \mathbb{E}_{\mu_1^u}[\psi]\right] \\
&= \mathbb{E}_{u\sim\mu_1}\left[\langle \mathbf{1}, \phi\rangle_{\mu_1^u} \cdot \langle \mathbf{1}, \psi\rangle_{\mu_1^u}\right].
\end{aligned}$$

□

Proof of Theorem 3.2.1. Fix an arbitrary $\phi \in \mathbb{R}^{\mathscr{U}}$. Using Lemma 3.2.5, we may decompose the quadratic form of $\mathcal{Q}_\mu$ as

$$\begin{aligned}
\langle \phi, \mathcal{Q}_\mu\phi\rangle_{\mu_1} &= \mathbb{E}_{u\sim\mu_1}\langle \phi, \mathcal{Q}_{\mu^u}\phi\rangle_{\mu_1^u} \\
&\leq \mathbb{E}_{u\sim\mu_1}\left[\boldsymbol{\lambda}_2\left(\mathcal{Q}_{\mu^u}\right) \cdot \left(\langle \phi, \phi\rangle_{\mu_1^u} - \langle \phi, \left(\mathbf{1}\otimes\mu_1^u\right)\phi\rangle_{\mu_1^u}\right) + \langle \phi, \left(\mathbf{1}\otimes\mu_1^u\right)\phi\rangle_{\mu_1^u}\right] \\
&\qquad\qquad \text{(Lemma 1.4.20)} \\
&\leq \boldsymbol{\lambda} \cdot \mathbb{E}_{u\sim\mu_1}\left[\langle \phi, \phi\rangle_{\mu_1^u}\right] + (1-\boldsymbol{\lambda}) \cdot \mathbb{E}_{u\sim\mu_1}\left[\langle \mathbf{1}, \phi\rangle_{\mu_1^u}^2\right] \\
&= \boldsymbol{\lambda} \cdot \langle \phi, \phi\rangle_{\mu_1} + (1-\boldsymbol{\lambda}) \cdot \langle \mathcal{Q}_\mu\phi, \mathcal{Q}_\mu\phi\rangle_{\mu_1}
\end{aligned}$$

In particular, taking ϕ to be a right eigenvector of $\mathcal{Q}_\mu$ with eigenvalue $\lambda_2(\mathcal{Q}_\mu)$, we see that

$$\begin{aligned}\lambda_2(\mathcal{Q}_\mu)\cdot\langle\phi,\phi\rangle_{\mu_1} &= \langle\phi,\mathcal{Q}_\mu\phi\rangle_{\mu_1}\\ &\le \lambda\cdot\langle\phi,\phi\rangle_{\mu_1}+(1-\lambda)\cdot\langle\mathcal{Q}_\mu\phi,\mathcal{Q}_\mu\phi\rangle_{\mu_1}\\ &= \left(\lambda+(1-\lambda)\cdot\lambda_2(\mathcal{Q}_\mu)^2\right)\cdot\langle\phi,\phi\rangle_{\mu_1}.\end{aligned}$$

This tells us that $\lambda_2(\mathcal{Q}_\mu)\le\lambda+(1-\lambda)\cdot\lambda_2(\mathcal{Q}_\mu)^2$, which is equivalent to the inequality

$$\lambda_2(\mathcal{Q}_\mu)\cdot(1-\lambda_2(\mathcal{Q}_\mu))\le\lambda\cdot(1+\lambda_2(\mathcal{Q}_\mu))\cdot(1-\lambda_2(\mathcal{Q}_\mu)).$$

Since $\mathcal{Q}_\mu$ is connected, $\lambda_2(\mathcal{Q}_\mu)<1$ by Fact 1.4.18, and we may cancel the $1-\lambda_2(\mathcal{Q}_\mu)$ from both sides. Rearranging then yields the desired inequality. □

3.3 The Mihail-Vazirani Conjecture

In this section, we positively resolve the Mihail-Vazirani Conjecture on the edge expansion of the *bases exchange graph*. For an n-element rank-r matroid $\mathcal{M}=(\mathscr{U},\mathscr{X})$, the bases exchange graph $G_\mathcal{M}$ is a unweighted undirected simple graph whose vertices correspond to bases of $\mathcal{M}$ and where two bases B,B' are connected by an edge if and only if $|B\triangle B'|=2$ (or, equivalently, $|B\cap B'|=r-1$). This graph is significant in polyhedral combinatorics, as it is precisely the 1-*skeleton graph of the matroid basis polytope*. To state the conjecture, we first define edge expansion.

Definition 19 (Edge Expansion)**.** *Let $G=(V,E)$ be an unweighted undirected simple graph. For $S\subseteq V$, we define the **edge expansion of** S as*

$$\mathsf{h}(S)\stackrel{\mathsf{def}}{=}\frac{|E(S,\overline{S})|}{\min\{|S|,|\overline{S}|\}}.$$

*We define the **edge expansion of** G as $\mathsf{h}(G)\stackrel{\mathsf{def}}{=}\min_{\emptyset\subsetneq S\subsetneq V}\mathsf{h}(S)$.*

Conjecture 3 ([MV89])**.** *Let $\mathcal{M}=(\mathscr{U},\mathscr{X})$ be a matroid. Then $\mathsf{h}(G_\mathcal{M})\ge 1$.*[4]

As one can imagine, the bases exchange graph $G_\mathcal{M}$ is closely related to the bases exchange walk, whose spectral gap we analyzed in the preceding section. Using Theorem 3.1.1, we settle the conjecture in the affirmative.

Theorem 3.3.1 ([Ana+19])**.** *Conjecture 3 is true.*

To prove this theorem, we first introduce another quantity related to edge expansion called *conductance*, which will be useful when considering appropriately weighted version of the bases exchange graph.

Definition 20 (Conductance)**.** *Let $G=(V,E,c)$ be an undirected graph with nonnegative edge weights $c:E\to\mathbb{R}_{\ge0}$. For $S\subseteq V$, we define the **conductance of** S as*

$$\Phi(S)\stackrel{\mathsf{def}}{=}\frac{c(S,\overline{S})}{\min\{\mathrm{Vol}(S),\mathrm{Vol}(\overline{S})\}}=\frac{\sum_{e\in E(S,\overline{S})}c(e)}{\min\left\{\sum_{v\in S}c(v),\sum_{v\notin S}c(v)\right\}}$$

*where recall $c(v)=\sum_{e\in E:e\sim v}c(e)$ is the weighted degree of the vertex $v\in V$. We define the **conductance of** G as $\Phi(G)\stackrel{\mathsf{def}}{=}\min_{\emptyset\subsetneq S\subsetneq V}\Phi(S)$.*

Similarly, if P is a Markov chain which is reversible w.r.t. a probability distribution μ over a finite state space Ω, then we define the conductance of P to be $\Phi(\mathsf{P})\stackrel{\mathsf{def}}{=}\Phi(G_\mathsf{P})$, where recall G_P is the underlying weighted graph of P (see Section 1.4.5).

One important feature of conductance is that it is closely related to spectral gaps of Markov chains. In particular, we have the classic *(discrete) Cheeger Inequalities* from spectral graph theory [AM85; Alo86], which are the discrete analog of Cheeger's Inequalities for manifolds [Che70]. We use the following more general version for Markov chains.

[4] As previously mentioned, Mihail-Vazirani actually posed the same conjecture but for the 1-skeleton graph of *any* polytope whose vertices have $\{0,1\}$-coordinates, including matroid basis polytopes. However, this more general conjecture remains open, and is beyond the scope of this thesis; again, see [Kai04] for other special cases which have been positively resolved, and [KLT22] for recent evidence against the most general version of this conjecture.

Theorem 3.3.2 ((Discrete) Cheeger Inequalities; see e.g. [SJ89; LS88]). *Let $G = (V, E, w)$ be an undirected graph with nonnegative edge weights $w : E \to \mathbb{R}_{\geq 0}$, and let $\mathsf{P}_G = D_G^{-1} A_G$ be the reversible Markov kernel of the corresponding simple random walk on the vertices of G. Then we have the following inequalities:*

$$\frac{\Phi(G)^2}{2} \leq \gamma\left(\mathsf{P}_G\right) \leq 2 \cdot \Phi(G)$$

Proof of Theorem 3.3.1. Observe that since μ is the uniform distribution, $\mathcal{P}_\mu$ is symmetric, and so $\mathcal{P}_\mu$ can be viewed as the weighted adjacency matrix of a weighted graph $H_{\mathcal{M}}$ whose vertices correspond to bases, and the edge between two bases τ, τ' has weight $\mathcal{P}_\mu(\tau, \tau') = \mathcal{P}_\mu(\tau', \tau)$. By Cheeger's Inequality (see Theorem 3.3.2) and Theorem 3.1.1, we have that

$$\Phi(H_{\mathcal{M}}) \geq \frac{\gamma\left(\mathcal{P}_\mu\right)}{2} \geq \frac{1}{2r}.$$

On the other hand, since $\mathcal{P}_\mu$ is a Markov chain, the weighted degree of every basis in $H_{\mathcal{M}}$ is precisely 1, so we have

$$\Phi(S) = \frac{\sum_{\tau \in S, \tau' \notin S} \mathcal{P}_\mu(\tau, \tau')}{\min\left\{|S|, |\overline{S}|\right\}}.$$

for every subset S of bases of $\mathcal{M}$. If we can show that $\mathcal{P}_\mu(\tau, \tau') \leq \frac{1}{2r}$ for every pair of distinct, neighboring τ, τ', then we would be done, since

$$\frac{1}{2r} \leq \Phi(H_{\mathcal{M}}) \leq \Phi(S) \leq \frac{1}{2r} \cdot \frac{\left|E(S, \overline{S})\right|}{\min\left\{|S|, |\overline{S}|\right\}} = \frac{1}{2r} \cdot \mathsf{h}(S)$$

holds for every S.

Fix two arbitrary bases τ, τ'. If τ, τ' are distinct, neighboring bases, then $|\tau \cap \tau'| = r - 1$. It follows that

$$\mathcal{P}_\mu\left(\tau, \tau'\right) = \frac{\mu(\tau')}{r^2 \cdot \mu_{r-1}\left(\tau \cap \tau'\right)} \leq \frac{1}{2r}$$

since $\mu_{r-1}\left(\tau \cap \tau'\right) \geq \frac{1}{r}\left(\mu(\tau) + \mu(\tau')\right) = \frac{2}{r}\mu(\tau')$, as desired. □

3.4 When the Vanilla Trickle-Down Theorem Fails

We saw earlier in this chapter how Oppenheim's vanilla trickle-down method (see Theorem 3.2.1) works beautifully in the setting of matroids (see Section 3.2.1). Indeed, as of this writing, this is the only known (elementary) method to obtain spectral independence for the uniform distribution over the bases of a matroid.[5] This influential method has also had a number of important applications in recent constructions of bounded-degree high-dimensional expanders [KO18; KO20a; OP22]. However, by itself without modifications, the method runs into severe barriers for most other classical sampling applications. We already saw one such example in Section 2.6.

In the hardcore model, the worst case 1-dimensional conditional distributions are essentially Gibbs distributions for the hardcore model on a graph with two vertices connected by a single edge. Via explicit computation, the local random walk of such distributions has second eigenvalue $\frac{\lambda}{1+\lambda}$. If $\lambda \geq \Omega(1)$, then Theorem 3.2.1 can only be applied $O(1)$ many rounds before it stops yielding meaningful bounds. It can't be used to analyze the spectral independence for any conditional distribution of nontrivial dimension. One would need $\lambda \leq \frac{1}{n}$ in order for Theorem 3.2.1 to yield anything useful. However, we already know via simple coupling techniques from the classical theory of Markov chains that $\lambda \leq \frac{1}{\Delta - 1}$ suffices to establish rapid mixing, where Δ is the maximum degree of the underlying graph. This is discrepancy is severe, particularly for bounded-degree graphs.

This unfortunately is a rather general issue with applying Theorem 3.2.1 directly to distributions often encountered in sampling-type applications. In Chapter 4, we will see one way to remedy this situation. There, we will discuss an application to sampling proper q-colorings when q depends linearly on the maximum degree Δ, whereas a direct application of Theorem 3.2.1 only yields rapid mixing when the number of colors depends linearly on the number of vertices n.

[5] Historically, spectral independence for matroids was first established by connecting it with log-concavity of the bases generating polynomial (see Chapter 5), which is then established using the recently developed *combinatorial Hodge theory* due to [AHK18]; see [HW17; AOV21].

Chapter 4

The Matrix Trickle-Down Method

In Section 3.4, we saw an example of a class of high-dimensional discrete distributions for which Oppenheim's Trickle-Down Theorem fails to yield any interesting bounds on the spectral independence of the distribution. Unfortunately, this defect is present for many other distributions arising in statistical physics, machine learning, etc. In this chapter, we attempt to overcome this obstacle by developing a more sophisticated trickle-down technique which uses *nonuniform bounds* on the entire spectrum of the local random walks $\mathcal{Q}_\mu$. This generalizes Oppenheim's influential result, and will be achieved using matrix inequalities. More broadly, this gives another method to establish spectral independence for high-dimensional discrete distributions and local spectral expansion for (weighted) high-dimensional simplicial complexes.

Our main case study in this chapter will be to sampling proper colorings in graphs. We show that for any $\epsilon > 0$, the natural Glauber dynamics mixes rapidly and generates a random proper *edge-coloring* of a graph (equivalently, proper vertex-coloring of the line graph) with maximum degree Δ whenever the number of colors is at least $q \geq \left(\frac{10}{3} + \epsilon\right)\Delta$ and Δ is at least some constant depending only on ϵ. For edge-colorings, this improves upon prior work [Vig00; Che+19], which show rapid mixing when $q \geq \left(\frac{11}{3} - \epsilon_0\right)\Delta$, where $\epsilon_0 \approx 10^{-5}$ is a small fixed constant.

The results in this chapter are based on [ALO22]. We note that the techniques in this chapter have been further developed and refined in a recent work of Abdolazimi–Oveis Gharan [AO22].

4.1 Sampling Proper Vertex-Colorings and Edge-Colorings of Graphs

We begin by stating the main problem we tackle and the algorithmic results we establish in this chapter. We also discuss the history and significance of the problem of sampling colorings in graphs.

Theorem 4.1.1 (Proper Edge-Colorings)**.** *Let $G = (V, E)$ be a graph of maximum degree Δ. For any $0 < \epsilon \leq \frac{1}{10}$ such that $\frac{\log^2 \Delta}{\Delta} \leq \frac{\epsilon^3}{15}$, and any collection of color lists $\mathcal{L} = \{\mathcal{L}(e)\}_{e \in E}$ satisfying $|\mathcal{L}(e)| \geq \deg(e) + \left(\frac{4}{3} + 4\epsilon\right)\Delta$, where $\deg(e)$ is the number of neighbors of e in the line graph of G, the spectral gap of the Glauber dynamics for sampling proper $\mathcal{L}$-edge-colorings on G is $\Omega\left(n^{-O(1/\epsilon)}\right)$ and the mixing time is $O\left(n^{O(1/\epsilon)}\right)$. Furthermore, if $\Delta \leq O(1)$, then the modified and standard log-Sobolev constants are $\Omega_{\epsilon,\Delta}(1/n)$.*

We remark that our general mixing time bound has no dependence on Δ or q. So, the algorithm runs in polynomial time even for graphs of unbounded degree.

The problem of (approximately) counting and sampling proper colorings in graphs is a fundamental question in the field of counting and sampling which has puzzled researchers for decades. Its study goes back to the 1990s [Jer95; SS97; Vig00], with applications to statistical physics. The uniform distribution over proper colorings is also known as the Gibbs distribution of the *antiferromagnetic q-state Potts model at zero temperature*. For $q \leq \Delta$, there is no FPRAS to approximately count proper q-colorings (at least when q is even) unless NP = RP, even when the graph is triangle-free and Δ-regular [GŠV15].

On the other hand, it is not hard to see that when $q \geq \Delta + 2$, the Glauber dynamics is irreducible and reversible w.r.t. the uniform distribution over all proper q-colorings of G. It is

a folklore conjecture that this threshold $q \geq \Delta + 2$ is sufficient for the Glauber dynamics to be rapidly mixing, and that efficient approximate counting and sampling algorithms exist as long as $q \geq \Delta + 1$.

Conjecture 4. *Let $G = (V, E)$ be a graph with maximum degree Δ. Then there exists an* FPRAS *for approximately counting proper q-colorings of G whenever $q \geq \Delta+1$. Furthermore, if $q \geq \Delta+2$, then the Glauber dynamics for approximately sampling uniformly random proper q-colorings of G mixes in $O(n \log n)$ steps.*

This conjecture, when combined with the hardness result of [GŠV15], provides an example of a *computational phase transition*; there is a threshold for q, based on Δ, which precisely delineates between the computationally tractable and intractable regimes. However, despite significant attempts, we are still very far from proving this conjecture. An example of an important model where such a computational phase transition for sampling was established is discussed in Chapter 7.

To this date, the best known result for general graphs is due Chen-Delcourt-Moitra-Perarnau-Postle [Che+19], who show that the Glauber dynamics mixes in polynomial time whenever $q \geq \left(\frac{11}{6} - \epsilon_0\right)\Delta$ for a small fixed constant $\epsilon_0 \approx 10^{-5}$; this slightly improves on the classical works of Jerrum [Jer95] and Vigoda [Vig00] which establish polynomial mixing times when $q > 2\Delta$ and $q > \frac{11}{6}\Delta$, respectively. Besides Glauber dynamics, there are two other methods of attack for counting and sampling graph colorings, namely Weitz's elegant algorithmic framework based on correlation decay [Wei06] and Barvinok's interpolation method [Sok01; GK12; LY13; LSS19; Ben+21].

Another fascinating recent line of work tries to circumvent the $\frac{11}{6}\Delta$ barrier by focusing on graphs satisfying *local sparsity* conditions, e.g. graphs which have large *girth* [HV03; Mol04; HV05; FV06; FV07; Dye+13; HVV15; Che+21d; Fen+21]. These results typically exploit (strong) *correlation decay* properties, which roughly speaking, say that if we color a vertex v with a color c, then the marginal probability of coloring a "far away" vertex u with a color c' does not change much. Although it is conjectured that the uniform distribution over proper q-vertex-colorings exhibits correlation decay, more formally known as *strong spatial mixing*, for $q \geq \Delta + 1$, to this date, we are lacking techniques to establish such a statement (see e.g., [GMP05; Yin14; GKM15; Eft+19]). We note that these local sparsity assumptions are typically very strong, as it is known that even triangle-free graphs can be colored with as little as $O\left(\frac{\Delta}{\log \Delta}\right)$ many colors [Joh96].

We study random proper edge-colorings of graphs precisely because this goes *against* this trend on locally sparse graphs. Edge-colorings can equivalently be seen as a vertex-colorings of *line graphs*, which are very *dense* locally; they contain induced cliques of size $\Omega(\Delta)$. To the best of our knowledge, the only previous result on sampling edge-colorings which goes substantially beyond the $\frac{11}{6}\Delta$ barrier is the recent work of Delcourt-Heinrich-Perarnau [DHP20], which shows that the Glauber dynamics mixes rapidly when the underlying graph is a tree and $q \geq \Delta+1$. Note that for a graph with maximum degree Δ, the maximum degree of the line graph can be as large as 2Δ. Therefore, with the $\left(\frac{11}{6} - \epsilon_0\right)\Delta$ result of [Che+19], one would need $q \geq \left(\frac{11}{3} - \frac{\epsilon_0}{2}\right)\Delta$ to guarantee polynomial mixing for all edge-coloring instances. In our main theorem we prove that this barrier can be broken for sampling edge-coloring of any graph with maximum degree Δ.

The second result of this chapter applies to proper vertex-colorings of trees. Specifically, we show that for any list vertex-coloring instance, where G is a tree with maximum degree Δ and the size of the color list of every vertex v is at least $\deg(v) + \epsilon\Delta$ for $\epsilon = \Omega\left(\frac{\log \Delta}{\sqrt{\Delta}}\right)$, the Glauber dynamics mixes rapidly.

Theorem 4.1.2. *Let $G = (V, E)$ be a tree of maximum degree Δ. For any $0 < \epsilon \leq 1$ such that $\frac{\log^2 \Delta}{\Delta} \leq \frac{\epsilon^2}{100}$ and any collection of color lists $\mathcal{L} = \{\mathcal{L}(v)\}_{v \in V}$ satisfying $|\mathcal{L}(v)| \geq \deg(v) + \epsilon\Delta$, the spectral gap of the Glauber dynamics for sampling proper $\mathcal{L}$-vertex-list-colorings on G is $\Omega\left(n^{-O(1/\epsilon)}\right)$, therefore the mixing time is $O\left(n^{O(1/\epsilon)}\right)$. Furthermore, if $\Delta \leq O(1)$, the modified and standard log-Sobolev constants are $\Omega_{\epsilon,\Delta}(1/n)$.*

The above theorem, although it is not as strong as [MSW07], shows that the Glauber dynamics mixes rapidly even when we have a list coloring problem on a tree. Furthermore it gives a possible avenue to exploit our techniques to prove that Glauber dynamics mixes rapidly on any graph when $q \geq (1 + \epsilon)\Delta$. We expect that upon further investigation our techniques can be coupled with the extensive literature on random proper colorings of graphs with large girth to break the $\left(\frac{11}{6} - \epsilon_0\right)\Delta$ barrier.

As discussed before, nearly all prior results showing rapid mixing of the Glauber dynamics for this problem use variants of the coupling method or the correlation decay property. Instead, our strategy is to establish spectral independence for the uniform distribution over proper vertex-colorings/edge-colorings. Historically, this was first done for graph colorings [Che+21d; Fen+21] using the *correlation decay* property in statistical physics, extending previous work for two-state spin systems [ALO21; CLV20]. More specifically, [Che+21d; Fen+21] showed that the correlation decay results of [GKM15] give $O(1)$-spectral independence bounds for proper colorings of triangle-free graphs when $q > \alpha\Delta$, where $\alpha \approx 1.763$ is a constant; their result can also be extended to cover all graphs when $q > 2\Delta$. They concluded rapid mixing of the Glauber dynamics in this regime, a result that seems difficult to obtain using coupling arguments.

However, despite the power of this approach, the main difficulty is that obtaining correlation decay for proper colorings is extremely challenging. To circumvent this, we develop and use a more advanced version of Oppenheim's vanilla Trickle-Down Theorem (see Theorem 3.2.1). More specifically, as alluded to earlier, we instead use matrices to simultaneously control all eigenvalues of the local random walks in the induction. As such, we call this a *matrix trickle-down method*. We provide sufficient conditions on these matrices for the induction to go through, and carefully construct them to prove the above results on sampling proper colorings.

Our next theorem is the main technical result of this chapter. Like Oppenheim's original Trickle-Down Theorem, it will be more convenient state everything directly in terms of the spectral gap of the local random walks $\mathcal{Q}_\mu$, rather than spectral independence and influence matrices. Again, by Lemma 2.3.3 these are completely equivalent. Throughout this chapter, for convenience, we write $\pi_\tau \overset{\text{def}}{=} \mu_1^\tau$ viewed as a vector in $\mathbb{R}^{\mathscr{U}}$ (or a probability distribution over $\mathscr{U}$) and $\Pi_\tau \overset{\text{def}}{=} \operatorname{diag}(\pi_\tau) \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$; if $\tau = \emptyset$, we drop the subscript.

Theorem 4.1.3 (Iterated Matrix Trickle-Down Method; [ALO22])**.** *Let μ be a totally connected probability distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and positive integer $n \geq 1$. Suppose $\{B_\tau \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}\}_{\tau\in\mathcal{X}(\leq n-2)}$ is a family of symmetric matrices satisfying the following:*

1. ***Base Case:*** *For every feasible τ with $|\tau| = n-2$, we have the spectral inequality*

$$\Pi_\tau \mathcal{Q}_{\mu^\tau} - 2\pi_\tau\pi_\tau^\top \preceq B_\tau \preceq \frac{1}{5}\Pi_\tau.$$

2. ***Recursive Condition:*** *For every feasible τ with $|\tau| = n-k$ where $k \geq 3$, B_τ satisfies*

$$B_\tau \preceq \frac{k-1}{3k-1}\Pi_\tau \quad \text{and} \quad \mathbb{E}_{x\sim\pi_\tau}[B_{\tau+x}] \preceq B_\tau - \frac{k-1}{k-2}B_\tau\Pi_\tau^{-1}B_\tau.$$

Then $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \rho\left(\Pi_\tau^{-1}B_\tau\right)$ for all feasible τ with $|\tau| \leq n-2$, where ρ represents the spectral radius. In particular, μ is a $(\lambda_0,\ldots,\lambda_{n-2})$-local spectral expander with $\lambda_k \leq \max_{\tau\in\mathcal{X}(k)}\rho(\Pi_\tau^{-1}B_\tau)$.

Remark 17. Oppenheim's vanilla Trickle-Down Theorem (see Theorem 3.2.1 and Corollary 3.2.2) is a special case of our result by taking the matrices B_τ to be a multiple of $\operatorname{diag}(\mu_\tau)$. We justify this formally through a quick calculation in Section 4.2 below.

Remark 18. It also turns out for our applications to proper colorings we will need a slight extension of the above theorem. However, one should take the above theorem as the heart of our technical contributions in this chapter. See Theorem 4.2.3 below and the surrounding discussion for more details on the slight extension.

We will sometimes refer to the B_τ matrices as *bounding matrices*. Typically, it is not difficult to construct some family of matrices $\{B_\tau\}_{\tau\in\mathcal{X}(\leq n-2)}$ satisfying the assumptions of Theorem 4.1.3. The key challenge is choosing the B_τ in a way such that one can bound $\rho(\Pi_\tau^{-1}B_\tau) \leq O\left(\frac{1}{n-|\tau|}\right)$.

Our second key insight is that the matrices B_τ can be designed to have convenient *sparsity patterns* depending on μ, which allow for straightforward bounds on $\rho(\Pi_\tau^{-1}B_\tau)$. For instance, in our application to proper colorings, our matrices B_τ will have rows and columns corresponding to vertex-color pairs $vc = (v,c)$ where $v \in V$ and $c \in [q]$, and they will be supported on the "proper coloring constraints", namely pairs uc, vc' of vertex-color pairs where $u \sim v$ in G and $c = c'$. We demonstrate the usefulness of this approach to sampling proper colorings in graphs below. As these matrix constructions can be rather involved, we will first start with simple constructions which yield weaker results, and then progressively increase the complexity.

4.1.1 Some Additional Linear Algebraic Preliminaries

Fact 4.1.4. *For any symmetric matrix $A \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ where $A_{i,j} \neq 0$ only for $i,j \in S \subseteq \mathscr{U}$, we have $A \preceq \|A\|_\infty \mathsf{Id}^S$.*

Fact 4.1.5. *For rectangular matrices $A, B \in \mathbb{R}^{m\times n}$ and positive $\epsilon > 0$, we have the inequalities $AB^\top + BA^\top \preceq \epsilon AA^\top + \frac{1}{\epsilon}BB^\top$ and $(A+B)(A+B)^\top \preceq (1+\epsilon)AA^\top + \left(1+\frac{1}{\epsilon}\right)BB^\top$.*

Proof. Since $\epsilon > 0$, we can write

$$0 \preceq \left(\sqrt{\epsilon}A - \frac{1}{\sqrt{\epsilon}}B\right)\left(\sqrt{\epsilon}A - \frac{1}{\sqrt{\epsilon}}B\right)^\top = \epsilon AA^\top + \frac{1}{\epsilon}BB^\top - AB^\top - BA^\top.$$

Rearranging yields the first inequality. Adding $AA^\top + BB^\top$ to both sides yields the second inequality. □

Lemma 4.1.6. *Let $A, B \in \mathbb{R}^{n\times n}$ be symmetric matrices such that $A\cdot(\mathsf{Id} - \alpha A) \preceq B\cdot(\mathsf{Id}-\alpha B)$ for a positive real number $\alpha > 0$. If $A, B \preceq \frac{1}{2\alpha}\cdot\mathsf{Id}$, then $A \preceq B$. Note that we crucially do not require $A, B \succeq 0$.*

Proof. It suffices to prove the claim when $\alpha = 1$, since the general claim then follows by replacing A, B with $\alpha A, \alpha B$, respectively.

First, observe the matrix map $M \mapsto M(I-M)$ is a bijection between $\left\{M \in \mathbb{R}^{n\times n} : M \preceq \frac{1}{2}\mathsf{Id}\right\}$ and $\left\{T \in \mathbb{R}^{n\times n} : T \preceq \frac{1}{4}\mathsf{Id}\right\}$ with inverse

$$T \mapsto \frac{1}{2}\mathsf{Id} - \left(\frac{1}{4}\mathsf{Id} - T\right)^{1/2}. \tag{4.1}$$

The way to see this is via the eigendecomposition. If $M = \sum_{i=1}^n \lambda_i \varphi_i\varphi_i^\top$ for an orthonormal eigendecomposition $\{\varphi_i\}_{i=1}^n$ with corresponding eigenvalues $\{\lambda_i\}_{i=1}^n$, then $M(I-M) = \sum_{i=1}^n \lambda_i(1-\lambda_i)\varphi_i\varphi_i^\top$. Hence, to prove this claim, it suffices to show that the real function $x \mapsto x(1-x)$ is a bijection between $(-\infty, 1/2]$ and $(-\infty, 1/4]$. To see this, observe that the quadratic $x(1-x) = \lambda$ has roots $x = \frac{1}{2} \pm \left(\frac{1}{4} - \lambda\right)^{1/2}$, and since we enforced that $\lambda \leq \frac{1}{4}$, we must choose x to be the smaller root, i.e. $\frac{1}{2} - \left(\frac{1}{4} - \lambda\right)^{1/2} \leq \frac{1}{2}$ gives the inverse function.

Knowing this explicit inverse function, we now return to the proof of the lemma. Since $A, B \preceq \frac{1}{2}I$, we may apply the inverse Eq. (4.1) to $A(I-A)$ (resp. $B(I-B)$) to recover A (resp. B). Hence, to prove the claim, it suffices to establish *operator monotonicity* of Eq. (4.1). A quick calculation reveals that this is equivalent to operator monotonicity of $M \mapsto \sqrt{M}$ for positive semidefinite $M \in \mathbb{R}^{n\times n}$, which is well-known and follows for instance by using the Löwner-Heinz Theorem [Löw34]. □

4.2 A General Matrix Trickle-Down Method

Our goal in this section is to prove Theorem 4.1.3 and its extension Theorem 4.2.3. One of our main insights is to replace the hypothesis $\lambda_2\left(\mathcal{Q}_{\mu^i}\right) \leq \lambda$ in Theorem 3.2.1, which merely provides a uniform bound on all nontrivial eigenvalues of $\mathcal{Q}_\mu$, with a matrix bound "$\mathcal{Q}_{\mu^i} \preceq_{\pi_i} \Pi_i^{-1}B_i$". The hope is that the matrix B_i itself can be easily bounded, and simultaneously provide information on where the "bad" eigenspaces of $\mathcal{Q}_{\mu^i}$ are. So, roughly speaking, although many of the 1-dimensional conditional distributions μ^τ may have large $\lambda_2\left(\mathcal{Q}_{\mu^\tau}\right)$, by carefully choosing the B_τ, one can "average out" these bad eigenspaces to show that the eigenvalues of $\mathcal{Q}_{\mu^\sigma}$ are small for smaller σ. We formalize this as follows.

Theorem 4.2.1 (One-Step Matrix Trickle-Down Theorem)**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and positive integer $n \geq 1$. Suppose the following hold:*

1. ***Connectivity:*** *$\lambda_2\left(\mathcal{Q}_\mu\right) < 1$, i.e. the local walk $\mathcal{Q}_\mu$ is irreducible (i.e. connected).*
2. ***Matrix Bound for Conditional Local Walks:*** *There is a family of symmetric matrices $\{B_x\}_{x\in\mathscr{U}}$ such that*

$$\Pi_x\mathcal{Q}_{\mu^x} - \alpha\pi_x\pi_x^\top \preceq B_x \preceq \frac{1}{2\alpha+1}\Pi_x$$

for all $x \in \mathscr{U}$.

Then the local walk $\mathcal{Q}_\mu$ satisfies the spectral bound $\Pi\mathcal{Q}_\mu - \left(2 - \frac{1}{\alpha}\right)\pi\pi^\top \preceq B$, and in particular $\lambda_2(\mathcal{Q}_\mu) \leq \rho(\Pi^{-1}B)$, where B is any symmetric matrix satisfying $B \preceq \frac{1}{2\alpha}\Pi$ and $\mathbb{E}_{x\sim\pi}[B_x] \preceq B - \alpha B\Pi^{-1}B$.

Note that by induction, Theorem 4.1.3 follows as an immediate consequence of this generalized one-step trickle-down result. The strange-looking factor of $\frac{k-1}{3k-1}$ arises from the fact that the recursion $\alpha_k = 2 - \frac{1}{\alpha_{k-1}}$ with base case $\alpha_2 = 2 = \frac{2}{1}$ is solved by taking $\alpha_k = \frac{k}{k-1}$ for $k \geq 2$, from which it follows that $\frac{1}{2\alpha_k+1} = \frac{k-1}{3k-1}$.

To see that the above theorem generalizes Theorem 3.2.1, note that if $\lambda_2(\mathcal{Q}_{\mu^x}) \leq \lambda \leq 1/2$ for all $x \in \mathscr{U}$ then $B_x = \lambda\Pi_x$, $\alpha = 1 - \lambda$, and $B = \frac{\lambda}{1-\lambda}\Pi$ satisfies the assumptions of the above theorem. In particular,

$$\Pi_x\mathcal{Q}_{\mu^x} - (1-\lambda)\pi_x\pi_x^\top \preceq \lambda\Pi_x = B_x \qquad (\lambda_2(\mathcal{Q}_{\mu^x}) \leq \lambda \text{ and } \pi_x\mathcal{Q}_{\mu^x} = \pi_x)$$
$$B_x = \lambda\Pi_x \leq \frac{1}{2(1-\lambda)+1}\Pi_x = \frac{1}{2\alpha+1}\Pi_x \qquad (\lambda \leq 1/2)$$

verifies the matrix bound for the conditional local walks, while

$$B = \frac{\lambda}{1-\lambda}\Pi \preceq \frac{1}{2(1-\lambda)}\Pi = \frac{1}{2\alpha}\Pi \qquad (\lambda \leq 1/2)$$
$$\begin{aligned}\mathbb{E}_{x\sim\pi}[B_x] = \lambda \cdot \mathbb{E}_{x\sim\pi}[\Pi_x] = \lambda\Pi &= \frac{\lambda}{1-\lambda}\Pi - (1-\lambda)\left(\frac{\lambda}{1-\lambda}\right)^2\Pi \\ &= B - \alpha B\Pi^{-1}B\end{aligned}$$

verifies the hypothesis on B. So, assuming connectivity of $\mathcal{Q}_\mu$, we get $\lambda_2(\mathcal{Q}_\mu) \leq \rho(\Pi^{-1}B) = \frac{\lambda}{1-\lambda}$ from Theorem 4.2.1 as desired.

Let us now prove Theorem 4.2.1. To do this, we use the following lemma.

Lemma 4.2.2. *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and positive integer $n \geq 1$. Suppose for a symmetric matrix B and an $\alpha \geq 1/2$, the matrix inequalities $\Pi\mathcal{Q}_\mu - \left(2 - \frac{1}{\alpha}\right)\pi\pi^\top \preceq \frac{1}{2\alpha}\Pi$ hold and*

$$\Pi\mathcal{Q}_\mu - \alpha\Pi\mathcal{Q}_\mu^2 \preceq B - \alpha B\Pi^{-1}B \tag{4.2}$$

Then we have the bound $\Pi\mathcal{Q}_\mu - \left(2 - \frac{1}{\alpha}\right)\pi\pi^\top \preceq B$.

Proof. Our goal is to apply Lemma 4.1.6 to suitably chosen A, B. Define $Q = \mathcal{Q}_\mu - \left(2 - \frac{1}{\alpha}\right)\mathbf{1}\pi^\top$. A quick calculation shows that $Q - \alpha Q^2 = \mathcal{Q}_\mu - \alpha\mathcal{Q}_\mu^2$, and so by multiplying both sides of Eq. (4.2) by $\Pi^{-1/2}$, we see that Eq. (4.2) is equivalent to

$$\Pi^{1/2}Q\Pi^{-1/2} - \alpha\Pi^{1/2}Q^2\Pi^{-1/2} \preceq \Pi^{-1/2}B\Pi^{-1/2} - \alpha\Pi^{-1/2}B\Pi^{-1}B\Pi^{-1/2}$$

Taking $A = \Pi^{1/2}Q\Pi^{-1/2}$ and $B = \Pi^{-1/2}B\Pi^{-1/2}$, we see by assumption that A, B are symmetric matrices satisfying $A, B \preceq \frac{1}{2\alpha}I$ and $A(I - \alpha A) \preceq B(I - \alpha B)$. It follows by Lemma 4.1.6 that $A \preceq B$, which is equivalent to $\Pi\mathcal{Q}_\mu - \left(2 - \frac{1}{\alpha}\right)\pi\pi^\top = \Pi Q \preceq B$ as desired. □

With this lemma in hand, let us now prove Theorem 4.2.1.

Proof of Theorem 4.2.1. The conclusion follows immediately from Lemma 4.2.2, and so it suffices to verify the conditions of the lemma. By assumption, we already have $B \preceq \frac{1}{2\alpha}\Pi$. Furthermore, $\Pi_x\mathcal{Q}_{\mu^x} - \alpha\pi_x\pi_x^\top \preceq B_x \preceq \frac{1}{2\alpha+1}\Pi_x$ implies that $\lambda_2(\mathcal{Q}_{\mu^x}) \leq \frac{1}{2\alpha+1}$. Since $\lambda_2(\mathcal{Q}_\mu) < 1$, by Theorem 3.2.1 (the original Trickle-Down Theorem), $\lambda_2(\mathcal{Q}_\mu) \leq \frac{1}{2\alpha}$. Combined with the inequality $2 - \frac{1}{\alpha} \geq 1 - \frac{1}{2\alpha}$, which holds since $\alpha \geq 1/2$, it follows that $\Pi\mathcal{Q}_\mu - \left(2 - \frac{1}{\alpha}\right)\pi\pi^\top \preceq \frac{1}{2\alpha}\Pi$.

All that remains is to verify Eq. (4.2). Observe that Lemma 3.2.5 translated into matrix identities implies that

$$\Pi\mathcal{Q}_\mu = \mathbb{E}_{x\sim\pi}[\Pi_x\mathcal{Q}_{\mu^x}] \tag{4.3}$$

and

$$\Pi\mathcal{Q}_\mu^2 = \mathbb{E}_{x\sim\pi}\left[\pi_x\pi_x^\top\right]. \tag{4.4}$$

It follows that

$$\begin{aligned}
\Pi\mathcal{Q}_\mu &= \mathbb{E}_{x\sim\pi}\left[\Pi_x \mathcal{Q}_{\mu^x}\right] && \text{(Eq. (4.3))}\\
&\preceq \mathbb{E}_{x\sim\pi}\left[\alpha\pi_x\pi_x^\top + B_x\right] && \text{(Assumption)}\\
&= \alpha\Pi\mathcal{Q}_\mu^2 + \mathbb{E}_{x\sim\pi}\left[B_x\right] && \text{(Eq. (4.4))}\\
&\preceq \alpha\Pi\mathcal{Q}_\mu^2 + B - \alpha B\Pi^{-1}B && \text{(Assumption)}
\end{aligned}$$

Rearranging, we obtain that $\Pi\mathcal{Q}_\mu - \alpha\Pi\mathcal{Q}_\mu^2 \preceq B - \alpha B\Pi^{-1}B$ as desired. □

4.2.1 A Slight Extension of Theorem 4.1.3

Here, we prove an extension of the matrix trickle-down method to take into account when our distributions factor as independent products of smaller complexes. This will be useful in the context of proper colorings when the input graph is broken into several connect components by coloring some of the vertices.

Theorem 4.2.3. *Let μ be a totally connected probability distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and positive integer $n \geq 1$. Suppose $\{B_\tau \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}\}_{\tau\in\mathcal{X}(\leq n-2)}$ is a family of symmetric matrices satisfying the following:*

1. ***Base Case:*** *For every feasible τ with $|\tau| = n-2$, we have the spectral inequality*

$$\Pi_\tau \mathcal{Q}_{\mu^\tau} - 2\pi_\tau\pi_\tau^\top \preceq B_\tau \preceq \frac{1}{5}\Pi_\tau.$$

2. ***Recursive Condition:*** *For every feasible τ with $|\tau| = n-k$ where $k \geq 3$, the following hold.*

   - *Suppose the probability distribution μ_τ^k over size-k subsets of $\mathcal{V} \subseteq \mathscr{U}$ factors as an independent product $\bigotimes_{i=1}^{\ell}\nu_i$ of distributions $\nu_1,\dots,\nu_\ell$, where each ν_i is a distribution over size-k_i subsets of $\mathcal{V}_i$ and $\mathcal{V}_1 \sqcup \cdots \sqcup \mathcal{V}_\ell$ is a partition of $\mathcal{V}$. In this case, we require that for every $\xi \in \mathcal{X}_\tau(k)$,*

$$B_\tau = \bigoplus_{1\leq i\leq \ell : n_i \geq 1} \frac{k_i\cdot(k_i-1)}{k\cdot(k-1)}\cdot B_{\tau\cup\xi_{-i}},$$

     where $\xi_{-i} = \xi \setminus \mathcal{V}_i$ for all $1 \leq i \leq \ell$.
   - *If μ_τ^k does not factorize nontrivially, then B_τ directly satisfies both*

$$B_\tau \preceq \frac{k-1}{3k-1}\Pi_\tau \quad \text{and} \quad \mathbb{E}_{x\sim\pi_\tau}\left[B_{\tau\cup\{x\}}\right] \preceq B_\tau - \frac{k-1}{k-2}B_\tau\Pi_\tau^{-1}B_\tau. \tag{4.5}$$

Then $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \rho\left(\Pi_\tau^{-1}B_\tau\right)$ for all feasible τ with $|\tau| \leq n-2$, where ρ represents the spectral radius. In particular, μ is a $(\lambda_0,\dots,\lambda_{n-2})$-local spectral expander with $\lambda_k \leq \max_{\tau\in\mathcal{X}(k)}\rho(\Pi_\tau^{-1}B_\tau)$.

Proof. Apply Theorem 4.2.1 and Lemma 4.2.4 (see below) inductively. □

4.2.2 Independent Products of Probability Distributions

The following lemma gives a simple block-diagonal bounding matrix for a product distribution $\mu\otimes\nu$ given bounding matrices for the two smaller distributions μ,ν.

Lemma 4.2.4. *Let μ,ν be probability distributions on $\binom{\mathscr{U}}{n}$ and $\binom{\mathcal{V}}{m}$, respectively, where $\mathscr{U},\mathcal{V}$ are disjoint finite sets and $0 \leq n \leq |\mathscr{U}|, 0 \leq m \leq |\mathcal{V}|$ are nonnegative integers. Let $\mu\otimes\nu$ be the product distribution over $\binom{\mathscr{U}\sqcup\mathcal{V}}{n+m}$ defined by $(\mu\otimes\nu)(\tau\sqcup\sigma) = \mu(\tau)\cdot\nu(\sigma)$ for all $\tau\in\binom{\mathscr{U}}{n}, \sigma$ i. Then we have the identity*

$$\mathcal{Q}_{\mu\otimes\nu} - \frac{n+m}{n+m-1}\cdot\mathbf{1}\otimes(\mu\otimes\nu)_1 = \begin{bmatrix} \frac{n-1}{n+m-1}\left(\mathcal{Q}_\mu - \frac{n}{n-1}\cdot\mathbf{1}\otimes\mu_1\right) & 0 \\ 0 & \frac{m-1}{n+m-1}\left(\mathcal{Q}_\nu - \frac{m}{m-1}\cdot\mathbf{1}\otimes\nu_1\right)\end{bmatrix}.$$

In particular, if B_μ, B_ν are symmetric matrices satisfying

$$\operatorname{diag}(\mu_1)\cdot\mathcal{Q}_\mu - \frac{n}{n-1}\cdot\mu_1\otimes\mu_1 \preceq B_\mu$$
$$\operatorname{diag}(\nu_1)\cdot\mathcal{Q}_\nu - \frac{m}{m-1}\cdot\nu_1\otimes\nu_1 \preceq B_\nu$$

then

$$\operatorname{diag}((\mu\otimes\nu)_1)\cdot\mathcal{Q}_{\mu\otimes\nu} - \frac{n+m}{n+m-1}\cdot(\mu\otimes\nu)_1\otimes(\mu\otimes\nu)_1 \preceq \begin{bmatrix} \frac{n(n-1)}{(n+m)(n+m-1)}B_\mu & 0 \\ 0 & \frac{m(m-1)}{(n+m)(n+m-1)}B_\nu \end{bmatrix}.$$

Proof. This is an immediate consequence of Lemma 2.2.1, using Lemma 2.3.3 to translate between local walks and influence matrices. Note that

$$(\mu\otimes\nu)_1 = \left[\frac{n}{n+m}\mu_1, \frac{m}{n+m}\nu_1\right] \quad\text{and}\quad \operatorname{diag}\left((\mu\otimes\nu)_1\right) = \begin{bmatrix} \frac{n}{n+m}\operatorname{diag}(\mu_1) & 0 \\ 0 & \frac{m}{n+m}\operatorname{diag}(\nu_1) \end{bmatrix}.$$

These claims can also be proved via direct calculation. □

A natural example where such products of probability distributions arise is in the uniform distribution over proper vertex-colorings of a *disconnected graph*. Say $G=(V,E)$ is a graph with n vertices and suppose G consists of ℓ maximal connected components $G[U_1],\dots,G[U_\ell]$ where $U_1\sqcup\cdots\sqcup U_\ell$ is a partition of V. Since proper vertex-coloring of G is the disjoint union of proper vertex-colorings of each component $G[U_i]$, the uniform distribution over proper vertex-colorings of G is the independent product, over all $1\le i\le\ell$, of the uniform distribution over proper vertex-colorings of $G[U_i]$.

Suppose we associate a matrix $A_\tau\in\mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ to every nonempty feasible τ with $|\tau|\le n-2$ (e.g. a bounding matrix) and assume that for any $1\le i\le\ell$, when τ_{-i} and σ_{-i} are two arbitrary proper vertex-colorings of all components $G[U_j]$ for $j\neq i$, then $A_{\tau_{-i}}=A_{\sigma_{-i}}$. We associate a block-diagonal matrix

$$f_\otimes\left(\mathcal{X},\{A_\tau\}_{\emptyset\subsetneq\tau\in\mathcal{X}(\le n-3)}\right)\stackrel{\mathsf{def}}{=}\sum_{1\le i\le\ell:|U_i|\neq1}A_{\tau_{-i}}. \tag{4.6}$$

In the case of proper edge-colorings of G, which corresponds to vertex-colorings of the line graph of G, the above definition also makes sense, and we will use it in our analysis.

4.3 Vertex Coloring

Fix a positive integer q, a n-vertex graph $G=(V,E)$, and a function $\mathcal{L}$ which maps each vertex $v\in V$ to a subset $\mathcal{L}(v)$ of $[q]$. We call the elements of $[q]$ *colors* and we say $(G,\mathcal{L})$ a *(vertex-)list-coloring instance*. The colors in $\mathcal{L}(v)$ are the colors *available to* v. For every $u,v\in V$ and every color $c\in[q]$, we write $u\sim_c v$ when $u\sim v$ and $c\in\mathcal{L}(u)\cap\mathcal{L}(v)$.

Definition 21 (β-Extra (Vertex-)List-Coloring Instance)**.** *We say a (vertex-)list-coloring instance $(G,\mathcal{L})$ is a β-**extra instance** if for each $v\in V$, $|\mathcal{L}(v)|\ge\beta+\Delta(v)$.*

We call a configuration (or assignment) $\sigma:V\to[q]$ a *$\mathcal{L}$-(vertex-list-)coloring* of G if $\sigma(v)\in\mathcal{L}(v)$ for all $v\in V$; we say σ is *proper* if $\sigma(u)\neq\sigma(v)$ whenever $u\sim v$. When it is clear from context, we say σ is a proper coloring to mean it is a proper $\mathcal{L}$-vertex-list-coloring. We say τ is proper *partial coloring* on $U\subset V$ when it is a proper $\mathcal{L}|_U$-vertex-list-coloring for the induced subgraph $G[U]$.

When the graph G and color lists $\mathcal{L}$ are clear from context, we write $\mu=\mu_{G,\mathcal{L}}$ for the uniform distribution over proper $\mathcal{L}$-vertex-list-coloring of $G=(V,E)$. Since we may view proper list-colorings σ as sets of vertex-color pairs (v,c), which we denote by vc for convenience, μ may be cast as a probability distribution over $\binom{\mathscr{U}}{n}$, where $\mathscr{U}$ is the set of all vertex-color pairs. $\mathcal{X}=\mathcal{X}_{G,\mathcal{L}}$ is then the collection of all proper partial colorings, with $\mathcal{X}(k)$ being the collection of all partial colorings on exactly k vertices, for all $0\le k\le n$. For a proper partial coloring on $U\subset V$, a vertex $v\in V\setminus U$, and a color $c\in\mathcal{L}(v)$, define

$$p(vc\mid\tau)\stackrel{\mathsf{def}}{=}\Pr_{\sigma\sim\mu}\left[\sigma(v)=c\mid\sigma(u)=\tau(u),\forall u\in U\right].$$

Recall that π_τ is the shorthand notation we will use in this chapter for the marginal distribution μ_1^τ over all vertex-color pairs. The entries of π_τ are then given by the $p(vc \mid \tau)$ scaled by $\frac{1}{n}$.

Given a proper partial coloring τ, pinning (or conditioning on) τ yields a smaller list-coloring instance $(G_\tau, \mathcal{L}_\tau)$. Here, the graph $G_\tau = G[V_\tau]$ is the induced subgraph on the subset V_τ of vertices which are not colored (or pinned) by τ, while $\mathcal{L}_\tau : V_\tau \to [q]$ assigns to each v the list of *remaining* colors available to v after pinning τ. In other words, $\mathcal{L}_\tau(v) \subseteq \mathcal{L}(v)$ is the subset of colors $c \in \mathcal{L}(v)$ such that no neighbor of v is colored c under τ. We will also let $\mathscr{U}_\tau$ denote the set of all remaining valid vertex-color pairs which can be added to τ; in other words, $\mathscr{U}_\tau \stackrel{\text{def}}{=} \{uc : u \in V_\tau, c \in \mathcal{L}_\tau(u)\}$.

It will be convenient notationally to let $\ell_\tau(v) \stackrel{\text{def}}{=} |\mathcal{L}_\tau(v)|$ denote the size of the color lists. To roughly quantify how much two (usually neighboring) vertices constrain each other, we define $\ell_\tau(u,v) \stackrel{\text{def}}{=} |\mathcal{L}_\tau(u) \cap \mathcal{L}_\tau(v)|$. Generalizing our notation $u \sim_c v$, we write $u \sim_{\tau,c} v$ for vertices u, v if $u \sim v$ and $c \in \mathcal{L}_\tau(u) \cap \mathcal{L}_\tau(v)$. Finally, for $U \subseteq V \setminus V_\tau$, let $\tau \mid_U \stackrel{\text{def}}{=} \{vc \in \tau : v \in U\}$ denote the partial coloring obtained by *restricting* τ to U.

Our aim is to apply Theorem 4.2.3 to bound the second eigenvalue of the transition probability matrix of the local walks, and then use the local-to-global theorems (e.g. Theorems 2.3.1 and 10.0.1) to bound the mixing time of the Glauber dynamics.

4.3.1 Diagonal Matrix Bounds

To better demonstrate the essence of our approach, we start by restricting our attention to when the bounding matrices $\{B_\tau\}_{\tau \in \mathcal{X}(\leq n-2)}$ in Theorem 4.2.3 are *diagonal matrices*. Using diagonal matrix bounds, we analyze the Glauber dynamics for $(1+\epsilon)\Delta$-extra vertex-list-coloring instances.

Theorem 4.3.1. *Suppose $(G, \mathcal{L})$ is a $(1+\epsilon)\Delta$-extra vertex-list-coloring instance for an $0 < \epsilon \leq 1$ such that $\frac{\ln(\Delta)+2}{\Delta} \leq \frac{\epsilon^2}{40}$. Then for the uniform distribution $\mu = \mu_{G,\mathcal{L}}$ over proper $\mathcal{L}$-colorings of G, for every $2 \leq k \leq n$ and every (proper) partial $\mathcal{L}$-coloring τ on $n-k$ vertices, we have $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{5/2\epsilon}{k-1}$.*

Combined with the local-to-global theorems (e.g. Theorems 2.3.1 and 10.0.1), this yields a mixing time of $O(n \log n)$ for bounded-degree graphs and $n^{O(1/\epsilon)}$ in general, at least in this setting where we have at least $(1+\epsilon)\Delta$ additional colors available to each vertex. Again, we emphasize that this mixing result in itself is not new; a simple coupling argument can already recover $O(n \log n)$ mixing for $(\Delta+1)$-extra vertex-list-coloring instances. However, we will see later on how our proof technique can be used to obtain new mixing results for sampling proper list-edge-colorings which, to the best of our knowledge, cannot be recovered via simple coupling arguments.

To prove the above statement, our strategy is to first construct a diagonal matrix F_τ for each partial coloring τ on $n-2$ vertices such that $\Pi_\tau \mathcal{Q}_{\mu^\tau} \preceq 2\pi_\tau \pi_\tau^\top + \Pi_\tau F_\tau$, which is the base case of Theorem 4.2.3. Then, instead of guessing a good choice of F_τ for partial colorings τ on fewer than $n-2$ vertices, we apply Theorem 4.2.3 to $B_\tau = \frac{\Pi_\tau F_\tau}{k-1}$ to derive recursive conditions on the entries of the diagonal matrices F_τ. This then yields $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau)}{k-1}$, which is nice since the numerator is simply the maximum entry of F_τ.

To "solve" the base case, i.e. find F_τ partial colorings τ on $n-2$ vertices, we prove a general proposition that is also useful for approaches that employ non-diagonal matrix bounds.

Proposition 4.3.2 (Colorings on Two Vertices (Base Case))**.** *Let $(G, \mathcal{L})$ be a vertex-list-coloring instance. For every pair of distinct vertices $u, v \in V$ and every (proper) partial coloring τ on $V \setminus \{u, v\}$ (so that $V_\tau = \{u, v\}$), we have the following bounds:*

- *Case $uv \notin E$:* $\Pi_\tau \mathcal{Q}_{\mu^\tau} - 2\pi_\tau \pi_\tau^\top \preceq 0$.
- *Case $uv \in E$:*

$$\Pi_\tau \mathcal{Q}_{\mu^\tau} - 2\pi_\tau \pi_\tau^\top \preceq \sqrt{\Pi_\tau} \cdot \widetilde{B}_\tau \cdot \sqrt{\Pi_\tau}, \tag{4.7}$$

where $\widetilde{B}_\tau$ is a block diagonal matrix with a 2×2 block $\widetilde{B}_\tau^c$ indexed by $\{uc, vc\}$ for every color c given by

$$\widetilde{B}_\tau^c \stackrel{\text{def}}{=} \begin{pmatrix} \frac{1}{(\ell_\tau(u)-1)(\ell_\tau(v)-1)} & \frac{-1}{\sqrt{(\ell_\tau(u)-1)(\ell_\tau(v)-1)}} \\ \frac{-1}{\sqrt{(\ell_\tau(u)-1)(\ell_\tau(v)-1)}} & \frac{1}{(\ell_\tau(u)-1)(\ell_\tau(v)-1)} \end{pmatrix}, \tag{4.8}$$

and all other entries are 0.

Proof. In the case $uv \notin E$, there is no interaction and μ^τ is a product distribution; all correlations are zero. In this case, $\mathcal{Q}_{\mu^\tau} = \mathbf{1} \otimes \pi_\tau$ so $\Pi_\tau \mathcal{Q}_{\mu^\tau} - 2\pi_\tau \pi_\tau^\top = -\pi_\tau \pi_\tau^\top \preceq 0$. The nontrivial case is when $uv \in E$.

For convenience, we drop τ from all notation in the proof. Write $\widetilde{B} = \widetilde{B}_d + \widetilde{B}_o$, where $\widetilde{B}_d \stackrel{\text{def}}{=} \operatorname{diag}(\widetilde{B})$ and $\widetilde{B}_o \stackrel{\text{def}}{=} \operatorname{offdiag}(\widetilde{B})$. First observe that for $c \in \mathcal{L}(u)$,

$$\pi(uc) = \begin{cases} \frac{\ell(v)-1}{2(\ell(u)\ell(v)-\ell(u,v))}, & \text{if } c \in \mathcal{L}(u) \cap \mathcal{L}(v) \\ \frac{\ell(v)}{2(\ell(u)\ell(v)-\ell(u,v))}, & \text{otherwise} \end{cases} \tag{4.9}$$

and a similar identity holds for any $c \in \mathcal{L}(v)$. Also, observe that

$$\Pi \mathcal{Q}_{\mu^\tau} = \frac{J - J^u - J^v}{2(\ell(u)\ell(v) - \ell(u,v))} + \sqrt{\Pi} \cdot \widetilde{B}_o \cdot \sqrt{\Pi},$$

where J, J^u, J^v are the all-ones matrix, all-ones matrix on uc rows/columns, and all-ones matrix on vc rows/columns, respectively. So, subtracting $\widetilde{B}_o$ from both sides of Eq. (4.7) and multiplying by $2(\ell(u)\ell(v) - \ell(u,v))$ it is enough to show

$$J - J^u - J^v - 4(\ell(u)\ell(v) - \ell(u,v))\pi\pi^\top \preceq 2(\ell(u)\ell(v) - \ell(u,v))\sqrt{\Pi}\widetilde{B}_d\sqrt{\Pi} \stackrel{\text{def}}{=} N_d \tag{4.10}$$

Write $\ell = \ell(v)\mathbf{1}^u + \ell(u)\mathbf{1}^v$. Also, let $s \in \mathbb{R}^{\ell(u)+\ell(v)}$ where $s(xc) = 1$ if $c \in \mathcal{L}(u) \cap \mathcal{L}(v)$ and $s(xc) = 0$ otherwise, for $x \in \{u, v\}$. Then, by Fact 4.1.5 we can write,

$$\begin{aligned} 4(\ell(u)\ell(v) - \ell(u,v))\pi\pi^\top &= \frac{(\ell - s)(\ell - s)^\top}{\ell(u)\ell(v) - \ell(u,v)} \underbrace{\succeq}_{\textit{Fact 4.1.5}} \frac{\ell\ell^\top + ss^\top - \frac{1}{2}\ell\ell^\top - 2ss^\top}{\ell(u)\ell(v) - \ell(u,v)} \\ &\succeq \frac{\ell\ell^\top}{2\ell(u)\ell(v)} - \frac{ss^\top}{\ell(u)\ell(v) - \ell(u,v)} \end{aligned}$$

Plugging this into Eq. (4.10) it is enough to show that

$$J - J^u - J^v + \frac{ss^\top}{\ell(u)\ell(v) - \ell(u,v)} = \mathbf{1}^u\mathbf{1}^{v\top} + \mathbf{1}^v\mathbf{1}^{u\top} + \frac{ss^\top}{\ell(u)\ell(v) - \ell(u,v)} \preceq \frac{\ell\ell^\top}{2\ell(u)\ell(v)} + N_d \tag{4.11}$$

First, observe that by another application of Fact 4.1.5,

$$\ell(v)^2\mathbf{1}^u\mathbf{1}^{u\top} + \ell(u)^2\mathbf{1}^v\mathbf{1}^{v\top} \succeq \ell(u)\ell(v)(\mathbf{1}^u\mathbf{1}^{v\top} + \mathbf{1}^v\mathbf{1}^{u\top}).$$

So,

$$\frac{\ell\ell^\top}{2\ell(u)\ell(v)} = \frac{(\ell(v)\mathbf{1}^u + \ell(u)\mathbf{1}^v)(\ell(v)\mathbf{1}^u + \ell(u)\mathbf{1}^v)^\top}{2\ell(u)\ell(v)} \succeq \mathbf{1}^u\mathbf{1}^{v\top} + \mathbf{1}^v\mathbf{1}^{u\top}$$

Let $\mathsf{Id}^\cap \in \mathbb{R}^{(\ell(u)+\ell(v))\times(\ell(u)+\ell(v))}$ be the identity matrix only on entries xc, xc where $x \in \{u, v\}$ and $c \in \mathcal{L}(u) \cap \mathcal{L}(v)$. Finally, Eq. (4.11) simply follows from the fact that

$$\frac{ss^\top}{\ell(u)\ell(v) - \ell(u,v)} \preceq \frac{\ell(u,v)}{\ell(u)\ell(v) - \ell(u,v)}\mathsf{Id}^\cap \preceq N_d$$

where the first inequality uses that the only nonzero rows of $ss^\top$ correspond to a common color and the sum of the entries of any such row is exactly $\ell(u,v)$ and the last inequality uses that $\frac{\ell(u,v)}{\ell(u)\ell(v)-\ell(u,v)} \leq \frac{1}{\max\{\ell(u),\ell(v)\}-1}$ and that $N_d(uc, uc) = \frac{1}{\ell(u)-1}, N_d(vc, vc) = \frac{1}{\ell(v)-1}$ if $c \in \mathcal{L}(u) \cap \mathcal{L}(v)$ and it is zero otherwise. □

Note that in the above proposition, if the instance has β extra colors, then $\widetilde{B}_\tau \preceq \left(\frac{1}{\beta} + \frac{1}{\beta^2}\right)\mathsf{Id}^\tau$ where Id^τ is the identity matrix on $\mathscr{U}_\tau$, the set of remaining vertex-color pairs which can be added to τ. This gives us the base case diagonal matrices F_τ. Now, using Theorem 4.2.3, we derive a recursive set of sufficient conditions on the family $\{F_\tau\}_{\tau \in \mathcal{X}(\leq n-2)}$ to get $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau)}{k-1}$ for all (proper) partial colorings τ.

Proposition 4.3.3. *Let $(G, \mathcal{L})$ be a β-extra vertex-list-coloring instance. Suppose $\{F_\tau \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}\}_{\tau\in\mathcal{X}(\le n-2)}$ is a family of diagonal matrices, each supported on $\mathscr{U}_\tau\times\mathscr{U}_\tau$, such that $F_\tau = f_\otimes\left(\mathcal{X}_\tau, \{F_{\tau\cup\sigma}\}_{\emptyset\subsetneq\sigma\in\mathcal{X}_\tau(\le n-|\tau|-3)}\right)$ if G_τ is disconnected and otherwise:*

1. *For all partial colorings τ on $n-2$ vertices: $F_\tau(vc,vc) = \frac{1}{\beta} + \frac{1}{\beta^2}$ for all $vc \in \mathscr{U}_\tau$.*
2. *For all partial colorings τ on $n-k$ vertices where $k \ge 3$: $F_\tau \preceq \frac{(k-1)^2}{3k-1}\mathsf{Id}^\tau$ and for all $vc \in \mathscr{U}_\tau$,*

$$\sum_{uc'\in\mathscr{U}_{\tau+vc}} p\left(uc' \mid \tau\cup vc\right)\cdot F_{\tau+uc'}(vc,vc) \le (k-2)F_\tau(vc,vc) - F_\tau(vc,vc)^2.$$

Then, for all $2 \le k \le n$ and partial colorings τ on $n-k$ vertices, $\lambda_2(\mathcal{Q}_{\mu^\tau}) \le \frac{\rho(F_\tau)}{k-1}$.

Proof. We prove that the conditions of Theorem 4.2.3 hold for $B \overset{\mathsf{def}}{=} \frac{\Pi_\tau F_\tau}{k-1}$ for any partial coloring τ on $n-k$ vertices. The desired condition holds for any partial coloring τ on $n-2$ vertices by Proposition 4.3.2. Now, let $k \ge 3$. First assume that G_τ is disconnected with maximal connected components $G_\tau[U_1], \dots, G_\tau[U_\ell]$. We can write $\mu = \bigotimes_{i=1}^{\ell}\nu_i$, where ν_i is the uniform distribution over proper vertex-list-colorings on $G_\tau[U_i]$. For a $\xi \in \mathcal{X}_\tau(k)$, recall $\xi_{-i} = \xi \setminus \xi\mid_{U_i}$. Therefore, if we write $k_i = |U_i|$ for each $1 \le i \le \ell$, noting that $k = \sum_{i=1}^{\ell} k_i$, then

$$\begin{aligned}
\sum_{1\le i\le\ell : k_i\ge 1} \frac{k_i(k_i-1)}{k(k-1)} B_{\tau\cup\xi_{-i}} &= \sum_{1\le i\le\ell : k_i\ge 1} \frac{k_i(k_i-1)}{k(k-1)} \frac{\Pi_{\tau\cup\xi_{-i}}}{k_i} F_{\tau\cup\xi_{-i}} && \text{(Definition of } B_{\tau\cup\xi_{-i}}\text{)}\\
&= \sum_{1\le i\le\ell : k_i\ge 1} (\Pi_\tau)^{\mathscr{U}_{\tau\cup\xi_{-i}}} \frac{F_{\tau\cup\xi_{-i}}}{k-1} && (*)\\
&= \frac{\Pi_\tau F_\tau}{k-1} = B_\tau. && \text{(Definition of } F_\tau\text{)}
\end{aligned}$$

as desired; note that for $(*)$, we crucially use that the components $G_\tau[U_i], G_\tau[U_j]$ have no edges between them for $i \ne j$, since this implies that $\Pi_{\tau\cup\xi_{-i}}$ agrees with Π_τ on all vertex-colors pairs uc with that $u \in U_j$ for some $j \ne i$ (up to normalization by a constant).

Now, assume that G_τ is connected. Note that since each entry of F_τ is at most $\frac{(k-1)^2}{3k-1}$, we have $B_\tau \preceq \frac{k-1}{3k-1}\Pi_\tau$. Therefore, it only remains to show that $\mathbb{E}_{vc\sim\pi_\tau}\left[B_{\tau+vc}\right] \preceq B_\tau - \frac{k-1}{k-2}B_\tau\Pi_\tau^{-1}B_\tau$. This is equivalent to showing that

$$\Pi_\tau^{-1}\mathbb{E}_{uc'\sim\pi_\tau}\left[\Pi_{\tau+uc'}\frac{F_{\tau+uc'}}{k-2}\right] \preceq \frac{F_\tau}{k-1} - \frac{F_\tau^2}{(k-2)(k-1)}.$$

One can check that

$$\mathbb{E}_{uc'\sim\pi_\tau}\left[\Pi_\tau^{-1}\Pi_{\tau+uc'}\frac{F_{\tau+vc}}{k-2}\right](vc,vc) = \frac{\sum_{uc'\in\mathscr{U}_{\tau+vc}} p(uc'\mid\tau+vc)F_{\tau+uc'}(vc,vc)}{(k-1)(k-2)}.$$

Therefore, it is enough that

$$\frac{\sum_{uc'\in\mathscr{U}_{\tau+vc}} p(uc'\mid\tau+vc)F_{\tau+uc'}(vc,vc)}{(k-1)(k-2)} \le \frac{F_\tau(vc,vc)}{k-1} - \frac{F_\tau(vc,vc)^2}{(k-1)(k-2)},$$

which holds by assumption. □

Now, to complete the proof of Theorem 4.3.1 it only remains to find $\{F_\tau\}_{\tau\in\mathcal{X}(\le n-2)}$ that satisfies the above conditions. The proof can be found in the Section 4.6. Let us remark why we need the assumption $\beta > \Delta$ in this proof. Consider the worst case example, where G is a complete graph with $\Delta+1$ vertices. In that case, by symmetry, $F_\tau(vc,vc) = \frac{1}{\beta} + \frac{1}{\beta^2}$ for all partial colorings on $\Delta-1$ vertices, and every matrix F_τ is a multiple of identity on $\mathscr{U}_\tau\times\mathscr{U}_\tau$, i.e. $F_\tau = f(k)\cdot\mathsf{Id}^\tau$ for some function $f : [\Delta] \to \mathbb{R}_{\ge 0}$. So, the conditions on F_τ reduces to the following system of inequalities:

$$\begin{aligned}
(k-1)f(k-2) &\le (k-2)f(k-1) - f(k-1)^2 \quad \forall\, 3 \le k \le \Delta\\
f(1) &= \frac{1}{\beta} + \frac{1}{\beta^2}.
\end{aligned}$$

It is not hard to see that such a system does not have a solution up to $k = \Delta+1$ when $\beta \le \Delta$.

4.4 List-Colorings for Trees via Non-Diagonal Bounding Matrices

By allowing the bounding matrices $\{B_\tau\}_{\tau\in\mathcal{X}(\leq n-2)}$ in Theorem 4.2.3 to be non-diagonal matrices, one can hope to get a tighter result. In this section, for any constant $\epsilon > 0$ we analyze the Glauber dynamics for $\epsilon\Delta$-extra vertex-list-coloring instances when the graph is a *tree*.

Theorem 4.4.1. *Let $\epsilon > 0$ be an arbitrary constant and consider a $\epsilon\Delta$-extra vertex-list-coloring instance $(G, \mathcal{L})$ such that G is a tree and $\frac{\ln^2(\Delta)}{\Delta} \leq \frac{\epsilon^2}{100}$. Then for the uniform distribution $\mu = \mu_{G,\mathcal{L}}$ over proper $\mathcal{L}$-colorings of G, for every $2 \leq k \leq n$ and every partial $\mathcal{L}$-coloring τ on $n-k$ vertices, we have $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\frac{1}{20}+\frac{1}{\epsilon}}{k-1}$.*

For any $k \geq 2$ and partial $\mathcal{L}$-coloring τ on $n-k$ vertices, assume that B_τ is of the form

$$B_\tau = \Pi_\tau \frac{F_\tau}{k-1} + \sqrt{\Pi_\tau}\frac{A_\tau}{k-1}\sqrt{\Pi_\tau}, \tag{4.12}$$

for a diagonal matrix F_τ and a hollow matrix A_τ. The goal is to find F_τ and A_τ such that B_τ satisfies the conditions of Theorem 4.2.3. This is easily doable if $k = 2$ by Proposition 4.3.2. A natural approach is to define A_τ, for a partial $\mathcal{L}$-coloring τ on at most $n-3$ vertices, such that

$$\sqrt{\Pi_\tau}\frac{A_\tau}{k-1}\sqrt{\Pi_\tau} = \mathbb{E}_{vc\sim\pi_\tau}\sqrt{\Pi_{\tau+vc}}\frac{A_{\tau+vc}}{k-2}\sqrt{\Pi_{\tau+vc}}, \tag{4.13}$$

when G_τ is connected. Note that the following definition is not restricted to trees.

Definition 22 (Hollow Matrices A_τ for Vertex Colorings on Trees)**.** *Given a vertex-list-coloring instance $(G, \mathcal{L})$ and the associated uniform distribution $\mu = \mu_{G,\mathcal{L}}$ over proper $\mathcal{L}$-colorings of G, define $\{A_\tau\}_{\tau\in\mathcal{X}(\leq n-2)}$ as follows: Let $A_\tau \overset{\text{def}}{=} f_\otimes(\mathcal{X}, \{A_{\tau\cup\sigma}\}_{\emptyset\subsetneq\sigma\in X_\tau(n-|\tau|-3)})$ if G_τ is disconnected and otherwise,*

1. *For any partial $\mathcal{L}$-coloring τ on $n-2$ vertices, say $G_\tau = (\{u, v\}, \{uv\})$; define $A_\tau \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ to be a hollow block-diagonal matrix with a block for every color such that $A_\tau(uc, vc) = A_\tau(vc, uc) = \frac{-1}{\sqrt{(\ell_\tau(u)-1)(\ell_\tau(v)-1)}}$, for $c \in \mathcal{L}_\tau(u) \cap \mathcal{L}_\tau(v)$, and all other entries are 0.*
2. *For any partial $\mathcal{L}$-coloring τ on $n-k$ vertices where $k \geq 3$, let*

$$A_\tau \overset{\text{def}}{=} \frac{k-1}{k-2}\sqrt{\Pi_\tau^{-1}}\left(\mathbb{E}_{vc\sim\pi_\tau}\sqrt{\Pi_{\tau+vc}}A_{\tau+vc}\sqrt{\Pi_{\tau+vc}}\right)\sqrt{\Pi_\tau^{-1}}. \tag{4.14}$$

Observe that A_τ is symmetric and hollow. Furthermore, its nonzero entries correspond to $u \sim_{\tau,c} v$, and when G_τ is connected,

$$A_\tau(vc, uc) = \frac{1}{k-2}\sum_{wc'\in\mathscr{U}_\tau : w\neq u,v}\sqrt{p(wc' \mid \tau + uc)p(wc' \mid \tau + vc)}A_{\tau\cup wc'}(uc, vc).$$

We can bound entries of A_τ for any $\tau \in \mathcal{X}(\leq n-3)$ as follows.

Proposition 4.4.2 (Entry Bounds for A_τ)**.** *Consider a β-extra vertex-list-coloring instance $(G, \mathcal{L})$. For any partial $\mathcal{L}$-coloring τ on at most $n-2$ vertices, and every $uc, vc \in \mathscr{U}_\tau$ such that $u \sim_{\tau,c} v$, we have the bounds*

$$-\frac{1}{\beta} \leq A_\tau(uc, vc) \leq 0.$$

Proof. Fix τ. We prove the claim by induction on $k = |V_\tau|$. It clearly holds for $k = 2$ by definition.

For $k > 2$ and $uc, vc \in \mathscr{U}_\tau$, we have

$$\begin{aligned}
A_\tau(uc, vc) &= \frac{1}{k-2} \sum_{w \in V_\tau : w \neq u,v} \sum_{c' \in \mathcal{L}_\tau(w)} \sqrt{p(wc' \mid \tau + uc) p(wc' \mid \tau + vc)} A_{\tau + wc'}(vc, uc) \\
&\geq \frac{1}{k-2} \sum_{w \in V_\tau : w \neq u,v} \sum_{c' \in \mathcal{L}_\tau(w)} \sqrt{p(wc' \mid \tau + uc) p(wc' \mid \tau + vc)} \cdot \frac{-1}{\beta} \\
&\qquad \text{(Inductive Hypothesis)} \\
&\geq -\frac{1}{k-2} \cdot \frac{1}{\beta} \sum_{w \in V_\tau : w \neq u,v} \left(\sum_{c' \in \mathcal{L}_\tau(w)} p(wc' \mid \tau + vc) \right) \left(\sum_{c' \in \mathcal{L}_\tau(w)} p(wc' \mid \tau + uc) \right) \\
&\qquad \text{(Cauchy-Schwarz)} \\
&\geq -\frac{1}{\beta} \cdot \frac{1}{k-2} \sum_{w \in V_\tau : w \neq u,v} 1 \\
&= -\frac{1}{\beta}.
\end{aligned}$$

One can further see that $A_\tau(uc, vc) \leq 0$ follows from the induction hypothesis. □

Now, we apply Theorem 4.2.3 to derive sufficient conditions on the family $\{F_\tau\}_{\tau \in \mathcal{X}(\leq n-2)}$ to get $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau + A_\tau)}{k-1}$ for all partial $\mathcal{L}$-colorings τ on at most $n-2$ vertices.

Proposition 4.4.3. *Let $(G, \mathcal{L})$ be a β-extra vertex-list-coloring instance such that G is a tree, and let $\mu = \mu_{G,\mathcal{L}}$ be the uniform distribution over proper $\mathcal{L}$-colorings of G. Root the tree G at an arbitrary vertex $r \in V$. For $v \neq r$, we write $a(v)$ to denote the immediate ancestor of v, i.e. the parent of v. Let $\{F_\tau \in \mathbb{R}^{\mathscr{U} \times \mathscr{U}}\}_{\tau \in \mathcal{X}(\leq n-2)}$ be a family of diagonal matrices supported on $\mathscr{U}_\tau \times \mathscr{U}_\tau$ such that $F_\tau = f_\otimes(X_\tau, \{F_{\tau \cup \sigma}\}_{\emptyset \subsetneq \sigma \in X_\tau(\leq n - |\tau| - 3)})$ if G_τ is disconnected and otherwise:*

1. *For all partial colorings τ on $n-2$ vertices: F_τ is defined as $F_\tau(vc, vc) = \frac{1}{\beta^2}$ for $vc \in \mathscr{U}_\tau$.*
2. *For all partial colorings τ on $n-k$ vertices where $k \geq 3$: $F_\tau \preceq \left(\frac{(k-1)^2}{3k-1} - \frac{1}{\beta} \right) \mathsf{Id}^\tau$, and for all $vc \in \mathscr{U}_\tau$,*

$$\sum_{uc' \in \mathscr{U}_{\tau + vc}} p(uc' \mid \tau + vc) F_{\tau + uc'}(vc, vc) \leq (k-2) F_\tau(vc, vc) - 2F_\tau(vc, vc)^2 - \mathfrak{g}_\tau(vc), \quad (4.15)$$

where $\mathfrak{g}_\tau(vc) = \frac{4\Delta_\tau(v)}{\beta^2}$ if v is the (induced) root of the rooted subtree G_τ, and $\mathfrak{g}_\tau(vc) = \frac{4(\Delta_\tau(v) + \Delta_\tau(a(v)) - 1)}{\beta^2}$ otherwise.

Then, for all $k \geq 2$ and proper $\mathcal{L}$-colorings τ on $n-k$ vertices, $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau + A_\tau)}{k-1}$, where A_τ is defined in Definition 22.

Proof. We prove that the conditions of Theorem 4.2.3 hold for $B_\tau \stackrel{\text{def}}{=} \Pi_\tau \frac{F_\tau}{k-1} + \sqrt{\Pi_\tau} \frac{A_\tau}{k-1} \sqrt{\Pi_\tau}$ for $\tau \in \mathcal{X}(\leq n-3)$. Note that the desired condition holds for any proper $\mathcal{L}$-coloring τ on exactly $n-2$ vertices. Now, take $k \geq 2$. Assume G_τ is disconnected. Using the definition of A_τ and our assumption about F_τ, the proof of this case is similar to what we argued in Proposition 4.3.3.

Now, assume that G_τ is connected. Note that by Proposition 4.4.2, the absolute value of every off-diagonal entry of A_τ is at most $\frac{1}{\beta}$ and that there are at most $k-1$ nonzero entries per row. Therefore, $\sqrt{\Pi_\tau} A_\tau \sqrt{\Pi_\tau} \preceq \frac{1}{\beta} \Pi_\tau$. Since each entry of F_τ is at most $\frac{(k-1)^2}{3k-1} - \frac{1}{\beta}$, we have $B_\tau \preceq \frac{k-1}{3k-1} \Pi_\tau$. Therefore, it only remains to show that $\mathbb{E}_{vc \sim \pi_\tau}[B_{\tau + vc}] \preceq B_\tau - \frac{k-1}{k-2} B_\tau \Pi_\tau^{-1} B_\tau$. This is equivalent to showing that

$$\sqrt{\Pi_\tau^{-1}} \mathbb{E}_{vc \sim \pi_\tau} \left[\Pi_{\tau + vc} \frac{F_{\tau + vc}}{k-2} + \sqrt{\Pi_{\tau + vc}} \frac{A_{\tau + vc}}{k-2} \sqrt{\Pi_{\tau + vc}} \right] \sqrt{\Pi_\tau^{-1}} \preceq \frac{F_\tau}{k-1} + \frac{A_\tau}{k-1} - \frac{(F_\tau + A_\tau)^2}{(k-1)(k-2)}. \quad (4.16)$$

We starting by proving an upper bound on A_τ^2. Define $A_{\tau,\mathsf{even}}(uc,vc) = A_{\tau,\mathsf{even}}(vc,uc) \stackrel{\mathsf{def}}{=} A(uc,vc)$ for all $uc, vc \in \mathscr{U}_\tau$ such that v is at an *even distance* from the root of G_τ and $a(u) = v$, and let other entries be 0. Define $A_{\tau,\mathsf{odd}} \stackrel{\mathsf{def}}{=} A_{\tau,\mathsf{even}} - A_{\tau,\mathsf{odd}}$. By Fact 4.1.5,

$$A_\tau^2 \preceq 2A_{\tau,\mathsf{even}}^2 + 2A_{\tau,\mathsf{odd}}^2.$$

Furthermore, $A_{\tau,\mathsf{even}}^2(uc,vc) = A_{\tau,\mathsf{even}}^2(vc,uc) \neq 0$ only if $u = v$ and v is at an even distance from the root, or when $a(u) = a(v)$ and u, v are at an odd distance from the root. A similar fact holds for $A_{\tau,\mathsf{odd}}^2$. Therefore, if we let $\mathfrak{g}_\tau(wc) = 0$ for $wc \notin \mathscr{U}_\tau$, we get

$$4\left(A_{\tau,\mathsf{even}}^2 + A_{\tau,\mathsf{odd}}^2\right) \preceq \mathrm{diag}(\mathfrak{g}_\tau),$$

where we applied Fact 4.1.4, and used Proposition 4.4.2 to bound the absolute value of the entries of $A_{\tau,\mathsf{even}}^2$ and $A_{\tau,\mathsf{odd}}^2$. Therefore, by Fact 4.1.5, the right-hand side of Eq. (4.16) is bounded by

$$\frac{F_\tau}{k-1} + \frac{A_\tau}{k-1} - \frac{2F_\tau^2 + 2A_\tau^2}{(k-1)(k-2)} \succeq \frac{F_\tau}{k-1} + \frac{A_\tau}{k-1} - \frac{2F_\tau^2 + \mathrm{diag}(\mathfrak{g}_\tau)}{(k-1)(k-2)}. \tag{4.17}$$

On the other hand, by the definition of A_τ (see Eq. (4.14)), the left-hand side of Eq. (4.16) is equal to

$$\mathbb{E}_{uc' \sim \pi_\tau}\left[\Pi_\tau^{-1}\Pi_{\tau+uc'}\frac{F_{\tau+uc'}}{k-2}\right] + \frac{A_\tau}{k-1}.$$

Furthermore, for $vc \in \mathscr{U}_\tau$,

$$\mathbb{E}_{uc' \sim \pi_\tau}\left[\Pi_\tau^{-1}\Pi_{\tau+uc'}\frac{F_{\tau+uc'}}{k-2}\right](vc,vc) = \frac{\sum_{uc' \in \mathscr{U}_{\tau+vc}} p(uc' \mid \tau + vc) F_{\tau+uc'}(vc,vc)}{(k-1)(k-2)}.$$

Combining this with Eq. (4.17), the desired inequality in Eq. (4.16) follows from the assumption (see Eq. (4.15)). □

Finally, with this in hand, we prove Theorem 4.4.1 similar to what we did for Theorem 4.3.1. The proof can be found in Section 4.6.

4.5 Edge Colorings

In this section, we build all of the techniques needed to prove the flagship result of this chapter on sampling edge colorings. First, we set up some notation, most of which will be highly analogous to our notation for vertex colorings. Fix a positive integer q, a n-vertex graph $G = (V, E)$ with m edges, and a function $\mathcal{L}$ which maps each *edge* $e \in E$ to a subset $\mathcal{L}(e) \subseteq [q]$ of colors. Since we only consider edge colorings in this section, we overload notation call the pair $(G, \mathcal{L})$ an *edge-list-coloring instance*.

For a vertex v and an edge e, we write $e \sim_c v$ when $e \sim v$ (i.e. e is incident to v) and $c \in \mathcal{L}(e)$. Furthermore, for any $e, f \in E$, we write $e \sim_c f$ when $e \sim f$ (i.e. e, f share an endpoint) and $c \in \mathcal{L}(e) \cap \mathcal{L}(f)$.

Definition 23 (β-Extra (Edge-)List-Coloring Instance)**.** *We say an (edge-)list-coloring instance $(G, \mathcal{L})$ is a β-**extra instance** if for each $e \in E$, $|\mathcal{L}(e)| \geq \beta + \Delta_G(e)$.*

An assignment $\sigma : E \to [q]$ is a *$\mathcal{L}$-(edge-list-)coloring* of G if $\sigma(e) \in \mathcal{L}(e)$ for all $e \in E$; we say σ is *proper* if $\sigma(e) \neq \sigma(f)$ whenever $e \sim f$. When it is clear from context, we say σ is a proper coloring to mean it is a proper $\mathcal{L}$-edge-list-coloring. We say τ is proper *partial coloring* on $F \subset E$ when it is a proper $\mathcal{L}\,|_F$-edge-list-coloring for the subgraph (V, F). Again, we may view a proper coloring as a set of edge-color pairs (e, c), which we denote by ec for simplicity of notation. We denote the uniform distribution over proper $\mathcal{L}$-edge-list-colorings of G by $\mu = \mu_{G,\mathcal{L}}$, which may be viewed as a probability distribution over $\binom{\mathscr{U}}{m}$ where $\mathscr{U}$ is the set of all edge-color pairs. $\mathcal{X} = \mathcal{X}_{G,\mathcal{L}}$ is then the collection of all proper partial colorings, with $\mathcal{X}(k)$ being the collection of partial colorings on exactly k vertices, for all $0 \leq k \leq m$. For a proper partial coloring on $F \subset E$ and $e \in E \setminus F$, define

$$p(ec \mid \tau) \stackrel{\mathsf{def}}{=} \Pr_{\sigma \sim \mu}\left[\sigma(e) = c \mid \sigma(f) = \tau(f), \forall f \in F\right].$$

As before, our aim is to apply Theorem 4.2.3 to bound the second eigenvalue of the transition probability matrix of the local walks, and then apply the local-to-global theorems (e.g. Theorems 2.3.1 and 10.0.1) to get a bound for the mixing time of the Glauber dynamics. The following is the main theorem of this section.

Theorem 4.5.1. *Let $(G, \mathcal{L})$ be a $\left(\frac{4}{3}+4\epsilon\right)\Delta$-extra edge-list-coloring instance for some $0 < \epsilon \le \frac{1}{10}$ such that $\frac{\ln^2(\Delta)}{\Delta} \le \frac{\epsilon^3}{15}$. Then for the uniform distribution $\mu = \mu_{G,\mathcal{L}}$ over proper $\mathcal{L}$-edge-colorings of G, for every $2 \le k \le m$ and every (proper) partial $\mathcal{L}$-edge-coloring τ on $m-k$ edges, we have $\lambda_2(\mathcal{Q}_{\mu^\tau}) \le \frac{\epsilon+\frac{1}{\epsilon}}{k-1}$.*

We remark that our analysis here is not tight and we expect that the factor of $4/3$ can be improved with a more careful analysis.

We proceed by introducing some additional notations and definitions; again, most of these are straightforward analogs of the notations and definitions used for vertex-colorings. Given a proper partial coloring $\tau \in \mathcal{X}$, let $E_\tau \subseteq E$ be the set of edges which remain uncolored under τ. Let $G_\tau = (V, E_\tau)$ and $\Delta_\tau(\cdot)$ be the degree function of G_τ. Similarly, if $e = \{u, v\}$, define $\Delta_\tau(e)$ to be number of edges in G_τ that share an endpoint with e, i.e. $\Delta_\tau(e) \stackrel{\text{def}}{=} \Delta_\tau(u) + \Delta_\tau(v) - 2$. We define $\mathcal{L}_\tau(e) \stackrel{\text{def}}{=} \{c \in \mathcal{L}(e) : ec \in \mathscr{U}_\tau\}$ to be the set of *remaining* colors available to e after pinning τ. We write $\mathscr{U}_\tau$ to denote the set of all remaining valid edge-color pairs which can be added to τ; in other words, $\mathscr{U}_\tau \stackrel{\text{def}}{=} \{ec : e \in E_\tau, c \in \mathcal{L}_\tau(e)\}$. For convenience, define $\ell_\tau(e) \stackrel{\text{def}}{=} |\mathcal{L}_\tau(e)|$ and $\ell_\tau(e,f) \stackrel{\text{def}}{=} |\mathcal{L}_\tau(e) \cap \mathcal{L}_\tau(f)|$. Furthermore, we write $e \sim_{\tau,c} v$ when $e \sim v$ and $c \in \mathcal{L}_\tau(e)$. Similarly, we write $e \sim_{\tau,c} f$ to mean $e \sim f$ and $c \in \mathcal{L}_\tau(e) \cap \mathcal{L}_\tau(f)$.

Finally, for any matrix $B \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$, define the *restriction* $B^v \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ *of* B *to* $v \in V$ as $B^v(ec, fc) \stackrel{\text{def}}{=} B(ec, fc)$ for any $e, f \sim v$, and 0 on all other entries. Similarly, define the *restriction* $B^c \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ *of* B *to* $c \in [q]$ as $B^c(ec, fc) \stackrel{\text{def}}{=} B(ec, fc)$ for all $e \sim_c f$, and 0 on all other entries.

Now, similar to our approach to vertex-coloring for trees, for any $k \ge 2$ and any partial coloring $\tau \in \mathcal{X}(m-k)$ on $m-k$ edges, assume that B_τ is of the form

$$B_\tau = \Pi_\tau \frac{F_\tau}{k-1} + \sqrt{\Pi_\tau}\frac{A_\tau}{k-1}\sqrt{\Pi_\tau}, \tag{4.18}$$

for a diagonal matrix F_τ and a hollow matrix A_τ. The goal is again to find F_τ and A_τ such that B_τ satisfies the conditions of Theorem 4.2.3. Again, for $k = 2$, Proposition 4.3.2 gives us such matrices. However, for $k \ge 3$, as opposed to what we did for vertex-colorings of trees, we let $\sqrt{\Pi_\tau}\frac{A_\tau}{k-1}\sqrt{\Pi_\tau}$ deviate from $\mathbb{E}_{ec\sim\pi_\tau}\left[\sqrt{\Pi_{\tau+ec}}\frac{A_{\tau+ec}}{k-2}\sqrt{\Pi_{\tau+ec}}\right]$ in order to control the growth of F_τ.

Definition 24 (Hollow Matrices $A_{\tau,\epsilon}$ for Edge Colorings)**.** *Fix $\epsilon > 0$. Given a β-extra edge-list-coloring instance $(G, \mathcal{L})$ with the associated uniform distribution $\mu = \mu_{G,\mathcal{L}}$ over proper $\mathcal{L}$-edge-colorings of G, define $\{A_{\tau,\epsilon}\}_{\tau\in\mathcal{X}(\le m-2)}$ as follows: Let $A_{\tau,\epsilon} \stackrel{\text{def}}{=} f_\otimes(\mathcal{X}, \{A_{\tau\cup\sigma,\epsilon}\}_{\emptyset\subsetneq\sigma\in X_\tau(m-|\tau|-3)})$ if the line graph of G_τ is disconnected and otherwise:*

1. *For any partial $\mathcal{L}$-edge-coloring on $m-2$ edges, define $A_{\tau,\epsilon} \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ to be a hollow block diagonal matrix with a block for every color such that*

$$A_{\tau,\epsilon}(ec, fc) = A_{\tau,\epsilon}(fc, ec) \stackrel{\text{def}}{=} -\frac{1}{\sqrt{(\ell_\tau(e)-1)(\ell_\tau(f)-1)}},$$

   for $e, f \in E_\tau$ and any $c \in \mathcal{L}_\tau(e) \cap \mathcal{L}_\tau(f)$, and all other entries are 0.

2. *For any partial $\mathcal{L}$-coloring τ on $m-k$ edges where $k \ge 3$, let*

$$A_{\tau,\epsilon} \stackrel{\text{def}}{=} \bar{A}_{\tau,\epsilon} + \frac{\text{offdiag}(S_{\tau,\epsilon})}{k-2}, \tag{4.19}$$

   where $\bar{A}_{\tau,\epsilon}$ and $S_{\tau,\epsilon}$ are defined as follows:

$$\bar{A}_{\tau,\epsilon} \stackrel{\text{def}}{=} \frac{k-1}{k-2}\sqrt{\Pi_\tau^{-1}}\left(\mathbb{E}_{gc\sim\pi_\tau}\sqrt{\Pi_{\tau+gc}}A_{\tau+gc,\epsilon}\sqrt{\Pi_{\tau+gc}}\right)\sqrt{\Pi_\tau^{-1}}, \tag{4.20}$$

$$S^v_{\tau,\epsilon} \stackrel{\text{def}}{=} \begin{cases} 4(1+\epsilon)\left(\left(\bar{A}^{+,v}_{\tau,\epsilon}\right)^2 + \left(\bar{A}^{-,v}_{\tau,\epsilon}\right)^2\right), & \text{if } \Delta_\tau(v) \le \frac{\beta}{4(1+\epsilon)} \\ 2(1+\epsilon)\left(\bar{A}^{v}_{\tau,\epsilon}\right)^2, & \text{otherwise} \end{cases} \tag{4.21}$$

and $S_{\tau,\epsilon} = \sum_{v\in V} S^v_{\tau,\epsilon}$.

Observe that all three matrices $\bar{A}_{\tau,\epsilon}, S_{\tau,\epsilon}, A_{\tau,\epsilon}$ are symmetric and hollow. Furthermore, the nonzero entries of these matrices correspond to $e, f \sim_{\tau,c} v$, for some $v \in V$ and when the line graph of G_τ is connected,

$$\bar{A}_{\tau,\epsilon}(ec, fc) = \frac{1}{k-2} \sum_{gc' \in \mathscr{U}_\tau : g \neq e,f} \sqrt{p(gc' \mid \tau + ec)p(gc' \mid \tau + fc)} A_{\tau+gc',\epsilon}(ec, fc). \tag{4.22}$$

When it is clear from context, we drop ϵ from the subscripts of the matrices defined above.

We begin our analysis by trying to establish some control on the entries of these matrices, much like how we first proved Proposition 4.4.2 in our analysis of vertex colorings on trees. However, the case of edge colorings is much more involved, as evidenced by the significantly more complicated construction of $A_{\tau,\epsilon}$. We will need the following lemma.

Lemma 4.5.2. *For any partial coloring $\tau \in \mathcal{X}$ on at most $m-2$ edges and every $ec, fc \in \mathscr{U}_\tau$, if the line graph of G_τ is connected, then*

$$\operatorname*{avg}_{g\in E_\tau : g\neq e,f} \min_{c'\in\mathcal{L}_\tau(g)} A_{\tau+gc',\epsilon}(ec, fc) \le \bar{A}_{\tau,\epsilon}(ec, fc) \le \operatorname*{avg}_{g\in E_\tau : g\neq e,f} \max_{c'\in\mathcal{L}_\tau(g)} A_{\tau+gc',\epsilon}(ec, fc)$$

Proof. For $k \ge 2$, let $\tau \in \mathcal{X}(m-k)$ be a partial $\mathcal{L}$-edge-coloring on $m-k$ edges. The claim clearly holds for $k = 2$ by definition. For $k > 2$ and $ec, fc \in \mathscr{U}_\tau$, we have

$$\begin{aligned}
&\bar{A}_\tau(ec, fc) \\
&= \frac{1}{k-2} \sum_{g\in E_\tau : g\neq e,f} \sum_{c'\in\mathcal{L}_\tau(g)} \sqrt{p(gc' \mid \tau + ec)p(gc' \mid \tau + fc)} A_{\tau+gc'}(ec, fc) \\
&\le \frac{1}{k-2} \sum_{g\in E_\tau : g\neq e,f} \left(\sum_{c'\in\mathcal{L}_\tau(g)} \sqrt{p(gc' \mid \tau + ec)p(gc' \mid \tau + fc)}\right) \cdot \left(\max_{c'\in\mathcal{L}_\tau(g)} A_{\tau+gc'}(ec, fc)\right) \\
&\le \frac{1}{k-2} \sum_{g\in E_\tau : g\neq e,f} \left(\left(\sum_{c'\in\mathcal{L}_\tau(e)} p(gc' \mid \tau + ec)\right)\left(\sum_{c'\in\mathcal{L}_\tau(g)} p(gc' \mid \tau + fc)\right)\right)\left(\max_{c'\in\mathcal{L}_\tau(g)} A_{\tau+gc'}(ec, fc)\right) \\
&\qquad\qquad \text{(Cauchy-Schwarz)} \\
&\le \frac{1}{k-2} \sum_{g\in E_\tau : g\neq e,f} \max_{c'\in\mathcal{L}_\tau(g)} A_{\tau+gc'}(ec, fc) \\
&= \operatorname*{avg}_{g\in E_\tau : g\neq e,f} \max_{c'\in\mathcal{L}_\tau(g)} A_{\tau+gc'}(ec, fc).
\end{aligned}$$

The other side of the inequality follows from a similar argument. □

As alluded to earlier, in order to find diagonal matrices $\{F_\tau\}_{\tau\in\mathcal{X}(\le m-2)}$ such that $\{B_\tau\}_{\tau\in\mathcal{X}(\le m-2)}$, as defined by Eq. (4.18), satisfies the conditions of Theorem 4.2.3, we would need to prove some bounds on the entries of $\{A_{\tau,\epsilon}\}_{\tau\in\mathcal{X}(\le m-2)}$ and $\{S_{\tau,\epsilon}\}_{\tau\in\mathcal{X}(\le m-2)}$. This is what we do next. The next two claims combined form an analog of Proposition 4.4.2 in our analysis for vertex colorings on trees.

Proposition 4.5.3. *Suppose $(G, \mathcal{L})$ is a β-extra edge-list-coloring instance, where $\beta \ge \left(\frac{4}{3} + 4\epsilon\right)\Delta$ for an $0 < \epsilon \le \frac{1}{10}$ such that $2\epsilon^{-2} \le \Delta$. For any partial coloring $\tau \in \mathcal{X}$ on at most $m-2$ edges, the matrix A_τ defined in Definition 24 satisfies the following: For every vertex $v \in V$, every color c, and every $e, f \sim_{\tau,c} v$,*

(i) if $\Delta_\tau(v) \le \frac{\beta}{4(1+\epsilon)}$, then $-\frac{1}{\beta} \le A_{\tau,\epsilon}(ec, fc) \le 4(1+\epsilon)\frac{\Delta_\tau(v)-2}{\beta^2}$,

(ii) otherwise, if $\Delta_\tau(v) \ge \frac{\beta}{4(1+\epsilon)}$, then $|A_{\tau,\epsilon}(ec, fc)| \le \frac{1}{(3/2)\beta - 2(1+2\epsilon)\Delta_\tau(v)}$.

Proof. Fix a vertex v. We prove the claim inductively for every pair of edges incident to v.

Case Item (i): Let τ be any partial coloring on $m-k$ edges for $k \geq 2$. We go by induction on $\Delta_\tau(v)+k$. We start with the base case, which is when $\Delta_\tau(v)+k=4$, i.e. $\Delta_\tau(v)=k=2$. It is easy to see that, for any color c and $e, f \sim_{\tau,c} v$, we have $-\frac{1}{\beta} \leq A_\tau(ec, fc) \leq 0$, by definition. Now, we prove the claim for $k \geq 2$ and $2 \leq \Delta_\tau(v) \leq \frac{\beta}{4(1+\epsilon)}$ such that $k+\Delta_\tau(v) \geq 5$. If the line graph of G_τ is not connected, then the statement trivially holds. Otherwise, by Lemma 4.5.2, for any color c and any $e, f \sim_{\tau,c} v$, we can write

$$\begin{aligned}
\bar{A}_\tau(ec, fc) &\leq \frac{\Delta_\tau(v)-2}{k-2} \max_{gc' \in \mathscr{U}_\tau : g \sim v, g \neq e,f} A_{\tau+gc'}(ec, fc) + \frac{k-\Delta_\tau(v)}{k-2} \max_{gc' \in \mathscr{U}_\tau : g \not\sim v} A_{\tau+gc'}(ec, fc) \\
&\leq \frac{\Delta_\tau(v)-2}{k-2} 4(1+\epsilon)\frac{\Delta_\tau(v)-3}{\beta^2} + \frac{k-\Delta_\tau(v)}{k-2} 4(1+\epsilon)\frac{\Delta_\tau(v)-2}{\beta^2} \\
&\qquad \text{(Inductive Hypothesis)} \\
&= \frac{4(1+\epsilon)(\Delta_\tau(v)-2)(k-3)}{\beta^2(k-2)} \qquad (4.23) \\
&\leq \frac{1}{\beta}, \qquad (4.24)
\end{aligned}$$

where in the final step we use $\Delta_\tau(v) \leq \frac{\beta}{4(1+\epsilon)}$. Similarly,

$$\bar{A}_\tau(ec, fc) \geq -\frac{\Delta_\tau(v)-2}{k-2} \min_{\substack{gc' \in \mathscr{U}_\tau \\ g \sim v, g \neq e,f}} A_{\tau+gc'}(ec, fc) - \frac{k-\Delta_\tau(v)}{k-2} \min_{\substack{gc' \in \mathscr{U}_\tau \\ g \not\sim v, g \neq e,f}} A_{\tau+gc'}(ec, fc) \geq -\frac{1}{\beta}. \qquad (4.25)$$

Therefore, by Eq. (4.21)

$$\begin{aligned}
S_\tau(ec, fc) &= 4(1+\epsilon) \sum_{\substack{g \sim_{\tau,c} v \\ g \neq e,f}} \left[\bar{A}^+_\tau(ec, gc)\bar{A}^+_\tau(gc, fc) + \bar{A}^-_\tau(ec, gc)\bar{A}^-_\tau(gc, fc)\right] \\
&\leq \frac{4(1+\epsilon)(\Delta_\tau(v)-2)}{\beta^2}, \qquad (4.26)
\end{aligned}$$

where the last inequality follows by Eqs. (4.24) and (4.25) and that v has at most $\Delta_\tau(v)-2$ incident edges that can be colored by c, other than e, f. So Eqs. (4.25) and (4.26), we get $A_\tau(ec, fc) = \bar{A}_\tau(ec, fc) + \frac{S_\tau(ec,fc)}{k-2} \geq -\frac{1}{\beta}$. Similarly, Eqs. (4.24) and (4.26) gives

$$A_\tau(ec, fc) \leq \frac{4(1+\epsilon)(\Delta_\tau(v)-2)(k-3)}{\beta^2(k-2)} + \frac{4(1+\epsilon)(\Delta_\tau(v)-2)}{k-2}\frac{1}{\beta^2} = 4(1+\epsilon)\frac{\Delta_\tau(v)-2}{\beta^2}.$$

Case Item (ii): Let τ be any partial coloring on $m-k$ edges for $k \geq 2$. We again go by induction on $\Delta_\tau(v)+k$. The base case is when $\Delta_\tau(v)=k=\frac{\beta}{4(1+\epsilon)}$, which we already proved in case Item (i) (note that we always have $k \geq \Delta_\tau(v)$). Now, we prove the claim for $\Delta_\tau(v) > \frac{\beta}{4(1+\epsilon)}$ (and $k \geq \Delta_\tau(v)$). If the line graph of G_τ is disconnected, then the statement trivially holds. Otherwise, for any color c and any $e, f \sim_{\tau,c} v$, we can write

$$\begin{aligned}
\left|\bar{A}_\tau(ec, fc)\right| &= \frac{\Delta_\tau(v)-2}{k-2} \max_{gc' \in \mathscr{U}_\tau : g \sim v, g \neq e,f} \left|A_{\tau+gc'}(ec, fc)\right| + \frac{k-\Delta_\tau(v)}{k-2} \max_{gc' \in \mathscr{U}_\tau : g \not\sim v} \left|A_{\tau+gc'}(ec, fc)\right| \\
&\leq \frac{\Delta_\tau(v)-2}{k-2}\frac{1}{1.5\beta - 2(1+2\epsilon)(\Delta_\tau(v)-1)} + \frac{k-\Delta_\tau(v)}{k-2}\frac{1}{1.5\beta - 2(1+2\epsilon)\Delta_\tau(v)} \qquad (4.27) \\
&\leq \frac{1}{1.5\beta - 2(1+2\epsilon)\Delta_\tau(v)}, \qquad (4.28)
\end{aligned}$$

where the second to last inequality follows by the inductive hypothesis. Furthermore, by Eq. (4.21),

$$|S_\tau(ec, fc)| = 2(1+\epsilon)\left|\sum_{g \sim_{\tau,c} v : g \neq e,f} \bar{A}_\tau(ec, gc)\bar{A}_\tau(gc, fc)\right| \leq 2(1+\epsilon)\frac{\Delta_\tau(v)-2}{(1.5\beta - 2(1+2\epsilon)\Delta_\tau(v))^2}.$$

where the inequality follows by Eq. (4.28) and that v has at most $\Delta_\tau(v)-2$ incident edges other than e, f that can be colored by c. Recall $A_\tau = \bar{A}_\tau + \frac{S_\tau}{k-2}$. So, the above inequality with Eq. (4.27) gives

$$\begin{aligned}|A_\tau(ec,fc)| &\le \frac{\Delta_\tau(v)-2}{k-2}\cdot\frac{1}{1.5\beta-2(1+2\epsilon)(\Delta_\tau(v)-1)} + \frac{k-\Delta_\tau(v)}{k-2}\cdot\frac{1}{1.5\beta-2(1+2\epsilon)\Delta_\tau(v)}\\ &\quad + \frac{2(1+\epsilon)}{k-2}\cdot\frac{\Delta_\tau(v)-2}{(1.5\beta-2(1+2\epsilon)\Delta_\tau(v))^2}\\ &\le \frac{1}{1.5\beta-2(1+2\epsilon)\Delta_\tau(v)},\end{aligned}$$

where in the second inequality we used that $\epsilon \le \frac{1}{10}, \Delta \ge 2\epsilon^{-2}$, and that

$$1.5\beta - 2(1+2\epsilon)\Delta_\tau(v) \ge 1.5\left(\frac{4}{3}+4\epsilon\right)\Delta - 2(1+2\epsilon)\Delta = 6\epsilon\Delta - 4\epsilon\Delta \ge 0.$$

□

Corollary 4.5.4. *Fix a β-extra edge-list-coloring instance $(G,\mathcal{L})$, where $\beta = (\frac{4}{3}+4\epsilon)\Delta$ for an $0<\epsilon\le\frac{1}{10}$ such that $2\epsilon^{-2}\le\Delta$. Let $\tau\in\mathcal{X}$ be a partial $\mathcal{L}$-edge-coloring on at most $m-k$ edges for k to be specified in a moment. Then the following bounds hold.*

(i) For every $k\ge 2$, $v\in V$, and $e,f\sim_{\tau,c} v$, $|A_{\tau,\epsilon}(ec,fc)|\le\frac{1}{2\epsilon\Delta}\le\frac{\epsilon}{4}$.

(ii) For every $k\ge 3$, $v\in V$, and $e,f\sim_{\tau,c} v$, $S_{\tau,\epsilon}(ec,fc)\le\frac{(1+\epsilon)(\Delta_\tau(v)-2)}{2\epsilon^2\Delta^2}$.

(iii) For every $k\ge 3$, and $ec\in\mathscr{U}_\tau$, $S_{\tau,\epsilon}(ec,ec)\le\frac{(1+\epsilon)(\Delta_\tau(e)-2)}{2\epsilon^2\Delta^2}$.

Proof. First, we verify Item (i). Using Proposition 4.5.3, when $\Delta_\tau(v)\le\frac{\beta}{4(1+\epsilon)}$, we have $|A_\tau(ec,fc)|\le\frac{1}{\beta}$, and when $\Delta_\tau(v)>\frac{\beta}{4(1+\epsilon)}$, we have

$$|A_\tau(ec,fc)| \le \frac{1}{1.5(4/3+4\epsilon)\Delta-2(1+2\epsilon)\Delta}\le\frac{1}{2\epsilon\Delta},$$

where we used $\Delta_\tau(v)\le\Delta$. So, by Lemma 4.5.2, we get $\left|\bar{A}_{\tau,\epsilon}(ec,fc)\right|\le\frac{1}{2\epsilon\Delta}$ for any $\tau\in\mathcal{X}(\le m-3)$.

Now, we verify Item (ii). If $\Delta_\tau(v)\le\frac{\beta}{4(1+\epsilon)}$, then by Eq. (4.21),

$$\begin{aligned}S_\tau(ec,fc) &= 4(1+\epsilon)\sum_{g\sim_{\tau,c}v:g\ne e,f}\bar{A}^{+,v}_\tau(ec,gc)\bar{A}^{+,v}_\tau(gc,fc)+\bar{A}^{-,v}_\tau(ec,gc)\bar{A}^{-,v}_\tau(gc,fc)\\ &\le 4(1+\epsilon)\sum_{g\sim_{\tau,c}v:g\ne e,f}\max_{hc'\in\mathscr{U}_\tau,h\ne e,g}|A_{\tau+hc'}(ec,gc)|\max_{hc'\in\mathscr{U}_\tau,h\ne f,g}|A_{\tau+hc'}(gc,fc)|\\ &\qquad\qquad\text{(Using Lemma 4.5.2)}\\ &\le 4(1+\epsilon)(\Delta^c_\tau(v)-2)\frac{1}{\beta^2}.\end{aligned}$$

Otherwise, if $\Delta_\tau(v)\ge\frac{\beta}{4(1+\epsilon)}$, with a similar use of Lemma 4.5.2,

$$S_\tau(ec,fc) = 2(1+\epsilon)\sum_{g\sim_{\tau,c}v:g\ne e,f}\bar{A}^v_\tau(ec,gc)\cdot\bar{A}^v_\tau(gc,fc)\le 2(1+\epsilon)(\Delta^c_\tau(v)-2)\left(\frac{1}{2\epsilon\Delta}\right)^2,$$

where the first inequality uses Item (i). Finally, Item (ii) follows from $\frac{4(1+\epsilon)}{\beta^2}\le\frac{1+\epsilon}{2\epsilon^2\Delta^2}$.

It remains to prove Item (iii). For a vertex u, let $\alpha(u)=4(1+\epsilon)$ if $\Delta_\tau(u)\le\frac{\beta}{4(1+\epsilon)}$ and $\alpha(u)=2(1+\epsilon)$ otherwise. By an argument similar to Item (ii),

$$\begin{aligned}S_\tau(ec,ec) &\le \alpha(u)\sum_{f\sim_{\tau,c}u:f\ne e}\max_{gc'\in\mathscr{U}_\tau:g\ne e,f}|A_{\tau+gc'}(ec,fc)|^2\\ &\qquad+\alpha(v)\sum_{f\sim_{\tau,c}v:f\ne e}\max_{gc'\in\mathscr{U}_\tau:g\ne e,f}|A_{\tau+gc'}(ec,fc)|^2\\ &\le(\Delta_\tau(u)+\Delta_\tau(v)-2)\max\left\{\frac{4(1+\epsilon)}{\beta^2},\frac{2(1+\epsilon)}{4\epsilon^2\Delta^2}\right\}\\ &\le\frac{(\Delta_\tau(e)-2)(1+\epsilon)}{2\epsilon^2\Delta^2}\qquad\qquad(\Delta_\tau(e)=\Delta_\tau(u)+\Delta_\tau(v)-2)\end{aligned}$$

This completes the proof. □

The following lemma is a crucial part of our proof as it will help us control the term $B_\tau \Pi_\tau^{-1} B_\tau$ in Eq. (4.5) effectively.

Lemma 4.5.5. *Consider a graph $G = (V, E)$, and some weight function $w : E \to \mathbb{R}_{\geq 0}$. Let $A \in \mathbb{R}_{\geq 0}^{E \times E}$ be the weighted adjacency matrix of its line graph. Then*

$$A^2 \preceq 2 \sum_{v \in V} (A^v)^2,$$

where $A^v(e, f) = A(e, f)$ if $e, f \sim v$ and 0 otherwise.

Proof. It is enough to show that for all $x \in \mathbb{R}^E$, $x^\top A^2 x \leq 2 \sum_{v \in V} x^\top (A^v)^2 x$. We have

$$x^\top A^2 x = \|Ax\|_2^2 = \sum_{e \in E} (Ax)(e)^2 = \sum_{e \in E} \langle A_e, x \rangle^2$$

where A_e is the row indexed by e. Now, let $e = \{u, v\} \in E$. We can write $\langle A_e, x \rangle = \langle A_e^u, x \rangle + \langle A_e^v, x \rangle$. Therefore, by an application of Fact 4.1.5,

$$\sum_{e \in E} \langle A_e, x \rangle^2 \leq 2 \sum_{e = \{u,v\} \in E} \left[\langle A_e^u, x \rangle^2 + \langle A_e^v, x \rangle^2 \right] = 2 \sum_{v \in V} \sum_{e \sim v} (A^v x)(e)^2 = 2 \sum_{v \in V} x^\top (A^v)^2 x.$$

□

Now, we apply Theorem 4.2.3 to derive sufficient conditions on the family $\{F_\tau\}_{\tau \in \mathcal{X}(\leq m-2)}$ to get $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau + A_\tau)}{k-1}$ for all $\tau \in \mathcal{X}(\leq m-2)$.

Proposition 4.5.6. *Let $(G, \mathcal{L})$ be a $\left(\frac{4}{3} + 4\epsilon\right)\Delta$-extra edge-list-coloring instance such that $0 \leq \epsilon \leq \frac{1}{10}$ and $\Delta \geq 2\epsilon^{-2}$, and let $\mu = \mu_{G,\mathcal{L}}$ be the associated uniform distribution over proper $\mathcal{L}$-edge-colorings of G. Suppose that $\{F_\tau \in \mathbb{R}^{\mathscr{U} \times \mathscr{U}}\}_{\tau \in \mathcal{X}(\leq m-2)}$ is a family of diagonal matrices supported on $\mathscr{U}_\tau \times \mathscr{U}_\tau$ such that $F_\tau = f_\otimes(X_\tau, \{F_{\tau \cup \sigma}\}_{\emptyset \subsetneq \sigma \in X_\tau(\leq m - |\tau| - 3)})$ if the line graph of G_τ is connected and otherwise:*

1. *For all partial $\mathcal{L}$-edge-colorings on $m-2$ edges: F_τ is defined as $F_\tau(ec, ec) = \frac{1}{\left(\frac{4}{3}+4\epsilon\right)^2 \Delta^2} = \frac{1}{\beta^2}$ for $ec \in \mathscr{U}_\tau$ and 0 on all other entries.*

2. *For all partial $\mathcal{L}$-colorings τ on $m-k$ edges where $k \geq 3$: $F_\tau \preceq \left(\frac{(k-1)^2}{3k-1} - \frac{1}{2\epsilon\Delta}\right) \mathsf{Id}^\tau$, and for any $ec \in \mathscr{U}_\tau$*

$$\sum_{gc' \in \mathscr{U}_{\tau + ec}} p(gc' \mid \tau + ec) F_{\tau + gc'}(ec, ec) \leq (k-2) F_\tau(ec, ec) - \left(\frac{2+\epsilon}{\epsilon}\right) F_\tau(ec, ec)^2 - \mathfrak{g}_\tau(ec), \tag{4.29}$$

where $\mathfrak{g}_\tau(ec) = \frac{(1+\epsilon)\Delta_\tau(e)}{2\epsilon^2\Delta^2} + \frac{(1+\epsilon)^2(2+3\epsilon+\epsilon^2)}{\epsilon^5\Delta^2}$.

Then for every $k \geq 2$ and every partial $\mathcal{L}$-edge-coloring τ on $m-k$ edges, $\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau + A_\tau)}{k-1}$, where A_τ is defined as in Definition 24.

Proof. We prove that the conditions of Theorem 4.2.3 hold for $\{B_\tau\}_{\tau \in \mathcal{X}(\leq m-2)}$ defined as follows:

$$B_\tau \stackrel{\text{def}}{=} \Pi_\tau \frac{F_\tau}{k-1} + \sqrt{\Pi_\tau} \frac{A_\tau}{k-1} \sqrt{\Pi_\tau}, \quad \forall \tau \in \mathcal{X}, k = m - |\tau| \geq 2.$$

Note that the condition of the theorem holds for any τ with $k = 2$. So, we prove the statement for τ when $k \geq 3$. Assume the line graph of G_τ is disconnected. Using the definition of A_τ and our assumption about F_τ, the proof of this case is similar to what we argued in Proposition 4.3.3.

Now, assume that the line graph of G_τ is connected. Note that by Corollary 4.5.4, the absolute value of every off-diagonal entry of A_τ is at most $\frac{1}{2\epsilon\Delta}$ and that there are at most $k-1$ nonzero entries per row. Therefore, $\sqrt{\Pi_\tau} \frac{A_\tau}{k-1} \sqrt{\Pi_\tau} \preceq \frac{1}{2\epsilon\Delta} \Pi_\tau$. Combined with the bound on entries of

diagonal matrix F_τ, this implies that $B_\tau \preceq \frac{k-1}{3k-1}\Pi_\tau$. Therefore, it only remains to show that $\mathbb{E}_{gc\sim\pi_\tau} B_{\tau+gc} \preceq B_\tau - \frac{k-1}{k-2} B_\tau \Pi_\tau^{-1} B_\tau$. This is equivalent to showing that

$$\sqrt{\Pi_\tau^{-1}}\mathbb{E}_{gc\sim\pi_\tau}\left[\Pi_{\tau+gc}\frac{F_{\tau+gc}}{k-2} + \sqrt{\Pi_{\tau+gc}}\frac{A_{\tau+gc}}{k-2}\sqrt{\Pi_{\tau+gc}}\right]\sqrt{\Pi_\tau^{-1}} \preceq \frac{F_\tau}{k-1} + \frac{A_\tau}{k-1} - \frac{(F_\tau+A_\tau)^2}{(k-2)(k-1)}. \tag{4.30}$$

We proceed by first proving a lower bound on the right-hand side. By two applications of Fact 4.1.5, we can write

$$\begin{aligned}(F_\tau + A_\tau)^2 &\preceq \left(1+\frac{2}{\epsilon}\right)F_\tau^2 + \left(1+\frac{\epsilon}{2}\right)A_\tau^2 \\ &= \left(1+\frac{2}{\epsilon}\right)F_\tau^2 + \left(1+\frac{\epsilon}{2}\right)\left(\bar{A}_\tau + \frac{\operatorname{offdiag}(S_\tau)}{k-2}\right)^2 \\ &\preceq \left(1+\frac{2}{\epsilon}\right)F_\tau^2 + (1+\epsilon)\bar{A}_\tau^2 + \frac{\left(3+\epsilon+\frac{2}{\epsilon}\right)\operatorname{offdiag}(S_\tau)^2}{(k-2)^2}.\end{aligned} \tag{4.31}$$

We proceed by finding a diagonal matrix to upper bound $\bar{A}_\tau^2$. For any color $c \in [q]$, $\bar{A}_\tau^c$ is the weighted adjacency matrix of a line graph. Therefore, by Lemma 4.5.5, $(\bar{A}_\tau^c)^2 \preceq 2\sum_{v\in V}\left(\bar{A}_\tau^{c,v}\right)^2$. Since $\bar{A}_\tau^2 = \sum_{c\in[q]}\left(\bar{A}_\tau^c\right)^2$, we get that

$$\bar{A}_\tau^2 \preceq 2\sum_{v\in V}\left(\bar{A}_\tau^v\right)^2 \preceq 4\sum_{v\in V}\left(\left(\bar{A}_\tau^{+,v}\right)^2 + \left(\bar{A}_\tau^{-,v}\right)^2\right).$$

where in the second inequality we used Fact 4.1.5. Therefore, by definition of S_τ (see Eq. (4.21)),

$$(1+\epsilon)\bar{A}_\tau^2 \preceq S_\tau = (k-2)(A_\tau - \bar{A}_\tau) + \operatorname{diag}(S_\tau).$$

So, by Eq. (4.31), we can lower bound the right-hand side of Eq. (4.30) as follows

$$\begin{aligned}\frac{F_\tau}{k-1} + \frac{A_\tau}{k-1} - \frac{(F_\tau+A_\tau)^2}{(k-1)(k-2)} \succeq \frac{F_\tau}{k-1} + \frac{\bar{A}_\tau}{k-1} - \frac{(1+\frac{2}{\epsilon})F_\tau^2}{(k-1)(k-2)} \\ - \frac{\operatorname{diag}(S_\tau)}{(k-1)(k-2)} - \frac{\left(3+\epsilon+\frac{2}{\epsilon}\right)\operatorname{offdiag}(S_\tau)^2}{(k-1)(k-2)^3}.\end{aligned}$$

On the other hand, by definition of $\bar{A}_\tau$ (see Eq. (4.20)), the left-hand side of Eq. (4.30) is equal to

$$\mathbb{E}_{gc\sim\pi_\tau}\left[\Pi_\tau^{-1}\Pi_{\tau+gc}\frac{F_{\tau+gc}}{k-2}\right] + \frac{\bar{A}_\tau}{k-1},$$

and

$$\mathbb{E}_{gc\sim\pi_\tau}\left[\Pi_\tau^{-1}\Pi_{\tau+gc}\frac{F_{\tau+gc}}{k-2}\right](ec,ec) = \sum_{gc'\in\mathscr{U}_{\tau+ec}} p(gc' \mid \tau+ec)F_{\tau+gc'}(ec,ec).$$

Comparing this with the assumption (see Eq. (4.29)), and letting $\mathfrak{g}_\tau(ec) = 0$ for all $ec \notin \mathscr{U}_\tau$, it is enough to show that

$$\operatorname{diag}(\mathfrak{g}_\tau) \succeq \operatorname{diag}(S_\tau) + \frac{\left(3+\epsilon+\frac{2}{\epsilon}\right)\operatorname{offdiag}(S_\tau)^2}{(k-2)^2}.$$

First, notice that

$$\frac{\operatorname{offdiag}(S_\tau)^2}{(k-2)^2} \preceq \frac{\|\operatorname{offdiag}(S_\tau)\|_\infty^2\,\mathsf{Id}^\tau}{(k-2)^2} \preceq \frac{4(1+\epsilon)^2(\Delta-2)^2(\Delta-1)^2}{4\epsilon^4\Delta^4(k-2)^2}\mathsf{Id}^\tau \preceq \frac{(1+\epsilon)^2}{\epsilon^4\Delta^2}\mathsf{Id}^\tau,$$

where the second inequality is by Fact 4.1.4, noting that by Item (ii) of Corollary 4.5.4, every off-diagonal entry of S_τ is at most $\frac{(1+\epsilon)(\Delta-2)}{2\epsilon^2\Delta^2}$ and that there are at most $2(\Delta-1)$ nonzero entries per row. Finally, the statement follows from Item (iii) of Corollary 4.5.4 which shows $S_\tau(ec,ec) \le \frac{(1+\epsilon)\Delta_\tau(e)}{2\epsilon^2\Delta^2}$ for any $ec \in \mathscr{U}_\tau$. □

With this in hand, we prove Theorem 4.5.1.

Proof of Theorem 4.5.1. For any $ec \in \mathscr{U}_\tau$ with $e = uv$, define

$$F_\tau(ec, ec) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \Delta_\tau(e) = 0 \\ f_1(\Delta_\tau(g)), & \text{if } \Delta_\tau(e) = 1, g \sim e \\ f_2(\Delta_\tau(e)), & \text{if } \Delta_\tau(e) \geq 2 \end{cases}$$

where $f_1(i) \stackrel{\text{def}}{=} \frac{1}{\left(\frac{4}{3}+4\epsilon\right)^2\Delta^2} + \frac{(4\epsilon^{-5}+0.6\epsilon^{-2})\sum_{k=1}^{i-1}\frac{1}{k}}{\Delta^2}$ for any $i \geq 2$, and $f_2(i) \stackrel{\text{def}}{=} \frac{5\epsilon^{-5}\ln\Delta + (4\epsilon^{-5}+\epsilon^{-2}i)\sum_{k=1}^{i-1}\frac{1}{k}}{\Delta^2}$ for $i \geq 2$. We prove that this satisfies the conditions of Proposition 4.5.6. Then, the statement follows from the fact that

$$\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau + A_\tau)}{k-1} \leq \frac{\epsilon + \frac{1}{\epsilon}}{k-1}.$$

Here, the last inequality follows by Eq. (4.32) below combined with the fact that every entry of A_τ is at most $\frac{1}{2\epsilon\Delta}$ (by Corollary 4.5.4) and that every row of A_τ has at most $2(\Delta-1)$ nonzero entries.

The base case condition for partial $\mathcal{L}$-edge-colorings on $m-k$ vertices with $k=2$ holds by definition, as $f_1(1) = \frac{1}{\left(\frac{4}{3}+4\epsilon\right)^2\Delta^2}$. Assume that $k \geq 3$ and $\tau \in \mathcal{X}(m-k)$. Similar to the proof of Section 4.6, when the line graph of G_τ is disconnected, the condition holds. Now, assume that the line graph of G_τ is connected. It follows that, for all $1 \leq i \leq 2\Delta$,

$$f_2(i) \leq \frac{9\epsilon^{-5}(\ln(\Delta)+1)}{\Delta^2} + \frac{2\epsilon^{-2}(\ln(\Delta)+2)}{\Delta} \underset{\frac{\ln^2(\Delta)}{\Delta} \leq \frac{\epsilon^3}{15}}{\leq} \frac{3\epsilon}{100} + \frac{2\epsilon}{15} \underset{\epsilon \leq 0.1, k \geq 2}{\leq} \frac{(k-1)^2}{3k-1} - \frac{1}{2\epsilon\Delta}. \quad (4.32)$$

A similar inequality holds for f_1 and $1 \leq i \leq 2\Delta$. It remains to check the condition Eq. (4.29) in Proposition 4.5.6. We need to show that for any $ec \in \mathscr{U}_\tau$,

$$\sum_{fc' \in \mathscr{U}_{\tau+ec}} p(fc' \mid \tau + ec) F_{\tau+fc'}(ec, ec) \preceq (k-2)F_\tau(ec, ec) - \left(1 + \frac{2}{\epsilon}\right) F_\tau(ec, ec)^2 - \mathfrak{g}_\tau(ec),$$

for

$$\mathfrak{g}_\tau(ec) = \frac{(1+\epsilon)\Delta_\tau(e)}{2\epsilon^2\Delta^2} + \frac{(1+\epsilon)^2(2+3\epsilon+\epsilon^2)}{\epsilon^5\Delta^2} \underset{\epsilon \leq 0.1}{\leq} \frac{0.6\Delta_\tau(e)\epsilon^{-2}}{\Delta^2} + \frac{3\epsilon^{-5}}{\Delta^2}.$$

Case 1: $\Delta_\tau(e) = 1, g \sim_\tau e$ Since the line graph of G_τ is connected and the number of uncolored edges under τ is at least 3, $\Delta_\tau(g) \geq 2$. So, it is enough to show that,

$$\begin{aligned} \sum_{fc' \in \mathscr{U}_\tau} p(fc' \mid \tau + ec) F_{\tau+fc'}(ec, ec) &= (\Delta_\tau(g)-1) f_1(\Delta_\tau(g)-1) + (k - \Delta_\tau(g) - 1) f_1(\Delta_\tau(g)) \\ &\leq (k-2) f_1(\Delta_\tau(g)) - \left(1 + \frac{2}{\epsilon}\right) f_1^2(\Delta_\tau(g)) - \frac{0.6\epsilon^{-2} + 3\epsilon^{-5}}{\Delta^2}. \end{aligned} \quad (4.33)$$

Now, note that

$$\begin{aligned} &(k-2)f_1(\Delta_\tau(g)) - (\Delta_\tau(g)-1)f_1(\Delta_\tau(g)-1) - (k - \Delta_\tau(g) - 1) f_1(\Delta_\tau(g)) \\ &= (\Delta_\tau(g)-1)(f_1(\Delta_\tau(g)) - f_1(\Delta_\tau(u)-1)) = \frac{0.6\epsilon^{-2} + 4\epsilon^{-5}}{\Delta^2}. \end{aligned}$$

Furthermore,

$$\left(1 + \frac{2}{\epsilon}\right) f_1(\Delta_\tau(g))^2 \underset{\epsilon \leq 0.1}{\leq} \frac{2.1}{\epsilon} \left(\frac{5\epsilon^{-5}\ln\Delta}{\Delta^2}\right)^2 \underset{\frac{\ln^2(\Delta)}{\Delta} \leq \frac{\epsilon^3}{15}}{\leq} \frac{\epsilon^{-5}}{\Delta^2}.$$

Putting these together, we get Eq. (4.33).

Case 2: $\Delta_\tau(e) \geq 2$ For convenience in writing the recursion, let $f_2(1) = \frac{5\epsilon^{-5}\ln(\Delta)}{\Delta^2}$. Following similar calculations, it is enough to show that

$$(\Delta_\tau(e)-1) f_2(\Delta_\tau(e)) - \Delta_\tau(e) f_2(\Delta_\tau(e)-1) \geq \left(1 + \frac{2}{\epsilon}\right) f_2(\Delta_\tau(e))^2 + \frac{0.6\epsilon^{-2}\Delta_\tau(e) + 3\epsilon^{-5}}{\Delta^2}. \quad (4.34)$$

Note that in the left-hand side of the above equation, we should write $f_1(\Delta_\tau(g))$ if $\Delta_\tau(e)-1=1$ and g is the only remaining neighbour of e, but since $f_1(i)\le \frac{5\epsilon^{-5}\ln(\Delta)}{\Delta^2}=f_2(1)$ for all $1\le i\le 2\Delta$ the above inequality is valid. Note that, by definition

$$(\Delta_\tau(e)-1)f_2(\Delta_\tau(e))-\Delta_\tau(e)f_2(\Delta_\tau(e)-1)=\frac{\epsilon^{-2}\Delta_\tau(e)+4\epsilon^{-5}}{\Delta^2}.$$

Furthermore, $\frac{\ln^2(\Delta)}{\Delta}\ge\frac{\epsilon^3}{15}$ and $\epsilon\le 0.1$ imply that $\ln\Delta\ge 10$, and we can write

$$\begin{aligned}\left(1+\frac{2}{\epsilon}\right)f_2(\Delta_\tau(e))^2 &\le \frac{2.1\cdot(\ln\Delta+2)^2}{\epsilon\Delta^4}\cdot\left(\epsilon^{-2}\Delta_\tau(e)+9\epsilon^{-5}\right)^2 && (\epsilon\le 0.1)\\ &\le \frac{2.1\cdot(\ln\Delta+2)^2}{\epsilon\Delta^4}\cdot\left(1.2\cdot(\epsilon^{-2}\Delta_\tau(e))^2+6\cdot(9\epsilon^{-5})^2\right) && (\text{Fact 4.1.5})\\ &\le \frac{0.4\epsilon^{-2}\Delta_\tau(e)+\epsilon^{-5}}{\Delta^2}. && (\ln\Delta\ge 10,\ \tfrac{\ln^2(\Delta)}{\Delta}\le\tfrac{\epsilon^3}{15})\end{aligned}$$

This finishes the proof of Eq. (4.34). □

4.6 Unfinished Proofs from Sections 4.3 and 4.4

Proof of Theorem 4.3.1. For each $\tau\in\mathcal{X}$ with $n-|\tau|\ge 2$, let $F_\tau\in\mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ be a diagonal matrix supported on $\mathscr{U}_\tau\times\mathscr{U}_\tau$ defined as follows: for any $vc\in\mathscr{U}_\tau$,

$$F_\tau(vc,vc)\overset{\text{def}}{=}\begin{cases}0, & \text{if } \Delta_\tau(v)=0,\\ f_1(\Delta_\tau(u)), & \text{if } \Delta_\tau(v)=1 \text{ and } u\sim_\tau v,\\ f_2(\Delta_\tau(v)), & \text{if } \Delta_\tau(v)\ge 2,\end{cases}$$

where $f_1(i)=\frac{1}{(1+\epsilon)\Delta}+\frac{1+2\sum_{j=1}^{i-1}\frac{1}{j}}{(1+\epsilon)^2\Delta^2}$ for $i\ge 1$ and $f_2(i)=\frac{i}{(1+\frac{\epsilon}{2})\Delta-(i-1)-\frac{4}{\epsilon}\sum_{j=1}^{i-1}\frac{1}{j}}$ for $i\ge 2$. We show that the conditions of Proposition 4.3.3 hold for $\{F_\tau\}_{\tau\in\mathcal{X}(\le n-2)}$. Then, the statement follows from the fact that $\rho(F_\tau)\le\frac{5}{2\epsilon}$. This is true because $\frac{\ln(\Delta)+2}{\Delta}\le\frac{\epsilon^2}{40}$ implies that for any $1\le i\le\Delta$, the denominator of $f_2(i)$ is at least $\frac{2\epsilon}{5}\Delta$ and thus $f_1(i)\le f_2(i)\le\frac{5}{2\epsilon}$.

The condition for partial $\mathcal{L}$-colorings τ on $n-k$ vertices for $k=2$ holds by definition. Assume $k\ge 3$. When G_τ is disconnected, one can check the condition holds because of the fact that the degrees of vertices of a connected component do not change by removing vertices from other connected components of the graph. Now, assume that G_τ is connected. Note that for every $v\in V_\tau$, we have

$$\frac{(k-1)^2}{3k-1}\ge\frac{\Delta_\tau(v)}{5}\underset{\frac{\ln(\Delta)+2}{\Delta}\le\frac{\epsilon^2}{40}}{\ge}\frac{8}{\epsilon^2}\underset{f_2(\Delta_\tau(v))\le\frac{5}{2\epsilon},\epsilon\le 1}{\ge}f_2(\Delta_\tau(v))\ge f_1(\Delta_\tau(v)).$$

Therefore, it is enough to show that $\sum_{uc'\in\mathscr{U}_{\tau+vc}}p(uc'\mid\tau+vc)F_{\tau+uc'}(vc,vc)\preceq(k-2)F_\tau(vc,vc)-F_\tau(vc,vc)^2$ for any $vc\in\mathscr{U}_\tau$.

Case 1: $\Delta_\tau(v)=1$, and $u\sim_\tau v$ Since G_τ is connected and $n-|\tau|\ge 3$, $\Delta_\tau(u)\ge 2$. We have

$$\sum_{wc'\in\mathscr{U}_\tau}p(wc'\mid\tau+vc)F_{\tau+wc'}(vc,vc)=(\Delta_\tau(u)-1)f_1(\Delta_\tau(u)-1)+(k-\Delta_\tau(u)-1)f_1(\Delta_\tau(u)).$$

On the other hand, $\frac{\ln(\Delta)+2}{\Delta}\le\frac{\epsilon^2}{40}$ and $\epsilon\le 1$ imply that $\frac{1+2\sum_{j=1}^{i-1}\frac{1}{j}}{(1+\epsilon)\Delta}\le\frac{1}{20(1+\epsilon)}$ for any $1\le i\le\Delta$. Therefore,

$$F_\tau(vc,vc)^2=f_1(\Delta_\tau(u))^2\le\left(\frac{1}{(1+\epsilon)\Delta}+\frac{1}{20(1+\epsilon)\Delta}\right)^2\le\frac{2}{(1+\epsilon)^2\Delta^2}.$$

Therefore $(k-2)F_\tau(vc,vc)-F_\tau(vc,vc)^2\ge(k-2)f_1(\Delta_\tau(u))-\frac{2}{(1+\epsilon)^2\Delta^2}$ and thus it is enough to show that

$$(\Delta_\tau(u)-1)\left(f_1\left(\Delta_\tau(u)\right)-f_1\left(\Delta_\tau(u)-1\right)\right)\ge\frac{2}{(1+\epsilon)^2\Delta^2}.$$

But this inequality holds with equality.

Case 2: $\Delta_\tau(v) \geq 2$ One can check that $\frac{\ln(\Delta)+1}{\Delta} \leq \frac{\epsilon^2}{40}$ implies $\frac{1}{(1+\epsilon)\Delta} + \frac{2\sum_{j=1}^{\Delta-1}\frac{1}{j}+1}{(1+\epsilon)^2\Delta^2} \leq \frac{1}{(1+\frac{\epsilon}{2})\Delta}$. For convenience define $f_2(1) = \frac{1}{1+\frac{\epsilon}{2}}$ and notice that $f_1(i) \leq f_2(1)$ for any $1 \leq i \leq \Delta$. We want to show that

$$\sum_{uc' \in \mathscr{U}_\tau} p(uc' \mid \tau + vc) F_{\tau+uc'}(vc, vc) = \Delta_\tau(v) f_2(\Delta_\tau(v) - 1) + (k - \Delta_\tau(v) - 1) f_2(\Delta_\tau(v))$$
$$\leq (k-2) f_2(\Delta_\tau(v)) - f_2(\Delta_\tau(v))^2.$$

Let $i \stackrel{\text{def}}{=} \Delta_\tau(v)$. We have

$$(i-1)f_2(i) - i f_2(i-1) = \frac{i(i-1) + \frac{4}{\epsilon} i}{\left(\left(1+\frac{\epsilon}{2}\right)\Delta - (i-1) - \frac{4}{\epsilon}\sum_{j=1}^{i-1}\frac{1}{j}\right)\left(\left(1+\frac{\epsilon}{2}\right)\Delta - (i-2) - \frac{4}{\epsilon}\sum_{j=1}^{i-2}\frac{1}{j}\right)}.$$

Therefore

$$(i-1)f_2(i) - i f_2(i-1) - f_2(i)^2 = \frac{\left(\frac{4}{\epsilon} - 1\right) i \left(\left(1+\frac{\epsilon}{2}\right)\Delta - (i-1) - \frac{4}{\epsilon}\sum_{j=1}^{i-1}\frac{1}{j}\right) - i^2\left(1 + \frac{4}{\epsilon}\frac{1}{i-1}\right)}{\left(\left(1+\frac{\epsilon}{2}\right)\Delta - (i-1) - \frac{4}{\epsilon}\sum_{j=1}^{i-1}\frac{1}{j}\right)^2\left(\left(1+\frac{\epsilon}{2}\right)\Delta - (i-2) - \frac{4}{\epsilon}\sum_{j=1}^{i-2}\frac{1}{j}\right)}.$$

The denominator is positive for $1 \leq i \leq \Delta$ and for the numerator we have

$$\left(\frac{4}{\epsilon} - 1\right) i \left(\left(1+\frac{\epsilon}{2}\right)\Delta - (i-1) - \frac{4}{\epsilon}\sum_{j=1}^{i-1}\frac{1}{j}\right) - i^2\left(1 + \frac{4}{\epsilon}\frac{1}{i-1}\right)$$
$$\geq \left(1 - \frac{\epsilon}{2}\right) i\Delta - \frac{16i}{\epsilon^2}(\ln(\Delta) + 1) - \frac{4i^2}{\epsilon(i-1)}. \qquad (i \leq \Delta)$$

Canceling out an i, and using $\frac{\ln(\Delta)+2}{\Delta} \leq \frac{\epsilon^2}{40}$ and $\epsilon \leq 1$ the right-hand side is nonnegative. □

Proof of Theorem 4.4.1. As before we make the tree rooted at an arbitrary vertex r. For any $k \geq 2$, any partial coloring τ on $m-k$ vertices, let $F_\tau \in \mathbb{R}^{\mathscr{U}\times\mathscr{U}}$ be a diagonal matrix supported on $\mathscr{U}_\tau \times \mathscr{U}_\tau$ defined as follows: For any $vc \in \mathscr{U}_\tau$,

$$F_\tau(vc, vc) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \Delta_\tau(v) = 0, \\ f_1(\Delta_\tau(u)), & \text{if } \Delta_\tau(v) = 1, u \sim v, \text{ and } v \text{ is the root of a component of } G_\tau \\ f_2(\Delta_\tau(v)), & \text{if } \Delta_\tau(v) \geq 2 \text{ and } v \text{ is the root of a component of } G_\tau \\ f_3(\Delta_\tau(v), \Delta_\tau(a(v))), & \text{if } \Delta_\tau(v) \geq 1 \text{ and } a(v) \in V_\tau \end{cases}$$

where $a(v)$ is the immediate ancestor of v in G, and $f_1(i) \stackrel{\text{def}}{=} \frac{5\left(\sum_{j=1}^{i-1}\frac{1}{j}\right)+1}{\epsilon^2\Delta^2}$ for $i \geq 1$, $f_2(i) \stackrel{\text{def}}{=} \frac{5\left(\ln(\Delta)+1+i\sum_{j=1}^{i-1}\frac{1}{j}\right)}{\epsilon^2\Delta^2}$ for $i \geq 2$, and $f_3(i,j) \stackrel{\text{def}}{=} \frac{5\left(\ln(\Delta)+1+j+i\sum_{k=1}^{i-1}\frac{1}{k}\right)}{\epsilon^2\Delta^2}$ for $i, j \geq 1$. We prove that the conditions of Proposition 4.4.3 hold for $\{F_\tau\}_{\tau \in \mathcal{X}(\leq n-2)}$. Then, the statement follows from the fact that

$$\lambda_2(\mathcal{Q}_{\mu^\tau}) \leq \frac{\rho(F_\tau + A_\tau)}{k-1} \leq \frac{\rho(F_\tau) + \frac{1}{\epsilon}}{k-1} + \frac{\frac{1}{20} + \frac{1}{\epsilon}}{k-1},$$

where the second to last inequality follows by (4.35) below combined with the fact that every entry of A_τ is at most $\frac{1}{\epsilon\Delta}$ and that every row of A_τ has at most Δ nonzero entries. The last inequality uses that since $\frac{\ln^2(\Delta)}{\Delta} \leq \frac{\epsilon^2}{100}$ and $\epsilon \leq 1$, for all $1 \leq i \leq \Delta$

$$f_1(i), f_2(i), f_3(i) \leq \frac{5(\ln(\Delta) + 2)}{\epsilon^2\Delta} \leq \frac{1}{20}. \qquad (4.35)$$

The base case condition for partial $\mathcal{L}$-colorings on $n-k$ vertices where $k = 2$ holds by definition. Assume $k \geq 3$. Similar to the proof of Theorem 4.3.1, when G_τ is disconnected, the condition holds. Now, assume G_τ is connected. By Eq. (4.35), for all $vc \in \mathscr{U}_\tau$, $F_\tau(vc, vc) \leq \frac{1}{20} \leq \frac{(k-1)^2}{3k-1} - \frac{1}{\beta}$. Therefore, it is enough to show that for any $vc \in \mathscr{U}_\tau$,

$$\sum_{wc' \in \mathscr{U}_{\tau+vc}} p(wc' \mid \tau + vc) F_{\tau+wc'}(vc, vc) \preceq (k-2) F_\tau(vc, vc) - 2F_\tau(vc, vc)^2 - \mathfrak{g}_\tau(vc),$$

for $\mathfrak{g}_\tau(vc)$ defined in Proposition 4.4.3.

Case 1: $\Delta_\tau(v) = 1$ and v is the root of a component of G_τ. Let u be the only neighbor of v. Since G_τ is connected and the number of uncolored vertices under τ is at least 3, $\Delta_\tau(u) \geq 2$. We need to show that

$$\begin{aligned}&\sum_{wc' \in \mathscr{U}_{\tau+vc}} p(wc' \mid \tau + vc) F_{\tau+wc'}(vc, vc)\\ &= (\Delta_\tau(u) - 1) f_1(\Delta_\tau(u) - 1) + (k - \Delta_\tau(u) - 1) f_1(\Delta_\tau(u))\\ &\leq (k-2) f_1(\Delta_\tau(u)) - 2 f_1(\Delta_\tau(u))^2 - \frac{4}{\epsilon^2 \Delta^2}.\end{aligned} \tag{4.36}$$

But,

$$\begin{aligned}&(k-2) f_1(\Delta_\tau(u)) - (\Delta_\tau(u) - 1) f_1(\Delta_\tau(u) - 1) - (k - \Delta_\tau(u) - 1) f_1(\Delta_\tau(u))\\ &= (\Delta_\tau(u) - 1)(f_1(\Delta_\tau(u)) - f_1(\Delta_\tau(u) - 1)) = \frac{5}{\epsilon^2 \Delta^2}.\end{aligned}$$

Furthermore, one can see that $\frac{\epsilon^2}{100}$ implies that $2 f_1(\Delta_\tau(u))^2 \leq \frac{50(\ln(\Delta)+2)^2}{\epsilon^4 \Delta^4} \leq \frac{1}{\epsilon^2 \Delta^2}$. This completes the proof of Eq. (4.36).

Case 2: $\Delta_\tau(v) \geq 2$, and v is the root of a component of G_τ. Note that f_1 is bounded above by $\frac{5(\ln(\Delta)+1)}{\epsilon^2 \Delta^2}$. For convenience in writing the recursion, let $f_2(1) = \frac{5(\ln(\Delta)+1)}{\epsilon^2 \Delta^2}$. Following similar calculations, it is enough to show that

$$(\Delta_\tau(v) - 1) f_2(\Delta_\tau(v)) - (\Delta_\tau(v)) f_2(\Delta_\tau(v) - 1) \geq 2 f_2(\Delta_\tau(v))^2 + \frac{4\Delta_\tau(v)}{\epsilon^2 \Delta^2}. \tag{4.37}$$

But one can see that, by definition

$$(\Delta_\tau(v) - 1) f_2(\Delta_\tau(v)) - (\Delta_\tau(v)) f_2(\Delta_\tau(v) - 1) = \frac{5\Delta_\tau(v)}{\epsilon^2 \Delta^2},$$

and

$$2 f_2(\Delta_\tau(v))^2 \leq \frac{50(\Delta_\tau(v)+2)^2 \ln^2(\Delta)}{\epsilon^4 \Delta^4} \underset{\Delta \geq 100}{\leq} \frac{55 \Delta_\tau(v)^2 \ln^2(\Delta)}{\epsilon^4 \Delta^4} \underset{\frac{\ln^2(\Delta)}{\Delta} \leq \frac{\epsilon^2}{100}}{\leq} \frac{0.55 \Delta_\tau(v)^2}{\epsilon^2 \Delta^3} \leq \frac{\Delta_\tau(v)}{\epsilon^2 \Delta^2}. \tag{4.38}$$

This finishes the proof of Eq. (4.37).

Case 3: $\Delta_\tau(v) \geq 1$ and $a(v) \in V_\tau$. For convenience in writing the recursion, for $1 \leq i \leq \Delta$, let $f_3(i, 0) = \frac{5\left(\ln(\Delta)+1+i\sum_{k=1}^{i-1} \frac{1}{k}\right)}{\epsilon^2 \Delta^2}$ and note that $\max_{1 \leq j \leq \Delta} f_1(j) \leq f_2(i) \leq f_3(i, 0)$. Similar to the previous cases, after simplifying the recursion, one can see that it is enough that for $i = \Delta_\tau(v)$ and $j = \Delta_\tau(a(v))$,

$$(i-1) f_3(i,j) - i f_3(i-1, j) + (j-1)(f_3(i,j) - f_3(i, j-1)) \geq 2 f_3(i,j)^2 + \frac{4i}{\epsilon^2 \Delta^2} + \frac{4(j-1)}{\epsilon^2 \Delta^2}.$$

Now, note that

$$(i-1) f_3(i,j) - i f_3(i-1, j) + (j-1)(f_3(i,j) - f_3(i, j-1)) = \frac{5i}{\epsilon^2 \Delta^2} + \frac{5(j-1)}{\epsilon^2 \Delta^2}.$$

Furthermore,

$$2 f_3(i)^2 = 2\left(f_2(i)^2 + \frac{5(j-1)}{\epsilon^2 \Delta^2}\right)^2 \leq 2.5 f_2(i)^2 + \frac{250(j-1)^2}{\epsilon^4 \Delta^4} \underset{\Delta \geq 100\epsilon^{-2}}{\leq} \frac{i}{\epsilon^2 \Delta^2} + \frac{j-1}{\epsilon^2 \Delta^2},$$

where the last inequality uses the calculations in Eq. (4.38). This finishes the proof. □

Chapter 5

The Geometry of Polynomials I: Log-Concavity

In this chapter, we introduce an extremely useful alternative viewpoint for all of the concepts we have seen so far. This viewpoint is based around encoding our high-dimensional discrete probability distributions μ as *multivariate polynomials*, and studying μ using analytic and algebraic properties of these polynomials. In this chapter, we focus on an analytic property called *log-concavity*, which is featured prominently in the study of continuous optimization and convex geometry. We'll show that 0-spectral independence is in fact equivalent to log-concavity. On the one hand, this yields new examples of distributions which are 0-spectrally independent and a new method for establishing 0-spectral independence. On the other hand, this enables us to use Oppenheim's Trickle-Down Method from Chapter 3 to establish log-concavity of a polynomial. In the next chapter, we will study how *zero-freeness*, a purely algebraic property, connects with spectral independence, a probabilistic property; see Chapter 6. The contents of this chapter are largely based on [Ana+19; Ana+18a].

We begin by defining how we will encode a probability distribution into a polynomial. The following definition is fundamental to this chapter and Chapter 6.

Definition 25 ((Multivariate) Generating Polynomial)**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$, where $\mathscr{U}$ is a finite ground set and $0 \leq n \leq |\mathscr{U}|$ is an integer. Define the* ***multivariate generating polynomial of*** *μ by*

$$g_\mu\left(x_i : i \in \mathscr{U}\right) \stackrel{\text{def}}{=} \sum_{\tau \in \binom{\mathscr{U}}{n}} \mu(\tau) x^\tau$$

Note that g_μ is a multiaffine and n-homogeneous.

Remark 19. Conversely, given a nonzero n-homogeneous multiaffine polynomial $g(x) = \sum_{S \in \binom{\mathscr{U}}{n}} c_S x^S$ with nonnegative coefficients, we can construct a corresponding high-dimensional discrete probability distribution μ_g over $\binom{\mathscr{U}}{n}$ by defining $\mu_g(S) \propto c_S$. The multivariate generating polynomial of this distribution is equal to g up to scaling by a constant.

Remark 20. g_μ is also sometimes called the *multivariate partition function of* μ, particularly when its coefficients aren't normalized to sum to 1 (and the polynomial isn't homogenized). For instance, the multivariate partition function of the uniform distribution over independent sets of a graph G is

$$\mathscr{Z}_G(\lambda_v : v \in V) = \sum_{I \subseteq V \text{ independent}} \prod_{v \in I} \lambda_v$$

whereas multivariate generating polynomial of this distribution is

$$g_\mu(x_v, y_v : v \in V) = \frac{1}{\mathscr{Z}_G(\mathbf{1})} \sum_{I \subseteq V \text{ independent}} \prod_{v \in I} x_v \prod_{v \in V \setminus I} y_v$$

Clearly, $\mathscr{Z}_G(\boldsymbol{\lambda}) = \mathscr{Z}_G(\mathbf{1}) \cdot g_\mu|_{x=\boldsymbol{\lambda}, y=\mathbf{1}}$. Recall $\mathscr{Z}_G$ is also known as the (multivariate) partition function of the Gibbs distribution of the hardcore gas model on G.

As previously mentioned, our goal is to connect analytic properties of the multivariate generating polynomial with spectral independence. In this chapter, we focus on log-concavity, a heavily studied notion in optimization and geometry.

Definition 26 (Log-Concave Polynomials)**.** *Let $f(x)$ be a nonzero multivariate polynomial in the variables $x = (x_i : i \in \mathscr{U})$ with nonnegative coefficients; f need not be homogeneous nor multiaffine. We say f is **log-concave at a point** $x \in \mathbb{R}^{\mathscr{U}}_{>0}$ if the Hessian of $\log f$*

$$\nabla^2 \log f = \frac{f \cdot (\nabla^2 f) - (\nabla f)(\nabla f)^\top}{f^2}$$

is negative semidefinite at x. Now, let $K \subseteq \mathbb{R}^{\mathscr{U}}_{>0}$. We have the following definitions in increasing order of strength.

- *We say f is log-concave on K if f is log-concave at each point $x \in K$.*
- *We say f is **strongly log-concave on** K if for any $k \geq 0$, and any sequence of elements $i_1, \ldots, i_k \in \mathscr{U}$, the polynomial $(\partial_{i_1} \cdots \partial_{i_k} f)(x)$ is log-concave on K. [Gur09; Gur10]*
- *We say f is **completely log-concave on** K if for any $k \geq 0$, and any sequence of nonzero nonnegative vectors $a_1, \ldots, a_k \in \mathbb{R}^{\mathscr{U}}_{\geq 0}$, the polynomial $(\partial_{a_1} \cdots \partial_{a_k} f)(x)$ is log-concave on K. [AOV21]*

It is not hard to see from the definition that, for any fixed d and n, the set of polynomials of degree at most d in n variables that are log-concave on $\mathbb{R}^n_{\geq 0}$ is closed in the Euclidean topology on $\mathbb{R}[x_1, \ldots, x_n]_{\leq d}$. Completely log-concave polynomials were introduced in [AOV21] based on similar notions of strongly log-concave and Alexandrov–Fenchel polynomials first studied in [Gur09; Gur10].

Remark 21. An equivalent characterization of log-concavity over a convex set $K \subseteq \mathbb{R}^n_{\geq 0}$ is to require that for every $u, v \in K$ and every $0 \leq \lambda \leq 1$, we have the inequality

$$\log f(\lambda u + (1-\lambda)v) \geq \lambda \log f(u) + (1-\lambda) \log f(v)$$

or equivalently,

$$f(\lambda u + (1-\lambda)v) \geq f(u)^\lambda f(v)^{1-\lambda}. \tag{5.1}$$

Example 8 (Volume Polynomials). Let $K_1, \ldots, K_m \subseteq \mathbb{R}^n$ be a sequence of convex bodies, i.e. compact convex sets with nonempty interior. Then it is well-known that the multivariate function

$$x = (x_1, \ldots, x_m) \mapsto \mathrm{Vol}\,(x_1 K_1 + \cdots + x_m K_m)$$

on $x \in \mathbb{R}^m_{\geq 0}$ agrees with a n-homogeneous multivariate polynomial with nonnegative coefficients called the *volume polynomial*. Here, the symbol $+$ denotes the *Minkowski sum*, defined as $K + L \stackrel{\mathsf{def}}{=} \{u + v : u \in K, v \in L\}$ for two subsets $K, L \subseteq \mathbb{R}^m$, and $\lambda K = \{\lambda u : u \in K\}$ denotes the *dilation* of K by a nonnegative scalar $\lambda \geq 0$. The classical *Brunn-Minkowski Inequality* exactly says that this polynomial is log-concave on all of $\mathbb{R}^m_{\geq 0}$. If fact, using the stronger *Alexandrov-Fenchel Inequalities*, one can show that this polynomial is completely log-concave on all of $\mathbb{R}^m_{\geq 0}$.

Example 9 (Determinantal Polynomials). In a manner very similar to Example 8, let $A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ be symmetric positive semidefinite matrices. Then the function

$$x = (x_1, \ldots, x_m) \mapsto \det\,(x_1 A_1 + \cdots + x_m A_m)$$

is also a n-homogeneous multivariate polynomial with nonnegative coefficients. This polynomial is known to be *real stable*, and hence completely log-concave on all of $\mathbb{R}^m_{\geq 0}$; see Section 5.7 for a more in-depth discussion of the connection between real stability and log-concavity.

The following theorem gives the first bridge between analytic properties of the generating polynomial g_μ and useful probabilistic properties of μ.

Theorem 5.0.1 (Log-Concavity $\iff$ 0-Spectral Independence)**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ where $\mathscr{U}$ is a finite ground set and $0 \leq n \leq |\mathscr{U}|$ is an integer. Then μ is $(0, \ldots, 0)$-spectrally independent (equivalently, μ is a $(0, \ldots, 0)$-local spectral expander) if and only if g_μ is strongly log-concave (at $\mathbf{1}$).*

Remark 22. There is nothing special about $\mathbf{1}$. For a vector $x \in \mathbb{R}^n_{\geq 0}$, if we define a new probability measure $x * \mu$ by $(x * \mu)(\tau) \propto \mu(\tau) \cdot \prod_{i \in \tau} x_i$, then strong log-concavity of g_μ at x is equivalent to strong log-concavity of $g_{x*\mu}$ at $\mathbf{1}$, which is equivalent to $(0, \ldots, 0)$-spectral independence for the measure $x * \mu$. Here, $x \in \mathbb{R}^n_{\geq 0}$ is often called an *external field* in the language of statistical physics, or *exponential tilt* in the language of probability theory.

Remark 23. The bridge goes further. In particular, [Ali+21] extended this result by showing that log-concavity of the composition $g_\mu \circ x^\alpha$ at $x \in \mathbb{R}^n_{\geq 0}$ for a parameter $0 \leq \alpha \leq 1$ is equivalent to $\left(\frac{1}{\alpha} - 1\right)$-spectral independence for $x * \mu$.

In light of Theorem 5.0.1, we say μ is a *discrete log-concave distribution* if g_μ is strongly log-concave (or equivalently, μ is $(0, \ldots, 0)$-spectrally independent).

Example 10 (Bases Generating Polynomials of Matroids). As we saw in Chapter 3 (see Theorem 3.1.2), the uniform distribution over bases of an n-element rank-r matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ is $(0, \ldots, 0)$-spectrally independent. By Theorem 5.0.1, this is equivalent to saying that the *bases generating polynomial of* $\mathcal{M}$

$$g_{\mathcal{M}}(x_i : i \in \mathscr{U}) \stackrel{\text{def}}{=} \sum_{\mathcal{B} \subseteq \mathscr{U} \text{ basis}} x^{\mathcal{B}}$$

is strongly log-concave at $\mathbf{1}$. Using the same proof of Theorem 3.1.2 combined with Theorem 5.0.2 below, one can show that $g_{\mathcal{M}}$ is actually completely log-concave on all of $\mathbb{R}^n_{\geq 0}$. This was already shown previously by [AOV21] using the beautiful *combinatorial Hodge theory* for matroids introduced in [AHK18], and further developed/extended in [HW17; Eur20; Bra+20b; Bra+20a; BES21; HSW21; ADH21; ADH22; Ber+22; EHL22]; see also [Bak18; Huh18; Huh22].

Theorem 5.0.1 further motivates the study of log-concave polynomials, which are interesting in their own right, although historically, the study of log-concave polynomials long preceded the invention of spectral independence. We will see in Section 5.5 how log-concave polynomials can be used to settle the strongest version of Mason's Conjecture in matroid theory and algebraic combinatorics. For the moment, we first discuss characterizations of log-concave polynomials. The following is a "polynomial" analog of connectivity for the local walks.

Definition 27 (Indecomposability)**.** *We say a multivariate polynomial $f(x)$ with variables $x = (x_i : i \in \mathscr{U})$ is* ***decomposable*** *if there exists a proper subset $\emptyset \subsetneq S \subsetneq \mathscr{U}$ of variables as well as nonzero polynomials $g \in \mathbb{K}[x_i : i \in S]$ and $h \in \mathbb{K}[x_i : i \in \mathscr{U} \setminus S]$ such that $f = g + h$. Otherwise, f is* ***indecomposable****. Equivalently, if we form an undirected graph $G_f = (\mathscr{U}, E)$ with vertex set $\mathscr{U}$ and an edge $\{i, j\} \in E$ if and only if $\partial_i \partial_j f \neq 0$, then f is indecomposable if and only if G_f is connected.*

Imposing indecomposability for all partial derivatives is called *total indecomposability*, and is akin to total connectivity for probability distribution (see Definition 15). With this notion in hand, we have the following theorem.

Theorem 5.0.2 (Characterization of Strong and Complete Log-Concavity)**.** *Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be a d-homogeneous (not necessarily multiaffine) polynomial with nonnegative coefficients. Then the following are equivalent.*

1. *f is completely log-concave at $w \in \mathbb{R}^n_{>0}$.*
2. *f is strongly log-concave at $w \in \mathbb{R}^n_{>0}$.*
3. *f satisfies both of the following conditions:*
   - ***Total Indecomposability:*** *For every multi-index $\alpha \in \mathbb{Z}^n_{\geq 0}$ with $|\alpha| \leq d - 2$, the polynomial $\partial^\alpha f$ is indecomposable (or identically zero). Note this condition is independent of the choice of $w \in \mathbb{R}^n_{>0}$.*
   - ***Log-Concavity in Degree-2:*** *For every multi-index $\alpha \in \mathbb{Z}^n_{\geq 0}$ with $|\alpha| = d - 2$, the polynomial $\partial^\alpha f$ is log-concave at $w \in \mathbb{R}^n_{>0}$ (or identically zero).*

We close the introduction to this chapter with the following proposition, which shows different equivalent ways of formulating log-concavity and can be helpful for intuition.

Proposition 5.0.3 ([AOV21; Ana+18a]). *Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be a d-homogeneous polynomial with nonnegative coefficients and $d \geq 2$. Fix a point $v \in \mathbb{R}^n_{\geq 0}$ s.t. $f(v) > 0$ and let $Q = \nabla^2 f(v)$. Then the following are equivalent:*

1. *f is log-concave at v.*
2. *The quadratic form $x \mapsto x^\top Q x$ is negative semidefinite on $(Qv)^\perp$.*
3. *The quadratic form $x \mapsto x^\top Q x$ is negative semidefinite on $(Qw)^\perp$ for every $w \in \mathbb{R}^n_{\geq 0}$ s.t. $Qw \neq 0$.*
4. *The quadratic form $x \mapsto x^\top Q x$ is negative semidefinite on some linear subspace of dimension-$(n-1)$.*
5. *The matrix $(v^\top Qv)Q - (Qv)(Qv)^\top$ is negative semidefinite.*
6. *Q has at most one positive eigenvalue.*
7. *$\partial_v f$ is log-concave at v (assuming $d \geq 3$).*

Remark 24. Whenever f has nonnegative coefficients and degree at least 2, $(\nabla^2 f)(x)$ has at least one positive entry for any $x \in \mathbb{R}^n_{>0}$. Hence, one can see (via, for example, the variational characterization of eigenvalues Theorem 1.4.1 along with a test vector with positive entries) that $(\nabla^2 f)(x)$ must have at least one strictly positive eigenvalue. Thus, whenever we write "at most one positive eigenvalue", we also mean it has "exactly one positive eigenvalue".

Proof of Proposition 5.0.3. By applying Euler's Homogeneous Function Theorem (see Fact 1.4.9 and Corollary 1.4.10) to f and $\partial_i f$, respectively, we see that $Qv = (d-1) \cdot \nabla f(v) \neq 0$ and $v^\top Qv = d(d-1) \cdot f(v) > 0$. The Hessian of $\log f$ at v then equals

$$\nabla^2 \log f(v) = \left(\frac{f \cdot \nabla^2 f - \nabla f \nabla f^\top}{f^2} \right) \Big|_{x=v} = d(d-1) \frac{v^\top Qv \cdot Q - \frac{d}{d-1}(Qv)(Qv)^\top}{(v^\top Qv)^2}. \tag{5.2}$$

(Item 1 $\implies$ Item 2) If f is log-concave at v, then the Hessian of $\log f$ at v is negative semidefinite, meaning that the quadratic form $x \mapsto x^\top \left(\nabla^2 \log f(v)\right) x$ is negative semidefinite globally. Using Eq. (5.2) above, when restricted to the linear space $(Qv)^\perp = \left\{x \in \mathbb{R}^n : x^\top Qv = 0\right\}$, this quadratic form simplifies to $x \mapsto \frac{d(d-1)}{v^\top Qv} \cdot x^\top Qx$, meaning that $x \mapsto x^\top Qx$ is negative semidefinite on this linear space.

(Item 2 $\implies$ Item 4) Since Qv is nonzero, $(Qv)^\perp$ is a linear subspace with dimension-$(n-1)$.

(Item 4 $\implies$ Item 6) If Q has at least 2 positive eigenvalues with linearly independent u_1, u_2, then $w^\top Qw > 0$ for any nonzero w in the span of u_1, u_2, which has dimension-2. Since any $(n-1)$-dimensional linear subspace V must intersect the span of u_1, u_2 nontrivially, Q cannot be negative semidefinite on V.

(Item 6 $\implies$ Item 5) Suppose Q has at most one positive eigenvalue. Let $w \in \mathbb{R}^n$ and consider the $n \times 2$ matrix P with columns v and w. Then

$$P^\top QP = \begin{bmatrix} v^\top Qv & v^\top Qw \\ w^\top Qv & w^\top Qw \end{bmatrix}.$$

Let $\lambda_2 \leq \lambda_1$ denote the eigenvalues of $P^\top QP$. Since Q has at most one positive eigenvalue, so does $P^\top QP$ by Lemma 1.4.6; in particular, $\lambda_2 \leq 0$. On the other hand, since the diagonal entry $v^\top Qv$ is positive, $P^\top QP$ has at least one nonnegative eigenvalue, i.e. $\lambda_1 \geq 0$. It follows that

$$w^\top \left((v^\top Qv) \cdot Q - (Qv)(Qv)^\top \right) w = \det \left(P^\top QP \right) = \lambda_1 \lambda_2 \leq 0.$$

Since $w \in \mathbb{R}^n$ was arbitrary, we conclude that $(v^\top Qv)Q - (Qv)(Qv)^\top$ is negative semidefinite.

(Item 5 $\implies$ Item 1) Suppose $(v^\top Qv)Q - (Qv)(Qv)^\top$ is negative semidefinite. Further subtracting $\frac{1}{d-1}(Qv)(Qv)^\top$ and scaling by the positive number $\frac{d(d-1)}{(v^\top Qv)^2}$ results in $\nabla^2 \log f(v)$ by Eq. (5.2) above, which must therefore also be negative semidefinite.

(Item 3 $\iff$ Item 4) Note that both conditions depend only on the matrix Q. If $w^\top Qw \neq 0$, then we can apply the equivalence (Item 2 $\iff$ Item 4) for the point w and the quadratic polynomial $q(x) = \frac{1}{2} x^\top Qx$, whose Hessian at any point is the matrix Q. Since Q is nonzero, any

point $w \in \mathbb{R}^n_{\geq 0}$ for which $w^\top Q w = 0$ can be obtained as a limit of points $u \in \mathbb{R}^n_{\geq 0}$ with $u^\top Q u \neq 0$. Since being negative semidefinite on a hyperplane $(Qu)^\perp$ is a closed condition on the set of $u \in \mathbb{R}^n$ with $Qu \neq 0$, this completes the proof.

(Item 1 $\iff$ Item 7) For $d \geq 3$, $\partial_v f$ is homogeneous of degree ≥ 2. Euler's Homogeneous Function Theorem (see Fact 1.4.9 and Corollary 1.4.10) applied to $\partial_i \partial_j f$ shows that the Hessian of $\partial_v f$ at v is a scalar multiple of the Hessian of f at v, namely $(d-2)\nabla^2 f(v)$. Thus by the equivalence (Item 1 $\iff$ Item 4), $\partial_v f$ is log-concave at v if and only if f is. □

5.1 0-Spectral Independence and Strong Log-Concavity

Our goal in this section is to prove Theorem 5.0.1. We will need the following intermediate results.

Lemma 5.1.1 (Derivatives of the Generating Polynomial)**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ where $\mathscr{U}$ is a finite ground set and $0 \leq n \leq |\mathscr{U}|$ is an integer. Then for every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k$, we have the identities*

$$(\partial^\sigma \log g_\mu)(\mathbf{1}) = (\partial^\sigma g_\mu)(\mathbf{1}) = \Pr_{\tau \sim \mu}[\tau \supseteq \sigma] = \binom{n}{k} \cdot \mu_k(\sigma).$$

Furthermore, the generating polynomial of the conditional distribution μ^σ is precisely

$$g_{\mu^\sigma} = \frac{\partial^\sigma g_\mu}{(\partial^\sigma g_\mu)(\mathbf{1})}.$$

Proof. The first identity follows immediately by inspection, since applying the differential operator ∂^σ annihilates any monomial $\mu(\tau)x^\tau$ if $\tau \not\supseteq \sigma$, and reduces the monomial $\mu(\tau)x^\tau$ to $\mu^\sigma(\tau \setminus \sigma)x^{\tau \setminus \sigma}$ if $\tau \supseteq \sigma$ (using the fact that $\mu^\sigma(\tau \setminus \sigma) = \mu(\tau)$). The second claim follows by a nearly identical line of reasoning; note the division by $(\partial^\sigma g_\mu)(\mathbf{1})$ is for normalization purposes (so that the coefficients sum to 1). □

Remark 25. More generally, if $\gamma = (\gamma_i)_{i \in \mathscr{U}} \in \mathbb{R}^{\mathscr{U}}_{\geq 0}$ is a collection of "external fields", then we have the identity

$$\gamma^\sigma \cdot (\partial^\sigma \log g_\mu)(\gamma) = \Pr_{\tau \sim \mu}[\tau \supseteq \sigma] = \binom{n}{k} \cdot (\gamma * \mu)_k(\sigma) \tag{5.3}$$

via the same argument. In general, the derivatives of the logarithm of g_μ (simply known as the *logarithmic derivatives of* g_μ) are more intimately related to marginal probabilities, since its derivative yields the correct normalization of $g_\mu(\gamma)$ in the denominator. In this case, we do not need to assume that the coefficients of g_μ sum to 1 (which we assumed in the equality $(\partial^\sigma g_\mu)(\mathbf{1}) = \Pr_{\tau \sim \mu}[\tau \supseteq \sigma]$ above).

Corollary 5.1.2 ((Normalized) Hessian of the Generating Polynomial)**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ where $\mathscr{U}$ is a finite ground set and $0 \leq n \leq |\mathscr{U}|$ is an integer. Then for every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k$, we have the identities*

$$\begin{aligned} \mu^\sigma_1 &= \frac{1}{n-k} \nabla g_{\mu^\sigma}(\mathbf{1}) \\ \mathcal{Q}_{\mu^\sigma} &= \frac{1}{n-k-1} \operatorname{diag}\left(\nabla g_{\mu^\sigma}(\mathbf{1})\right)^{-1} \nabla^2 g_{\mu^\sigma}(\mathbf{1}) \\ &= \frac{1}{n-k-1} \operatorname{diag}\left(\nabla \partial^\sigma g_\mu(\mathbf{1})\right)^{-1} \nabla^2 \partial^\sigma g_\mu(\mathbf{1}). \end{aligned} \tag{5.4}$$

Proof. This follows immediately by combining Lemma 5.1.1 and Fact 2.3.2. □

Proof of Theorem 5.0.1. Since differentiating the generating polynomial corresponds to conditioning (by Lemma 5.1.1), it suffices to prove that g_μ is log-concave at $\mathbf{1}$ if and only if $\lambda_2(\mathcal{Q}_\mu) \leq 0$. Since $\mathcal{Q}_\mu = \frac{1}{n-1} \operatorname{diag}\left(\nabla g_\mu(\mathbf{1})\right)^{-1} \nabla^2 g_\mu(\mathbf{1})$ by Corollary 5.1.2, the equivalence between log-concavity and $\lambda_2(\mathcal{Q}_\mu) \leq 0$ follows by combining Proposition 5.0.3 (Item 1 $\iff$ Item 6) with Lemma 1.4.8. □

Remark 26. One can actually also show that the (one-sided) influence matrix $\mathcal{I}$ is simply a normalization of the Hessian of $\log g_\mu$. In particular, we have

$$\mathcal{I}_\mu - I = \operatorname{diag}(\nabla g_\mu(\mathbf{1}))^{-1} \nabla^2 \log g_\mu(\mathbf{1}).$$

We conclude this subsection with a revised dictionary mapping out the synonymous terminology used in different research communities, where we have updated Table 2.1 to include a column for the geometry of polynomials.

Probability Theory	**High-Dimensional Expansion**	**Geometry of Polynomials**
Distribution μ	(Weighted) Simplicial Complex	Generating Polynomial g_μ
Support $\operatorname{supp}(\mu)$	Facets/Maximal Faces	Support $\operatorname{supp}(g_\mu)$
Cardinality of $\tau \in \operatorname{supp}(\mu)$	Dimension of τ $(+1)$	Degree of $\tau \in \operatorname{supp}(g_\mu)$
Homogeneous	Pure	Homogeneous
Conditioning	Link/Localization	Differentiation
Total Connectivity	Connectivity of All Links	Total Indecomposability
Marginal Probabilities	Localized Weight Function	(Logarithmic) Derivatives
Influence Matrix	Local Random Walk	(Normalized Log-)Hessian
Spectral Independence	Local Spectral Expansion	(Fractional) Log-Concavity

Table 5.1: Dictionary Between High-Dimensional Discrete Probability Distributions, Multivariate Polynomials, and High-Dimensional Expanders

5.2 Reduction to Quadratics: Proof of Theorem 5.0.2

The most interesting and nontrivial portion of this theorem is that total indecomposability and log-concavity for degree-2 derivatives implies completely log-concavity, which is the strongest, and requires log-concavity for every directional derivative and every degree. We establish this implication inductively. We crucially use the following proposition, which essentially gives the inductive step.

Proposition 5.2.1. *For $d \geq 3$, let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be an indecomposable d-homogeneous polynomial with nonnegative coefficients. Fix $w \in \mathbb{R}^n_{>0}$. If $\partial_i f$ is log-concave at w for every $i = 1, \ldots, n$, then so is $\partial_v f$ for every $v \in \mathbb{R}^n_{\geq 0}$.*

Proof of Theorem 5.0.2. Clearly, complete log-concavity implies strong log-concavity at w. Furthermore, strong log-concavity implies log-concavity for degree-2. To see that strong log-concavity implies total indecomposability, observe that if f is decomposable, i.e. $f = g + h$ for two nonzero d-homogeneous polynomials $g \in \mathbb{R}[x_i : i \in S], h \in \mathbb{R}[x_i : i \in [n] \setminus S]$ and a nonempty proper subset $\emptyset \subsetneq S \subsetneq [n]$ of variables, then we may write $\nabla^2 f$ as the block matrix

$$\nabla^2 f = \begin{bmatrix} \nabla^2 g & 0 \\ 0 & \nabla^2 h \end{bmatrix}.$$

Since f, g, h are nonzero and have nonnegative coefficients, $\nabla^2 g, \nabla^2 h$ each have at least one positive eigenvalue, so $\nabla^2 f$ must have at least two positive eigenvalues, contradicting log-concavity of f. The same reasoning applies to each partial derivative $\partial^\alpha f$ of f. Thus, all that remains is to show that total indecomposability plus log-concavity for degree-2 derivatives imply complete log-concavity.

We induct on $d = \deg(f)$. The case $d = 2$ is immediate, so let $d \geq 3$. First, we claim that for any positive vector $v \in \mathbb{R}^n_{>0}$, $\partial_v f$ is also indecomposable. To see this, suppose $\emptyset \subsetneq S \subsetneq [n]$ is some subset of the variables. Since f is indecomposable, f contains a monomial of the form $c_\alpha z^\alpha$, where $\operatorname{supp}(\alpha)$ has nonempty intersection with both S and $[n] \setminus S$, and $c_\alpha > 0$. Since f is homogeneous of degree $d \geq 3$ and v has strictly positive entries, $\partial_v(c_\alpha z^\alpha)$ also has a monomial whose support has nonempty intersection with both S and $[n] \setminus S$. Hence, so does $\partial_v f$. As this holds for all $\emptyset \subsetneq S \subsetneq [n]$, we conclude that $\partial_v f$ is also indecomposable.

Let $v_1, \ldots, v_k \in \mathbb{R}^n_{\geq 0}$, which we may assume are all nonzero, as otherwise there is nothing to prove. Our goal is to show $\partial_{v_1} \cdots \partial_{v_k} f$ is log-concave at w. If $k \geq d - 1$, then $\partial_{v_1} \cdots \partial_{v_k} f$ is either

identically zero or linear with nonnegative coefficients, in which case it is log-concave on $\mathbb{R}^n_{\geq 0}$, so assume $0 \leq k \leq d-2$. By Proposition 5.0.3 (Item 7 $\implies$ Item 1), to show that f is log-concave at w, it suffices to show that $\partial_w f$ is log-concave at w. This reduces the case $k = 0$ to the case $k = 1$.

Now suppose $1 \leq k \leq d-2$. By induction, $\partial_j f$ is completely log-concave at w for all $j = 1, \ldots, n$, and hence $\partial_{v_1} \cdots \partial_{v_{k-1}} \partial_j f = \partial_j \partial_{v_1} \cdots \partial_{v_{k-1}} f$ is log-concave at w. Since $\partial_{v_1} \cdots \partial_{v_{k-1}} f$ is indecomposable and has degree $d - k + 1 \geq 3$, it follows from Proposition 5.2.1 that $\partial_{v_1} \cdots \partial_{v_k} f$ is log-concave at w. □

All that remains is to prove Proposition 5.2.1. We will provide two proofs. Our first proof goes through an extension of the argument used to prove Oppenheim's Trickle-Down Theorem (see Theorem 3.2.1). This was the argument used in [Ana+19].

First Proof of Proposition 5.2.1. By a limiting argument, we may assume $v \in \mathbb{R}^n_{>0}$ without loss of generality. Define vectors $\pi_v, \pi_1, \ldots, \pi_n \in \mathbb{R}^n$ by

$$\pi_v(j) \stackrel{\text{def}}{=} \sum_{k=1}^n v_j v_k \cdot \partial_j \partial_k \partial_v f(w) = v_j \cdot \partial_j \partial_v^2 f(w)$$
$$\pi_i(j) \stackrel{\text{def}}{=} \sum_{k=1}^n v_j v_k \cdot \partial_j \partial_k \partial_i f(w) = v_j \cdot \partial_i \partial_j \partial_v f(w),$$

and consider the random walk matrices

$$Q_v \stackrel{\text{def}}{=} \operatorname{diag}(\pi_v)^{-1} \cdot \operatorname{diag}(v) \left(\nabla^2 \partial_v f(w)\right) \operatorname{diag}(v)$$
$$Q_i \stackrel{\text{def}}{=} \operatorname{diag}(\pi_i)^{-1} \cdot \operatorname{diag}(v) \left(\nabla^2 \partial_i f(w)\right) \operatorname{diag}(v),$$

which are self-adjoint w.r.t. the inner products $\langle \phi, \psi \rangle_v \stackrel{\text{def}}{=} \phi^\top \operatorname{diag}(\pi_v) \psi$ and $\langle \phi, \psi \rangle_i \stackrel{\text{def}}{=} \phi^\top \operatorname{diag}(\pi_i) \psi$, respectively. Note that $\pi_v = \sum_{i=1}^n v_i \cdot \pi_i$. This is highly reminiscent of Eq. (5.4), and are analogs of the "local walks". To show log-concavity of $\partial_v f$ at w, it suffices to show that Q_v has at most one positive eigenvalue, by Proposition 5.0.3 and Lemma 1.4.8.

Since $\nabla^2 \partial_v f(w) = \sum_{i=1}^n v_i \nabla^2 \partial_i f(w)$, by linearity, we have that

$$\langle \phi, Q_v \phi \rangle_v = \sum_{i=1}^n v_i \cdot \langle \phi, Q_i \phi \rangle_i$$

for every test vector $\phi \in \mathbb{R}^n$. Since $\partial_i f$ is log-concave, $\nabla^2 \partial_i f(w)$ has at most one positive eigenvalue by Proposition 5.0.3, whence Q_i has at most one positive eigenvalue by Lemma 1.4.8. By Lemma 1.4.20, we deduce the upper bound

$$\langle \phi, Q_v \phi \rangle_v \leq \sum_{i=1}^n v_i \cdot \frac{\langle \mathbf{1}, \phi \rangle_i^2}{\langle \mathbf{1}, \mathbf{1} \rangle_i} \underset{(*)}{=} \langle Q_v \phi, Q_v \phi \rangle_v.$$

If we can justify the final equality $(*)$, then taking ϕ to be a right eigenvector of Q_v corresponding to the second largest eigenvalue $\lambda_2(Q_v)$, we have the inequality $\lambda_2(Q_v) \leq \lambda_2(Q_v)^2$. Since $\partial_v f$ is indecomposable, Q_v is connected, and so $\lambda_2(Q_v) < 1$. Combined with $\lambda_2(Q_v) \leq \lambda_2(Q_v)^2$ yields $\lambda_2(Q_v) \leq 0$ as desired.

All the remains is to justify $(*)$, which we do now. Observe that

$$v_i \cdot \langle \mathbf{1}, \phi \rangle_i = \sum_{j=1}^n v_i v_j \cdot \partial_i \partial_j \partial_v f(w) \cdot \phi(j) = \left(\operatorname{diag}(v) \left(\nabla^2 \partial_v f(w)\right) \operatorname{diag}(v) \phi\right)(i)$$
$$v_i \cdot \langle \mathbf{1}, \mathbf{1} \rangle_i = \sum_{j=1}^n v_i v_j \cdot \partial_i \partial_j \partial_v f(w) = v_i \cdot \partial_i \partial_v^2 f(w) = \pi_v(i).$$

It follows that

$$\begin{aligned}\sum_{i=1}^n v_i \cdot \frac{\langle \mathbf{1}, \phi\rangle_i^2}{\langle \mathbf{1}, \mathbf{1}\rangle_i} &= \sum_{i=1}^n v_i \cdot \langle \mathbf{1}, \mathbf{1}\rangle_i \cdot \left(\frac{v_i \cdot \langle \mathbf{1}, \phi\rangle_i}{v_i \cdot \langle \mathbf{1}, \mathbf{1}\rangle_i}\right)^2 \\ &= \sum_{i=1}^n \pi_v(i) \cdot \left(\frac{\left(\operatorname{diag}(v)\left(\nabla^2 \partial_v f(w)\right)\operatorname{diag}(v)\phi\right)(i)}{\pi_v(i)}\right)^2 \\ &= \sum_{i=1}^n \pi_v(i) \cdot (Q_v\phi)(i)^2 \\ &= \langle Q_v\phi, Q_v\phi\rangle_v,\end{aligned}$$

yielding $(*)$ as desired. □

Our second proof of Proposition 5.2.1 works more directly with polynomials, and in particular, leverages the many equivalent formulations of log-concavity from Proposition 5.0.3. This was the argument used in [Ana+18a]. We start with the following lemma, which allows us to iteratively add partial derivatives to form directional derivatives while preserving log-concavity along the way.

Lemma 5.2.2. *Let $f, g \in \mathbb{R}[x_1, \ldots, x_n]$ be homogenous polynomials with nonnegative coefficients satisfying $\partial_u f = \partial_v g \neq 0$ for some vectors $u, v \in \mathbb{R}^n_{\geq 0}$. If f and g are log-concave at $w \in \mathbb{R}^n_{>0}$, then so is $f + g$.*

Proof. The assumption that $\partial_u f = \partial_v g \neq 0$ implies that f and g have the same degree d. If $d = 1$, then $f + g$ is a linear form with nonnegative coefficients, which is automatically log-concave on $\mathbb{R}^n_{\geq 0}$. Now suppose $d \geq 2$. If $Q_f = \nabla^2 f(w)$ and $Q_g = \nabla^2 g(w)$, then $\partial_u f = \partial_v g$ implies that for each $i = 1, \ldots, n$,

$$(Q_f u)_i = (\partial_i \partial_u f)(w) = (\partial_i \partial_v g)(w) = (Q_g v)_i,$$

showing that $Q_f u = Q_g v$. Since $\partial_u f$ has nonnegative coefficients and is not identically zero, we also have that $\partial_u f(w) \neq 0$, meaning that $Q_f u \neq 0$. By Proposition 5.0.3 (Item 1 $\implies$ Item 3) and log-concavity of f and g at w, each quadratic form $x \mapsto x^\top Q_f x$, $x \mapsto x^\top Q_g x$ is negative semidefinite on $(Q_f u)^\perp = (Q_g v)^\perp$. It follows that their sum $x \mapsto x^\top (Q_f + Q_g)x$ given by the matrix $Q_f + Q_g = \nabla^2(f + g)\big|_{x=w}$ is also negative semidefinite on this $(n-1)$-dimensional linear space, so by Proposition 5.0.3 (Item 4 $\implies$ Item 1), $f + g$ is log-concave at w. □

Second Proof of Proposition 5.2.1. If $\partial_i f$ is identically zero for some i, then we can consider f as a polynomial in the other variables. Without loss of generality, we can assume that $\partial_i f$ is nonzero for all i, and if necessary relabel $x_1, \ldots, x_n$ so that that for every $2 \leq j \leq n$, there exists $i < j$ for which $\partial_i \partial_j f \neq 0$. Such a relabeling is possible by indecomposability (for instance, one can take a breadth-first or depth-first traversal of the graph with vertex set $[n]$ and edges $\{i, j\}$ corresponding to pairs satisfying $\partial_i \partial_j f \neq 0$).

Again, fix $v \in \mathbb{R}^n_{\geq 0}$. If $v = 0$, there is nothing to prove, so assume v has at least one positive coordinate. Without loss of generality, say, by permuting the variables, we may assume that $v = (v_1, \ldots, v_k, 0, \ldots, 0)$ for some $k \geq 1$ and $v_1, \ldots, v_k > 0$. We show by induction on $1 \leq t \leq k$ that for any $1 \leq t \leq k$, $\sum_{i=1}^t v_i \partial_i f$ is log-concave at w. The case $t = 1$ follows by assumption while the case $t = k$ yields the claim. For $1 \leq t < k$, let u denote the truncation of v to its first t coordinates, i.e. $u = (v_1, \ldots, v_t, 0, \ldots, 0)$, and let a denote the vector $v_{t+1}\mathbf{1}_{\{t+1\}}$. By induction, both $\partial_u f$ and $\partial_a f = v_{t+1}\partial_{t+1} f$ are log-concave at w, and

$$\partial_a \partial_u f = \partial_u \partial_a f = \sum_{i=1}^t v_i v_{t+1} \partial_i \partial_{t+1} f.$$

Since the coefficients of each summand are positive and $\partial_i \partial_{t+1} f$ is nonzero for some $1 \leq i \leq t$, this sum is also nonzero. Then by Lemma 5.2.2, $\partial_u f + \partial_a f = \sum_{i=1}^{t+1} v_i \partial_i f$ is log-concave at w. For $t = k - 1$, this is exactly $\partial_v f$. □

5.3 Closure Properties of Log-Concave Polynomials

Much like the theory of real stability (see Section 5.7.1), the collection of log-concave polynomials also enjoys many natural and useful *closure properties*, i.e. they admit a rich class of operations which preserve log-concavity. This section is devoted to the study of some of these operations. For instance, one such operation called *polarization* is particularly useful in developing new fast mixing Markov chains for sampling which previously had not been studied. Indeed, many of the algorithmic results in Chapter 12 are based on this operation. A more complete theory of the closure properties of log-concave polynomials was eludicated in [BH20].

One of the basic operations that preserves (strong) log-concavity is an affine change of coordinates. This was first proved in [AOV21], but for completeness we include the proof here.

Lemma 5.3.1 ([AOV21])**.** *If $f \in \mathbb{R}[z_1, \dots, z_n]$ is completely log-concave on $\mathbb{R}^n_{\geq 0}$ and $T : \mathbb{R}^m \to \mathbb{R}^n$ is an affine transformation such that $T(\mathbb{R}^m_{\geq 0}) \subseteq \mathbb{R}^n_{\geq 0}$, then $f(T(y_1, \dots, y_m)) \in \mathbb{R}[y_1, \dots, y_m]$ is completely log-concave on $\mathbb{R}^m_{\geq 0}$.*

Proof. First, we prove that if f is log-concave on $\mathbb{R}^n_{\geq 0}$, then $f \circ T = f(T(y_1, \dots, y_m))$ is log-concave on $\mathbb{R}^m_{\geq 0}$. By assumption, for any $u, v \in \mathbb{R}^m_{\geq 0}$, we have $T(u), T(v) \in \mathbb{R}^n_{\geq 0}$. Thus for any $0 < \lambda < 1$,

$$f(T(\lambda u + (1-\lambda)v)) = f(\lambda T(u) + (1-\lambda)T(v)) \geq f(T(u))^{\lambda} f(T(v))^{1-\lambda}.$$

Therefore $f \circ T$ is log-concave on $\mathbb{R}^m_{\geq 0}$.

Now suppose that f is completely log-concave and let $v_1, \dots, v_k \in \mathbb{R}^m_{\geq 0}$. Since $T(\mathbb{R}^m_{\geq 0}) \subseteq \mathbb{R}^n_{\geq 0}$ and T is affine, $T(x) = Ax + b$ for some $A \in \mathbb{R}^{n \times m}_{\geq 0}$ and $b \in \mathbb{R}^n_{\geq 0}$. In particular, $Av_1, \dots, Av_k \in \mathbb{R}^n_{\geq 0}$, which means that $\partial_{Av_1} \cdots \partial_{Av_k} f$ is log-concave over $\mathbb{R}^n_{\geq 0}$. By the Chain Rule for differentiation, we have

$$\partial_{v_1} \cdots \partial_{v_k} (f \circ T) = (\partial_{Av_1} \cdots \partial_{Av_k} f) \circ T.$$

Since composition with T preserves log-concavity, this polynomial is log-concave over $\mathbb{R}^m_{\geq 0}$. Since this holds for all k and all $v_1, \dots, v_k \in \mathbb{R}^m_{\geq 0}$, $f \circ T$ is completely log-concave over $\mathbb{R}^m_{\geq 0}$ as desired. □

Additional basic strong log-concavity preservers are the following.

Proposition 5.3.2. *The following basic operations preserve strong log-concavity.*

1. *(Coefficient-Wise Limits) If $\{p_t\}_{t=1}^{\infty} \subseteq \mathbb{R}[z_1, \dots, z_n]$ is a sequence of strongly log-concave polynomials on $\mathbb{R}^n_{\geq 0}$ converging coefficient-wise to some fixed polynomial $p \in \mathbb{R}[z_1, \dots, z_n]$, then p is strongly log-concave on $\mathbb{R}^n_{\geq 0}$.*
2. *(External Fields) If $p \in \mathbb{R}[z_1, \dots, z_n]$ is a strongly log-concave polynomial on $\mathbb{R}^n_{\geq 0}$ and $v \in \mathbb{R}^n_{\geq 0}$ is an external field, then the polynomial $p(v_1 z_1, \dots, v_n z_n)$ is also strongly log-concave on $\mathbb{R}^n_{\geq 0}$.*
3. *(Products) If $p, q \in \mathbb{R}[z_1, \dots, z_n]$ are homogeneous strongly log-concave polynomials on $\mathbb{R}^n_{\geq 0}$, then so is their product $p \cdot q$.*

Proof. For Item 1, note that coefficient-wise convergence implies that the Hessian of p_t evaluated at any point converges to the Hessian of p evaluated at that point. To handle derivatives, we simply note that for any collection of vectors $v_1, \dots, v_k \in \mathbb{R}^n_{\geq 0}$, the polynomials $\partial_{v_1} \cdots \partial_{v_k} p_t$ converge coefficient-wise to $\partial_{v_1} \cdots \partial_{v_k} p$ as $t \to \infty$. Item 2 is an immediate consequence of Lemma 5.3.1 by taking the affine transformation $T : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ to be $T(z) = \mathrm{diag}(v) \cdot z$.

For Item 3, first note that it is straightforward to verify that $p(z) \cdot q(w) \in \mathbb{R}[z_1, \dots, z_n, w_1, \dots, w_n]$ is strongly log-concave on $\mathbb{R}^{2n}_{\geq 0}$. This can be done by direct inspection. The desired claim then follows again from Lemma 5.3.1 by applying the linear transformation

$$T(z, w) = \begin{bmatrix} I & 0 \\ I & 0 \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} z \\ z \end{bmatrix},$$

whose composition with $p(z) \cdot q(w)$ is precisely the desired product $p(z) \cdot q(z)$. □

We will also need the polarization operation, which converts a homogeneous polynomial which isn't multiaffine, into an "equivalent" homogeneous polynomial which is multiaffine. This operation is well-known to preserve real stability of a polynomial [BB09b] (see Section 5.7.1 for a discussion of real stable polynomials). It was shown in [BH20] that this operations also preserves strong log-concavity. This will have a number of useful algorithmic consequences below.

Proposition 5.3.3 (Polarization; [BH20]). *For an element κ of $\mathbb{N}^n$, let*

$$\mathbb{R}_\kappa[z_1,\dots,z_n] = \{\textit{polynomials in } \mathbb{R}[z_1,\dots,z_n] \textit{ of degree at most } \kappa_i \textit{ in } z_i,\ \forall 1 \le i \le n\}$$
$$\mathbb{R}^a_\kappa[z_{ij}] = \{\textit{multiaffine polynomials in } \mathbb{R}[z_{ij}]_{1\le i\le n, 1\le j\le \kappa_i}\}.$$

*The **polarization map** $\prod_\kappa^\uparrow$ is a linear map that sends the monomial $z^\alpha = \prod_{i=1}^n z_i^{\alpha_i}$ to the product*

$$\frac{1}{\binom{\kappa}{\alpha}} \prod_{i=1}^n e_{\alpha_i}\left(\{z_{ij}\}_{1\le j\le \kappa_i}\right)$$

where $e_{\alpha_i}(\{z_{ij}\}_{1\le j\le\kappa_i})$ is the elementary symmetric polynomial of degree α_i in the variables $\{z_{ij}\}_{1\le j\le\kappa_i}$ and $\binom{\kappa}{\alpha} = \prod_{i=1}^n \binom{\kappa_i}{\alpha_i}$. If $g \in \mathbb{R}_\kappa[z_1,\dots,z_n]$ is strongly log-concave, then $\prod_\kappa^\uparrow(g)$ is also strongly log-concave.

The following lemma is also useful. It says for multiaffine homogeneous polynomials, log-concavity on $\mathbb{R}^n_{\ge 0}$ is equivalent to the seemingly stronger notions of strong log-concavity and complete log-concavity on $\mathbb{R}^n_{\ge 0}$.

Lemma 5.3.4. *Let $g \in \mathbb{R}[z_1,\dots,z_n]$ be a multiaffine homogeneous polynomial with nonnegative coefficients. If g is log-concave on all of $\mathbb{R}^n_{\ge 0}$, then it is completely log-concave on all of $\mathbb{R}^n_{\ge 0}$.*

Proof. For a multiaffine polynomial g, its partial derivatives can be obtained as

$$\partial_1 g = \lim_{t\to\infty} \frac{g(t, z_2, \dots, z_n)}{t}.$$

This shows that the derivatives of a multiaffine log-concave polynomial are limits of log-concave polynomials, which themselves are log-concave. □

5.3.1 Geometric Scaling of Coefficients

Finally, we show that the class of homogeneous multiaffine strongly log-concave polynomials is closed under raising all coefficients to a fixed exponent less than 1. This was first shown in [Ana+19], and has applications to sampling from *"smoothed" determinantal point processes* which we further discuss below. Intriguingly, this operation does *not* preserve real stability.

Theorem 5.3.5. *Let $f(x_1,\dots,x_n) = \sum_{S\subseteq[n]} c_S \prod_{i\in S} x_i$ be a degree-k homogeneous multiaffine with nonnegative coefficients. If f is strongly log-concave at a point $v \in \mathbb{R}^n_{>0}$, then so is the polynomial $f_\alpha(x_1,\dots,x_n) \stackrel{\text{def}}{=} \sum_{S\subseteq[n]} c_S^\alpha \prod_{i\in S} x_i$, for every $0 \le \alpha \le 1$.*

As our primary application, we use the above theorem to design a sampling algorithm for "smoothed" determinantal point processes. A *determinantal point process (DPP)* on a set of elements $[n]$ is a probability distribution $\mu : 2^{[n]} \to \mathbb{R}_{\ge 0}$ identified by a symmetric positive semidefinite matrix $L \in \mathbb{R}^{n\times n}$, where for any $S \subseteq [n]$ we have

$$\mu(S) \propto \det(L_S),$$

where L_S is the principal sub-matrix of L indexed by the elements of S. Determinantal point processes are fundamental to the study of a variety of tasks in machine learning, including text summarization, image search, news threading, and diverse feature selection [see, e.g., KT12]. A *k-determinantal point process (k-DPP)* is a determinantal point process conditioned on the sets S having size k.

Given a positive semidefinite matrix L, let μ be the corresponding k-DPP. We have

$$g_\mu(x) \propto \sum_{S\in\binom{[n]}{k}} \det(L_S) \cdot \prod_{i\in S} x_i.$$

It turns out the above polynomial is real stable, which in particular implies various useful negative dependence properties for the distribution μ as shown in [BBL09]. Since real stability implies strong log-concavity by Theorem 5.7.4, combining this with Theorem 5.3.5 above yields the following log-concavity result.

Corollary 5.3.6 ("Smoothed" k-DPP)**.** *For every positive semidefinite matrix $L \succeq 0$ and exponent $0 \leq \alpha \leq 1$, the polynomial*

$$\sum_{S \in \binom{[n]}{k}} \det(L_S)^\alpha \prod_{i \in S} x_i$$

is strongly log-concave.

Previously, Anari, Oveis Gharan, and Rezaei [AOR16] showed that a natural Markov chain with the Metropolis rule mixes rapidly and generates a random sample from k-DPP distributions, corresponding to the case $\alpha = 1$. By combining Corollary 5.3.6 with Theorems 2.3.1 and 5.0.1, we are able to extend efficient sampling to all "smoothed" k-DPP distributions, where for any set S, $\mu(S) \propto \det(L_S)^\alpha$ for some $0 \leq \alpha \leq 1$, again via a simple Markov chain. The weights $\det(L_S)^\alpha$ may be thought of as a way to interpolate between two extremes for selecting diverse data points. When $\alpha = 0$, we recover the uniform distribution; increasing α then increases the distribution's preference for diverse sets.

We also note that for $\alpha = 1/2$, it is known that Corollary 5.3.6 follows from the Brunn-Minkowski Theorem applied to appropriately defined *zonotopes*.[1] For $\alpha = 1$ as mentioned earlier, the above polynomial is actually real stable, and hence strongly log-concave. Theorem 5.3.5 gives a unified proof that all of these polynomials are strongly log-concave. This operation of "exponentiating" such probability distributions was further studied in [MSJ18].

Proof of Theorem 5.3.5. If $k = 0, 1$, the claim is obvious so assume $k \geq 2$. The claim is obvious when $\alpha = 1$, and the case $\alpha = 0$ follows by taking coefficient-wise limits as $\alpha \to 0$. Hence, we will also assume $0 < \alpha < 1$. Finally, we will assume that all coefficients c_S are strictly positive. The result for general strongly log-concave polynomials then follows by taking coefficient-wise limits.

By Theorem 5.0.2, we must prove that $\partial_T f_\alpha$ is log-concave for every $T \in \binom{[n]}{k-2}$. By Proposition 5.0.3, this is equivalent to showing that $\nabla^2 \partial_T f_\alpha$ has at most one positive eigenvalue. Note that since all coefficients are strictly positive, total indecomposability is immediately satisfied. Observe that we may concisely write

$$\nabla^2 \partial_T f = \left[c_{T \cup \{i,j\}}\right]_{ij}$$
$$\nabla^2 \partial_T f_\alpha = \left[c^\alpha_{T \cup \{i,j\}}\right]_{ij}$$

As $\nabla^2 \partial_T f$ has at most one positive eigenvalue, and all entries are nonnegative, we may write

$$\nabla^2 \partial_T f = vv^\top - A$$

for some vector $v \in \mathbb{R}^n$ and a positive semidefinite matrix A. Note that since $\nabla^2 \partial_T f$ has strictly positive entries, the Perron-Frobenius Theorem (see Theorem 1.4.4) tells us that the entries of v are strictly positive. In particular, $c_{T \cup \{i,j\}} = v_i v_j - A_{ij} > 0$ where $v_i v_j > 0$. Our goal is to write

$$c^\alpha_{T \cup \{i,j\}} = (v_i v_j - A_{ij})^\alpha = v_i^\alpha v_j^\alpha \left(1 - \frac{A_{ij}}{v_i v_j}\right)^\alpha$$

and then Taylor expand $\left(1 - \frac{A_{ij}}{v_i v_j}\right)^\alpha$. Consider the function $\varphi_\alpha(x) = (1-x)^\alpha$, whose Taylor expansion about zero is

$$\sum_{\ell=0}^{\infty} \frac{\prod_{j=0}^{\ell-1}(\alpha - j)}{\ell!} \cdot (-1)^\ell x^\ell = \sum_{\ell=0}^{\infty} \left(\prod_{j=0}^{\ell-1} \frac{\alpha - j}{1+j}\right) \cdot (-1)^\ell x^\ell = 1 - \sum_{\ell=1}^{\infty} \left(\prod_{j=0}^{\ell-1} \left|\frac{\alpha - j}{1+j}\right|\right) \cdot x^\ell,$$

where for the last equality, we crucially use the fact that $0 < \alpha < 1$. The interval of convergence of this power series contains $(-1, 1)$, since if $a_\ell = (-1)^\ell \prod_{j=0}^{\ell-1} \frac{\alpha - j}{1+j}$, then

$$\left|\frac{a_{\ell+1} x^{\ell+1}}{a_\ell x^\ell}\right| = |x| \cdot \left|\frac{\alpha - \ell}{1+\ell}\right| \to |x| \quad \text{as} \quad \ell \to \infty$$

[1] A zonotope is a Minkowski sum of finitely many, say centrally symmetric, segments $[-v, v] \subseteq \mathbb{R}^n$.

gives a radius of convergence of 1 by the Ratio Test. We verify that this power series also converges at $x \in \{-1, 1\}$ so that we have convergence on the entire closed interval $[-1, 1]$. We can then apply this to $\left(1 - \frac{A_{ij}}{v_i v_j}\right)^\alpha$, since $\nabla^2 \partial_T f$ having nonnegative entries and A being positive semidefinite imply that

$$\det(A_{S,S}) = A_{ii}A_{jj} - A_{ij}^2 \geq 0 \implies |A_{ij}| \leq \sqrt{A_{ii}A_{jj}} = v_i v_j$$

This shows $\frac{A_{ij}}{v_i v_j} \in [-1, 1]$.

To show that the power series converges at $x \in \{-1, 1\}$, it suffices to prove absolute convergence, i.e. that $\sum_{\ell=1}^\infty \prod_{j=0}^{\ell-1} \left|\frac{\alpha - j}{1+j}\right| < \infty$. For this, we first bound each term for $k > 1$ as

$$\prod_{j=0}^{\ell-1} \left|\frac{\alpha - j}{1+j}\right| = \frac{\alpha}{\ell} \cdot \frac{\prod_{j=1}^{\ell-1}(j-\alpha)}{\prod_{j=1}^{\ell-1} j} = \frac{\alpha}{\ell} \prod_{j=1}^{\ell-1} \left(1 - \frac{\alpha}{j}\right) \leq \frac{\alpha}{\ell} \cdot e^{-\alpha \sum_{j=1}^{\ell-1} \frac{1}{j}} \leq \frac{\alpha}{\ell} \cdot e^{-\alpha \log(\ell-1)} < \frac{\alpha}{(\ell-1)^{1+\alpha}}$$

Hence,

$$\sum_{\ell=1}^\infty \prod_{j=0}^{\ell-1} \left|\frac{\alpha - j}{1+j}\right| < \alpha + \alpha \sum_{\ell>1}^\infty \frac{1}{(\ell-1)^{1+\alpha}} < \infty$$

where we crucially use that $\alpha > 0$.

Having verified that the power series is valid for every entry of our matrix, we have

$$\begin{aligned}
\nabla^2 \partial_T f_\alpha &= \left[v_i^\alpha v_j^\alpha\right]_{ij} \odot \left(xx^\top - \sum_{\ell=1}^\infty \left(\prod_{j=0}^{\ell-1} \left|\frac{\alpha - j}{1+j}\right|\right) \cdot \left[\frac{A_{ij}}{v_i v_j}\right]^{\odot \ell}\right) \\
&= \underbrace{\left[v_i^\alpha v_j^\alpha\right]_{ij}}_{(1)} - \underbrace{\sum_{\ell=1}^\infty \left(\prod_{j=0}^{\ell-1} \left|\frac{\alpha - j}{1+j}\right|\right) \cdot \left(\left[v_i^\alpha v_j^\alpha\right]_{ij} \odot \left[\frac{A_{ij}}{v_i v_j}\right]^{\odot \ell}\right)}_{(2)}.
\end{aligned}$$

Here, we recall that $A \odot B$ denotes the Hadamard product of A, B, where $(A \odot B)_{ij} = A_{ij}B_{ij}$. Similarly, $A^{\odot \ell}$ denotes the ℓ-iterated Hadamard product of A with itself.

All we must do is prove that (1) and (2) are both positive semidefinite, and that (1) is rank-1. Observe that

$$[v_i^\alpha v_j^\alpha]_{ij} = [v_i^\alpha]_i \cdot [v_i^\alpha]_i^\top \qquad \left[\frac{1}{v_i v_j}\right]_{ij} = \left[\frac{1}{v_i}\right]_i \cdot \left[\frac{1}{v_i}\right]_i^\top$$

This tells us (1) is positive semidefinite and rank-1. For the second, observe that

$$\left[\frac{A_{ij}}{v_i v_j}\right]_{ij} = A \odot \left[\frac{1}{v_i v_j}\right]_{ij}$$

As $A \succeq 0$ by assumption, this matrix is positive semidefinite by the Schur Product Theorem (see Theorem 1.4.3). Again, inductively applying the Schur Product Theorem, we have $\left[v_i^\alpha v_j^\alpha\right]_{ij} \odot \left[\frac{A_{ij}}{v_i v_j}\right]^{\odot \ell} \succeq 0$ for every ℓ. As (2) is a nonnegative linear combination of positive semidefinite matrices, it is positive semidefinite. □

Remark 27. Note that this operation does not preserve complete log-concavity when f is not assumed to be multiaffine. For example, consider the degree-2 bivariate polynomial $f(x, y) = ax^2 + bxy + cy^2$, where $a, b, c > 0$. Here,

$$\nabla^2 f = \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix}$$

so log-concavity amounts to $\det(\nabla^2 f) = 4ac - b^2 \leq 0$, i.e. $b^2 \geq 4ac$. Now, raise each coefficient to the power α. Then,

$$\nabla^2 f_\alpha = \begin{bmatrix} 2a^\alpha & b^\alpha \\ b^\alpha & 2c^\alpha \end{bmatrix}$$

so log-concavity amounts to $\det(\nabla^2 f_\alpha) = 4a^\alpha c^\alpha - b^{2\alpha} \leq 0$, i.e. $b^2 \geq 4^{1/\alpha}ac$. Clearly, as one decreases α to 0, this inequality gets stronger, which certainly isn't implied by log-concavity of f. The problem lies in the fact that when you differentiate a monomial that contains variables with multiplicities, you will obtain "factorial coefficients" which are not raised to the power α. The operation must be modified appropriately to take this into account; see [BH20] for further discussion.

5.4 The Random Cluster Model and the Tutte Polynomial

Another application of this theory is estimating the partition function of the *random cluster model* from statistical mechanics due to Fortuin and Kasteleyn [For71; FK72; For72a; For72b]. For an n-element rank-r matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ and parameters $0 \leq p \leq 1, q \geq 0$, the *random cluster measure* on $2^{\mathscr{U}}$ is defined by

$$\mu^{\mathsf{RC}}_{\mathcal{M},p,q}(\tau) \propto q^{r+1-\mathrm{rank}(\tau)} p^{|\tau|}(1-p)^{n-|\tau|}, \quad \forall \tau \subseteq \mathscr{U} \tag{5.5}$$

with associated partition function

$$\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}(p,q) \stackrel{\mathsf{def}}{=} \sum_{\tau \subseteq \mathscr{U}} q^{r+1-\mathrm{rank}(\tau)} p^{|\tau|}(1-p)^{n-|\tau|}. \tag{5.6}$$

We refer interested readers to the monograph by Grimmett [Gri09], and references therein, for further information on the random cluster model. Typically, one considers the special case where $\mathcal{M}$ is a graphic matroid, in which case the exponent of q is simply the number of connected components of S.

In fact, the polynomial $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}$ is closely related to the *Tutte polynomial*

$$\mathcal{T}_{\mathcal{M}}(x,y) \stackrel{\mathsf{def}}{=} \sum_{\tau \subseteq \mathscr{U}} (x-1)^{r-\mathrm{rank}(\tau)} (y-1)^{|\tau|-\mathrm{rank}(\tau)}$$

in combinatorics. Indeed, we can write

$$\mathcal{T}_{\mathcal{M}}(x,y) = \frac{1}{(x-1)(y-1)^{r+1}} \cdot \mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}\left(y-1, (x-1)(y-1)\right).$$

Hence, an FPRAS for estimating $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}(p,q)$ for $p \geq 0$ and $0 \leq q \leq 1$ gives an FPRAS for estimating $\mathcal{T}_{\mathcal{M}}(x,y)$ in the region described by the inequalities $y \geq 1$ and $0 \leq (x-1)(y-1) \leq 1$.

There is an extensive literature on hardness of exact computation and inapproximability of the Tutte polynomial and the partition function of the random cluster model. It is known that exact computation of the Tutte polynomial for a graph is $\#\mathsf{P}$-hard at all points (x,y) except at $(1,1), (-1,-1), (0,-1), (-1,0)$, along the hyperbola $(x-1)(y-1) = 1$, and for planar graphs, along the hyperbola $(x-1)(y-1) = 2$ [JVW90; Ver91; Wel94]. In the realm of inapproximability, it is known that even for planar graphs, there is no FPRAS to approximate the Tutte polynomial for $x > 1, y < -1$ or $y > 1, x < -1$ assuming $\mathsf{NP} \neq \mathsf{RP}$ [GJ08; GJ12b]. Furthermore, there is no FPRAS for estimating the partition function $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}$ of the random cluster model on general graphic matroids when $q > 2$, nor is there an FPRAS for $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}$ at $q = 2$ for general binary matroids, unless there is an FPRAS for counting independent sets in a bipartite graph [GJ12a; GJ13; GJ14].

Prior to our work [Ana+19], one could only compute $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}$ when $q = 1$ (trivial), or when $q = 2$ due to the close connection with the ferromagnetic Ising model [JS93; GJ18]. Our next result gives an FPRAS which estimates $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}(p,q)$ for any $0 < q \leq 1$ and $0 \leq p \leq 1$. We note that the random cluster measure for $0 < q < 1$ has historically been difficult to analyze since it lacks *monotonicity*, unlike the case when $q \geq 1$. The celebrated *Fortuin–Kasteleyn–Ginibre (FKG) Inequality* does not hold for $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ when $0 < q < 1$, and establishing negative association properties for $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ remains an outstanding open problem in this regime. We will establish our FPRAS by showing that $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ is spectrally independent; in particular, the correlations in $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ can be controlled spectrally, and they satisfy a form of *spectrally negative correlations*, a property already discussed previously in Chapter 3.

Theorem 5.4.1 (Local Markov Chain for Full Random Cluster when $0 < q \leq 1$)**.** *Let $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ be an n-element matroid with rank function* $\mathrm{rank} : 2^{\mathscr{U}} \to \mathbb{Z}_{\geq 0}$*, and fix parameters $0 < q \leq 1$ and*

$0 \le p \le 1$. *Then there is a simple local Markov chain for sampling from the random cluster measure* $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ *which updates the assignments of at most two elements of* $\mathscr{U}$ *in each step and mixes in* $O(n \log n)$*-steps.*

Remark 28. One might ask about the mixing rate for the classically studied *Glauber dynamics*. We provide some preliminary analysis of this chain by showing $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ is $\frac{1}{\sqrt{q}}$-spectrally independent whenever $0 < q \le 1$, independent of p. By Theorem 2.3.1, this implies mixing in $O\left(n^{2+\frac{1}{\sqrt{q}}}\right)$-steps, where n is the number of elements in the ground set of the matroid. We combining with some additional techniques of [Ana+22c], we can shave a factor of n off the mixing time. We refer the interested reader to Appendix F.3.

It was also brought to our attention that in an independent work, Guo–Mousa have shown $O_{p,q}(n^2 \log n)$-mixing of the Glauber dynamics, where only the constant depends on the parameters p, q. Their starting point is Theorem 5.4.1 (which Guo–Mousa proved independently of our work). They then compare the Dirichlet form of the local Markov chain in Theorem 5.4.1 with the Dirichlet form of the *squared/two-step* Glauber dynamics $\mathcal{P}^2_\mu$ by bounding the worst-case ratio between transition probabilities. Their techniques imply that the modified log-Sobolev constant of $\mathcal{P}^2_\mu$ satisfies the lower bound $\varrho(\mathcal{P}^2_\mu) \ge \frac{1}{\left(1+\frac{p}{q(1-p)}\right)\cdot n^2}$. We refer interested readers to [Mou22].

We further show that the random cluster measures conditioned on having a fixed number of edges is also spectrally independent, and hence yields rapid mixing of a simple Markov chain for sampling from these distributions.

Theorem 5.4.2 (Spectral Independence for Level-k Random Cluster when $0 < q \le 1$; [Ana+19])**.** *Let* $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ *be an* n*-element matroid with rank function* $\operatorname{rank} : 2^{\mathscr{U}} \to \mathbb{Z}_{\ge 0}$*, and fix a parameter* $0 < q \le 1$*. Then for every* $0 \le k \le n$*, the distribution* $\mu = \mu^{\mathsf{RC}}_{\mathcal{M},k,q}$ *defined by*

$$\mu(\tau) \propto q^{-\operatorname{rank}(\tau)}, \quad \forall \tau \in \binom{\mathscr{U}}{k}$$

is $(0, \dots, 0)$*-spectrally independent. In particular, the down-up walk for* $\mu^{\mathsf{RC}}_{\mathcal{M},k,q}$ *mixes rapidly.*

Again, via standard reductions from approximate counting to approximate sampling, Theorem 5.4.2 gives an FPRAS for estimating the level-k partition function

$$\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M},k,q} = \sum_{\tau \in \binom{\mathscr{U}}{k}} q^{-\operatorname{rank}(\tau)}$$

as well as the full partition function $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M}}(p,q)$ defined in Eq. (5.6).

We now proceed to prove Theorem 5.4.2; as it is slightly more technical, we relegate the proof of Theorem 5.4.1 to Appendix F.3. To give a unified treatment of these two results, we will use the following theorem due to Brändén-Huh [BH18].

Theorem 5.4.3 ([BH18]; see also [Ana+19; Ana+18a])**.** *Let* $\mathcal{M} = ([n], \mathcal{X})$ *be a rank-*r *matroid with ground set* $[n]$ *and rank function* $\operatorname{rank} : 2^{\mathscr{U}} \to \mathbb{Z}_{\ge 0}$*. For every fixed parameter* $0 < q \le 1$*, the* n*-homogeneous multivariate polynomial*

$$h_{\mathcal{M},q}(y, x_1, \dots, x_n) \stackrel{\mathsf{def}}{=} \sum_{\sigma \subseteq [n]} y^{n-|\sigma|} q^{-\operatorname{rank}(\sigma)} \prod_{i \in \sigma} x_i \tag{5.7}$$

is completely log-concave on all of $\mathbb{R}^{n+1}_{\ge 0}$*.*

Historically, it was first shown in [Ana+19] that the multivariate partition function for the level-k random cluster model

$$\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M},k,q}(x_1, \dots, x_n) \stackrel{\mathsf{def}}{=} \sum_{\tau \in \binom{\mathscr{U}}{k}} q^{-\operatorname{rank}(\tau)} \prod_{i \in \tau} x_i = \frac{1}{(n-k)!}\, \partial_y^{n-k} h_{\mathcal{M},q}\big|_{y=0} \tag{5.8}$$

is strongly log-concave at $\mathbf{1}$. This was done using Oppenheim's Trickle-Down Theorem (see Theorem 3.2.1), and immediately yields Theorem 5.4.2.

Shortly after, we showed in [Ana+18a] that the *homogenized independence polynomial*

$$h_{\mathcal{M}}(y, x_1, \ldots, x_n) \stackrel{\text{def}}{=} \sum_{\sigma \in \mathcal{X}} y^{n-|\sigma|} \prod_{i \in \sigma} x_i = \lim_{q \to 0} q^n \cdot h_{\mathcal{M},q}\left(\frac{y}{q}, x_1, \ldots, x_n\right) \tag{5.9}$$

is completely log-concave on $\mathbb{R}^{n+1}_{\geq 0}$, where the limit is taken coefficient-wise. We then used this to resolve the strongest form of Mason's Conjecture [Mas72]; we discuss this in more detail later in Section 5.5. Independently of [Ana+18a], Brändén-Huh [BH18; BH20] proved Theorem 5.4.3 and also used it to resolve the strongest form of Mason's Conjecture [Mas72]. As Eqs. (5.8) and (5.9) show, Theorem 5.4.3 is the most complete and unified result, although the proofs of log-concavity for all of these polynomials are nearly identical. For completeness, we give a proof of Theorem 5.4.3 in Section 5.4.1 below. For now, we use this log-concavity result to complete the proof of Theorem 5.4.2. Again, we leave the proof of Theorem 5.4.1 to Appendix F.3.

Proof of Theorem 5.4.2. Eq. (5.8) combined with Theorem 5.4.3 shows that $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M},k,q}(x_1, \ldots, x_n)$ is strongly log-concave on all of $\mathbb{R}^n_{\geq 0}$ (and, in particular, at $\mathbf{1}$). Since $\mathcal{Z}^{\mathsf{RC}}_{\mathcal{M},k,q}$ is equal to the multivariate generating polynomial of $\mu^{\mathsf{RC}}_{\mathcal{M},k,q}$ up to normalization by a constant, applying Theorem 5.0.1 then finishes the proof. □

5.4.1 Log-Concavity of the Homogenized Random Cluster Partition Function

In this subsection, we prove Theorem 5.4.3 using Theorem 5.0.2.

Proof of Theorem 5.4.3. First, observe that for any $\tau \subseteq [n]$,

$$\partial_x^\tau h_{\mathcal{M},q} = h_{\mathcal{M}/\tau,q} \tag{5.10}$$

where recall that $\mathcal{M}/\tau$ is the matroid contraction of $\mathcal{M}$ w.r.t. τ. This is because for any $\sigma \subseteq [n] \setminus \tau$, $\operatorname{rank}_{\mathcal{M}}(\tau) + \operatorname{rank}_{\mathcal{M}/\tau}(\sigma) = \operatorname{rank}_{\mathcal{M}}(\tau \sqcup \sigma)$. Hence, it suffices to verify the conditions of Theorem 5.0.2 Item 3 for all matroids $\mathcal{M}$ and only derivatives w.r.t. y.

We first verify indecomposability for $\partial_y^k h_{\mathcal{M},q}$ for every $0 \leq k \leq n-2$. From Definition 27, this is equivalent to showing that the underlying graph of the matrix $\nabla^2 \partial_y^k h_{\mathcal{M},q}$ is connected, whose vertices are labeled by the variables of $\partial_y^k h_{\mathcal{M},q}$. It is straightforward to check that for every $i \in [n]$, there is a monomial in $\partial_y^k h_{\mathcal{M},q}$ which contains the term yx_i. Hence, the underlying graph of $\nabla^2 \partial_y^k h_{\mathcal{M},q}$ contains a star centered at the variable y as a subgraph, which immediately implies connectivity. As this holds for all matroids $\mathcal{M}$ and any $0 \leq k \leq n-2$, we have total indecomposability of $h_{\mathcal{M},q}$ for every matroid $\mathcal{M}$.

Now, we verify log-concavity for the quadratics obtained via differentiating $h_{\mathcal{M},q}$. Again by Eq. (5.10), it suffices to show that for every $n \geq 1$ and every matroid $\mathcal{M}$ with n ground elements, the quadratic polynomial

$$\begin{aligned}
\partial_y^{n-2} h_{\mathcal{M},q}(x, y) &= \frac{n!}{2} \cdot y^2 + (n-1)! \cdot \sum_{i \in [n]} q^{-\operatorname{rank}(i)} y x_i + (n-2)! \cdot \sum_{\{i,j\} \in \binom{[n]}{2}} q^{-\operatorname{rank}(i,j)} x_i x_j \\
&\propto \binom{n}{2} \cdot y^2 + (n-1) \cdot \sum_{i \in [n]} q^{-\operatorname{rank}(i)} y x_i + \sum_{\{i,j\} \in \binom{[n]}{2}} q^{-\operatorname{rank}(i,j)} x_i x_j
\end{aligned} \tag{5.11}$$

is log-concave on $\mathbb{R}^{n+1}_{\geq 0}$. Let Q denote the Hessian of this quadratic polynomial, which we note is a constant matrix, independent of $x \in \mathbb{R}^n, y \in \mathbb{R}$. Its entries are given by

$$Q = \nabla^2 \partial_y^{n-2} h_{\mathcal{M},q} = \begin{bmatrix} n(n-1) & v \\ v & B \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}_{\geq 0}$$

where $v \in \mathbb{R}^n_{\geq 0}$ is the vector with entries $v_i = q^{-\operatorname{rank}(i)}$, and $B \in \mathbb{R}^{n \times n}_{\geq 0}$ is the matrix with zeros on the diagonal and $B_{ij} = q^{-\operatorname{rank}(i,j)}$ for $i \neq j$. Since this Hessian doesn't depend on the evaluation,

it suffices to show log-concavity at, say, the point $a = (1, 0, \ldots, 0) \in \mathbb{R}^{n+1}$. By Proposition 5.0.3 (Item 1 $\iff$ Item 5), this happens if and only if the matrix

$$\left(a^\top Q a\right) Q - (Qa)(Qa)^\top = n(n-1)Q - (n-1)^2 \begin{bmatrix} n \\ v \end{bmatrix} \begin{bmatrix} n \\ v \end{bmatrix}^\top = (n-1) \begin{bmatrix} 0 & 0 \\ 0 & nB - (n-1)vv^\top \end{bmatrix}$$

is negative semidefinite. Thus it suffices to show that $nB - (n-1)vv^\top$ is negative semidefinite.

First, consider the matrix $B - vv^\top$. For $i \neq j$, we have that the entry $B_{ij} = q^{-\operatorname{rank}(i,j)}$ in B is not equal to the entry $v_i v_j = q^{-\operatorname{rank}(i)} q^{-\operatorname{rank}(j)}$ in $vv^\top$ if and only if i, j are distinct nonloops in $\mathcal{M}$ which are parallel to each other. In this case, $\operatorname{rank}(i,j) = \operatorname{rank}(i) = \operatorname{rank}(j) = 1$, so the former is q^{-1} while the latter is q^{-2}. As $\mathcal{M}$ is a matroid, the matroid partition property (see Fact 3.2.3) tells us that the nonloops of $\mathcal{M}$ may be partitioned into equivalence classes of parallel elements $S_1 \sqcup \cdots \sqcup S_k$. This lets us rewrite the matrix B as

$$B = vv^\top - \sum_{\ell \text{ loop}} v_\ell^2 \mathbf{1}_\ell \mathbf{1}_\ell^\top - \sum_{j=1}^{k} v_{S_j} v_{S_j}^\top$$

where each $v_{S_j} \in \mathbb{R}^n$ is the restriction of v to S_j, i.e. formed by zeroing out all entries of v corresponding to elements not in S_j. It follows that we may write

$$nB - (n-1)vv^\top = vv^\top - n \cdot \left(\sum_{\ell \text{ loop}} v_\ell^2 \mathbf{1}_\ell \mathbf{1}_\ell^\top + \sum_{j=1}^{k} v_{S_j} v_{S_j}^\top \right).$$

Written in this form, we can now check that $nB - (n-1)vv^\top$ is negative semidefinite. Let $x \in \mathbb{R}^n$ be arbitrary, and consider

$$x^\top \left(nB - (n-1)vv^\top\right) x = \left(v^\top x\right)^2 - n \cdot \left(\sum_{\ell \text{ loop}} \left(v_\ell \mathbf{1}_\ell^\top x\right)^2 + \sum_{j=1}^{k} \left(v_{S_j}^\top x\right)^2 \right).$$

Since $S_1, \ldots, S_k$ partition the nonloops of $\mathcal{M}$, $v = \sum_{\ell \text{ loop}} v_\ell \mathbf{1}_\ell + \sum_{j=1}^{k} v_{S_i}$. Now, for any real numbers $u_1, \ldots, u_t$, the Cauchy-Schwarz Inequality applied to the vectors $\mathbf{1}$ and $u = [u_1, \ldots, u_t]$ implies that $\left(\sum_{i=1}^{t} u_i\right)^2 \leq t \cdot \sum_{i=1}^{t} u_i^2$. This then gives

$$\left(v^\top x\right)^2 = \left(\sum_{\ell \text{ loop}} v_\ell \mathbf{1}_\ell^\top x + \sum_{i=1}^{\ell} v_{S_i}^\top x \right)^2 \leq (k + \#\{\text{loops}\}) \cdot \left(\sum_{\ell \text{ loop}} \left(v_\ell \mathbf{1}_\ell^\top x\right)^2 + \sum_{i=1}^{\ell} \left(v_{S_i}^\top x\right)^2 \right).$$

Since the number of equivalence classes k of nonloops of $\mathcal{M}$ plus the number of loops is at most n, it follows that $x^\top \left(nB - (n-1)vv^\top\right) x \leq 0$ for all $x \in \mathbb{R}^n$. We conclude that $nB - (n-1)vv^\top$ is negative semidefinite, and so by Proposition 5.0.3, $\partial_y^{n-2} h_{\mathcal{M},q}$ is log-concave on $\mathbb{R}_{\geq 0}^{n+1}$ as desired. $\square$

5.5 Mason's Ultra-Log-Concavity Conjecture

In this section, we take a brief detour away from Markov chain analysis to give a short, self-contained proof of the strongest version of Mason's conjecture [Mas72]. This was an important conjecture in matroid theory and algebraic combinatorics which had resisted attack for nearly 50 years.

Conjecture 5 (Mason's Conjecture [Mas72])**.** *For a matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ on n elements with $\mathfrak{f}_k = |\mathcal{X}_k|$ independent sets of size k, we have the following inequalities for every $1 < k < n$:*

1. $\mathfrak{f}_k^2 \geq \mathfrak{f}_{k-1} \cdot \mathfrak{f}_{k+1}$ *(log-concavity)*
2. $\mathfrak{f}_k^2 \geq \left(1 + \frac{1}{k}\right) \cdot \mathfrak{f}_{k-1} \cdot \mathfrak{f}_{k+1}$
3. $\mathfrak{f}_k^2 \geq \left(1 + \frac{1}{k}\right) \cdot \left(1 + \frac{1}{n-k}\right) \cdot \mathfrak{f}_{k-1} \cdot \mathfrak{f}_{k+1}$ *(ultra log-concavity)*

Note that Items 1 to 3 are written in order of increasing strength. The weaker property of *unimodality* was previously conjectured by Welsh [Wel71].

At the heart of our proof we show that for any matroid, the homogenization of the generating polynomial of its independent sets is completely log-concave. Combining this with the previously stated closure properties of completely log-concave polynomials along with the fact that homogeneous bivariate strongly log-concave polynomials have ultra log-concave coefficients then finishes the proof.

[Huh12; HK12] used algebro-geometric methods to establish related conjectures for realizable matroids. Building on this, [Len13] established Item 1 for realizable matroids. A major breakthrough was achieved in [AHK18], which proved Item 1 and related conjectures for general matroids by developing a combinatorial Hodge theory for matroids. Extending their techniques, [HSW21] managed to prove Item 2. However, prior to [Ana+18a; BH18], the strongest form of Mason's conjecture, Item 3, was only proven to hold when $n \leq 11$ or $k \leq 5$ [KN11]. We refer to [Sey75; Dow80; Mah85; Zha85; HS89; AS16] for other partial results on Mason's conjecture.

We prove the following, which was also independently established in [BH18].

Theorem 5.5.1 ([Ana+18a; BH18])**.** *For a matroid $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ on n elements with $\mathfrak{f}_k = |\mathcal{X}_k|$ independent sets of size k, the sequence $\mathfrak{f}_0, \mathfrak{f}_1, \ldots, \mathfrak{f}_n$ is ultra log-concave. That is, for every $1 < k < n$,*

$$\left(\frac{\mathfrak{f}_k}{\binom{n}{k}}\right)^2 \geq \frac{\mathfrak{f}_{k-1}}{\binom{n}{k-1}} \cdot \frac{\mathfrak{f}_{k+1}}{\binom{n}{k+1}}$$

Remark 29. As our analysis shows, one can actually replace n with the number of *parallel classes* of $\mathcal{M}$, which may be strictly smaller than n. This leads to a stronger inequality.

Related Prior Work It is well-known that the uniform distribution over all spanning trees of a graph is negatively correlated and more generally negatively associated, see [Pem00] for background. This fact more generally extends to *regular* matroids. Prior to our work, many researchers tried to approach Mason's conjecture through the lens of negative correlation [SW75; Wag08; BBL09; KN10; KN11]. However, for many matroids the uniform distribution on bases is not negatively correlated and furthermore, negative correlation does not necessarily imply log-concavity of its rank sequences, see e.g. [Wag08].

Again, consider the bases generating polynomial

$$g_{\mathcal{M}}(x) = \sum_{\mathcal{B} \subseteq \mathscr{U} \text{ basis}} \prod_{i \in \mathcal{B}} x_i$$

Then the negative correlation property is equivalent to all off-diagonal entries of $\nabla^2 \log g_{\mathcal{M}}$ being nonpositive when evaluated at $\mathbf{1}$, i.e.

$$(\nabla^2 \log g_{\mathcal{M}}(\mathbf{1}))_{i,j} = g_{\mathcal{M}}(\mathbf{1}) \cdot \partial_i \partial_j g_{\mathcal{M}}(\mathbf{1}) - \partial_i g_{\mathcal{M}}(\mathbf{1}) \cdot \partial_j g_{\mathcal{M}}(\mathbf{1}) \leq 0,$$

for all $1 \leq i, j \leq n$, $i \neq j$. This inequality holds for *regular matroids* (e.g. graphic matroids), but not necessarily for linear matroids. We refer interested readers to [HSW21] for a more detailed discussion of counterexamples to negative correlation in matroids. We note however that log-concavity of $g_{\mathcal{M}}$ does imply an *approximate* version of negative correlation, where one adds a factor 2 in front of the second term $\partial_i g_{\mathcal{M}}(\mathbf{1}) \cdot \partial_j g_{\mathcal{M}}(\mathbf{1})$.

In [AOV21] it was observed that for any matroid $\mathcal{M}$, the polynomial $g_{\mathcal{M}}$ is completely log-concave. This means that even though $\nabla^2 \log g_{\mathcal{M}}(\mathbf{1})$ can have positive entries, all of its eigenvalues are nonpositive, and the same holds for all partial derivatives of $g_{\mathcal{M}}$. As previously mentioned in Chapter 3, this property of the uniform distribution over bases of a matroid is called *spectral negative dependence*.

Independent and Subsequent Work Brändén–Huh independently established the strongest version of Mason's conjecture [BH18] using their framework of "Lorentzian polynomials", which are further explored in [BH20]. Again, this condition is defined differently from complete log-concavity, but the two are equivalent by Theorem 5.0.2. This is also proved in [BH20]; see Section 5.7.2 for further discussion. Backman-Eur-Simpson reprove some of the results of [AHK18] in this language [BES21].

Recently, Chen–Pak [CP21b; CP22] developed a very general machinery called *combinatorial atlases* for establishing log-concave inequalities through elementary linear algebra which led to new log-concavity results for posets, greedoids, and other combinatorial structures of interest. This machinery can also rederive Theorem 5.5.1, and avoids the use of polynomials. Finally, we mention a recent result of Brändén–Leake [BL21], which gives a proof of the Heron-Rota-Welsh conjectures based on an extension of Lorentzian polynomials to cones beyond the nonnegative orthant. While these conjectures were famously already settled by Adiprasito–Huh–Katz [AHK18], the techniques of [BL21] are elementary.

Our Approach The main ingredient of our proof of Theorem 5.5.1 is to show that the *homogenization* of the generating polynomial of all independent sets of a matroid is completely log-concave. This was established in [Ana+18a] directly using Theorem 5.0.2. Here, we directly appeal to Theorem 5.4.3, which was proved in [BH18] using essentially the same argument.

Theorem 5.5.2 (Log-Concavity of Homogenized Independence Polynomials; [Ana+18a])**.** *For any matroid $\mathcal{M} = ([n], \mathcal{X})$, the polynomial*

$$h_{\mathcal{M}}(x_1, \dots, x_n, y) \stackrel{\text{def}}{=} \sum_{\sigma \in \mathcal{X}} y^{n-|\sigma|} \prod_{i \in \sigma} x_i$$

is completely log-concave on $\mathbb{R}^{n+1}_{\geq 0}$.

Proof. Observe that $h_{\mathcal{M}}$ may be obtained from $h_{\mathcal{M},q}$ via the coefficient-wise limit as $q \to 0$; see Eq. (5.9). Hence, the claim follows from Theorem 5.4.3 combined with Proposition 5.3.2 Item 1. □

We then combine this with the following characterization of ultra log-concave sequences in terms of completely log-concave polynomials. This correspondence was first observed by Gurvits [Gur08], and we give a short proof for the sake of completeness in Section 5.5.1.

Proposition 5.5.3 ([Gur10])**.** *Let $\{c_k\}_{k=0}^n$ be a sequence of positive real numbers and consider the bivariate n-homogeneous polynomial $f(x, y) = \sum_{k=0}^n c_k x^k y^{n-k} \in \mathbb{R}[x, y]$. Then f is completely log-concave on $\mathbb{R}^2_{\geq 0}$ if and only if the sequence $c_0, \dots, c_n$ is ultra log-concave, i.e. for every $1 < k < n$,*

$$\left(\frac{c_k}{\binom{n}{k}}\right)^2 \geq \frac{c_{k-1}}{\binom{n}{k-1}} \cdot \frac{c_{k+1}}{\binom{n}{k+1}}. \tag{5.12}$$

With these two ingredients in hand, we proceed to give a short proof of Theorem 5.5.1.

Proof of Theorem 5.5.1. By Theorem 5.5.2, $h_{\mathcal{M}}(x_1, \dots, x_n, y)$ is completely log-concave on $\mathbb{R}^{n+1}_{\geq 0}$. By Lemma 5.3.1, since the image of $\mathbb{R}^2_{\geq 0}$ under the linear map $(x, y) \mapsto (x, \dots, x, y)$ is contained in $\mathbb{R}^{n+1}_{\geq 0}$, the polynomial

$$f_{\mathcal{M}}(x, y) \stackrel{\text{def}}{=} \sum_{k=0}^{r} \mathfrak{f}_k x^k y^{n-k} = h_{\mathcal{M}}(x, \dots, x, y)$$

is completely log-concave, where r is the rank of $\mathcal{M}$. Ultra log-concavity then follows by Proposition 5.5.3. □

5.5.1 Log-Concave Polynomials and Ultra Log-Concave Sequences

In this subsection, we give a short proof of Proposition 5.5.3, which recall establishes a formal connection between ultra log-concave sequences and log-concavity of the homogenization of their univariate generating polynomial. This was first observed and proved in [Gur09].

Proof of Proposition 5.5.3. Since the c_k are positive, f is totally indecomposable, and so by Theorem 5.0.2, complete log-concavity of f is equivalent to log-concavity of the quadratic polynomials $q_k(x, y) \stackrel{\text{def}}{=} \partial_x^{k-1} \partial_y^{n-k-1} f$ for every $1 < k < n$. Notice that for any $0 \leq m \leq n$,

$$\partial_x^m \partial_y^{n-m} f = (n-m)! \cdot m! \cdot c_m = n! \frac{c_m}{\binom{n}{m}}.$$

Using this for each $m \in \{k-1, k, k+1\}$, we can write the Hessian of q_k as

$$\nabla^2 q_k = \begin{bmatrix} \partial_x^2 q_k & \partial_x \partial_y q_k \\ \partial_x \partial_y q_k & \partial_y^2 q_k \end{bmatrix} = n! \cdot \begin{bmatrix} c_{k+1} \Big/ \binom{n}{k+1} & c_k \Big/ \binom{n}{k} \\ c_k \Big/ \binom{n}{k} & c_{k-1} \Big/ \binom{n}{k-1} \end{bmatrix}.$$

By Proposition 5.0.3, log-concavity of q_k is equivalent to $\nabla^2 q_k$ having exactly one positive eigenvalue (and hence, exactly one nonpositive eigenvalue). This is equivalent to its determinant being nonpositive, which is precisely the desired ultra log-concavity inequality:

$$0 \geq \det(\nabla^2 q_k) = (n!)^2 \left(\frac{c_{k-1}}{\binom{n}{k-1}} \cdot \frac{c_{k+1}}{\binom{n}{k+1}} - \left(\frac{c_k}{\binom{n}{k}} \right)^2 \right).$$

□

Remark 30. Without appealing to Theorem 5.0.2, one can also show that ultra log-concavity of $\{c_k\}_{k=0}^n$ implies complete log-concavity of f by using a bit of convex geometry. More specifically, by *Shephard's construction* [She60], there exists two convex bodies $K, L \subseteq \mathbb{R}^n$ such that $f(x, y) = \mathrm{Vol}(xK + yL)$. Complete log-concavity of f then follows from the Alexandrov-Fenchel Inequalities.

5.6 On the Support of Discrete Log-Concave Distributions

In this section, we study the support of discrete log-concave distribution, or equivalently, the support of multiaffine log-concave polynomials. We show that for any homogeneous multiaffine strongly log-concave polynomial p with nonnegative coefficients, $\mathrm{supp}(p)$ form the bases of a matroid. This characterization was proved independently in [BH20]. This is a generalization of a result of [Cho+04], which says that the support of a homogeneous multiaffine real stable polynomial with nonnegative coefficients form the bases of a matroid.

Theorem 5.6.1. *Let $p \in \mathbb{R}[z_1, \ldots, z_n]$ be a homogeneous multiaffine polynomial with nonnegative coefficients. If p is strongly/completely log-concave on $\mathbb{R}^n_{\geq 0}$, then $\mathrm{supp}(p)$ form the bases of a matroid.*

The main tool to prove Theorem 5.6.1 is to leverage closure properties of strongly log-concave polynomials, namely closure under taking external fields and limits. This allows us to "isolate" the edges in the associated *Newton polytope* $\mathsf{Newt}(p) \stackrel{\mathsf{def}}{=} \mathrm{conv}(\mathrm{supp}(p))$. We then show that log-concavity of the resulting polynomials associated to these edges implies that they must be of the form $\mathbf{1}_i - \mathbf{1}_j$ for some $i \neq j$ in $[n]$. This implies $\mathrm{supp}(p)$ form the bases of a matroid due to the geometric definition of matroid bases (see Theorem 1.4.22).

We will need the following operation on polynomials, which will be shown to preserve log-concavity. This is a standard operation in *tropical geometry*; see e.g. [RST05].

Definition 28 (Initial Form)**.** *Fix a polynomial $p(z) = \sum_{\kappa \in \mathbb{Z}^n_{\geq 0}} c_\kappa z^\kappa \in \mathbb{R}[z_1, \ldots, z_n]$. For an arbitrary vector $v \in \mathbb{R}^n$, we define the **degree of** p **w.r.t.** v as*

$$\deg_v(p) \stackrel{\mathsf{def}}{=} \sup_{x \in \mathsf{Newt}(p)} \langle x, v \rangle = \max_{\kappa \in \mathrm{supp}(p)} \langle \kappa, v \rangle .$$

*With this in hand, we define the **initial form of** p **w.r.t.** v as the polynomial*

$$\mathrm{In}_v(p)(z_1, \ldots, z_n) \stackrel{\mathsf{def}}{=} \sum_{\kappa : \langle \kappa, v \rangle = \deg_v(p)} c_\kappa z^\kappa.$$

Lemma 5.6.2 (Taking Initial Forms Preserves Log-Concavity)**.** *Suppose $p \in \mathbb{R}[z_1, \ldots, z_n]$ be a homogeneous polynomial with nonnegative coefficients. If p is strongly log-concave on $\mathbb{R}^n_{\geq 0}$, then so is $\mathrm{In}_v(p)$ for every $v \in \mathbb{R}^n$.*

Proof. First, observe that $\mathrm{In}_v(p)$ may be recovered as the coefficient-wise limit of the polynomials

$$p_t(z_1, \ldots, z_n) \stackrel{\mathsf{def}}{=} t^{-\deg_v(p)} \cdot p\left(t^{v_1} \cdot z_1, \ldots, t^{v_n} \cdot z_n\right) \text{ as } t \to +\infty.$$

Since strong log-concavity is closed under taking external fields by Proposition 5.3.2 Item 2, each p_t is strongly log-concave. Since $\mathrm{In}_v(p)$ is the limit of strongly log-concave polynomials, it is also strongly log-concave by Proposition 5.3.2 Item 1. □

With this tool in hand, we can now prove Theorem 5.6.1.

Proof of Theorem 5.6.1. Fix an arbitrary edge of the Newton polytope $\mathsf{Newt}(p)$, with endpoints $\mathbf{1}_S, \mathbf{1}_T \in \{0,1\}^n$. Pick any vector $v \in \mathbb{R}^n$ such that the linear functional $x \mapsto \langle x, v\rangle$ is maximize on this edge. In other words, pick v such that $\langle \kappa, v\rangle < \langle \mathbf{1}_S, v\rangle = \langle \mathbf{1}_T, v\rangle$ for all other $\kappa \in \mathrm{supp}(p)$ not equal to $\mathbf{1}_S, \mathbf{1}_T$. The resulting initial form is simply

$$\mathrm{In}_v(p) = c_S x^S + c_T x^T$$

where $c_S, c_T > 0$, which is strongly log-concave on $\mathbb{R}^n_{\geq 0}$ by Lemma 5.6.2. Due to the geometric definition of matroid bases given in Theorem 1.4.22, our goal is to show that the edge $\mathbf{1}_S - \mathbf{1}_T$ connecting $\mathbf{1}_S, \mathbf{1}_T$ is equal to $\mathbf{1}_i - \mathbf{1}_j$ for some pair of distinct elements $i, j \in [n]$. Since $\mathbf{1}_S - \mathbf{1}_T = \mathbf{1}_{S\setminus T} - \mathbf{1}_{T\setminus S}$, it suffices to show that $|S \setminus T| = |T \setminus S| = 1$.

Suppose for contradiction that this is not the case, i.e. assume there exist distinct elements $i, j \in T \setminus S$ (the case $i, j \in S \setminus T$ is analogous). Setting $z_k = 1$ in $\mathrm{In}_v(p)$ for all $k \neq i, j$ yields the polynomial $c_S + c_T \cdot z_i z_j$. We show this polynomial cannot be log-concave on $\mathbb{R}^2_{\geq 0}$, contradicting strong log-concavity of $\mathrm{In}_v(p)$, and hence of p itself. We compute the Hessian of the logarithm of $c_S + c_T \cdot z_i z_j$. We have that

$$\begin{aligned}\nabla^2 \log(c_S + c_T \cdot z_i z_j) &= \frac{1}{c_S + c_T \cdot z_i z_j} \cdot \begin{bmatrix} 0 & c_T \\ c_T & 0 \end{bmatrix} - \frac{1}{(c_S + c_T \cdot z_i z_j)^2} \cdot \begin{bmatrix} c_T z_j \\ c_T z_i \end{bmatrix} \begin{bmatrix} c_T z_j \\ c_T z_i \end{bmatrix}^\top \\ &= \frac{1}{(c_S + c_T \cdot z_i z_j)^2} \cdot \begin{bmatrix} -c_T^2 z_j^2 & c_S c_T \\ c_S c_T & -c_T^2 z_i^2 \end{bmatrix}.\end{aligned}$$

Now, observe that in the limit as $z_i, z_j \to 0$, this matrix converges to

$$\frac{1}{c_S^2} \cdot \begin{bmatrix} 0 & c_S c_T \\ c_S c_T & 0 \end{bmatrix}$$

which clearly is not negative semidefinite. Indeed, hitting the matrix on both sides with, say $\mathbf{1}$, yields a strictly positive quantity, since $c_S, c_T > 0$ by the assumption $\mathbf{1}_S, \mathbf{1}_T \in \mathrm{supp}(p)$. □

[BH20] independently proved an extension of Theorem 5.6.1, which says that the support of any homogeneous strongly log-concave polynomial (which need not be multiaffine) is M-convex in the sense of Definition 30. This can be done by using a proof very similar to the one above. In Appendix F.1, we show that strongly log-concave polynomials which satisfy a certain natural "partiteness" property must factorize as a product of linear forms. In probabilistic language, this says that distributions over discrete product spaces which are $(0, \ldots, 0)$-spectrally independent must be independent product distributions.

5.7 Connections with Other Notable Classes of Polynomials

In this section, we briefly discuss connections between log-concave polynomials and two other well-studied classes of polynomials in the literature.

5.7.1 Real Stable Polynomials

In this section, we mention connections between log-concave polynomials and *real stable polynomials*, which have gained significant attention in recent years due to several breakthroughs in quantum physics, operator theory, functional analysis, and algebraic graph theory. For instance, [MSS15b] used these polynomials to positive resolve the Kadison-Singer Problem in the foundations of quantum mechanics, and [MSS22] used them to obtain sharper restricted invertibility estimates. [MSS15a] used these polynomials to construct infinite families of (simple) Ramanujan graphs (see also [HPS18]). An alternative construction of Ramanujan graphs (allowing parallel edges) was given in [MSS15c], which again took advantage of real stability. This result was then made algorithmic by [Coh16]. See [Cho+04; Gur06a; Gur06b; Gur08; Brä10; PP14; AO17; CS22] for additional applications in combinatorics, including an elementary proof of the van der Waerden Conjecture, and [Ana+17; Ana+18b] for applications in game theory.

Definition 29 (Half-Plane Stable)**.** *For $0 \le \theta < 2\pi$ and $\epsilon \ge 0$, define the open (rotated and shifted) half-plane by*

$$\mathbb{H}_{\theta,\epsilon} \stackrel{\text{def}}{=} \left\{ e^{-i\theta} z : \Im(z) > \epsilon \right\} \subseteq \mathbb{C}.$$

*We say a polynomial $p(z_1, \ldots, z_n) \in \mathbb{C}[z_1, \ldots, z_n]$ is $\mathbb{H}_{\theta,\epsilon}$-**stable** if p does not have a root in $\mathbb{H}^n_{\theta,\epsilon}$. In particular, the zero polynomial is $\mathbb{H}_{\theta,\epsilon}$-stable.*

*We call $\mathbb{H}_{0,0}$ and $\mathbb{H}_{\frac{\pi}{2},0}$ the **upper half-plane** and **right half-plane**, respectively. We say p is **Hurwitz stable** if it is $\mathbb{H}_{\frac{\pi}{2},0}$-stable. We say p is **real stable** if it is $\mathbb{H}_{0,0}$-stable and has real coefficients. We observe that for homogeneous polynomials, if $\mathbb{H}_{\theta,0}$-stability holds for some angle θ, then it holds for all angles.*

*Finally, we say a distribution $\mu : 2^{[n]} \to \mathbb{R}_{\ge 0}$ is **strongly Rayleigh** if and only if its generating polynomial g_μ is real stable [see BBL09].*

The following classical facts give examples of real stable polynomials.

Theorem 5.7.1 (see e.g. [BBL09])**.** *For $0 \le k \le n$, the degree-k elementary symmetric polynomial in n variables $e_k(z_1, \ldots, z_n)$ is real stable.*

The following is sometimes colloquially referred to as the "mother of all real stable polynomials".

Theorem 5.7.2 (see e.g. [BB08; BB10])**.** *For symmetric positive semidefinite matrices $A_1, \ldots, A_n \in \mathbb{R}^{d\times d}$ and an arbitrary symmetric matrix $B \in \mathbb{R}^{d\times d}$ (which need not be positive semidefinite), the polynomial*

$$p(z_1, \ldots, z_n) = \det\left(B + \sum_{i=1}^n z_i A_i \right)$$

is real stable.

One of the primary reasons real stable polynomials are useful is that they admit a rich set of operations which preserve real stability. Beautiful and comprehensive treatments of such real stability preservers are given in [BB09a; BB09b; BB09c; BB10]. For our purposes, we will only need the following two operations.

Theorem 5.7.3 ([Wag11])**.** *If $p \in \mathbb{R}[z_1, \ldots, z_n]$ is real stable, then the following are also real stable:*

- *$p|_{z_i=a}$ for $a \in \mathbb{R}$ and $i \in [n]$ (Specialization).*
- *$\partial_i p$ for all $i \in [n]$ (Differentiation).*

Any univariate real-rooted polynomial with nonnegative coefficients admits a factorization of the form $C\prod_{i=1}^d (z + r_i)$ for some nonnegative real numbers $a, r_1, \ldots, r_d$, so its logarithm $\log C + \sum_{i=1}^d \log(z+r_i)$ is a sum of functions which are concave on the nonnegative real axis. It follows that such polynomials are log-concave on $\mathbb{R}_{\ge 0}$. Since real stable polynomials are a natural generalization of univariate *real-rooted* polynomials to the multivariate setting, it is natural to expect that real stable polynomials with nonnegative coefficients are also log-concave on the nonnegative orthant $\mathbb{R}^n_{\ge 0}$.

Theorem 5.7.4 ([Gül97]; see also [Går59])**.** *Let $p \in \mathbb{R}[z_1, \ldots, z_n]$ be a real stable polynomial with nonnegative coefficients. Then p is log-concave on $\mathbb{R}^n_{\ge 0}$.*

Roughly speaking, one can prove this by first showing that real stability is equivalent *hyperbolicity* w.r.t. directions in $\mathbb{R}^n_{>0}$, and then appealing to the aforementioned fact that univariate real-rooted polynomials with nonnegative coefficients are log-concave on the nonnegative real line. For inhomogeneous polynomials, one needs to first do a homogenization step. The first step is formalized as follows, which we use later on to show the equivalence between real stability and log-concavity for homogeneous degree-2 polynomials.

Lemma 5.7.5 (Real Stability & Hyperbolicity; [BB10])**.** *Let $p \in \mathbb{R}[z_1, \ldots, z_n]$ be a d-homogeneous polynomial with real coefficients. Then p is real stable if and only if for every positive direction $e \in \mathbb{R}^n_{>0}$ and every $x \in \mathbb{R}^n$, the univariate restriction $t \mapsto p(te + x)$ is real-rooted.*

Proof. We prove the contrapositive of both directions.

Suppose p is not real stable. Then there exists $z_1, \dots, z_n$ with $\Im z_1, \dots, \Im z_n > 0$ such that $p(z_1, \dots, z_n) = 0$. Define $e \in \mathbb{R}^n_{>0}$ and $x \in \mathbb{R}^n$ by $e_j = \Im z_j$ and $x_j = \Re z_j$ for all $j = 1, \dots, n$. Then the imaginary unit i is a root of the univariate restriction $t \mapsto p(te + x)$, which shows it is not real-rooted.

Now, suppose there exists $e \in \mathbb{R}^n_{>0}$ and $x \in \mathbb{R}^n$ such that the univariate restriction $t \mapsto p(te+x)$ is not real-rooted. Fix one such complex root t, and note that we may assume without loss of generality that $\Im t > 0$. This is because the coefficients of p are real, and so the roots of p come in complex conjugate pairs. It follows that $z_1, \dots, z_n$ defined by $z_j = te_j + x_j = (e_j \cdot \Re t + x_i) + i \cdot e_j \cdot \Im t$ for all $j = 1, \dots, n$ satisfies $\Im z_j > 0$ for all $j = 1, \dots, n$, since $e \in \mathbb{R}^n_{>0}$ and $\Im t > 0$. Furthermore, $p(z_1, \dots, z_n) = 0$ by definition of $z_1, \dots, z_n$ and t. Hence, p is not real stable. □

Finally, we mention that the class of homogeneous real stable polynomials with nonnegative coefficients truly is a proper subset of the class of homogeneous strongly log-concave polynomials with nonnegative coefficients. For instance, a well-known result of Brändén [Brä07] (answering a question posed by [Cho+04]) shows that the *Fano matroid*, a 7-element matroid representable over $\mathbb{F}_2$, is not the support of *any* real stable polynomial. In particular, its bases generating polynomial is not real stable. On the other hand, we know that the bases generating polynomial of any matroid is strongly log-concave.

5.7.2 Lorentzian Polynomials

Finally, in this subsection, we state the equivalence between strongly/completely log-concave polynomials and the independently developed notion of *Lorentzian polynomials* [BH20]. These polynomials have had a tremendous impact in combinatorics e.g. [EH20; BL21; Huh+22; FG22; BLP22]. Most notably, it has led to the resolution of several longstanding conjectures in algebraic combinatorics, including the strongest version of Mason's Conjecture [BH18] (this was done independently by [Ana+18a] using completely log-concave polynomials; see Section 5.5). When convenient, we will occasionally use this alternative notion.

To define Lorentzian polynomials, we first need to define a generalization of matroid bases called M-convex sets. These objects are fundamental to the study of discrete convexity. We refer interested readers to Murota's monograph [Mur03] for an in-depth treatment.

Definition 30 (M-convex Set; [Mur03])**.** *We say a subset $J \subseteq \mathbb{N}^n$ to be* M*-convex if it satisfies any one of the following equivalent conditions:*

- *For any $\alpha, \beta \in J$ and any index i satisfying $\alpha_i > \beta_i$, there is an index j satisfying $\alpha_j < \beta_j$ and $\alpha - \mathbf{1}_i + \mathbf{1}_j \in J$.*
- *For any $\alpha, \beta \in J$ and any index i satisfying $\alpha_i > \beta_i$, there is an index j satisfying $\alpha_j < \beta_j$ and $\alpha - \mathbf{1}_i + \mathbf{1}_j \in J$ and $\beta - \mathbf{1}_j + \mathbf{1}_i \in J$.*

We note that any M-convex set J must be a subset of

$$\Delta_n^d \stackrel{\text{def}}{=} \{\alpha \in \mathbb{N}^n : |\alpha| = d\}$$

for some fixed d. Conversely, for $d = 1$, any $J \subseteq \Delta_n^1$ is M-convex. One should view the conditions of M-convexity as an extension of the matroid basis exchange axioms to points in $\mathbb{N}^n$. Indeed, a subset of $\{0, 1\}^n$ is M-convex if and only if it forms the bases of a matroid.

Definition 31 (Lorentzian Polynomial; [BH20])**.** *Let $p \in \mathbb{R}[z_1, \cdots, z_n]$ be a homogeneous polynomial of degree-d with nonnegative coefficients. We say p is* ***Lorentzian*** *if either $d \leq 1$, or $d \geq 2$, $\operatorname{supp}(p)$ is* M*-convex, and $\partial^\alpha p$ is real stable for all α satisfying $|\alpha| = d - 2$.*

In light of Theorem 5.0.2, in particular Item 3, the equivalence between strongly/completely log-concave polynomials and Lorentzian polynomials isn't too surprising. The only apparent differences are total indecomposability versus M-convexity of $\operatorname{supp}(p)$, and log-concavity versus stability for degree-2 polynomials obtained via partial differentiation. The former discrepancy is resolved by the characterization of the support of homogeneous log-concave polynomials as M-convex sets (see [BH20]; see also Section 5.6). The latter discrepancy can be easily resolved from the following lemma.

Lemma 5.7.6. *Let p be a homogeneous degree-2 polynomial with nonnegative coefficients. Then p is real stable if and only if p is log-concave on $\mathbb{R}^n_{\geq 0}$.*

Proof. We may write $p(z) = z^\top Q z$, where $Q = \nabla^2 p \in \mathbb{R}^{n\times n}$ is a symmetric matrix with nonnegative entries, and $z = [z_1, \ldots, z_n]$ is the vector of variables of p. Note that since p is homogeneous of degree-2, it cannot be identically zero. Since it also has nonnegative coefficients, Q must have at least one strictly positive entry.

By Lemma 5.7.5, p being real stable is equivalent to saying that for every $e \in \mathbb{R}^n_{>0}$ and $x \in \mathbb{R}^n$, the univariate restriction $t \mapsto p(te+x) = t^2 \cdot e^\top Q e + 2t \cdot e^\top Q x + x^\top Q x$ is real-rooted. In particular, the discriminant of this polynomial must be nonnegative, i.e. we have the inequality

$$\left(e^\top Q x\right)^2 \geq \left(e^\top Q e\right) \cdot \left(x^\top Q x\right).$$

Rearranging, this is equivalent to saying that

$$x^\top \left(\left(e^\top Q e\right) Q - (Qe)(Qe)^\top\right) x \leq 0, \quad \forall x \in \mathbb{R}^n.$$

Thus, we have shown that real stability of p is equivalent to the matrix $\left(e^\top Q e\right) Q - (Qe)(Qe)^\top$ being negative semidefinite for every $e \in \mathbb{R}^n_{>0}$. By Proposition 5.0.3, this is equivalent to p being log-concave at every $e \in \mathbb{R}^n_{>0}$, which via a limiting argument, means p is log-concave on all of $\mathbb{R}^n_{\geq 0}$. □

For further discussion on the equivalence between strongly log-concave, completely log-concave, and Lorentzian homogeneous polynomials, see [BH20].

Finally, we mention a beautiful recent work of Brändén–Leake [BL21], which develops a theory of Lorentzian polynomials on convex cones other than the nonnegative orthant $\mathbb{R}^n_{\geq 0}$; see also [Dey22]. Their main application is to give an elementary "polynomial proof" of the conjectures of Heron, Rota, and Welsh, which say that the coefficients of the *characteristic polynomial* of every matroid form a log-concave sequence. These conjectures were famously first resolved in a groundbreaking work of Adiprasito–Huh–Katz [AHK18], which developed a sophisticated algebraic machinery called "Hodge theory for combinatorial geometries" (see also [Huh12; HK12; AS16] for interesting special cases resolved prior to [AHK18], [Bak18; Huh18; Huh22] for additional expository material, and [HW17; Eur20; Bra+20b; Bra+20a; BES21; HSW21; ADH21; ADH22; Ber+22; EHL22] for further developments).[2]

[2] A "combinatorial geometry" is essentially another term for *matroid*, although the two are not exactly the same; see [AHK18] for more details.

Chapter 6

The Geometry of Polynomials II: Stability and Zero-Freeness

> "The one contribution of mine that I hope will be remembered has consisted in just pointing out that all sorts of problems of combinatorics can be viewed as problems of location of the zeros of certain polynomials and in giving these zeros a combinatorial interpretation."
>
> – Gian-Carlo Rota [RS85]

In the previous Chapter 5, we saw the correspondence between high-dimensional discrete probability distributions μ and their multivariate generating polynomials g_μ. We saw how an *analytic property* of g_μ, namely (strong or complete) log-concavity, implies $(0, \dots, 0)$-spectral independence, a probabilistic property of the distribution μ. This is the strongest possible bound on the influence matrices of μ, and allowed us to show rapid mixing for simple, local Markov chains which samples from general discrete log-concave distributions. In this chapter, we further strengthen this connection between analytic properties of multivariate polynomials and spectral independence. We show how the structure of the *zeros* (or *roots*) of g_μ, a more algebraic property, can also be used to bound influences.

The connection between zeros of polynomials, probability, combinatorics and statistical physics has a long and rich history; see the quote above due to Rota as well as Chapter 1 for more discussion. The seminal theory of Lee and Yang [LY52] demonstrated that the accumulation of zeros of the partition function on systems growing in scale imply (physical) phase transitions. In other words, the zeros of the partition function can be used to detect the presence of a phase transition. Here, by a phase transition, we mean a sudden change in the properties of a system when some underlying parameter changes even only slightly.

In recent years, particularly with the invention and development of Barvinok's *polynomial interpolation* algorithm [Bar16b; Bar17a; Bar16a], which was further refined by Patel–Regts [PR17], the connection between zeros and phase transitions has been vastly strengthened. In particular, there is now also a beautiful and intricate correspondence between the locations of zeros and *computational phase transitions*. When there are no zeros, there are deterministic algorithms which approximate the partition function up to any desired (inverse polynomial) level of accuracy in quasipolynomial or even polynomial time. On the other hand, for many models (e.g. the hardcore gas model; see Chapter 7), the accumulation of zeros implies not only the interpolation method will fail, but no efficient algorithm (in the sense of FPRAS) exists for approximating the partition function unless NP = RP.

Roughly speaking, the polynomial interpolation method is a mathematically elegant approach which works in the following manner. To approximate the partition function of a model 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 only a *logarithmic number of terms*, which yields a (quasi-)polynomial time algorithm to approximate the partition function at λ. 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 has a number of significant consequences. For Gibbs distributions of spin systems on graphs with n vertices, the interpolation method generally has running time scaling as $O(n^C)$ where the constant C depends on the parameters of the model *as well as the maximum degree* Δ of the graph [PR17]. In particular, if Δ is unbounded, the interpolation method does not run in polynomial time; indeed, for general distributions, it only runs in quasi-polynomial time [Bar16a]. On the other hand, by going running the Glauber dynamics and going through spectral independence for the analysis, we can $O(n \log n)$-time sampling algorithms when $\Delta \leq O(1)$, and more generally, polynomial time sampling algorithms with exponent independent of Δ[1]. The content of this chapter is based on [CLV21b].

6.1 Techniques and Applications of Zero-Freeness

We begin by stating here three sample applications of the techniques in this chapter; further applications can be found in Appendix D. For a graph $G = (V, E)$, we say a vertex v is *covered* by a subset $F \subseteq E$ of edges if v is incident to at least one edge in F. The subset $F \subseteq E$ is called an *edge cover* if all vertices are covered by F. Note there is always a trivial edge cover by setting $F = E$. An FPRAS (fully polynomial randomized approximation scheme) was presented for counting the number of edge covers for 3-regular graphs [BR09]. In [LLL14], an FPTAS 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 Markov chains was presented in [HLZ16].

The correlation decay algorithm of [LLL14] was extended to weighted (partial) edge covers (with worse running time guarantees) in [LLZ14a]. 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 $F \subseteq E$ is associated with the weight $w(F) = \rho^{|\mathsf{unc}(F)|}\lambda^{|F|}$, where $\mathsf{unc}(F)$ denotes the set of vertices that are not covered by F. The Gibbs distribution over all subsets of edges is given by $\mu(F) \propto w(F)$. 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 [Guo+21]; see also [BCR20]. Combining the zero-free results in [Guo+21] with the technical results of this chapter, we immediately obtain an FPRAS using a simple Markov chain algorithm and with significantly faster running time guarantees.

Theorem 6.1.1 (Weighted Edge Covers)**.** *Let $\Delta \geq 3$ be an integer and let $\lambda > 0$, $\rho \in [0, 1]$ be reals. Then for any n-vertex graph $G = (V, E)$ of maximum degree $\Delta \leq O(1)$, 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 [JS93] presenting an FPRAS for the partition function of the ferromagnetic Ising model on any graph at any temperature. The Ising model on a graph $G = (V, E)$ is described by two parameters, the edge activity $\beta_{\mathsf{Ising}} > 0$ and the vertex activity $\lambda_{\mathsf{Ising}} > 0$. The Gibbs distribution of the Ising model is over all $\{+, -\}$ (or, equivalently, $\{0, 1\}$) spin assignments to vertices. Every configuration $\sigma : V \to \{+, -\}$ has density $\mu_{\mathsf{Ising}}(\sigma) \propto \beta_{\mathsf{Ising}}^{m(\sigma)} \lambda_{\mathsf{Ising}}^{|\sigma^{-1}(+)|}$, 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 $\beta_{\mathsf{Ising}} > 1$, 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 $F \subseteq E$ has weight $w(F) = \rho^{|\mathsf{odd}(F)|}\lambda^{|F|}$ where $\mathsf{odd}(F)$ is the set of odd-degree vertices in the subgraph (V, F), and $\mu(F) \propto w(F)$. The weighted even subgraphs model is related to the ferromagnetic Ising model by $\beta_{\mathsf{Ising}} = \frac{1+\lambda}{1-\lambda}$ and $\lambda_{\mathsf{Ising}} = \frac{1+\rho}{1-\rho}$, for which one can easily transform a subset of edges from μ to a sample from μ_{Ising} [GJ09]. Note that if $\rho = 0$, then μ is the distribution over all weighted (exact) even subgraphs, which corresponds to

[1] This is actually an apples and oranges comparison, since the interpolation method has additional desirable features which may compensate for its relatively poor running time. See Section 6.1.2 for a more precise discussion of the differences between our methods and the interpolation method.

the ferromagnetic Ising model without external fields (i.e. $\lambda_{\mathsf{Ising}} = 1$). In [JS93], a Markov chain algorithm is presented to sample weighted even subgraphs of an arbitrary (unbounded-degree) graph in time $O(m^3\mathsf{poly}(1/\rho))$, where m is the number of edges. In another direction, [LSS19] 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 Theorem [LY52]. As is common for this type of approach, the running time of [LSS19] 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 6.1.2 (Weighted Even Subgraphs)**.** *Let $\Delta \geq 3$ be an integer and let $\lambda > 0$, $\rho \in (0, 1]$ be reals. Then for any n-vertex graph $G = (V, E)$ of maximum degree $\Delta \leq O(1)$, the Glauber dynamics for sampling random weighted even subgraphs of G with parameters λ, ρ mixes in $Cn \log n$-steps where $C = C(\Delta, \lambda, \rho)$ is a constant independent of n.*

In particular, this gives an approximate sampling algorithm with running time $O(n \log n)$ for the ferromagnetic Ising model with edge activity $\beta_{\mathsf{Ising}} = \frac{1+\lambda}{1-\lambda}$ and vertex activity $\lambda_{\mathsf{Ising}} = \frac{1+\rho}{1-\rho}$.

Remark 31. In [JS93], 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 [JS93] is polynomial in $1/\rho$, and therefore this gives a $\mathsf{poly}(n)$ time algorithm for sampling weighted exact even subgraphs and hence for the ferromagnetic Ising model without external fields. Unfortunately, Theorem 6.1.2 cannot be used to obtain a sampler for $\rho = 0$ in the same way, since our bound on the mixing time of the Glauber dynamics (the constant C from Theorem 6.1.2) depends exponentially on $1/\rho$.

Finally, we simultaneously generalize [JS89; Dye+21; BCR20] 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 define the model more precisely. 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

$$\mu(\sigma) \propto \beta^{m_1(\sigma)}\gamma^{m_0(\sigma)}\lambda^{|\sigma^{-1}(1)|}, \quad \forall \sigma : E \to \{0, 1\} \tag{6.1}$$

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 6.1.3 (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 $\beta\gamma < 1$. Then for any n-vertex graph $G = (V, E)$ 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 Appendix D.

6.1.1 Zero-Freeness Implies Spectral Independence

We now state the main technical results of this chapter. To do this, we first precisely define the kinds of zero-free regions we will be working with.

Definition 32 (Stable Polynomial)**.** *Let $f(x_1, \ldots, x_n)$ be a multivariate polynomial. For $\Gamma_1, \ldots, \Gamma_n \subseteq \mathbb{C}$, we say f is $\prod_{i=1}^n \Gamma_i$-**stable** if $f(x_1, \ldots, x_n) \neq 0$ whenever $x_i \in \Gamma_i$ for all $i = 1, \ldots, n$. For simplicity, for $\Gamma \subseteq \mathbb{C}$, we say f is Γ-stable if f is $\prod_{i=1}^n \Gamma_i$-stable where $\Gamma_i = \Gamma$ for all $i = 1, \ldots, n$.*

Remark 32. One can of course consider zero-free regions which are much more general, i.e. any subset of $\mathbb{C}^n$. In this thesis, we will only consider the case where the zero-free region is a Cartesian product $\prod_{i=1}^n \Gamma_i$ of subsets $\Gamma_i \subseteq \mathbb{C}$.

Fix some ground set $\mathscr{U}$. To state our main result, we also recall that for $\boldsymbol{\gamma} \in \mathbb{R}^{\mathscr{U}}_{\geq 0}$ (e.g. $\boldsymbol{\gamma} = \mathbf{1}$), the distribution $\boldsymbol{\gamma} * \mu$ is the probability measure defined by $(\gamma * \mu)(\tau) \propto \mu(\tau) \cdot \prod_{i \in \tau} \gamma_i$. We say that we applied an *external field* $\boldsymbol{\gamma}$ to μ. The distribution $\boldsymbol{\gamma} * \mu$ is also sometimes called an *exponential tilt* of μ.

We prove the following.

Theorem 6.1.4 (Stability $\implies$ Spectral Independence, Homogeneous Version)**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for a finite ground set $\mathscr{U}$ and integer $0 \leq n \leq |\mathscr{U}|$. Let $\boldsymbol{\gamma} = (\gamma_i)_{i \in \mathscr{U}} \in \mathbb{R}^{\mathscr{U}}_{\geq 0}$ (e.g. $\boldsymbol{\gamma} = \mathbf{1}$), and let $\{\Gamma_i \subseteq \mathbb{C} : i \in \mathscr{U}\}$ be a collection of nonempty open connected subsets of $\mathbb{C}$. Assume g_μ is $\prod_{i \in \mathscr{U}} \Gamma_i$-stable, $\boldsymbol{\gamma} * \mu$ is $\mathfrak{b}$-marginally bounded, and $\boldsymbol{\gamma} \in \prod_{i \in \mathscr{U}} \Gamma_i$. Define*

$$\delta_i \stackrel{\mathsf{def}}{=} \frac{1}{\gamma_i} \cdot \operatorname{dist}(\gamma_i, \partial \Gamma_i) = \operatorname{dist}\left(1, \frac{1}{\gamma_i} \cdot \partial \Gamma_i\right), \quad \forall i \in \mathscr{U}$$
$$\delta \stackrel{\mathsf{def}}{=} \min_{i \in \mathscr{U}} \delta_i.$$

Then $\boldsymbol{\gamma} * \mu$ is η-spectrally independent with η satisfying the bound:

$$\eta \leq \frac{1}{\mathfrak{b}\delta^2}.$$

Furthermore, we have improved bounds in the following scenarios.

1. *If Γ_i is unbounded for all $i \in \mathscr{U}$, then*

$$\eta \leq \frac{4}{\delta}.$$

   Note that for this case, we do not require $\boldsymbol{\gamma} * \mu$ to be $\mathfrak{b}$-marginally bounded.

2. *If there exists $\gamma^* \in \mathbb{R}_{\geq 0}$ such that $\gamma_i \in (0, \gamma^*) \subseteq \Gamma_i$ for all $i \in \mathscr{U}$, then*

$$\eta \leq \frac{4}{\delta} \min\left\{\frac{1}{\mathfrak{b}} - 1, \frac{\gamma_{\max}}{\mathfrak{b} \cdot (\gamma^* - \gamma_{\max})} + 1\right\}$$

   where $\gamma_{\max} \stackrel{\mathsf{def}}{=} \max_{i \in \mathscr{U}} \gamma_i$.

For technical reasons, we will also need an analog of Theorem 6.1.4 which works under slightly weaker stability conditions for probability distributions over discrete product spaces; see Theorem 6.4.1 in Section 6.4 for the statement. We are not aware of a way to deduce this version directly from Theorem 6.1.4 (e.g. via a homogenization trick), although the proofs are nearly identical. Hence, to simplify exposition, we have relegated Theorem 6.4.1 to Section 6.4, as Theorem 6.1.4 and its proof provide the bulk of the ideas.

6.1.2 Relations with Previous Works

It was already known [FM92; AOR16] that *real stability*, i.e. Γ-stability where Γ is the upper half-plane, implies rapid mixing of a simple and local Markov chain for sampling from μ. This was not achieved using spectral independence, although it is well-known that real stability implies the *negative correlation* property and hence, spectral independence via e.g. Example 6. This stability requirement was significantly weakened in [Ali+21] to allow the regions Γ to be *sectors* $\{z = re^{i\omega} \in \mathbb{C} : |\omega| \leq \theta\}$ with constant aperture $0 < \theta \leq \pi/2$. They then used this to show rapid mixing for Markov chains sampling fixed-size matchings, commons bases of special kinds of matroids, and more. We refer the interested reader to [Ali+21] for further discussion. The techniques in this chapter build off of the ideas in [Ali+21]. Our technical results Theorems 6.1.4 and 6.4.1 strengthen theirs in the sense that 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.

To establish zero-free regions for our main applications, we utilize the approach in [Guo+21], which reduces the problem via Asano-Ruelle contractions [Asa70; Rue71] 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 [Wag09; BCR20] to establish zero-free regions, although their methods do not go through Asano–Ruelle contractions; see Section 6.6 for more details.

It was also shown in a sequence of papers [Bar16b; BS16; BS17a; Bar17a; Bar17b; Reg18] 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 ultimately requires a bound on the influence matrices 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 Appendix D.

Algorithmically, our results have several advantages over prior works utilizing zero-free regions. In particular, the polynomial interpolation method pioneered by Barvinok [Bar16a] 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 [PR17]. 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 additionally needs the zero-free region to contain a point at which the partition function is easily computable. On the other hand, our approach is fundamentally restricted to nonnegative real parameters, whereas Barvinok's approach can be extended to complex parameters. Furthermore, the interpolation method is *deterministic*, whereas our algorithms are randomized. It is an interesting question to what extent the currently known best randomized algorithms can be matched in performance by deterministic algorithms. We refer interested readers to [Bar16a] for more the interpolation method, and the complexity of computation partition functions more broadly.

6.2 Complex Analysis Background

We refer to subsets of the complex plane as *regions*.[2] Let $\Gamma \subseteq \mathbb{C}$ be a region. Denote the *complement* of Γ by $\Gamma^{\mathfrak{c}} \stackrel{\mathsf{def}}{=} \mathbb{C} \setminus \Gamma$, its *interior* by Γ°, its *closure* by $\overline{\Gamma}$, and its *boundary* by $\partial\Gamma$. We say Γ is *unbounded* if for any $M \in \mathbb{R}_{\geq 0}$ there exists $z \in \Gamma$ such that $|z| > M$; otherwise it is called *bounded*. We say Γ is *closed* if $\Gamma = \overline{\Gamma}$, *open* if $\Gamma = \Gamma^{\circ}$, and *compact* if Γ is closed and bounded (by the Heine-Borel Theorem). For $z \in \mathbb{C}$ let $\operatorname{dist}(z, \Gamma) \stackrel{\mathsf{def}}{=} \inf_{w \in \Gamma} |z - w|$ 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 $z + \Gamma = \Gamma + z \stackrel{\mathsf{def}}{=} \{z + w : w \in \Gamma\}$, $z\Gamma = \{zw : w \in \Gamma\}$, and $\Gamma^{-1} = (\Gamma \setminus \{0\})^{-1} \stackrel{\mathsf{def}}{=} \{w^{-1} : w \in \Gamma \setminus \{0\}\}$. For $\Gamma_1, \Gamma_2 \subseteq \mathbb{C}$, let $\Gamma_1 + \Gamma_2 \stackrel{\mathsf{def}}{=} \{z_1 + z_2 : z_1 \in \Gamma_1, z_2 \in \Gamma_2\}$ denote their *Minkowski sum* and $\Gamma_1 \cdot \Gamma_2 \stackrel{\mathsf{def}}{=} \{z_1 z_2 : z_1 \in \Gamma_1, z_2 \in \Gamma_2\}$ 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 $r \in \mathbb{R}_{\geq 0}$, 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 corresponding closed disk. We often simply write $\mathbb{D}$ for $\mathbb{D}(0, 1)$. Recalling Definition 29, for $0 \leq \theta < 2\pi$ and $\epsilon \geq 0$, we let

$$\mathbb{H}_{\theta,\epsilon} = \left\{e^{-i\theta} z : \Im(z) > \epsilon\right\} \subseteq \mathbb{C}.$$

denote the open (rotated and shifted) half-plane. A *circular region* in the complex plane is the interior or exterior of a disk, or an open half-plane.

In this chapter, we will primarily consider the open and closed left half-planes (or equivalently, the open and closed right-half planes) $\mathbb{H}_{-\frac{\pi}{2},\epsilon} = \{x + iy : x < -\epsilon\}$ and $\overline{\mathbb{H}}_{-\frac{\pi}{2},\epsilon} = \{x + iy : x \leq -\epsilon\}$. As such, just for this chapter, we will simply write $\mathbb{H}_{\epsilon}$ instead of $\mathbb{H}_{-\frac{\pi}{2},\epsilon}$, etc.

Let $\Gamma \subseteq \mathbb{C}$ be a nonempty open region. 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.

[2] This is slightly nonstandard, as a region is more commonly defined as a nonempty, open and simply connected subset of $\mathbb{C}$. It will be sufficient and convenient for our purposes for a region being any subset of $\mathbb{C}$.

A nonempty 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 and is called the *interior* while the other is unbounded and called the *exterior*. A nonempty 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.

Throughout, we select the principal branch for the complex functions $z \mapsto \log z$ and $z \mapsto z^{1/d}$. The following classical results from complex analysis will be useful.

Theorem 6.2.1 (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 6.2.2 (Riemann Mapping Theorem)**.** *Let $\Gamma \subseteq \mathbb{C}$ be a nonempty open simply connected region that is not $\mathbb{C}$. Then for every $z \in \Gamma$, there exists a unique biholomorphic mapping $f : \Gamma \to \mathbb{D}(0,1)$ such that*

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

Theorem 6.2.3 (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 6.2.4 (Multivariate Open Mapping Theorem, [KW17, Theorem 1.8.1])**.** *Let $n \geq 1$ be an integer and let $\mathcal{K} \subseteq \mathbb{C}^n$ be a nonempty 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.*

Theorem 6.2.5 (Hurwitz's 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.*

Lemma 6.2.6. *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 $\widetilde{\mathcal{S}}$ is a connected component of $\overline{\mathcal{S}}^{\mathfrak{c}}$, then $\widetilde{\mathcal{S}}$ is open and simply connected.*

Proof. Clearly $\widetilde{\mathcal{S}}$ is open and connected. If $\widetilde{\mathcal{S}}$ is not simply connected, then there exists a Jordan curve (simple closed curve) γ in $\widetilde{\mathcal{S}}$ whose interior region contains a point $z_0 \notin \widetilde{\mathcal{S}}$. 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 $\overline{\mathcal{S}}^{\mathfrak{c}}$ and thus $z_0 \notin \widetilde{\mathcal{S}}$ is connected to $\widetilde{\mathcal{S}}$ in $\overline{\mathcal{S}}^{\mathfrak{c}}$, contradicting to the assumption that $\widetilde{\mathcal{S}}$ is a connected component of $\overline{\mathcal{S}}^{\mathfrak{c}}$. 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 $\mathcal{P}$ in $\mathcal{S}$ connecting z and w. Note that $\mathcal{P}$ must intersect with γ, because the interior and exterior of γ are disconnected. This yields a contradiction since $\gamma \subseteq \widetilde{\mathcal{S}} \subseteq \mathcal{S}^{\mathfrak{c}}$ while $\mathcal{P} \subseteq \mathcal{S}$. □

6.3 Spectral Independence via Stability: The General Homogeneous Setting

As we saw from Lemma 5.1.1, Corollary 5.1.2, and Remark 25, there is an intimate relationship between $\nabla^2 \log g_\mu(\boldsymbol{\gamma})$ and $\mathcal{Q}_{\boldsymbol{\gamma} * \mu}$. The intuition is that the "farther" $\boldsymbol{\gamma}$ is away from the zeros of g_μ, the "smoother" $\log g_\mu$ is, and hence, the more bounded $\nabla^2 \log g_\mu(\boldsymbol{\gamma})$ will be. We will formalize this using the Schwarz-Pick Theorem, i.e. Theorem 6.2.1, following [Ali+21]. In particular, we

will construct a univariate holomorphic function $f : \mathbb{D}(0,1) \to \mathbb{D}(0,1)$ based roughly on $\log g_\mu$ such that $|f'(0)| \gtrsim_{\delta,\mathfrak{b}} \|\mathcal{I}_\mu\|_\infty$; since $|f'(0)| \leq 1$, this yields $\|\mathcal{I}_\mu\|_\infty \lesssim_{\delta,\mathfrak{b}} 1$ as desired.

This strategy was first devised in a beautiful work of [Ali+21]. There, the authors restricted attention to the case when the stability regions Γ_i are *sectors* in the complex plane with *aperture* $\theta \geq \Omega(1)$ and which are symmetric about the real axis. This includes the case of Hurwitz stable polynomials, i.e. polynomials which are stable w.r.t. the entire right half-plane of $\mathbb{C}$. In their setting, the univariate holomorphic functions f can be constructed without much effort by applying well-known conformal maps between such sectors and the unit disk.

However, here we are faced with a more challenging situation, since we try to establish spectral independence from an *arbitrary* collection of zero-free regions $\{\Gamma_i \subseteq \mathbb{C}\}_{i \in \mathscr{U}}$. There are no assumptions on the structure of the Γ_i beyond being nontrivially large, e.g. that they contain a ball around the point of interest with radius at least some constant independent of $n, |\mathscr{U}|$. So, designing good mappings from these stability regions to the unit disk is highly nontrivial. We achieve this by both carefully describing the regions which arise in our analysis, as well as utilizing tools from complex analysis, especially the Riemann Mapping Theorem (see Theorem 6.2.2).

We now make this more precise. We prove the following technical result, which we will show implies Theorem 6.1.4.

Proposition 6.3.1. *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for a finite ground set $\mathscr{U}$ and integer $0 \leq n \leq |\mathscr{U}|$. Let $\boldsymbol{\gamma} = (\gamma_i)_{i \in \mathscr{U}} \in \mathbb{R}^{\mathscr{U}}_{\geq 0}$ (e.g. $\boldsymbol{\gamma} = \mathbf{1}$), and let $\{\Gamma_i \subseteq \mathbb{C} : i \in \mathscr{U}\}$ be a collection of nonempty open connected subsets of $\mathbb{C}$. Assume g_μ is $\prod_{i \in \mathscr{U}} \Gamma_i$-stable, $\boldsymbol{\gamma} * \mu$ is $\mathfrak{b}$-marginally bounded, and $\boldsymbol{\gamma} \in \prod_{i \in \mathscr{U}} \Gamma_i$. Then for every $i \in \mathscr{U}$, we have the bound*

$$\sum_{j \in \mathscr{U} : j \neq i} |\mathcal{I}_{\boldsymbol{\gamma} * \mu}(i \to j)| \leq \min\left\{\frac{1}{p_i \delta^2}, \frac{4}{\delta} \operatorname{dist}(1, \mathcal{C}_i)\right\}$$

where for all $i \in \mathscr{U}$,

$$\delta_i \stackrel{\text{def}}{=} \frac{1}{\gamma_i} \cdot \operatorname{dist}(\gamma_i, \partial\Gamma_i) = \operatorname{dist}\left(1, \frac{1}{\gamma_i} \cdot \partial\Gamma_i\right), \quad \delta \stackrel{\text{def}}{=} \min_{i \in \mathscr{U}} \delta_i$$

$$p_i \stackrel{\text{def}}{=} \Pr_{\tau \sim \boldsymbol{\gamma} * \mu}[i \in \tau] \geq \mathfrak{b}$$

$$\mathcal{C}_i \stackrel{\text{def}}{=} -\frac{1}{p_i} \cdot \left(\frac{1}{\gamma_i}\Gamma_i - 1\right)^{-1} = \left\{-\frac{1}{p_i \cdot (z-1)} : z \in \left(\frac{1}{\gamma_i}\Gamma_i\right) \setminus \{1\}\right\}.$$

We prove this proposition in Section 6.3.1 below. All that remains to finish the proof of the theorem is to bound $\operatorname{dist}(1, \mathcal{C}_i)$. We have the following lemma.

Lemma 6.3.2. *For every $i \in \mathscr{U}$, we have the following:*

1. *If Γ_i is unbounded, then*

$$\operatorname{dist}(1, \mathcal{C}_i) \leq 1.$$

2. *If we define $\alpha_i \stackrel{\text{def}}{=} \inf(\Gamma_i \cap \mathbb{R}_{\geq 0}) \leq \gamma_i(1 - \delta_i)$ and $\beta_i \stackrel{\text{def}}{=} \sup(\Gamma_i \cap \mathbb{R}_{\geq 0}) \geq \gamma_i(1 + \delta_i)$, then*

$$\operatorname{dist}(1, \mathcal{C}_i) \leq \min\left\{\frac{\gamma_i}{p_i \cdot (\gamma_i - \alpha_i)} - 1, \frac{\gamma_i}{p_i \cdot (\beta_i - \gamma_i)} + 1\right\}.$$

With these tools in hand, we now prove Theorem 6.1.4.

Proof of Theorem 6.1.4. We have that

$$\eta \leq \|\mathcal{I}_{\boldsymbol{\gamma} * \mu}\|_\infty - 1 \leq \max_{i \in \mathscr{U}} \sum_{j \in \mathscr{U} : j \neq i} |\mathcal{I}_{\boldsymbol{\gamma} * \mu}(i \to j)| \leq \frac{1}{\mathfrak{b}\delta^2}$$

which establishes the first bound. If $0 \in \overline{\Gamma_i}$ and Γ_i is unbounded for all $i \in \mathscr{U}$, then by Lemma 6.3.2 and Item 1, we get an upper bound of $\frac{4}{\delta}$ instead. We now tackle the two remaining cases.

First, assume there exists $\gamma^* \in \mathbb{R}_{\geq 0}$ such that $\gamma_i \in (0, \gamma^*) \subseteq \Gamma_i$ for all $i \in \mathscr{U}$. Then $\alpha_i = 0$ and $\beta_i \geq \gamma^*$ for all $i \in \mathscr{U}$, which yields

$$\operatorname{dist}(1, \mathcal{C}_i) \leq \min\left\{\frac{1}{p_i} - 1, \frac{\gamma_i}{p_i \cdot (\gamma^* - \gamma_i)} + 1\right\}$$

and

$$\max_{i \in \mathscr{U}} \sum_{j \in \mathscr{U} : j \neq i} |\mathcal{I}_{\boldsymbol{\gamma} * \mu}(i \to j)| \leq \min\left\{\frac{1}{\mathfrak{b}} - 1, \frac{\gamma_{\max}}{\mathfrak{b} \cdot (\gamma^* - \gamma_{\max})} + 1\right\}$$

as desired. □

All that remains is to prove Proposition 6.3.1 and Lemma 6.3.2. We prove the latter now for simplicity.

Proof of Lemma 6.3.2. Informally, one could simply "plug in" $z = \infty$ to obtain a real number, and calculate its distance to 1 to establish Item 1. Similarly, one could "plug in" $z = \frac{\alpha_i}{\gamma_i}, \frac{\beta_i}{\gamma_i}$ to establish each bound in Item 2. However, since Γ_i is open, $\alpha_i, \beta_i, \infty \notin \Gamma_i$. Hence, to make the proof mathematically rigorous, in each case $\frac{\alpha_i}{\gamma_i}, \frac{\beta_i}{\gamma_i}, \infty$, one can take a monotone sequence $\{z_k\}_{k=1}^{\infty} \subseteq \left(\frac{1}{\gamma_i}\Gamma_i\right) \setminus \{1\}$ which converges to $\frac{\alpha_i}{\gamma_i}, \frac{\beta_i}{\gamma_i}, \infty$ respectively, plug in each z_k, and take $\liminf$. This is straightforward and we omit the details for brevity. □

6.3.1 Row-Wise Influence Bounds from Zero-Freeness: The General Homogeneous Setting

In this subsection, we prove Proposition 6.3.1. We carry out the strategy outlined in the opening discussion of Section 6.3. For every $i \in \mathscr{U}$, define the multivariate function

$$F_i(x_j : j \neq i) \stackrel{\text{def}}{=} \frac{\partial_{x_i} \log g_\mu(\boldsymbol{\gamma} \odot x)|_{x_i = 1}}{p_i} = \frac{1}{p_i} \cdot \frac{\gamma_i \cdot (\partial_i g_\mu)(\gamma_j \cdot x_j : j \neq i)}{g_\mu(\gamma_i; \gamma_j \cdot x_j : j \neq i)}. \tag{6.2}$$

We will need the following facts about the F_i.

Lemma 6.3.3. *The multivariate complex function $F_i : \prod_{j \in \mathscr{U} : j \neq i} \frac{1}{\gamma_j}\Gamma_j \to \mathbb{C}$ defined by Eq. (6.2) satisfies the following.*

1. *F_i is well-defined and holomorphic on $\prod_{j \in \mathscr{U} : j \neq i} \frac{1}{\gamma_j}\Gamma_j$. Furthermore, $F_i(\mathbf{1}) = 1$.*
2. *For every $j \in \mathscr{U}$ with $j \neq i$,*
$$(\partial_j F_i)(\mathbf{1}) = \mathcal{I}_{\boldsymbol{\gamma} * \mu}(i \to j).$$
3. *Suppose $F_i \not\equiv 1$. Then $1 \notin \overline{\mathcal{C}_i}$. Furthermore, if $\widetilde{\mathcal{C}_i}$ is the connected component of $\overline{\mathcal{C}_i}^{\mathfrak{c}}$ which contains 1, then $\widetilde{\mathcal{C}_i}$ is open and simply connected, and*
$$\operatorname{Im}(F_i) \subseteq \widetilde{\mathcal{C}_i}.$$

We prove Lemma 6.3.3 later, and first complete the proof of Proposition 6.3.1. Note that we may assume $F_i \not\equiv 1$, i.e. F_i is not constant, since otherwise, Lemma 6.3.3 Item 1 implies that $\mathcal{I}_{\boldsymbol{\gamma} * \mu}(i \to j)$ for all $j \in \mathscr{U}$ with $j \neq i$.

For each $j \in \mathscr{U}$ with $j \neq i$, let $\varphi_j : \mathbb{D}(0,1) \to \frac{1}{\gamma_j}\Gamma_j$ be a holomorphic function satisfying $\varphi_j(0) = 1$ and $\operatorname{sign}\left(\varphi_j'(0)\right) = \operatorname{sign}\left(\mathcal{I}_{\boldsymbol{\gamma} * \mu}(i \to j)\right)$; let $\varphi = (\varphi_j : j \neq i) : \mathbb{D}(0,1) \to \prod_{j \neq i} \frac{1}{\gamma_j}\Gamma_j$. Furthermore, let $\psi : \widetilde{\mathcal{C}_i} \to \mathbb{D}(0,1)$ be a holomorphic function such that $\psi'(1) \in \mathbb{R}_{\geq 0}$. Note that the restrictions $\varphi_j'(0) \in \mathbb{R}$ and $\psi'(1) \in \mathbb{R}_{\geq 0}$ can always be achieved, since we can replace $\varphi_j(z)$ by $\varphi_j(e^{i\theta_j} z)$, and replace $\psi(z)$ by $e^{i\theta}\psi(z)$ (for suitable angles θ_j, θ). We will build our functions in a

moment. First, observe that if we define $f_i \overset{\text{def}}{=} \psi \circ F_i \circ \varphi : \mathbb{D}(0,1) \to \mathbb{D}(0,1)$, then by the Chain Rule,

$$\begin{aligned} f_i'(0) &= \psi'(1) \sum_{j \in \mathscr{U} : j \neq i} \varphi_j'(0) \cdot (\partial_j F_i)(\mathbf{1}) \\ &= \psi'(1) \sum_{j \in \mathscr{U} : j \neq i} \varphi_j'(0) \cdot \mathcal{I}_{\gamma * \mu}(i \to j) && \text{(Lemma 6.3.3 Item 2)} \\ &\geq \psi'(1) \cdot \min_{j \in \mathscr{U} : j \neq i} \left\{ \left| \varphi_j'(0) \right| \right\} \cdot \sum_{j \in \mathscr{U} : j \neq i} \left| \mathcal{I}_{\gamma * \mu}(i \to j) \right|. \end{aligned}$$

Since $f_i : \mathbb{D}(0,1) \to \mathbb{D}(0,1)$ is holomorphic, by the Schwarz-Pick Theorem (see Theorem 6.2.1), we have $|f_i'(0)| \leq 1$, from which it follows that

$$\sum_{j \in \mathscr{U} : j \neq i} |\mathcal{I}_{\gamma * \mu}(i \to j)| \leq \frac{1}{\psi'(1) \cdot \min_{j \in \mathscr{U} : j \neq i} \left| \varphi_j'(0) \right|}.$$

We now construct our functions φ_j and ψ. Define $\varphi_j : \mathbb{D}(0,1) \to \frac{1}{\gamma_j}\Gamma_j$ by $\varphi_j(z) \overset{\text{def}}{=} 1 + s_{ij} \cdot \delta_j \cdot z$, where $s_{ij} = \operatorname{sign}(\mathcal{I}_{\gamma * \mu}(i \to j))$. Since $\operatorname{dist}\left(1, \frac{1}{\gamma_j}\partial\Gamma_j\right) = \delta_j$, the set $\frac{1}{\gamma_j}\Gamma_j$ contains the radius-δ_j open disk around 1, which precisely is the image of $\mathbb{D}(0,1)$ under φ_j. It follows immediately that $\min_{j \in \mathscr{U} : j \neq i} \left| \varphi_j'(0) \right| \geq \delta$, yielding an upper bound of

$$\sum_{j \in \mathscr{U} : j \neq i} |\mathcal{I}_{\gamma * \mu}(i \to j)| \leq \frac{1}{\delta \cdot \psi'(1)}.$$

To finish the proof, we construct two different choices of ψ, say $\psi_1, \psi_2 : \widetilde{\mathcal{C}}_i \to \mathbb{D}(0,1)$, such that

$$\frac{1}{\psi_1'(1)} \leq \frac{1}{p_i \cdot \delta} \tag{6.3}$$

$$\frac{1}{\psi_2'(1)} \leq 4 \cdot \operatorname{dist}(1, \mathcal{C}_i), \tag{6.4}$$

where each bound corresponds to an upper bound in the statement of the proposition.

We start with ψ_1, which is easier. Define $\psi_1(z) = p_i \cdot \delta_i \cdot z$. This is a well-defined holomorphic function from $\widetilde{\mathcal{C}}_i$ to $\mathbb{D}(0,1)$ because $\mathbb{D}(1, \delta_i) \subseteq \Gamma_i$ implies

$$\begin{aligned} \mathcal{C}_i = -\frac{1}{p_i} (\Gamma_i - 1)^{-1} \supseteq -\frac{1}{p_i} (\mathbb{D}(1, \delta_i) - 1)^{-1} = \frac{1}{p_i \cdot \delta_i} \cdot \overline{\mathbb{D}(0,1)}^{\mathfrak{c}} \\ \implies \widetilde{\mathcal{C}}_i \subseteq \overline{\mathcal{C}_i}^{\mathfrak{c}} \subseteq \frac{1}{p_i \cdot \delta_i} \cdot \mathbb{D}(0,1). \end{aligned}$$

Clearly, $\psi_1'(1) = p_i \cdot \delta_i \geq p_i \cdot \delta$, yielding the first upper bound in Eq. (6.3).

To construct ψ_2, observe that since $\widetilde{\mathcal{C}}_i$ is open and simply connected (by Lemma 6.3.3 Item 3), the Riemann Mapping Theorem (see Theorem 6.2.2) implies that there exists a (unique) biholomorphic mapping $\psi_2 : \widetilde{\mathcal{C}}_i \to \mathbb{D}(0,1)$ such that $\psi_2(1) = 0$ and $\psi_2'(1) \in \mathbb{R}_{\geq 0}$. Since ψ_2^{-1} is a bijective holomorphic function from $\mathbb{D}(0,1)$ to $\widetilde{\mathcal{C}}_i$ satisfying $\psi_2^{-1}(0) = 1$, by Koebe's One-Quarter Theorem (see Theorem 6.2.3), we have the bound

$$\frac{1}{4} \cdot \left| \left(\psi_2^{-1}\right)'(0) \right| \leq \operatorname{dist}\left(1, \partial\widetilde{\mathcal{C}}_i\right) \leq \operatorname{dist}(1, \mathcal{C}_i).$$

The Inverse Function Theorem then yields

$$\frac{1}{\psi_2'(1)} = \left| \left(\psi_2^{-1}\right)'(0) \right| \leq 4 \cdot \operatorname{dist}(1, \mathcal{C}_i)$$

which is the desired second upper bound from Eq. (6.3).

To complete the proof, we finally prove Lemma 6.3.3.

Proof of Lemma 6.3.3. Item 1 is immediate using zero-freeness of g_μ. Item 2 can be established via a calculation nearly identical to Lemma 5.1.1 and Remark 25. Hence, all that remains is to establish Item 3.

First, suffices to show the inclusion $\mathrm{Im}(F_i) \subseteq \mathcal{C}_i^{\mathfrak{c}}$. To see this, observe that since F_i is a non-constant holomorphic function and $\prod_{j\in\mathscr{U}} \frac{1}{\gamma_j}\Gamma_j$ is open and connected, by the Open Mapping Theorem (see Theorem 6.2.4), $\mathrm{Im}(F_i)$ is open and connected. Openness of $\mathrm{Im}(F_i)$ combined with $\mathrm{Im}(F_i) \subseteq \mathcal{C}_i^{\mathfrak{c}}$ implies $\mathrm{Im}(F_i) \subseteq (\mathcal{C}_i^{\mathfrak{c}})^\circ = \overline{\mathcal{C}_i}^{\mathfrak{c}}$, while connectivity of $\mathrm{Im}(F_i)$ combined with $1 = F_i(\mathbf{1}) \in \mathrm{Im}(F_i)$ implies $\mathrm{Im}(F_i)$ is contained in the connected component of $\overline{\mathcal{C}_i}^{\mathfrak{c}}$ which contains 1. $\overline{\mathcal{C}_i}^{\mathfrak{c}}$ is open and connected by definition. To show that it is simply connected, we appeal to Lemma 6.2.6.

We now show $\mathrm{Im}(F_i) \subseteq \mathcal{C}_i^{\mathfrak{c}}$. We prove this by contradiction. Assume there exists $\boldsymbol{\lambda} \in \prod_{j\in\mathscr{U}:j\neq i} \frac{1}{\gamma_j}\Gamma_j$ such that $F_i(\boldsymbol{\lambda}) \in \mathcal{C}_i$. Then there exists $z_i \in \frac{1}{\gamma_i}\Gamma_i$ with $z_i \neq 1$ such that

$$-\frac{1}{p_i \cdot (z_i - 1)} = F_i(\boldsymbol{\lambda}) = \frac{1}{p_i} \cdot \frac{\gamma_i \cdot (\partial_i g_\mu)(\boldsymbol{\gamma}_{-j} \odot \boldsymbol{\lambda})}{g_\mu(\gamma_i; \boldsymbol{\gamma}_{-j} \odot \boldsymbol{\lambda})}.$$

Rearranging, it follows that

$$g_\mu(\gamma_i; \boldsymbol{\gamma}_{-j} \odot \boldsymbol{\lambda}) - \gamma_i \cdot (\partial_i g_\mu)(\boldsymbol{\gamma}_{-i} \odot \boldsymbol{\lambda}) + z_i\gamma_i \cdot (\partial_i g_\mu)(\boldsymbol{\gamma}_{-i} \odot \boldsymbol{\lambda}) = 0.$$

But the left-hand side is precisely $g_\mu(\boldsymbol{\gamma} \odot \boldsymbol{\lambda}')$ where $\lambda'_j = \lambda_j$ for $j \neq i$ and $\lambda'_i = z_i$. This contradicts $\prod_{j\in\mathscr{U}} \Gamma_j$ stability of g_μ since $\boldsymbol{\gamma} \odot \boldsymbol{\lambda}' \in \prod_{j\in\mathscr{U}} \Gamma_j$. □

Remark 33. As we previously mentioned, this line of reasoning was first produced in [Ali+21], who specialized to the case when the stability regions Γ_i are open *sectors* in the right half-plane which are symmetric about the real axis (e.g. Hurwitz stable polynomials, where Γ_i is the entire right half-plane). Theorem 6.1.4 Item 1 recovers this special case, but holds at a much greater level of generality. If one has a more precise understanding of the stability regions Γ_i, then one could conceivably construct better maps φ, ψ which can yield better bounds.

6.4 Spectral Independence via Stability: Discrete Product Spaces

Our goal in this section is to establish an analog of Theorem 6.1.4, which we works under slightly weaker stability conditions for discrete product spaces such as the Boolean cube $\{0,1\}^n$. It is the same as Theorem 6.1.4 conceptually, and its proof follow a nearly identical line of argument. However, to be fully precise, we need it in some of our applications. We are not aware of a simple and direct method which recovers Theorem 6.4.1 from Theorem 6.1.4 (e.g. via a homogenization trick). To state it, let us first set up some additional notation which is specific to discrete product spaces.

Throughout this section, $\Sigma_1, \dots, \Sigma_n$ will denote finite alphabets (i.e. finite sets) with $|\Sigma_i| \geq 2$ for all i. We call $\prod_{i=1}^n \Sigma_i = \Sigma_1 \times \cdots \times \Sigma_n$ a *discrete product space*. Furthermore, for each $i = 1, \dots, n$, we will arbitrarily fix a *distinguished "reference" element* $0_i \in \Sigma_i$; for instance, a natural choice when $\Sigma_i = \{0,1\}$ is to take $0_i = 0$. When the choice of index i is clear from context, we often drop the subscript and just write 0.

We will be interested in probability distributions μ over $\prod_{i=1}^n \Sigma_i$, e.g. the Gibbs distribution of a graphical model like the hardcore gas model. Again, we may cast such a μ as a distribution over $\binom{\mathscr{U}}{n}$, where $\mathscr{U} \stackrel{\mathsf{def}}{=} \{(i, s_i) : i \in [n], s_i \in \Sigma_i\}$ is the collection of coordinate-assignment pairs. For convenience, we also write $\mathscr{U}_0 \stackrel{\mathsf{def}}{=} \{(i, 0_i) : i \in [n]\}$ for the collection of coordinate-assignment pairs which map coordinate i to the reference element 0_i; similarly, we write $\mathscr{U}_{-0} \stackrel{\mathsf{def}}{=} \mathscr{U} \setminus \mathscr{U}_0$ for the collection of coordinate-assignment pairs which exclude 0_i. For the purposes of this section, for a vector or collection $\boldsymbol{\gamma} = \{\gamma_{i,s}\}_{(i,s)\in\mathscr{U}}$ (perhaps taking values in $\mathbb{R}^{\mathscr{U}}$, $\mathbb{R}^{\mathscr{U}}_{\geq 0}$ or $\mathbb{C}^{\mathscr{U}}$), we write $\boldsymbol{\gamma}_{-0}$ for the subvector or subset with entries corresponding to $(i,s) \in \mathscr{U}_{-0}$; analogously, $\boldsymbol{\gamma}_0$ denotes the subvector or subset with entries corresponding to $(i,s) \in \mathscr{U}_0$.

We can of course study the homogeneous multivariate generating polynomial g_μ of μ and appeal to Theorem 6.1.4 by requiring there be a large zero-free region for every variable $x_{i,s}$ over all $i \in [n], s \in \Sigma_i$. In this section, we ask for something slightly weaker. For $\boldsymbol{\gamma}_0 = (\gamma_{i,s})_{(i,s)\in\mathscr{U}_0} \in \mathbb{R}^{\mathscr{U}_0}_{\geq 0}$,

consider the following multiaffine *inhomogeneous* polynomial

$$\begin{aligned}\tilde{g}_\mu(x) = \tilde{g}_{\mu,\boldsymbol{\gamma}_0}\left(x|_{\mathscr{U}_{-0}}\right) &\stackrel{\text{def}}{=} g_\mu(x)|_{x_{i,s}=\gamma_{i,s}\forall(i,s)\in\mathscr{U}_0} \\ &= \sum_{\sigma\in\prod_{i=1}^n \Sigma_i} \mu(\sigma) \prod_{i:\sigma(i)=0_i} \gamma_{i,0} \cdot \prod_{i:\sigma(i)\neq 0_i} x_{i,\sigma(i)}.\end{aligned} \tag{6.5}$$

For example, if μ is a probability distribution over $2^{[n]}$ (equivalently, over $\{0,1\}^n$) and $\boldsymbol{\gamma}_0 = \mathbf{1}$, then

$$g_\mu(x) = \sum_{S\subseteq[n]} \mu(S) \prod_{i\in S} x_{i,1} \prod_{i\notin S} x_{i,0}$$

while

$$\tilde{g}_\mu(x) = \sum_{S\subseteq[n]} \mu(S) \prod_{i\in S} x_i.$$

We have the following theorem.

Theorem 6.4.1 (Stability $\implies$ Spectral Independence, Discrete Product Version; [CLV21b])**.** *Let μ be a probability distribution over a discrete product space $\prod_{i=1}^n \Sigma_i$ for a positive integer $n \geq 1$. Let $\boldsymbol{\gamma} = (\gamma_{i,s})_{(i,s)\in\mathscr{U}} \in \mathbb{R}_{\geq 0}^{\mathscr{U}}$ (e.g. $\boldsymbol{\gamma} = \mathbf{1}$), and let $\{\Gamma_{i,s} \subseteq \mathbb{C} : (i,s) \in \mathscr{U}_{-0}\}$ be a collection of nonempty open connected subsets of $\mathbb{C}$. Assume $\tilde{g}_\mu = \tilde{g}_{\mu,\boldsymbol{\gamma}_0}$ is $\prod_{(i,s)\in\mathscr{U}_{-0}} \Gamma_{i,s}$-stable, $\boldsymbol{\gamma} * \mu$ is $\mathfrak{b}$-marginally bounded, and $\boldsymbol{\gamma}_{-0} \in \prod_{(i,s)\in\mathscr{U}_{-0}} \Gamma_{i,s}$. Define*

$$\begin{aligned}\delta_{i,s} &\stackrel{\text{def}}{=} \frac{1}{\gamma_{i,s}} \cdot \operatorname{dist}(\gamma_{i,s}, \partial\Gamma_{i,s}) = \operatorname{dist}\left(1, \frac{1}{\gamma_{i,s}} \cdot \partial\Gamma_{i,s}\right), \quad \forall (i,s) \in \mathscr{U}_{-0} \\ \delta &\stackrel{\text{def}}{=} \min_{(i,s)\in\mathscr{U}_{-0}} \delta_{i,s}.\end{aligned}$$

Then $\boldsymbol{\gamma} * \mu$ is η-spectrally independent with η satisfying the bound:

$$\eta \leq \frac{2}{\mathfrak{b}\delta^2}.$$

Furthermore, we have improved bounds in the following scenarios.

1. *For each $i = 1, \ldots, n$, let $\Gamma_{i,0} = \Gamma_{i,0_i}$ be the connected component of $\gamma_{i,0} \bigcap_{s\in\Sigma_i : s\neq 0_i} \frac{1}{\gamma_{i,s}}\Gamma_{i,s}$ that contains $\gamma_{i,0}$. If $\Gamma_{i,0}$ is unbounded and $0 \in \overline{\Gamma_{i,0}}$ for all $i = 1, \ldots, n$, then*

$$\eta \leq \frac{8}{\delta}.$$

   Note that for this case, we do not require $\boldsymbol{\gamma} * \mu$ to be $\mathfrak{b}$-marginally bounded.

2. *If there exists $\gamma^* \in \mathbb{R}_{\geq 0}$ such that $\gamma_{i,s} \in (0, \gamma^*) \subseteq \Gamma_{i,s}$ for all $(i,s) \in \mathscr{U}_{-0}$, then*

$$\eta \leq \frac{8}{\delta} \min\left\{\frac{1}{\mathfrak{b}} - 1, \frac{\gamma_{\max}}{\mathfrak{b}\cdot(\gamma^* - \gamma_{\max})} + 1\right\}$$

   where $\gamma_{\max} \stackrel{\text{def}}{=} \max_{(i,s)\in\mathscr{U}_{-0}} \gamma_{i,s}$.

3. *If there exists $\gamma^* \in \mathbb{R}_{\geq 0}$ such that $\gamma_{i,s} \in (\gamma^*, \infty) \subseteq \Gamma_{i,s}$ for all $(i,s) \in \mathscr{U}_{-0}$, then*

$$\eta \leq \frac{8}{\delta} \min\left\{\frac{1}{\mathfrak{b}} - 1, \frac{\gamma^*}{\mathfrak{b}\cdot(\gamma_{\min} - \gamma^*)} + 1\right\}$$

   where $\gamma_{\min} \stackrel{\text{def}}{=} \min_{(i,s)\in\mathscr{U}_{-0}} \gamma_{i,s}$.

Example 11. One might wonder if one can obtain a spectral independence bound which depend only on δ, even if Γ_i is bounded for all i. The following example shows this is not possible in general.

Consider the distribution μ over $\{0,1\}^n$ which assigns probability $1/2$ to $\emptyset$ and $[n]$, respectively. Up to scaling by a universal constant, the associated multiaffine nonhomogeneous polynomial is simply

$$\tilde{g}_\mu(x) = 1 + \prod_{i=1}^n x_i.$$

Clearly, $\tilde{g}_\mu$ is $\prod_{i=1}^n \Gamma_{i,1}$-stable with $\Gamma_{i,1} = \mathbb{D}$ for all $i = 1,\dots,n$, simply because whenever $x_1,\dots,x_n \in \mathbb{D}$, we have $|\prod_{i=1}^n x_i| = \prod_{i=1}^n |x_i| < 1$. Hence, if $\gamma_{i,1} \in [0,1]$ is bounded away from 0 and 1 by a universal constant for all $i = 1,\dots,n$, then $\delta \geq \Omega(1)$ independent of n. On the other hand, the distribution is clearly $(n-1)$-spectrally independent since the support of the local random walk $\mathcal{Q}_\mu$ is disconnected; one can also refer to Example 7 or carry out a direct computation of $\mathcal{I}_\mu$ (or Ψ_μ) and its eigenvalues. What saves the above theorem from this example is the fact that the best marginal lower bound $\mathfrak{b}$ one can establish for $(\mathbf{1},\boldsymbol{\gamma}) * \mu$ is exponentially small in n (when the $\gamma_{i,1}$ are bounded away from 0 and 1).

Let us first state and prove the corresponding analogs of Proposition 6.3.1 and Lemma 6.3.2. Again, the first gives a more generic bound on the total influence of a coordinate-assignment pair.

Proposition 6.4.2. *Let μ be a probability distribution over a discrete product space $\prod_{i=1}^n \Sigma_i$ for a positive integer $n \geq 1$. Let $\boldsymbol{\gamma} = (\gamma_{i,s})_{(i,s)\in\mathscr{U}} \in \mathbb{R}_{\geq 0}^{\mathscr{U}}$ (e.g. $\boldsymbol{\gamma} = \mathbf{1}$), and let $\{\Gamma_{i,s} \subseteq \mathbb{C} : (i,s) \in \mathscr{U}_{-0}\}$ be a collection of nonempty open connected subsets of $\mathbb{C}$. Assume $\tilde{g}_\mu = \tilde{g}_{\mu,\gamma_0}$ is $\prod_{(i,s)\in\mathscr{U}_{-0}} \Gamma_{i,s}$-stable, $\boldsymbol{\gamma} * \mu$ is $\mathfrak{b}$-marginally bounded, and $\boldsymbol{\gamma}_{-0} \in \prod_{(i,s)\in\mathscr{U}_{-0}} \Gamma_{i,s}$. Then for every $(i,s_i) \in \mathscr{U}$, we have the bound*

$$\sum_{(j,s_j)\in\mathscr{U}: j\neq i} |\mathcal{I}_{\boldsymbol{\gamma}*\mu}((i,s_i)\to(j,s_j))| \leq \min\left\{\frac{2}{p_{i,s_i}\delta^2}, \frac{8}{\delta}\operatorname{dist}(1,\mathcal{C}_{i,s_i})\right\}$$

where recall $\Gamma_{i,0} = \Gamma_{i,0_i}$ is the connected component of $\gamma_{i,0}\bigcap_{s\in\Sigma_i: s\neq 0_i} \frac{1}{\gamma_{i,s}}\Gamma_{i,s}$ for all $i = 1,\dots,n$, and

$$\begin{aligned}
\delta_{i,s} &\stackrel{\mathsf{def}}{=} \frac{1}{\gamma_{i,s}} \cdot \operatorname{dist}(\gamma_{i,s}, \partial\Gamma_{i,s}) = \operatorname{dist}\left(1, \frac{1}{\gamma_{i,s}}\cdot\partial\Gamma_{i,s}\right), \quad \forall (i,s) \in \mathscr{U}_{-0} \\
\delta &\stackrel{\mathsf{def}}{=} \min_{(i,s)\in\mathscr{U}_{-0}} \delta_{i,s} \\
p_{i,s} &\stackrel{\mathsf{def}}{=} \Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(i) = s] \geq \mathfrak{b}, \quad \forall (i,s) \in \mathscr{U} \\
\mathcal{C}_{i,s} &\stackrel{\mathsf{def}}{=} -\frac{1}{p_{i,s}} \cdot \left(\frac{1}{\gamma_{i,s}}\Gamma_{i,s} - 1\right)^{-1} = \left\{-\frac{1}{p_{i,s}\cdot(z-1)} : z \in \left(\frac{1}{\gamma_{i,s}}\Gamma_{i,s}\right)\setminus\{1\}\right\}, \quad s \neq 0_i \\
\mathcal{C}_{i,0} &\stackrel{\mathsf{def}}{=} \frac{1}{p_{i,0}}\left(\left(\frac{1}{\gamma_{i,0}}\Gamma_{i,0} - 1\right)^{-1} + 1\right) = \left\{\frac{z}{p_{i,0}\cdot(z-1)} : z \in \left(\frac{1}{\gamma_{i,0}}\Gamma_{i,0}\right)\setminus\{1\}\right\}.
\end{aligned}$$

The next lemma helps us bound the distances $\operatorname{dist}(1,\mathcal{C}_{i,s})$ appearing in Proposition 6.4.2.

Lemma 6.4.3. *For every $(i,s) \in \mathscr{U}$, we have the following:*

1. *If $s \neq 0_i$ and $\Gamma_{i,s}$ is unbounded, or $s = 0_i$ and $0 \in \overline{\Gamma_{i,s}}$, then*

$$\operatorname{dist}(1,\mathcal{C}_{i,s}) \leq 1.$$

2. *If we define $\alpha_{i,s} \stackrel{\mathsf{def}}{=} \inf(\Gamma_{i,s}\cap\mathbb{R}_{\geq 0})$ and $\beta_{i,s} \stackrel{\mathsf{def}}{=} \sup(\Gamma_{i,s}\cap\mathbb{R}_{\geq 0})$, then*

$$\operatorname{dist}(1,\mathcal{C}_{i,s}) \leq \min\left\{\frac{\gamma_{i,s}}{p_{i,s}\cdot(\gamma_{i,s}-\alpha_{i,s})} - 1, \frac{\gamma_{i,s}}{p_{i,s}\cdot(\beta_{i,s}-\gamma_{i,s})} + 1\right\}, \quad \forall i \in [n], s \in \Sigma_i - 0_i$$

and

$$\operatorname{dist}(1,\mathcal{C}_{i,0}) \leq \min\left\{\frac{\alpha_{i,0}}{p_{i,0}\cdot(\gamma_{i,0}-\alpha_{i,0})} + 1, \frac{\beta_{i,0}}{p_{i,0}\cdot(\beta_{i,0}-\gamma_{i,0})} - 1\right\}, \quad \forall i \in [n].$$

Note that $\alpha_{i,s} \leq \gamma_{i,s}(1-\delta_{i,s}), \beta_{i,s} \geq \gamma_{i,s}(1+\delta_{i,s})$ for all $(i,s) \in \mathscr{U}_{-0}$, and $\alpha_{i,0} \leq \gamma_{i,0}(1-\delta), \beta_{i,0} \geq \gamma_{i,0}(1+\delta)$ for all $i \in [n]$.

Proof of Theorem 6.4.1. One can follow the proof of Theorem 6.1.4 but apply Proposition 6.4.2 and Lemma 6.4.3 instead appropriately. We omit the details. □

Proof of Lemma 6.4.3. Similarly, one can follow the proof of Lemma 6.3.2. We omit the details. □

6.4.1 Row-Wise Influence Bounds from Zero-Freeness: Discrete Product Spaces

Our aim in this subection is to prove Proposition 6.4.2, following a similar argument to the proof of Proposition 6.3.1. Our analysis is only made slightly more complicated because we need to handle influences involving coordinate-assignment pairs (i, s) with $s = 0_i$.

We begin by first observing that for every $(i, s_i) \in \mathscr{U}$,

$$\sum_{(j,s_j)\in\mathscr{U}:j\neq i} |\mathcal{I}_{\boldsymbol{\gamma}*\mu}((i,s_i)\to(j,s_j))| \leq 2 \sum_{(j,s_j)\in\mathscr{U}:j\neq i, s_j\neq 0_j} |\mathcal{I}_{\boldsymbol{\gamma}*\mu}((i,s_i)\to(j,s_j))|. \tag{6.6}$$

This inequality follows simply because for every $(j, s) \in \mathscr{U}$,

$$\Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j) = s] = 1 - \sum_{s'\in\Sigma_j : s'\neq s} \Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j) = s'],$$

and the same holds under arbitrary conditionings. In particular, for every $j \neq i$,

$$\begin{aligned}
|\mathcal{I}_{\boldsymbol{\gamma}*\mu}((i,s_i)\to(j,0_j))| &= \left|\Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j)=0_j \mid \tau(i)=s_i] - \Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j)=0_j]\right| \\
&= \left|\sum_{s_j\neq 0_j} \Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j)=s_j \mid \tau(i)=s_i] - \sum_{s_j\neq 0_j} \Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j)=s_j]\right| \\
&\leq \sum_{s_j\neq 0_j} \left|\Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j)=s_j \mid \tau(i)=s_i] - \Pr_{\tau\sim\boldsymbol{\gamma}*\mu}[\tau(j)=s_j]\right| \\
&\qquad\qquad \text{(Triangle Inequality)} \\
&= \sum_{s_j\neq 0_j} |\mathcal{I}_{\boldsymbol{\gamma}*\mu}((i,s_i)\to(j,s_j))|
\end{aligned}$$

Summing over all $j \neq i$ then yields Eq. (6.6), which explains the extra factor of 2 when comparing Theorem 6.4.1 with Theorem 6.1.4. It also hints at why are able to get away with requiring stability for variables $x_{i,s}$ when $s \neq 0_i$, although this remains to be flushed out more precisely.

In light of Eq. (6.6) it suffices to show that for every $(i, s_i) \in \mathscr{U}$,

$$\sum_{(j,s_j)\in\mathscr{U}:j\neq i, s_j\neq 0_j} |\mathcal{I}_{\boldsymbol{\gamma}*\mu}((i,s_i)\to(j,s_j))| \leq \min\left\{\frac{1}{p_{i,s_i}\delta^2}, \frac{4}{\delta}\operatorname{dist}(1, \mathcal{C}_{i,s_i})\right\} \tag{6.7}$$

The case $s_i \neq 0_i$ can be handled identically to how Proposition 6.3.1 is proved, since we are then just looking at the influence between $(i, s_i), (j, s_j)$, both of which have variables in $\tilde{g}_\mu$ by $s_i \neq 0_i, s_j \neq 0_j$. So, we omit the details for this case.

We focus our analysis on the case of $s_i = 0_i$. Fix $i \in [n]$. We follow the proof of Proposition 6.3.1, but only highlight key differences. Define the multivariate function

$$F_{i,0}(x) \stackrel{\text{def}}{=} \frac{1}{p_{i,0}} \cdot \frac{\gamma_{i,0} \cdot \tilde{g}_{\mu^{i\leftarrow 0}}(\boldsymbol{\gamma}\odot x)}{\tilde{g}_\mu(\boldsymbol{\gamma}\odot x)}. \tag{6.8}$$

We need the following facts about $F_{i,0}$, which are analogs of Lemma 6.3.3.

Lemma 6.4.4. *The multivariate complex function $F_{i,0} : \prod_{(j,s)\in\mathscr{U}_{-0}:j\neq i} \frac{1}{\gamma_{j,s}}\Gamma_{j,s} \to \mathbb{C}$ defined by Eq. (6.8) satisfies the following.*

1. *$F_{i,0}$ is well-defined and holomorphic on $\prod_{(j,s)\in\mathscr{U}_{-0}:j\neq i} \frac{1}{\gamma_{j,s}}\Gamma_{j,s}$. Furthermore, $F_{i,0}(\mathbf{1}) = 1$.*
2. *For every $(j, s_j) \in \mathscr{U}$ with $j \neq i$ and $s_j \neq 0_j$,*

$$(\partial_{j,s_j} F_{i,0})(\mathbf{1}) = \mathcal{I}_{\boldsymbol{\gamma}*\mu}((i,0)\to(j,s_j)).$$

3. *Suppose* $F_{i,0} \not\equiv 1$. *Then* $1 \notin \overline{\mathcal{C}_{i,0}}$. *Furthermore, if* $\widetilde{\mathcal{C}}_{i,0}$ *is the connected component of* $\overline{\mathcal{C}_{i,0}}^{\mathbf{c}}$ *which contains* 1, *then* $\widetilde{\mathcal{C}}_{i,0}$ *is open and simply connected, and*

$$\mathrm{Im}(F_{i,0}) \subseteq \widetilde{\mathcal{C}}_{i,0}.$$

With this lemma in hand, the rest of the argument proceeds identically to the proof of Proposition 6.3.1, using essentially the same functions $\psi : \widetilde{\mathcal{C}}_{i,0} \to \mathbb{D}(0,1)$ and $\varphi_{j,s} : \mathbb{D}(0,1) \to \Gamma_{j,s}$ as in the proof of Proposition 6.3.1. We omit the details here. All the remains is to prove Lemma 6.4.4.

Proof of Lemma 6.4.4. Item 1 follows from stability of $\tilde{g}_\mu$, while Item 2 follows from direct calculations. We omit details here and refer the reader to the proof of Lemma 6.3.3.

For Item 3, again we first show that $\mathrm{Im}(F_{i,0}) \subseteq \mathcal{C}_{i,0}^{\mathbf{c}}$. Suppose for the sake of contradiction that $F_{i,0}(\boldsymbol{\lambda}) \in \mathcal{C}_{i,0}$ for some $\boldsymbol{\lambda} \in \prod_{(j,s)\in\mathscr{U}_{-0}:j\neq i} \frac{1}{\gamma_{j,s}}\Gamma_{j,s}$. Then there exists $z_{i,0} \in \frac{1}{\gamma_{i,0}}\Gamma_{i,0}$ such that

$$\frac{z_{i,0}}{p_{i,0}\cdot(z_{i,0}-1)} = F_{i,0}(\boldsymbol{\lambda}) = \frac{1}{p_{i,0}} \cdot \frac{\gamma_{i,0}\cdot \tilde{g}_{\mu^{i\leftarrow 0}}(\boldsymbol{\gamma}\odot\boldsymbol{\lambda})}{\tilde{g}_\mu(\boldsymbol{\gamma}\odot\boldsymbol{\lambda})},$$

where we extend $\boldsymbol{\lambda}$ to all of $\mathscr{U}_{-0}$ by taking $\lambda_{i,s} = 1$ for every $s \neq 0_i$. Rearranging, it follows that

$$\begin{aligned}
0 &= \gamma_{i,0}\cdot \tilde{g}_{\mu^{i\leftarrow 0}}(\boldsymbol{\gamma}\odot\boldsymbol{\lambda}) + z_{i,0}\cdot \tilde{g}_\mu(\boldsymbol{\gamma}\odot\boldsymbol{\lambda}) - z_{i,0}\cdot\gamma_{i,0}\cdot \tilde{g}_{\mu^{i\leftarrow 0}}(\boldsymbol{\gamma}\odot\boldsymbol{\lambda}) \\
&= \gamma_{i,0}\cdot \tilde{g}_{\mu^{i\leftarrow 0}}(\boldsymbol{\gamma}\odot\boldsymbol{\lambda}) + \sum_{s\in\Sigma_i : s\neq 0_i} z_{i,0}\gamma_{i,s}\cdot \tilde{g}_{\mu^{i\leftarrow s}}(\boldsymbol{\gamma}\odot\boldsymbol{\lambda}) \\
&= \tilde{g}_\mu(\boldsymbol{\gamma}\odot\boldsymbol{\lambda}')
\end{aligned}$$

where $\boldsymbol{\lambda}'$ is given by $\lambda'_{j,s} = \lambda_{j,s}$ for all $(j,s) \in \mathscr{U}$ with $j \neq i$, $\lambda'_{i,0_i} = \lambda_{i,0_i}$, and $\lambda'_{i,s} = z_{i,0}$ for all $s \neq 0_i$. Since $\lambda'_{i,s} = z_{i,0} \in \frac{1}{\gamma_{i,0}}\Gamma_{i,0} \subseteq \bigcap_{s\neq 0_i} \frac{1}{\gamma_{i,s}}\Gamma_{i,s}$ for all $s \in \Sigma_i - 0_i$ and $\lambda_{j,s} \in \frac{1}{\gamma_{j,s}}\Gamma_{j,s}$ for every $(j,s) \in \mathscr{U}_{-0}$ with $j \neq i$ by construction, $\boldsymbol{\gamma}\odot\boldsymbol{\lambda}' \in \prod_{(j,s)\in\mathscr{U}_{-0}}\Gamma_{j,s}$. Combined with $\tilde{g}_\mu(\boldsymbol{\gamma}\odot\boldsymbol{\lambda}') = 0$, this contradicts $\prod_{(j,s)\in\mathscr{U}_{-0}}\Gamma_{j,s}$-stability of $\tilde{g}_\mu$. □

6.5 Handling Conditional Distributions

To apply local-to-global theorems such as Theorem 2.3.1 to obtain rapid mixing results, one needs to bound the spectral independence of all conditional distributions. Hence, in order to apply Theorems 6.1.4 and 6.4.1 to all such conditional distributions, one naturally should demand that there be a corresponding zero-free region for the associated conditional multivariate generating polynomials/partition functions.

An example where this *fails* is the famous ferromagnetic Ising model without external field. The celebrated Lee–Yang Circle Theorem [LY52] says that the multivariate partition function is $\mathbb{D}$-stable and $\overline{\mathbb{D}}^{\mathbf{c}}$-stable. However, when a pinning is applied, particularly when some vertices are pinned to $+$ while some others are pinned to $-$, we do not have $\mathbb{D}$-stability nor $\overline{\mathbb{D}}^{\mathbf{c}}$-stability for the conditional partition function. To see this, notice that such a pinning can result in *inconsistent external fields* in the corresponding conditional distribution; some fields are < 1 (hence in $\mathbb{D}$) while others are > 1 (hence in $\overline{\mathbb{D}}^{\mathbf{c}}$), 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 (see, for instance, [GM07]).

However, in this section, we'll see that if the zero-free regions are unbounded, then we can get the same stability regions for all conditional distributions *for free*. This can greatly simplify some proofs, where one just needs to establish an unbounded stability region for the unconditional partition function.

Lemma 6.5.1 (Stability Under Conditioning, Homogeneous Version)**.** *Let* μ *be a probability distribution over* $\binom{\mathscr{U}}{n}$ *for a finite ground set* $\mathscr{U}$ *and integer* $0 \leq n \leq |\mathscr{U}|$. *Suppose its homogeneous multivariate generating polynomial* g_μ *is* $\prod_{i\in\mathscr{U}}\Gamma_i$*-stable where* Γ_i *is unbounded for all* $i \in \mathscr{U}$. *Then for every feasible* $\tau \subseteq \mathscr{U}$, *the generating polynomial* g_{μ^τ} *of the conditional distribution* μ^τ *is* $\prod_{i\in\mathscr{U}\setminus\tau}\Gamma_i$*-stable.*

For discrete product measures where we consider the inhomogeneous generating polynomial $\tilde{g}_\mu$ (see Eq. (6.5)), we will also demand that 0 be contained in the closure of the stability regions. The intuition for this will become clear when we give the proof below.

Lemma 6.5.2 (Stability Under Conditioning, Discrete Product Version; [CLV21b])**.** *Let μ be a probability distribution over a discrete product space $\prod_{i=1}^n \Sigma_i$ for a positive integer $n \geq 1$. Suppose its inhomogeneous multivariate generating polynomial $\tilde{g}_\mu$ is $\prod_{(i,s)\in\mathscr{U}_{-0}} \Gamma_{i,s}$-stable where $\Gamma_{i,s}$ is unbounded and $0 \in \overline{\Gamma_{i,s}}$ for all $(i,s) \in \mathscr{U}_{-0}$. Then for every feasible pinning τ on any subset of coordinates $S \subseteq [n]$, the inhomogeneous multivariate generating polynomial $\tilde{g}_{\mu^\tau}$ of the conditional distribution μ^τ is $\prod_{(i,s)\in\mathscr{U}_{-0}:i\notin S} \Gamma_{i,s}$-stable.*

Lemma 6.5.1 follows essentially from Lemma 5.1.1, which says that conditioning μ on τ corresponds to differentiating g_μ w.r.t. variables corresponding to elements in τ, and the fact that differentiation preserves stability for multiaffine polynomials. The extra condition that 0 be in the closure of the stability regions for Lemma 6.5.2 stems from having to handle the special reference elements $0_i \in \Sigma_i$, which have no variables in $\tilde{g}_\mu$. More specifically, for each $i = 1, \ldots, n$, plugging in 0 to $x_{i,s}$ for all $(i,s) \in \mathscr{U}_{-0}$ is equivalent to conditioning on $(i, 0_i)$. All we need is to ensure that these operations indeed preserve stability.

Lemma 6.5.3. *Let $p \in \mathbb{C}[z_1, \ldots, z_n]$ be a multiaffine polynomial, and assume p is $\prod_{i=1}^n \Gamma_i$-stable for nonempty open connected regions $\Gamma_1, \ldots, \Gamma_n \subseteq \mathbb{C}$. Then the following hold.*

1. *(Inversion) The polynomial $q_1(z_1, \ldots, z_n) \stackrel{\text{def}}{=} z_1 \cdot p(1/z_1, z_2, \ldots, z_n)$ is $\left(\Gamma_1^{-1} \times \prod_{\ell=2}^n \Gamma_\ell\right)$-stable;*
2. *(Specialization) If $0 \in \overline{\Gamma_1}$, then the polynomial $q_2(z_1, \ldots, z_n) \stackrel{\text{def}}{=} p(0, z_2, \ldots, z_n)$ is either $\prod_{\ell=2}^n \Gamma_\ell$-stable or identically zero;*
3. *(Differentiation) If Γ_1 is unbounded, then the polynomial $q_3(z_1, \ldots, z_n) \stackrel{\text{def}}{=} (\partial_1 p)(z_1, z_2, \ldots, z_n)$ is either $\prod_{\ell=2}^n \Gamma_\ell$-stable or identically zero.*

Proof. Let us first consider inversion. Suppose for contradiction that q_1 is not $\left(\Gamma_1^{-1} \times \prod_{\ell=2}^n \Gamma_\ell\right)$-stable. Then there exists $w_1 \in \Gamma_1^{-1}$ and $z_\ell \in \Gamma_\ell$ for $2 \leq \ell \leq n$ such that $q_1(w_1, z_2, \ldots, z_n) = 0$. Note that $w_1 = 1/z_1$ for some $z_1 \in \Gamma_1 \setminus \{0\}$. It follows that

$$0 = z_1 q_1(w_1, z_2, \ldots, z_n) = z_1 w_1 p\left(\frac{1}{w_1}, z_2, \ldots, z_n\right) = p(z_1, z_2, \ldots, z_n),$$

contradicting to the stability of p. Hence, we have the desired stability for q_1.

Next, consider specialization. Since Γ_1 is open and $0 \in \overline{\Gamma_1}$, there exists a sequence of complex numbers $\{\zeta_m\}_{m=1}^\infty \subseteq \Gamma_1$ such that $\lim_{m\to\infty} \zeta_m = 0$. 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 on p. Furthermore, the sequence $\{f_m\}_{m=1}^\infty$ converges to q_2 coefficient-wise, and hence uniformly on compact subsets; see, e.g., Lemma 33 in [Ali+21]. Hurwitz's Theorem (see Theorem 6.2.5) then implies that q_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 $q_1(z_1, z_2, \ldots, z_n) = z_1 p(\frac{1}{z_1}, z_2, \ldots, z_n)$ is $\left(\Gamma_1^{-1} \times \prod_{\ell=2}^n \Gamma_\ell\right)$-stable. Now observe that, for a multiaffine polynomial p, the derivative q_3 of p with respect to z_1 is the same as specialization of q_1 at $z_1 = 0$:

$$q_3(z_2, \ldots, z_n) = (\partial_1 p)(z_1, z_2, \ldots, z_n) = q_1(0, z_2, \ldots, z_n).$$

Hence, we immediately conclude from previous results that q_3 is either $\prod_{\ell=2}^n \Gamma_\ell$-stable or identically zero. □

Lemmas 6.5.1 and 6.5.2 are immediate consequences of Lemma 6.5.3.

Proof of Lemma 6.5.1. Observe that g_μ is multaffine. The lemma then follows from Lemma 5.1.1 and Lemma 6.5.3 Item 3. Notice that the conditional partition functions are never identically zero since pinnings are extendable to valid full configurations. □

Proof of Lemma 6.5.2. Observe that $\tilde{g}_\mu$ (see Eq. (6.5)) is multiaffine. The natural analog of Lemma 5.1.1 holds as well, where differentiating w.r.t. $z_{i,s}$ for $(i,s) \in \mathscr{U}_{-0}$ corresponds to conditioning on $\sigma(i) = s$. Conditioning on $\sigma(i) = 0_i$ is handled by setting $z_{i,s} = 0$ for all $s \in \Sigma_i \setminus \{0_i\}$. The lemma then follows from Lemma 6.5.3 Items 2 and 3. Again, notice that the conditional partition functions are never identically zero since pinnings are extendable to valid full configurations. □

6.6 Spectral Independence and Zero-Freeness 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

$$\mathcal{Z}_G(\boldsymbol{\lambda}) \stackrel{\text{def}}{=} \sum_{\sigma : E \to \{0,1\}} \prod_{v \in V} f_v\left(\left|\sigma_{E(v)}\right|\right) \prod_{e \in E : \sigma_e = 1} \lambda_e, \tag{6.9}$$

where $E(v)$ is the set of all edges adjacent to v, $\sigma_{E(v)}$ is the configuration restricted on $E(v)$, and $\left|\sigma_{E(v)}\right|$ 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 $\mathcal{Z}_G(\mathbf{1})$ yields the number of *matchings* (of any size) in G, that is, subsets of edges such that every vertex is incident to at most one selected edge.

- **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 $\mathcal{Z}_G(\mathbf{1})$ 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 $\mathcal{Z}_G(\mathbf{1})$ 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, e.g. consider the cycle)

- **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
$$f_v(k) = \begin{cases} \beta^{\binom{k}{2}} \beta^{\binom{d-k}{2}}, & \text{if } 0 \leq k \leq d; \\ 0, & \text{otherwise.} \end{cases}$$

In all of the above examples, prior works managed to show that the Glauber dynamics admits an inverse polynomial spectral gap ([JS89] for matchings, [HLZ16] for edge covers, [JS93] for weighted even subgraphs, and [Dye+21] for the Ising model in the antiferromagnetic $\beta < 1$ regime). Furthermore, all of these results were obtained via the canonical paths method [JS89], and its winding extension [McQ13]. 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 [CLV21a] (see Theorem 10.0.1 in Chapter 10), we obtain optimal mixing times and sub-Gaussian concentration results for these problems in the bounded-degree regime.

One of the convenient aspects of our approach is that establishing the required zero-root region for the complicated multivariate partition function can be boiled down to establishing stability for a bounded-degree *univariate* polynomial with coefficients derived from the local functions f_v. This was one of the main insights of [Wag09; Guo+21; BCR20]. 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

$$P_v(z) = \sum_{k=0}^{d} \binom{d}{k} f_v(k) z^k. \tag{6.10}$$

[Guo+21] showed using *Asano-Ruelle contractions* [Asa70; Rue71] 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 Eq. (6.9) is Γ-stable where $\Gamma = \left[-(\Phi^{\mathfrak{c}})^2\right]^{\mathfrak{c}}$. 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 exactly $\deg(v)$; under these assumptions, one can apply the famous Grace-Walsh-Szegö Coincidence Theorem to the local polynomials, see [Guo+21; BB09b]. A straightforward generalization of their techniques yields the following.

Theorem 6.6.1 ([Guo+21])**.** *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 $\mathcal{Z}_G(\boldsymbol{\lambda})$ is $\prod_{e \in E} \Gamma_e$-stable, where for each edge $e = \{u, v\}$, $\Gamma_e = (-\Phi_u^{\mathfrak{c}} \cdot \Phi_v^{\mathfrak{c}})^{\mathfrak{c}} \subseteq \mathbb{C}$.*

Using Theorem 6.6.1, [Guo+21] 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 6.1.1 to 6.1.3 build upon these zero-free results as well as Theorems 6.4.1 and 10.0.1 (note that we can obtain spectral independence for matchings from Theorems 6.4.1 and 6.6.1, which was already known in [CLV21a] with a better bound by correlation decay proofs; see Appendix B). Zero-free regions were also established for weighted edge covers and the antiferromagnetic Ising model on line graphs in [BCR20], using techniques from [Wag09].

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 $\mathbb{H}_\epsilon = \{x+iy : x < -\epsilon\}$ and $\overline{\mathbb{H}}_\epsilon = \{x+iy : x \leq -\epsilon\}$ for $\epsilon \in \mathbb{R}_{\geq 0}$.

Lemma 6.6.2 (Lemma 5 in [Guo+21])**.** *For $\epsilon > 0$, let $\Gamma = \left(-\overline{\mathbb{H}}_\epsilon^2\right)^{\mathfrak{c}}$. Then Γ contains $\mathbb{R}_{\geq 0}$, and for every $\lambda \in \mathbb{R}_{\geq 0}$ we have $\operatorname{dist}(\lambda, \partial\Gamma) = \lambda + \epsilon^2$ if $\lambda \in (0, \epsilon^2)$, and $\operatorname{dist}(\lambda, \partial\Gamma) = 2\epsilon\sqrt{\lambda}$ if $\lambda \in [\epsilon^2, \infty)$.*

For completeness, we provide a proof in Section 6.6.2. 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 6.1.1 to 6.1.3. Note that by Lemma 6.5.2 and Theorem 6.4.1, 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 6.1.1. By Theorem 10.0.1, it suffices to prove η-spectral independence for $\eta = O_{\Delta,\lambda,\rho}(1)$. By Theorem 6.4.1 and Lemma 6.5.2, it suffices to prove that the multivariate partition function Eq. (6.9) is Γ-stable, where $\Gamma \subseteq \mathbb{C}$ is an open connected region containing $\mathbb{R}_{\geq 0}$ 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 $\overline{\mathbb{H}}^{\mathfrak{c}}_{1/2}$-stable. Then by Theorem 6.6.1, the inversion of the weighted edge cover partition function $\mathcal{Z}_G(\boldsymbol{\lambda})$ is $\left(-\overline{\mathbb{H}}^2_{1/2}\right)^{\mathfrak{c}}$-stable, and therefore $\mathcal{Z}_G(\boldsymbol{\lambda})$ is Γ-stable for

$$\Gamma = \left[\left(-\overline{\mathbb{H}}^2_{1/2}\right)^{\mathfrak{c}}\right]^{-1} = \left[-\overline{\mathbb{D}}(-1,1)^2\right]^{\mathfrak{c}}.$$

This region Γ is also derived in [BCR20]. We remark that the region $-\overline{\mathbb{D}}(-1,1)^2$ is cardioid-shaped, and its complement Γ is an open connected region containing $\mathbb{R}_{\geq 0}$; see Lemma 3.9 and Figure 1 in [BCR20]. Hence, we have $\mathbb{R}_{\geq 0} \subseteq \Gamma$ and $\delta = \Omega_{\Delta,\lambda,\rho}(1)$ as desired. □

Proof of Theorem 6.1.2. We may assume $\rho \in (0,1)$ since if $\rho = 1$ then we get a trivial product distribution. Once again, by Theorem 10.0.1, it suffices to prove η-spectral independence for $\eta = O_{\Delta,\lambda,\rho}(1)$, and by Theorem 6.4.1 and Lemma 6.5.2, it suffices to prove that the multivariate partition function Eq. (6.9) is Γ-stable, where $\Gamma \subseteq \mathbb{C}$ is an open connected region containing $\mathbb{R}_{\geq 0}$ and $\delta = \frac{1}{\lambda}\operatorname{dist}(\lambda, \partial\Gamma) = \Omega_{\Delta,\lambda,\rho}(1)$.

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 $t_v \in \mathbb{R}_{\geq 0}$ is given by

$$t_v = \left(\frac{1+\rho}{1-\rho}\right)^{1/\deg(v)} > 1.$$

It follows that P_v is $\left[\overline{\mathbb{D}}\left(-\frac{t_v^2+1}{t_v^2-1}, \frac{2t_v}{t_v^2-1}\right)\right]^{\mathfrak{c}}$-stable. Then by Theorem 6.6.1, $\mathcal{Z}_G(\boldsymbol{\lambda})$ is $\prod_{e\in E}\Gamma_e$-stable, where for each edge $e = uv \in E$,

$$\Gamma_e = \left[-\overline{\mathbb{D}}\left(-\frac{t_u^2+1}{t_u^2-1}, \frac{2t_u}{t_u^2-1}\right) \cdot \overline{\mathbb{D}}\left(-\frac{t_v^2+1}{t_v^2-1}, \frac{2t_v}{t_v^2-1}\right)\right]^{\mathfrak{c}}.$$

In particular, $\mathcal{Z}_G(\boldsymbol{\lambda})$ is Γ-stable for

$$\Gamma = \left[-\overline{\mathbb{D}}\left(-\frac{t^2+1}{t^2-1}, \frac{2t}{t^2-1}\right)^2\right]^{\mathfrak{c}} \subseteq \Gamma_e, \quad \forall e \in E, \quad \text{where } t = \left(\frac{1+\rho}{1-\rho}\right)^{1/\Delta} > 1.$$

The region Γ is open and connected. Observe that we have $\Gamma \supseteq \left(-\overline{\mathbb{H}}^2_{\frac{t-1}{t+1}}\right)^{\mathfrak{c}}$. Hence, by Lemma 6.6.2 we have $\mathbb{R}_{\geq 0} \subseteq \Gamma$ and $\delta = \Omega_{\Delta,\lambda,\rho}(1)$ as desired. □

Proof of Theorem 6.1.3. By Theorem 10.0.1 it suffices to prove η-spectral independence for $\eta = O_{\Delta,\beta,\gamma,\lambda}(1)$. By Theorem 6.4.1 and Lemma 6.5.2 it suffices to prove that the multivariate partition function Eq. (6.9) is Γ-stable, where $\Gamma \subseteq \mathbb{C}$ is an open connected region containing $\mathbb{R}_{\geq 0}$ and $\delta = \frac{1}{\lambda}\operatorname{dist}(\lambda, \partial\Gamma) = \Omega_{\Delta,\beta,\gamma,\lambda}(1)$.

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 6.6.3 below (see Section 6.6.1 for the proof), all roots of these polynomials are strictly negative reals, i.e. they are contained in $(-\infty, -\epsilon_{\deg(v)}]$ for some constant $\epsilon_{\deg(v)} =$

$\epsilon_{\deg(v)}(\beta,\gamma) > 0$ depending only on $\deg(v), \beta, \gamma$. Then by Theorem 6.6.1, $\mathcal{Z}_G(\boldsymbol{\lambda})$ is $\prod_{e\in E}\Gamma_e$-stable, where for each edge $e = uv \in E$,

$$\Gamma_e = \left(-\overline{\mathbb{H}}_{\epsilon_{\deg(u)}} \cdot \overline{\mathbb{H}}_{\epsilon_{\deg(v)}}\right)^{\mathfrak{c}}.$$

In particular, $\mathcal{Z}_G(\boldsymbol{\lambda})$ is Γ-stable for $\Gamma = \left(-\overline{\mathbb{H}}_{\epsilon}^2\right)^{\mathfrak{c}}$ where $\epsilon = \min_{1\le d\le\Delta}\epsilon_d$ depends only on Δ, β, γ. The region Γ is open and connected, and by Lemma 6.6.2 it contains $\mathbb{R}_{\ge 0}$ and we have $\delta = \Omega_{\Delta,\beta,\gamma,\lambda}(1)$ as desired. □

6.6.1 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 6.1.3 above. We generalize a result due to [BCR20] 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 6.6.3 (Generalization of Lemma 4.3 in [BCR20])**.** *For every $\beta \ge 0, \gamma > 0$ with $\beta\gamma < 1$ and every positive integer $d \ge 1$, 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 6.6.4. *For every $\beta \ge 0$, $\gamma > 0$ and every positive integer $d \ge 1$, we have that*

$$P_{d+1}(z) = \gamma^d P_d(z/\gamma) + zP_d(\beta z).$$

Proof. We have

$$\begin{aligned}
P_{d+1}(z) &= \sum_{k=0}^{d+1}\underbrace{\binom{d+1}{k}}_{=\binom{d}{k}+\binom{d}{k-1}}\beta^{\binom{k}{2}}\gamma^{\binom{d+1-k}{2}}z^k \\
&= \sum_{k=0}^{d}\binom{d}{k}\beta^{\binom{k}{2}}\gamma^{\binom{d+1-k}{2}}z^k + \sum_{k=0}^{d}\binom{d}{k}\beta^{\binom{k+1}{2}}\gamma^{\binom{d-k}{2}}z^{k+1} \\
&= \sum_{k=0}^{d}\binom{d}{k}\beta^{\binom{k}{2}}\gamma^{\binom{d-k}{2}}\gamma^{d-k}z^k + z\sum_{k=0}^{d}\binom{d}{k}\beta^{\binom{k}{2}}\gamma^{\binom{d-k}{2}}\beta^k z^k \\
&= \gamma^d P_d(z/\gamma) + zP_d(\beta z).
\end{aligned}$$

□

Proof of Proposition 6.6.3. If $\beta = 0$ then P_d is linear 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 $r_i/r_{i+1} < \beta\gamma$ for all $1 \le i \le d-1$. 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 $d \ge 2$. By Lemma 6.6.4, 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 inductive hypothesis for P_d.

Now, we claim that for each $i = 2, \dots, d$, 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 $i = 2, \dots, d$.

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 there exists 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 have proved 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 $i = 1, \dots, d+1$. To finish the induction, we need to show that $s_i/s_{i+1} < \beta\gamma$ for all $i = 1, \dots, d$, which follows by $s_i/s_{i+1} < (\gamma r_i)/(r_i/\beta) = \beta\gamma$. □

6.6.2 Complement of Product of Shifted Half-Planes

In this subsection, we prove Lemma 6.6.2, which roughly describes the region $\left(-\overline{\mathbb{H}}_\epsilon^2\right)^\complement$, where recall $\mathbb{H}_\epsilon$ is the shifted left half-plane.

Proof of Lemma 6.6.2. It was shown in [Guo+21] that

$$\Gamma = \left(-\overline{\mathbb{H}}_\epsilon^2\right)^\complement = \left\{\rho e^{i\theta} : \rho < \frac{2\epsilon^2}{1-\cos\theta},\ 0 < \theta < 2\pi\right\}.$$

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

$$\begin{aligned}
& \rho < \frac{2\epsilon^2}{1-\cos\theta} \\
\iff \quad & \rho(1-\cos\theta) < 2\epsilon^2 \\
\iff \quad & \rho < x + 2\epsilon^2 \\
\iff \quad & x^2 + y^2 < (x+2\epsilon^2)^2 \\
\iff \quad & y^2 < 4\epsilon^2(x+\epsilon^2).
\end{aligned}$$

Therefore, we see that

$$\Gamma = \left\{x + iy : y^2 < 4\epsilon^2(x+\epsilon^2)\right\},$$

which clearly contains $\mathbb{R}_{\geq 0}$.

Furthermore, for $\lambda \in \mathbb{R}_+$ we have

$$\begin{aligned}
\operatorname{dist}(\lambda, \partial\Gamma) &= \inf_{z\in\partial\Gamma} |z-\lambda| \\
&= \inf_{(x,y)\in\mathbb{R}^2:\, y^2 = 4\epsilon^2(x+\epsilon^2)} \sqrt{(x-\lambda)^2 + y^2} \\
&= \inf_{x\in[-\epsilon^2,\infty)} \sqrt{(x-\lambda)^2 + 4\epsilon^2(x+\epsilon^2)} \\
&= \begin{cases} \lambda + \epsilon^2, & \lambda \in (0, \epsilon^2); \\ 2\epsilon\sqrt{\lambda}, & \lambda \in [\epsilon^2, \infty). \end{cases}
\end{aligned}$$

This establishes the lemma. □

Chapter 7

The Correlation Decay Method: Bridging Spatial and Temporal Mixing

This chapter is about spin systems in statistical physics. Here, one natural and extremely well-studied Markov chain, one which is described by the down-up walk, is called the *Glauber dynamics* or *Gibbs sampler*. One can imagine it as a natural way to describe how a (physical) system of interacting particles *evolves over time*. For such system, there is an intimate connection between *spatial mixing*, in the sense that random assignments to different sites decorrelate quickly in distance, and *temporal mixing*, in the sense that the successive random configurations generated by Glauber dynamics decorrelate quickly from the starting configuration as time evolves.

This is a remarkable phenomenon that was previously observed for spin systems in *lattices* such as $\mathbb{Z}^d$ [Wei04; Dye+04b; CP21a]. Unfortunately, those analyses are restricted to graphs satisfying certain "growth" conditions. In this chapter, we will show that correlation decay implies spectral independence, and hence, rapid mixing of the Glauber dynamics using our local-to-global theorems. Our analysis will extend this connection to all bounded-degree graphs, and gives another general purpose tool for studying the mixing time of the Glauber dynamics.

The connection between spatial mixing and algorithms actually goes much deeper. In the context of two-spin systems like the hardcore model, Weitz pioneered an alternative method for approximate counting and sampling which directly uses spatial mixing [Wei06]. This is known as the *correlation decay method*, and one of its striking features is that it is a completely deterministic method for estimating partition functions. There is now a long line of work directly using correlation decay as an algorithm for approximating the partition function of many important statistical physics models [Wei06; Bay+07; GK07; MS07; BG08; LLY12; GK12; LLY13; LY13; SSY13; Res+13; LLL14; LLZ14a; LWZ14; SST14; Sin+15; LL15b; LL15a; LYZ16; SYZ19; LSS20] (see also [Wei04; Sri14]).

Furthermore, it turns out that in a certain sense, the presence or absence of spatial mixing actually *determines* the complexity of approximate counting and sampling for many models[1], most notably the hardcore gas model. More specifically, the problem of approximating the partition function of the hardcore gas model on all bounded-degree graphs admits efficient algorithms if and only if the parameters of the model lie in the so-called *tree uniqueness regime*, i.e. the regime in which spatial mixing holds on the infinite Δ-regular tree. Weitz's algorithm furnishes one side of the implication, while the hardness results of [Sly10], and later [SS14; Gal+14; GŠV15; GŠV16], furnish the converse. This bidirectional connection holds for numerous other antiferromagnetic models such as the antiferromagnetic Ising model.

[1] For the *ferromagnetic* Ising model, spatial mixing is sufficient but *not* necessary for the existence of FPRAS for approximating the partition function. This was demonstrated in the seminal work of [JS93], which gives a Markov chain based FPRAS for approximating the partition function of the ferromagnetic Ising model at any temperature, despite the lack of spatial mixing at sufficiently low temperatures. Generally speaking, the absence of spatial mixing only implies hardness of approximation for *antiferromagnetic* models, since at low temperatures, there is typically some NP-hard combinatorial optimization problem lurking in the background which is hard to approximate.

7.1 The Hardcore Gas Model

Throughout this chapter, we focus on the *hardcore model* for simplicity, although many of our results extend to other spin systems in statistical physics such as the antiferromagnetic and ferromagnetic Ising models, and the monomer-dimer model; see [Fen+21; Che+21d] for a similar analysis for proper colorings on triangle-free graphs. Let us first recall the definition of the hardcore model. Fix a graph $G = (V, E)$ and recall an *independent set* is a subset of vertices $I \subseteq V$ such that no pair of vertices in I are connected by an edge. Fix a parameter $\lambda \geq 0$ (often called the *fugacity*), and define the Gibbs distribution $\mu_{G,\lambda}$ of the hardcore model over G by

$$\mu_{G,\lambda}(I) \propto \lambda^{|I|}, \quad \forall \text{ independent sets } I \subseteq V.$$

Equivalently, we may view $\mu_{G,\lambda}$ as a probability distribution over *feasible configurations* in $\{0,1\}^V$, where a configuration $\sigma : V \to \{0,1\}$ is feasible if $\{v \in V : \sigma(v) = 1\} \subseteq V$ forms an independent set. For such feasible configurations, we have $\mu_{G,\lambda}(\sigma) \propto \lambda^{\#\{v \in V : \sigma(v) = 1\}}$. The associated (univariate) partition function is given by

$$\mathcal{Z}_G(\lambda) \stackrel{\text{def}}{=} \sum_{I \subseteq V \text{ independent}} \lambda^{|I|}.$$

This is also sometimes referred to as the *(univariate) independence polynomial* of G.

The following is our main algorithmic result.

Theorem 7.1.1 (Rapid Mixing for Tree-Unique Hardcore Model)**.** *Let $G = (V, E)$ be a n-vertex graph with maximum degree Δ, and assume $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Then the Glauber dynamics for sampling from $\mu_{G,\lambda}$ mixes in $O_{\Delta,\delta}(n \log n)$-steps if $\Delta \leq O(1)$ and $n^{O(1/\delta)}$-steps in general.*

In the bounded-degree regime, the dependence on the maximum degree Δ and the gap δ scales as $\Delta^{O(\Delta^2/\delta)}$.

Subsequent Works: Following [ALO21; CLV20; CLV21a], [Bla+22] improved the dependence on the maximum degree Δ, and [JPV22] reduced the dependence on Δ from exponential to polynomial for bounding the spectral gap. [Che+21b] then established an $\Omega_\delta(1/n)$ lower bound on the spectral gap, independent of the maximum degree. [Ana+21b] then proved that a slightly modified version of the Glauber dynamics called the *balanced Glauber dynamics* mixes in $O_\delta(n \log n)$-steps, independent of the maximum degree. Finally, [Che+21a; CE22] independently proved that the standard Glauber dynamics mixes in $O_\delta(n \log n)$-steps, independent of Δ. The dependence on the gap δ in these last four results is $\exp(O(1/\delta))$. It is an interesting question to see if this can be reduced, perhaps to $\mathsf{poly}(1/\delta)$ as was done in [Eft+16] for graphs with additional structure.

These results, as well as Theorem 7.1.1, all rely on the following theorem, which establishes spectral independence for the hardcore model in the tree uniqueness regime. Theorem 7.1.1 then follows immediately by combining this with Theorems 2.3.1 and 10.0.1.

Theorem 7.1.2 (Spectral Independence for Tree-Unique Hardcore Model)**.** *Let $G = (V, E)$ be a n-vertex graph with maximum degree Δ, and assume $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Then the Gibbs distribution $\mu_{G,\lambda}$ of the hardcore model on G with fugacity λ is $(\eta_0, \ldots, \eta_{n-2})$-spectrally independent with*

$$\eta_k \leq \min\left\{\frac{32}{\delta}, \frac{\lambda}{1+\lambda}(n-k-1)\right\}$$

for all $k = 0, \ldots, n-2$.

In Appendix E, we give evidence that this $O(1/\delta)$-upper bound on the spectral independence is tight for the class of all bounded-degree graphs; see Theorem E.1.1. Of course, for special classes of graphs (e.g. lattices), one can go beyond $\lambda_c(\Delta)$, which is based purely on the maximum degree; see e.g. [Res+13].

7.1.1 From Correlation Decay to Spectral Independence: A High-Level Overview

At a very high level, our strategy is to take advantage of *correlation decay* properties of the hardcore model when $\lambda \leq (1-\delta)\lambda_c(\Delta)$.

Definition 33 (Weak/Strong Spatial Mixing)**.** *Let (V, d) be a finite metric space (e.g. the shortest path metric on a graph with vertex set V), and let μ be a probability distribution over a discrete product space $\prod_{v\in V}\Sigma_v$ for nonempty finite sets $\{\Sigma_v : v \in V\}$. We say μ exhibits **weak spatial mixing** w.r.t d with rate $\mathfrak{R} : \mathbb{R}_{\geq 0} \to [0, 1]$ if for every $v \in V$, every $s \in \Sigma_v$, every $S \subseteq V$ with $v \notin S$, and every pair of feasible boundary conditions ξ, ξ' on S, we have*

$$\left|\Pr_{\sigma\sim\mu}\left[\sigma(v) = s \,\middle|\, \sigma|_S = \xi\right] - \Pr_{\sigma\sim\mu}\left[\sigma(v) = s \,\middle|\, \sigma|_S = \xi'\right]\right| \leq \mathfrak{R}(d(v, S)). \tag{7.1}$$

*We say μ exhibits **strong spatial mixing** w.r.t. d with rate $\mathfrak{R}$ if we can replace the upper bound of $\mathfrak{R}(d(v, S))$ in Eq. (7.1) with $\mathfrak{R}(d(v, \triangle(\xi, \xi')))$, where $\triangle(\xi, \xi') \subseteq S$ is the subset of points on which ξ, ξ' differ. Equivalently, we say μ exhibits strong spatial mixing w.r.t. d with rate $\mathfrak{R}$ if weak spatial mixing w.r.t. d with rate $\mathfrak{R}$ holds for every conditional distribution of μ.*

Remark 34. Throughout, we focus on the case where d is the unweighted shortest path metric on a graph $G = (V, E)$ with vertex set V. We note there are many works which use different metrics [LLY13; Res+13], even in the settings where there is some underlying graph. We also note there are alternative forms of correlation decay based on computation trees that have been successfully used to obtain approximation algorithms [Bay+07; GK07].

We are typically interested in the case when $\mathfrak{R}$ is a function *decaying exponentially fast to* 0. For instance, it was shown in [Wei06] that when $\lambda \leq (1-\delta)\cdot\lambda_c(\Delta)$, the Gibbs distribution $\mu_{G,\lambda}$ of the hardcore model on a graph $G = (V, E)$ with maximum degree Δ satisfies strong spatial mixing with rate $\mathfrak{R}(t) \leq C \cdot (1 - O(\delta))^t$ for some constant C depending only on λ, Δ[2].

Weak spatial mixing with an exponentially fast decay rate $0 < \alpha < 1$ already says that the rows and columns of Ψ_μ decay exponentially fast in graph distance away from the diagonal. This intuitively should constrain the eigenvalues of Ψ_μ. Indeed, for graphs such as the integer lattice $G = \mathbb{Z}^d$, we have that the following holds for every vertex $r \in V$:

$$\sum_{v\neq r}|\Psi_\mu(r \to v)| \leq \sum_{k=1}^{\infty}|\mathbb{B}(r, k)|\cdot\mathfrak{R}(k) \lesssim_d \sum_{k=1}^{\infty} k^d \cdot \alpha^k \leq O_{d,\alpha}(1). \tag{7.2}$$

Here, recall $\mathbb{B}(r, k) = \{v \in V : \mathrm{dist}_G(r, v) \leq k\}$ denotes the (closed) ball of radius-k around r in the graph, and $0 < \alpha < 1$ is the rate of exponential decay. Since this holds for every $r \in V$, this would imply $O_d(1)$-spectral independence. To obtain bounds on the spectral independence parameter for all conditional distributions, one would then appeal to strong spatial mixing instead of weak spatial mixing by itself. Note that for spectral independence, we only need to understand the total sum of correlations between just pairs of vertices. This is in contrast to strong spatial mixing results, where one has to analyze the combined influence of *any* subset of vertices on another given vertex.

The crucial aspect of $\mathbb{Z}^d$ that we used in the above crude analysis was that the balls around each vertex only grow polynomially fast in radius, and so the exponentially fast decay of correlations completely overpowers this growth. However, most graphs, e.g. expander graphs, do not have this *subexponential growth* property, and so this analysis completely breaks down. Indeed, we will see that the rate of exponential decay α will be $1 - \Theta(\delta)$ when $\lambda < (1-\delta)\lambda_c(\Delta)$, which for general graphs is much slower than the rate at which the balls grow in volume w.r.t. distance. So we need a new approach.

It turns out, the way to strengthen this analysis is to *open up the proof of spatial mixing*, rather than treat it as a blackbox. More specifically, we implement the following two steps.

1. **Reduction to Trees:** Weitz establishes strong spatial mixing on graphs G of maximum degree Δ when $\lambda < \lambda_c(\Delta)$ by first reducing the problem on G to the same problem but on an associated *tree of self-avoiding walks*. This is a finite but exponentially large tree with the same maximum degree, and is motivated by the intuition that the *infinite Δ-regular tree* (also known as the *Bethe lattice with degree Δ*) should be the "worst case" out of all graphs of maximum degree Δ.

   We will do the same, by showing that for every vertex in the graph, the total absolute influence of that vertex is at most the total absolute influence of the root of the corresponding self-avoiding walk tree. This reduces the problem to showing that for every rooted tree, the total influence of the root is upper bounded by a constant independent of the size of the tree.

[2]When $\lambda = \lambda_c(\Delta)$, it is known that strong spatial mixing still holds with a rate function $\mathfrak{R}$ which decays to 0. However, the rate of decay is no longer exponentially fast. See [Wei06] for more details.

2. **Influence Bounds on Trees:** One additional motivation for reducing the analysis is to trees is that there is a recursive method for *computing* the marginal probabilities of vertices in a tree. Whether or not these *tree recursions* "contract" in a certain sense determines whether or not spatial mixing holds. The contractive properties of these tree recursions are precisely what we use to bound the total influence of the root node.

   Here, we leverage *potential functions* which help "amortize" the decay of correlations. This has become a standard tool in proving correlation decay and analyzing algorithms based on correlation decay. See [LLY12; LLY13; Res+13; SSY13; SST14; Sin+15] and references therein for instantiations of this powerful method. We also refer interested readers to [Sri14] for further discussion of the potential method.

One major advantage of using correlation decay to establish rapid mixing using spectral independence, as opposed to using the correlation decay algorithm, is that oftentimes, the type of correlation decay is much weaker, at least for multi-state spin systems. There, one typically needs a very strong type of correlation decay on an associated *computation tree* (see e.g. [GK07; GK12; LY13]), which in the case of two-state spin systems, is simply Weitz's self-avoiding walk tree. On the other hand, as [Fen+21; Che+21d] demonstrate for proper colorings on triangle-free graphs, weaker correlation decay results suffice to establish spectral independence and hence, rapid mixing of the Glauber dynamics.

We now proceed to describe the tree recursion and discuss the significance of the threshold $\lambda_c(\Delta)$.

7.1.2 The Tree Recursions and the Tree Uniqueness Threshold

Fix a tree T rooted at some vertex r. For a vertex v in T, let $\ell(v)$ denote its distance from the root r. We will sometimes refer to it as the "level" which contains v. For a level ℓ, let $L_r(\ell) = \{v \in T : \ell(v) = \ell\}$. For a vertex $u \in T$, we will write T_u for the subtree of T rooted at u.

A key tool we will need to analyze the hardcore model on trees is given by the *tree recurrence* (or *tree recursion*). To describe the tree recurrence, we consider a convenient change of variables w.r.t. the marginal probabilities, following [Wei06]. Fix a tree T arbitrarily rooted at some vertex $r \in T$. Recall that $\mu_{T,r}$ denotes the marginal distribution on $\{0,1\}$ for the root r, i.e. $\mu_{T,r}(1)$ is the probability that r is in a random independent set sampled from the Gibbs distribution of the hardcore model on T. We define the *marginal ratio* as

$$R_{T,r} \stackrel{\mathsf{def}}{=} \frac{\mu_{T,r}(1)}{\mu_{T,r}(0)} = \frac{\mu_{T,r}(1)}{1-\mu_{T,r}(1)}. \tag{7.3}$$

We drop the subscripts T, r when the tree and its root are clear from context.

Note that since $x \mapsto \frac{x}{1-x}$ is monotone on $[0,1]$, it is a bijection between $[0,1]$ and $[0,+\infty]$, and so the value of $R_{T,r}$ also uniquely determines the marginal probability $\mu_{T,r}(1)$. In particular, one could also have written the tree recursion purely in terms of the marginals $\Pr_{T,r}[r]$; however, this change of variables turns out to be significantly more convenient to work with.

With this notation in hand, we may write the *multivariate tree recurrence* for the hardcore model as

$$R_{T,r} = F(R_{T_u,u} : u \in L_r(1)) \qquad \text{where} \qquad F(R_1,\dots,R_d) \stackrel{\mathsf{def}}{=} \lambda \prod_{i=1}^{d} \frac{1}{R_i+1}. \tag{7.4}$$

Note that this tree recurrence naturally leads to a simple polynomial-time dynamic programming algorithm for exactly computing $\mathcal{Z}_G(\lambda)$ on any tree.

Fact 7.1.3 (Tree Recursion Derivatives)**.** *For all $i = 1,\dots,d$ and all $\boldsymbol{R} = (R_1,\dots,R_d) \in \mathbb{R}^d_{\geq 0}$,*

$$\partial_{R_i} F(\boldsymbol{R}) = -\lambda \cdot \frac{1}{(R_i+1)^2} \cdot \prod_{j \neq i} \frac{1}{R_j+1} = -\frac{F(\boldsymbol{R})}{R_i+1} \leq 0. \tag{7.5}$$

In particular, F is monotone decreasing in each coordinate.

In the case of a depth-ℓ *complete* d-ary tree rooted at r with no boundary conditions, all marginals of vertices at a fixed level of the tree are equal by symmetry. Thus, the only relevant

parameter here is the depth, and the tree recurrence simplifies to a *univariate tree recurrence* given by

$$f_d(R) \stackrel{\text{def}}{=} \lambda \left(\frac{1}{R+1}\right)^d \tag{7.6}$$

It turns out that $f_d(\cdot)$ has a *unique fixed point* which we call $\hat{R}_d$, i.e. $\hat{R}_d = \lambda \left(\frac{1}{\hat{R}_d+1}\right)^d$. One can see this as follows. Note that since f_d is strictly decreasing, $g_d(R) \stackrel{\text{def}}{=} f_d(R) - R$ is strictly decreasing, and satisfies $g_d(0) = f_d(r) = \lambda$ while $\lim_{R\to+\infty} g_d(R) = -\infty$. By the Intermediate Value Theorem, there exists $\hat{R}_d$ such that $g_d(\hat{R}_d) = 0$. Note this $\hat{R}_d$ is unique by strict monotonicity of g_d. Translating back into f_d, we have $\hat{R}_d$ is the unique fixed point of f_d, i.e. $f_d(\hat{R}_d) = \hat{R}_d$. It turns out the properties of this fixed point $\hat{R}_d$ and the behavior of f_d at $\hat{R}_d$ govern the correlation decay properties of the hardcore model, as we will see in the following subsection.

The way the threshold $\lambda_c(\Delta)$ is derived is by analyzing when $\left|f'_{\Delta-1}(\hat{R}_{\Delta-1})\right|$ is less than 1. It turns out the gap between $\left|f'_{\Delta-1}(\hat{R}_{\Delta-1})\right|$ and 1 governs the rate α in the definition of weak spatial mixing. [LLY13] quantified this in the following definition.

Definition 34 (Up-to-Δ Uniqueness; [LLY13])**.** *We say the hardcore model with parameter λ is* ***up-to-Δ unique*** *if $\left|f'_d\left(\hat{R}_d\right)\right| < 1$ for every $1 \le d < \Delta$, where $\hat{R}_d$ denotes the unique fixed point of f_d. Furthermore, we say λ is* ***up-to-Δ unique with gap*** *$0 < \delta < 1$ if $\left|f'_d\left(\hat{R}_d\right)\right| \le 1 - \delta$ for every $1 \le d < \Delta$.*

It is not hard to show that (for the hardcore model) up-to-Δ uniqueness with gap $0 < \delta < 1$ is equivalent to $\lambda \le (1 - \Theta(\delta)) \cdot \lambda_c(\Delta)$.

Lemma 7.1.4 (Gapped Up-to-Δ Uniqueness in the Hardcore Model)**.** *λ is up-to-Δ unique with gap $0 < \delta < 1$ if and only if $\lambda < (1 - \Theta(\delta))\lambda_c(\Delta)$.*

We supply a proof in Section 7.4. Hence, throughout the paper whenever one encounters the phrase "up-to-Δ unique with gap $0 < \delta < 1$", one may safely assume $\lambda \le (1 - \Theta(\delta))\lambda_c(\Delta)$.

We conclude this subsection with the following marginal bounds, which can be shown using the tree recursions.

Fact 7.1.5 (Hardcore Model Marginal Bounds)**.** *For every graph $G = (V, E)$, every vertex $r \in V$, and every boundary condition $\sigma_\Lambda : \Lambda \to \{0, 1\}$ where $\Lambda \subseteq V \setminus \{r\}$, we have the upper and lower bounds*

$$\frac{\lambda}{1+\lambda} \cdot \left(\frac{1}{1+\lambda}\right)^\Delta \le \mu^{\sigma_\Lambda}_{G,r}(1) \le \frac{\lambda}{1+\lambda}.$$

Proof. For convenience, we prove the case when there is no conditioning σ_Λ; the general case can be obtained just by deleting all vertices $v \in \Lambda$ s.t. $\sigma_\Lambda(v) = 0$ and deleting all closed neighborhoods $N[v]$ for all $v \in \Lambda$ s.t. $\sigma_\Lambda(v) = 1$.

Since $\sigma(r) = 1$ forces $\sigma(v) = 0$ for all $v \sim r$, we have

$$\begin{aligned}
\mu_{G,r}(1) &= \Pr_{\sigma\sim\mu}[\sigma(r) = 1, \sigma(v) = 0, \forall v \sim r] \\
&= \Pr_{\sigma\sim\mu}[\sigma(r) = 1 \mid \sigma(v) = 0, \forall v \sim r] \cdot \Pr_{\sigma\sim\mu}[\sigma(v) = 0, \forall v \sim r] \\
&= \frac{\lambda}{1+\lambda} \cdot \underbrace{\Pr_{\sigma\sim\mu}[\sigma(v) = 0, \forall v \sim r]}_{(*)}.
\end{aligned}$$

The upper bound just follows by bounding $(*)$ by 1. For the lower bound, if we order the vertices

of $N(r)$ arbitrarily as $v_1, \ldots, v_d$, then

$$\begin{aligned}(*) &= \prod_{j=1}^{d} \Pr_{\sigma\sim\mu}[\sigma(v_j) = 0 \mid \sigma(v_i) = 0, \forall 1 \le i \le j] \\ &= \prod_{j=1}^{d} \left(1 - \underbrace{\Pr_{\sigma\sim\mu}[\sigma(v_j) = 1 \mid \sigma(v_i) = 0, \forall 1 \le i \le j]}_{\le \frac{\lambda}{1+\lambda} \text{ using the upper bound we just proved}}\right) \\ &\ge \left(\frac{1}{1+\lambda}\right)^{d}.\end{aligned}$$

□

Remark 35. Note that when $\lambda < O(1/\Delta)$, e.g. when λ is up-to-Δ unique, the upper and lower bounds are within universal constant factors of each other since $\left(\frac{1}{1+\lambda}\right)^{\Delta} \ge \Omega(1)$ in this case.

Later on, we state a generalization of this in Fact A.6.1, whose proof is provided in Appendix A.6.

7.1.3 Related Prior Works on the Hardcore Model

We conclude this section with a discussion of the extensive prior work on this model.

Deterministic Approximate Counting Algorithms The question of building deterministic approximation algorithms for estimating $\mathcal{Z}_G(\lambda)$ on bounded degree graphs has been settled. [Wei06] proved that there is an FPTAS on graphs of maximum degree $\le \Delta$ whenever $\lambda < \lambda_c(\Delta)$. [HSV18] extends this result to estimating the multivariate independence polynomial, and [PR19] proves the existence of a zero-free region around $[0, \lambda_c(\Delta))$, which makes Barvinok's polynomial interpolation technique [Bar16a] applicable to estimating $\mathcal{Z}_G(\lambda)$; see [PR17]. We note that the Bethe approximation for estimating $\mathcal{Z}_G(\lambda)$ has also been studied in [Cha+11]. One important caveat is that the running time of Weitz's algorithm scales as $(n/\epsilon)^{C \log \Delta}$ where the approximation factor is $1 \pm \epsilon$ and the constant C depends polynomially on the gap δ (recall, $\lambda < (1-\delta)\lambda_c(\Delta)$). Barvinok's method also has a similar scaling, with an exponential dependence on Δ. This unfortunately appears to be a rather general feature of deterministic FPTAS for this and related problems.

Previous Mixing Results For studying the mixing time of the Glauber dynamics in the uniqueness regime, there has been a long line of work starting with [LV97; LV99; DG00; Vig01]. For general graphs, the state-of-the-art was given by [Vig01], which showed the Glauber dynamics mixes in $O(n \log n)$ steps when $\lambda < \frac{2}{\Delta-2}$. A more recent result of [Eft+16] shows that for any $0 < \delta < 1$, there is a $\Delta_0(\delta)$ such that for any $\Delta \ge \Delta_0(\delta)$ and $\lambda = (1-\delta)\lambda_c(\Delta)$, the Glauber dynamics mixes in $O(n \log n)$ steps for graphs with maximum degree Δ and girth ≥ 7.

More is known for restricted families of graphs. The hardcore distribution over independent sets of the line graph $L(G)$ of a graph G is equivalent to the monomer-dimer distribution over matchings of G itself. Here, the Glauber dynamics is known to mix in time $O(\lambda^3 mn^2 \log n)$ time [JS89]. [Bay+07; Sin+15] give deterministic FPTAS for this problem in the full range of λ on bounded-degree graphs. It is proved in [Wei06] that the Glauber dynamics mixes in $O(n^2)$ steps for any $\lambda < \lambda_c(\Delta)$ when the input graph has maximum degree $\le \Delta$ and satisfies subexponential growth. This encompasses, for instance, the integer lattices $\mathbb{Z}^d$. On such lattices, there is an equivalence between strong spatial mixing and optimal mixing of the Glauber dynamics [Dye+04b; Wei04]. [MSW03; Wei04; MSW07] obtained rapid mixing for trees, and [Hay06] obtained rapid mixing for planar graphs. For graphs of large girth, [HV05] studies the mixing time of the Glauber dynamics and [BG08] studies deterministic correlation decay algorithms. In the case of the square grid $\mathbb{Z}^2$, we have a more precise understanding [VVY13; Res+13; Bla+13; Bla+19]. [MS08b; MS13; SSY13; Sin+15] study the case of $G(n, d/n)$ random graphs, or more generally graphs with bounded connective constant.

Hardness Results On the hardness side, exact computation of $\mathcal{Z}_G(\lambda)$ is known to be #P-Hard [Val79; Vad95; Gre00], even for very restricted families of graphs [Vad02]. For hardness of approximation, [LV97] showed there exists a constant $c > 0$ such that there is no FPRAS for estimating $\mathcal{Z}_G(1)$ when $\lambda > c/\Delta$ unless NP = RP. For the case of evaluating $\mathcal{Z}_G(1)$, this was improved in [DFJ02], which showed that there is no FPRAS for estimating $\mathcal{Z}_G(1)$ on graphs with maximum degree exceeding 25 unless NP = RP. [DFJ02] further showed that the Glauber dynamics has exponential mixing time for $\Delta \geq 6$. [MWW07] provided further evidence the Markov chain techniques are likely to fail for sampling from the Gibbs distribution when $\lambda > \lambda_c(\Delta)$. These results were dramatically improved in the work of [Sly10] (and further refined by follow-up works [SS14; Gal+14; GŠV15; GŠV16]), which showed that unless NP = RP, there is no FPRAS for estimating $\mathcal{Z}_G(\lambda)$ on graphs of maximum degree $\leq \Delta$ when $\lambda > \lambda_c(\Delta)$.

Beyond the Hardcore Model There are also many works extending results for the hardcore model to general antiferromagnetic two-state spin systems. For antiferromagnetic Ising models in the uniqueness regime, there are FPTAS based on both correlation decay [SST14] (see also [ZLB11] for a special case) and polynomial interpolation [LSS19; Liu19; SS19]. Combined with algorithms for the hardcore model, these give FPTAS for all antiferromagnetic two-state spin systems via reductions described for instance in [SST14]. In a more direct fashion, [LLY12; LLY13] give deterministic correlation decay algorithms for all antiferromagnetic two-state spin systems up to the uniqueness threshold. [GJP03] analyze the corresponding Glauber dynamics via the path coupling method, but do not obtain rapid mixing in entire uniqueness regime. [GJP03] provide hardness of approximation for a certain range of edge activities β, γ. [SS14; Gal+14; GŠV15; GŠV16] extend these hardness results to all antiferromagnetic two-state spin systems in the nonuniqueness regime.

7.2 Weitz's Self-Avoiding Walk Tree

Weitz's self-avoiding walk tree is best defined in the broader context of *two-state spin systems*, of which the hardcore model and the Ising model are special cases. Everything we say in this section applies to all such two-state spin systems. Let $G = (V, E)$ be a graph, and for fixed parameters $0 \leq \beta \leq \gamma < \infty$ and external fields $\boldsymbol{\lambda} = \{\lambda_v : v \in V\} \in \mathbb{R}^V_{\geq 0}$, we define the Gibbs distribution $\mu = \mu_{G,\beta,\gamma}$ of the two-state spin system on G with parameters β, γ and external fields $\boldsymbol{\lambda}$ by

$$\mu(\sigma) \propto \beta^{m_0(\sigma)}\gamma^{m_1(\sigma)} \prod_{v \in V : \sigma(v) = 1} \lambda_v, \quad \forall \sigma : V \to \{0,1\}$$

where $m_s(\sigma) = \#\{uv \in E : \sigma(u) = \sigma(v) = s\}$ for all $s \in \{0,1\}$. We typically be interested in the case where all external fields are equal to some fixed $\lambda \in \mathbb{R}_{\geq 0}$, although nearly everything we say will also extend to the case of heterogeneous $\boldsymbol{\lambda}$.

The case $\beta = 0, \gamma = 1$ recovers the hardcore model, while the case $\beta = \gamma$ recovers the Ising model. The reader is welcome to take $\beta = 0, \gamma = 1$ in this section, as nothing will be lost; we write things in full generality here since it will be useful for later analyses. When $\beta\gamma > 1$, the system is *ferromagnetic*, i.e. the distribution puts more probability mass on configurations with more monochromatic edges. On the other hand, when $\beta\gamma < 1$, the system is *antiferromagnetic*, i.e. the system prefers more disagreements in the configuration. The associated *multivariate partition function* with variables $\boldsymbol{\lambda} = \{\lambda_v : v \in V\}$ is then given by

$$\mathcal{Z}_G(\boldsymbol{\lambda}) = \sum_{\sigma : V \to \{0,1\}} \beta^{m_0(\sigma)}\gamma^{m_1(\sigma)} \prod_{v \in V : \sigma(v) = 1} \lambda_v.$$

If $\Lambda \subseteq V$ is a subset of vertices and $\sigma_\Lambda : \Lambda \to \{0,1\}$ is a boundary condition, then the associated conditional Gibbs distribution over $\{0,1\}^{V \setminus \Lambda}$ is given by

$$\mu_G^{\sigma_\Lambda}(\sigma) \propto \beta^{m_0(\sigma|\sigma_\Lambda)}\gamma^{m_1(\sigma|\sigma_\Lambda)} \prod_{v \in V \setminus \Lambda : \sigma(v) = 1} \lambda_v, \quad \forall \sigma : V \setminus \Lambda \to \{0,1\} \tag{7.7}$$

where for $s \in \{0,1\}$, $m_s(\cdot \mid \sigma_\Lambda)$ denotes the number of edges such that both endpoints receive the spin s and at least one of them is in $V \setminus \Lambda$. The associated multivariate partition function conditional on σ_Λ is then defined as

$$\mathcal{Z}_G^{\sigma_\Lambda}(\boldsymbol{\lambda}) = \sum_{\sigma : V \setminus \Lambda \to \{0,1\}} \beta^{m_0(\sigma|\sigma_\Lambda)}\gamma^{m_1(\sigma|\sigma_\Lambda)} \prod_{v \in V \setminus \Lambda : \sigma(v) = 1} \lambda_v. \tag{7.8}$$

Of course, one can also view $\mu_G^{\sigma_\Lambda}$ as a distribution over $\{0,1\}^V$ by concatenating $\sigma_\Lambda : \Lambda \to \{0,1\}$ to any configuration $\sigma : V \setminus \Lambda \to \{0,1\}$. The multivariate partition function would then gain a multiplicative factor of $\beta^{m_0(\sigma_\Lambda)}\gamma^{m_1(\sigma_\Lambda)} \prod_{v\in\Lambda:\sigma(v)=1} \lambda_v$. We will primarily stick with Eq. (7.8).

Definition 35 (Self-Avoiding Walk Tree; [SS05], [Wei06])**.** *Fix a graph $G = (V, E)$ and a vertex $r \in V$. A **self-avoiding walk** of length ℓ in G beginning at r is a sequence of vertices $r = v_0, \dots, v_\ell$ such that $v_0, \dots, v_\ell$ are all distinct, and $v_i \sim v_{i-1}$ for each $i = 1, \dots, \ell$. The **self-avoiding walk tree** $\mathrm{T}_{\mathsf{SAW}}(G, r)$ is a rooted tree whose vertices correspond to walks $v_0, \dots, v_\ell$ such that either*

1. *$v_0, \dots, v_\ell$ itself is a self-avoiding walk, or*
2. *$v_0, \dots, v_{\ell-1}$ is a self-avoiding walk, and $v_\ell = v_i$ for some $i < \ell - 2$, i.e. v_ℓ closes a cycle; note that $v_\ell = v_{\ell-2}$ (backtracking) and $v_\ell = v_{\ell-1}$ (staying) are both prohibited.*

Two such walks are adjacent in $\mathrm{T}_{\mathsf{SAW}}(G, r)$ if and only if one extends the other.

Next, we specify boundary conditions. Specifically, for each vertex $v \in G$, we first order its neighbors arbitrarily (e.g. one could just totally order the vertices of G). Now consider a walk $v_0, \dots, v_\ell$ such that v_ℓ closes a cycle. Let $i < \ell - 2$ be such that $v_\ell = v_i$. We assign the vertex in T corresponding to the walk $v_0, \dots, v_\ell$ a spin value of

1. *0 if the neighbor v_{i+1} of v_i is larger than the neighbor $v_{\ell-1}$, and*
2. *1 if the neighbor v_{i+1} of v_i is smaller than the neighbor $v_{\ell-1}$.*

*These are the **structural boundary conditions**. For $\Lambda \subseteq V \setminus \{r\}$ and a valid partial configuration $\sigma_\Lambda \in \{0,1\}^\Lambda$, we define the self-avoiding walk tree $\mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ with conditioning σ_Λ by assigning the spin $\sigma_\Lambda(v)$ to every copy $\hat{v}$ of v in $\mathrm{T}_{\mathsf{SAW}}(G, r)$ and removing all descendants of $\hat{v}$, for each $v \in \Lambda$. These are **inherited boundary conditions**. Let $\mathcal{C}(v)$ denote the set of copies of v in $\mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$, i.e. vertices in T corresponding to walks which end at v. Let $\mathcal{F}(v) \subseteq \mathcal{C}(v)$ denote the set of **free** (i.e. unpinned) copies of v in $\mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$. We write $\hat{r}$ for the copy of r, i.e. the root of $\mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$.*

Finally, if $\mu_G^{\sigma_\Lambda}$ denotes the (conditional) Gibbs distribution over $\{0,1\}^{V\setminus\Lambda}$ of the two-spin system on G with external fields λ_v and boundary conditions $\sigma_\Lambda \in \{0,1\}^\Lambda$ on $\Lambda \subseteq V$ (see Eq. (7.7)), then for every $r \in V \setminus \Lambda$, we will write $\mu_T^{\sigma_\Lambda}$ for the (conditional) Gibbs distribution of the two-spin systems on $T = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ with the same parameters along with external fields $\lambda_{\hat{v}} = \lambda_v$ for all copies $\hat{v} \in \mathcal{C}(v)$ and every $v \in V$. Similarly, if $\mathcal{Z}_G^{\sigma_\Lambda}(\boldsymbol{\lambda})$ denotes the (conditional) multivariate partition function for μ_G (see Eq. (7.8)), then we will write $\mathcal{Z}_T^{\sigma_\Lambda}(\boldsymbol{\lambda})$ for the multivariate partition function for μ_T with the same variables $\boldsymbol{\lambda}$, i.e. $\lambda_{\hat{v}} = \lambda_v$ for all copies $\hat{v} \in \mathcal{C}(v)$ and every $v \in V$.

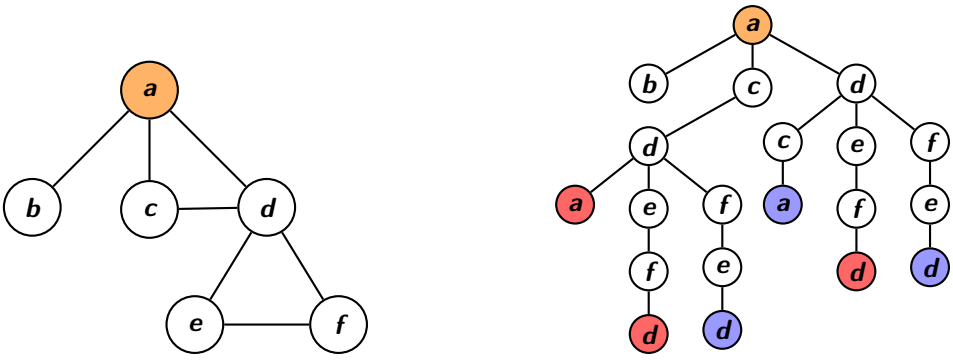
Note that T is a finite tree since any vertex in a self-avoiding walk can be visited at most once. Furthermore, pinned vertices in T come in two types. The first comes from τ_{SAW}, which arise from the cycle structure of G. In other words, pinnings of this type are "structural". The second comes from σ_{SAW}, which arise from fixed vertices in G. Pinnings of this second type are simply "copied" assignments. For convenience, whenever we consider a self-avoiding walk tree, we will implicitly assume that τ_{SAW} is part of any assignment, without writing it explicitly. In the case of the hardcore model, this is equivalent to simple throwing away all fixed vertices, and neighbors of vertices fixed to "in" (i.e. 0).

Fact 7.2.1. *Let $G = (V, E)$ be a graph, fix $r \in V$, and let $\sigma_\Lambda \in \{0,1\}^\Lambda$ be a boundary condition on $\Lambda \subseteq V \setminus \{r\}$. Let $T = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$.*

1. *For every $v \in V \setminus \Lambda$ and every copy $\hat{v} \in \mathcal{C}(v)$, either $\hat{v}$ is a leaf, or $\deg_T(\hat{v}) = \deg_G(v)$.*
2. *The maximum degree of T equals the maximum degree of $G[V \setminus \Lambda]$.*
3. *For a vertex $u \in G$, we have $\ell_T(u) = \mathrm{dist}_G(r, u)$.*

Theorem 7.2.2. *Let $G = (V, E)$ be a graph, $r \in V$ be a vertex, and $\sigma_\Lambda : \Lambda \to \{0,1\}$ be a boundary condition on $\Lambda \subseteq V \setminus \{r\}$. Let $T = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ be the self-avoiding walk tree of G rooted at $\hat{r}$ with inherited boundary conditions σ_Λ. Then there exists a polynomial $P_{G,r}^{\sigma_\Lambda} = P_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$ independent of λ_r such that*

$$\mathcal{Z}_T^{\sigma_\Lambda} = \mathcal{Z}_G^{\sigma_\Lambda} \cdot P_{G,r}^{\sigma_\Lambda}.$$



(a) Base graph G (b) $T = \mathrm{T}_{\mathsf{SAW}}(G, a)$

Figure 7.1: Red and blue vertices in T indicate the boundary condition τ_{SAW}, with red representing "in" and blue representing "out". These boundary conditions are considered "structural" as they depend only upon the cycle structure of the base graph G. Here, for each vertex in G, we order its neighbors reverse lexicographically.

Remark 36. We remark that [Ben18] proved a univariate version of Theorem 7.2.2 for the hardcore model, and [LSS19] 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 $P^{\sigma_\Lambda}_{G,r}$ 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.

Corollary 7.2.3 (Marginal and Influence Preservation for Self-Avoiding Walk Trees)**.** *Let $G = (V, E)$ be a graph, $r \in V$ be a vertex, and $\sigma_\Lambda : \Lambda \to \{0, 1\}$ be a boundary condition on $\Lambda \subseteq V \setminus \{r\}$. Let $T = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ be the self-avoiding walk tree of G rooted at $\hat{r}$ with inherited boundary conditions σ_Λ. Then we have that*

$$\mu^{\sigma_\Lambda}_{G,r} \equiv \mu^{\sigma_\Lambda}_{T,\hat{r}} \qquad \text{(Marginal Preservation; [Wei06])}$$

$$\Psi^{\sigma_\Lambda}_G(r \to v) = \sum_{\hat{v} \in \mathcal{F}(v)} \Psi^{\sigma_\Lambda}_T(\hat{r} \to \hat{v}), \quad \forall v \in V \setminus \Lambda, v \neq r \qquad \text{(Influence Preservation)}$$

Remark 37. The proof of Theorem 7.2.2 can be adapted to give a purely combinatorial proof of influence preservation in Corollary 7.2.3. Like in the proof of [Wei06, 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.

Proof. By Theorem 7.2.2,

$$\log \mathcal{Z}^{\sigma_\Lambda}_T = \log \mathcal{Z}^{\sigma_\Lambda}_G + \log P^{\sigma_\Lambda}_{G,r}$$

where $P^{\sigma_\Lambda}_{G,r}$ is independent of λ_r. Hence, differentiating both sides w.r.t. λ_r yields

$$\mu^{\sigma_\Lambda}_{G,r}(1) = \partial_{\lambda_r} \log \mathcal{Z}^{\sigma_\Lambda}_T = \partial_{\lambda_r} \log \mathcal{Z}^{\sigma_\Lambda}_G = \mu^{\sigma_\Lambda}_{T,\hat{r}}(1).$$

which is the first claim. By differentiating both sides again but w.r.t. λ_v, we obtain that the *covariances* are preserved, i.e.

$$\mathrm{Cov}^{\sigma_\Lambda}_G(r, v) = \sum_{\hat{v} \in \mathcal{F}(v)} \mathrm{Cov}^{\sigma_\Lambda}_T(\hat{r}, \hat{v})$$

where recall

$$\mathrm{Cov}^{\sigma_\Lambda}_G(r, v) = \Pr^{\sigma_\Lambda}_G[\sigma(r) = \sigma(v) = 1] - \Pr^{\sigma_\Lambda}_G[\sigma(r) = 1] \cdot \Pr^{\sigma_\Lambda}_G[\sigma(v) = 1]$$

and $\mathrm{Cov}^{\sigma_\Lambda}_T$ is defined analogously but w.r.t. $\mu^{\sigma_\Lambda}_T$. Normalizing (i.e. dividing) both sides by

$$\mu^{\sigma_\Lambda}_{G,r}(1) \cdot \mu^{\sigma_\Lambda}_{G,r}(0) = \mu^{\sigma_\Lambda}_{T,\hat{r}}(1) \cdot \mu^{\sigma_\Lambda}_{T,\hat{r}}(0)$$

then yields the claim. □

Before we give the proof of Theorem 7.2.2, let us introduce a couple more pieces of notation which will be convenient. For every $v \in V \setminus \Lambda$ and $s \in \{0,1\}$, we shall write $v \leftarrow s$ to represent the set of configurations such that $\sigma(v) = s$ (i.e., $\{\sigma : V \setminus \Lambda \to \{0,1\} : \sigma(v) = s\}$) and let $\mathcal{Z}_G^{\sigma_\Lambda}(v \leftarrow s) = \mathcal{Z}_G^{\sigma_\Lambda}(\boldsymbol{\lambda}; v \leftarrow s)$ be sum of all terms in $\mathcal{Z}_G^{\sigma_\Lambda}(\boldsymbol{\lambda})$ corresponding to configurations with $v \leftarrow s$. For concreteness and intuition, note that $\mathcal{Z}_G^{\sigma_\Lambda}(v \leftarrow s)$ and $\mathcal{Z}_G^{\sigma_\Lambda, v \leftarrow s}(\boldsymbol{\lambda})$ are the same up to a multiplicative factor corresponding to the contribution of the assignment $v \leftarrow s$[3]. We further extend this notation and write $\mathcal{Z}_G^{\sigma_\Lambda}(U \leftarrow \sigma_U)$ for every $U \subseteq V \setminus \Lambda$ and $\sigma_U : U \to \{0,1\}$. We adopt the same notations for the self-avoiding walk tree as well.

Proof of Theorem 7.2.2. We inductively show that there exists a polynomial $P_{G,r}^{\sigma_\Lambda} = P_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$, independent of λ_r, such that

$$\mathcal{Z}_T^{\sigma_\Lambda}(r \leftarrow 1) = \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 1) \cdot P_{G,r}^{\sigma_\Lambda} \quad \text{and} \quad \mathcal{Z}_T^{\sigma_\Lambda}(r \leftarrow 0) = \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 0) \cdot P_{G,r}^{\sigma_\Lambda}. \tag{7.9}$$

The high-level proof idea of Eq. (7.9) is similar to the corresponding result in [Wei06, Theorem 3.1]. Let m be the number of edges with at least one endpoint in $V \setminus \Lambda$. We use induction on m. When $m = 0$ the statement is trivial since $T = G$. Assume that Eq. (7.9) holds for all graphs and all conditionings with fewer than m edges. Suppose that the root r has d neighbors $v_1, \dots, v_d$. Define G' to be the graph obtained by replacing the vertex r with d vertices $r_1, \dots, r_d$ and then connecting $\{r_i, v_i\}$ for $1 \le i \le d$.

Consider first the case where $(G \setminus \{r\}) \setminus \Lambda$ is still connected. For each i, let $G_i = G' - r_i$. Define the two-state spin system on G_i with the same parameters $(\beta, \gamma, \boldsymbol{\lambda})$, plus an additional conditioning that the vertices $r_1, \dots, r_{i-1}$ are fixed to spin 0 while $r_{i+1}, \dots, r_d$ are fixed to spin 1; we denote this conditioning by σ_{U_i} with $U_i = \{r_1, \dots, r_d\} \setminus \{r_i\}$. Then, $T = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ can be generated by the following recursive procedure; see Fig. 7.2 for an illustration.

Algorithm: $\mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$

1. For each i, let $T_i = \mathrm{T}_{\mathsf{SAW}}(G_i, v_i; \sigma_\Lambda, \sigma_{U_i})$, where we include the extra conditioning σ_{U_i};
2. Let $T = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ be the union of r and $T_1, \dots, T_d$ by connecting $\{r, v_i\}$ for $1 \le i \le d$; output T.

For the purpose of proof, we also consider the two-state spin system on G' with the same parameters $(\beta, \gamma, \boldsymbol{\lambda})$, with an exception that we let the vertices $r_1, \dots, r_d$ have no fields (i.e. setting $\lambda_{r_i} = 1$ for $1 \le i \le d$ instead of λ_r).

We then observe that

$$\mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 1) = \lambda_r \cdot \mathcal{Z}_{G'}^{\sigma_\Lambda}(r_1 \leftarrow 1, \dots, r_d \leftarrow 1),$$

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

$$\mathcal{Z}_{G'}^{\sigma_\Lambda}(r_1 \leftarrow 0, \dots, r_{i-1} \leftarrow 0, r_i \leftarrow 1, \dots, r_d \leftarrow 1) = \beta \cdot \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 1) + \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 0).$$

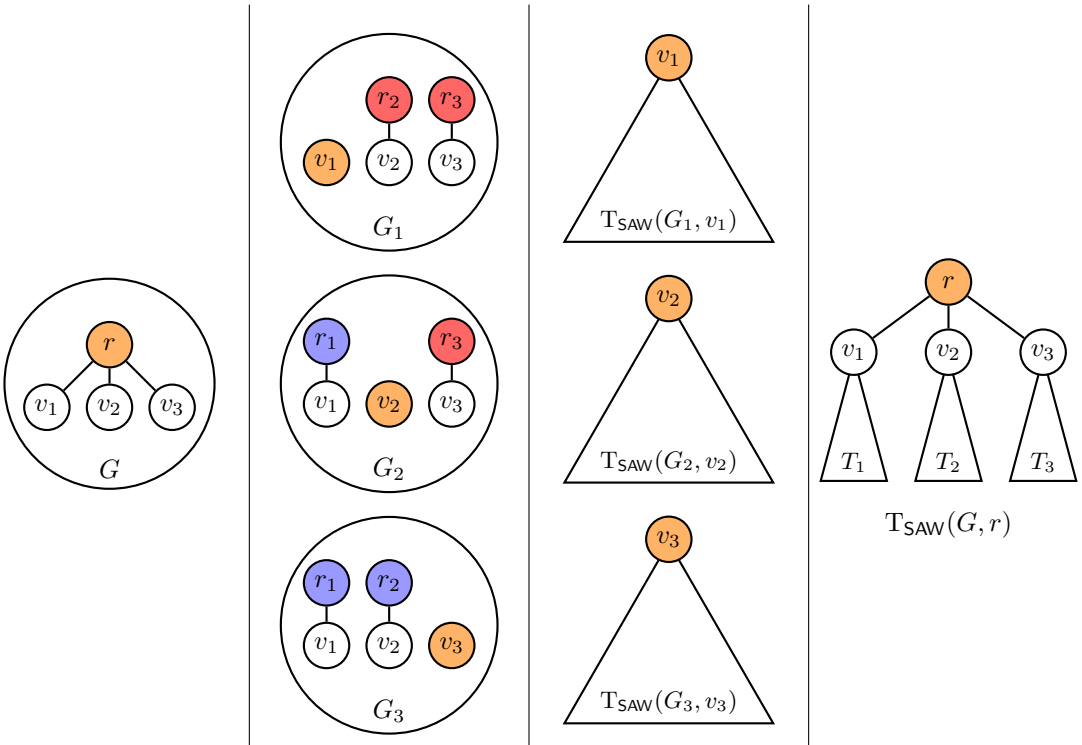
Notice that both sides are independent of the field λ_r. For the left side, all r_i do not have a field for the spin system on G'. For the right side, recall that we do not count the weight of fixed vertices for the conditional partition function for each G_i. Now define $Q_{G,r}^{\sigma_\Lambda} = Q_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$ by

$$Q_{G,r}^{\sigma_\Lambda} = \prod_{i=2}^{d} \mathcal{Z}_{G'}^{\sigma_\Lambda}(r_1 \leftarrow 0, \dots, r_{i-1} \leftarrow 0, r_i \leftarrow 1, \dots, r_d \leftarrow 1),$$

which is independent of λ_r. Then we get

$$\begin{aligned} \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 1) \cdot Q_{G,r}^{\sigma_\Lambda} &= \lambda_r \cdot \prod_{i=1}^{d} \mathcal{Z}_{G'}^{\sigma_\Lambda}(r_1 \leftarrow 0, \dots, r_{i-1} \leftarrow 0, r_i \leftarrow 1, \dots, r_d \leftarrow 1) \\ &= \lambda_r \cdot \prod_{i=1}^{d} \left(\beta \cdot \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 1) + \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 0)\right). \end{aligned}$$

[3] For the interested reader, this multiplicative factor is $\beta^{\#\{u \sim v : \sigma(u) = s\}}$ if $s = 0$ and $\gamma^{\#\{u \sim v : \sigma(u) = s\}} \lambda_v$ if $s = 1$.



(a) Start with the original graph G and "root" vertex $r \in G$.

(b) "Split" r into $d = \deg_G(r)$ copies and impose structural boundary conditions which "telescope". This produces d distinct graphs with distinct boundary conditions which we treat independently.

(c) Recursively build $T_i = \mathrm{T}_{\mathsf{SAW}}(G_i, v_i)$ starting with the unique neighbor v_i of copy r_i for each $i = 1, \ldots, \deg_G(r)$. This recursion terminates since we have reduced the number of unpinned vertices by one in the "split" step.

(d) Merge the (sub)trees built in the previous recursive step. In this figure, $T_i = \mathrm{T}_{\mathsf{SAW}}(G_i, v_i)$ for each $i = 1, 2, 3$.

Figure 7.2: This is a visualization of a single step of the recursive construction of Weitz's self-avoiding walk tree. Red/blue denote 1/0 or in/out, respectively. We highlight "root" vertices, i.e. vertices from which we begin self-avoiding walks, using the color orange. To save space, we only draw r and its neighbors, neglecting the rest of the graph and possible edges between the neighbors.

Using a similar argument, we also have

$$\begin{aligned}\mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 0) \cdot Q_{G,r}^{\sigma_\Lambda} &= \prod_{i=1}^{d} \mathcal{Z}_{G'}^{\sigma_\Lambda}(r_1 \leftarrow 0, \dots, r_i \leftarrow 0, r_{i+1} \leftarrow 1, \dots, r_d \leftarrow 1) \\ &= \prod_{i=1}^{d} \left(\mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 1) + \gamma \cdot \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 0) \right).\end{aligned}$$

Since we assume that $(G \setminus \{r\}) \setminus \Lambda$ is connected, the graph $G_i \setminus \Lambda$ 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

$$\mathcal{Z}_{T_i}^{\sigma_{\Lambda_i}}(r \leftarrow 1) = \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(r \leftarrow 1) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}} \quad \text{and} \quad \mathcal{Z}_{T_i}^{\sigma_{\Lambda_i}}(r \leftarrow 0) = \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(r \leftarrow 0) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}};$$

these polynomials are independent of λ_r since the conditional partition functions for the G_i 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

$$\begin{aligned}\mathcal{Z}_T^{\sigma_\Lambda}(r \leftarrow 1) &= \lambda_r \cdot \prod_{i=1}^{d} \left(\beta \cdot \mathcal{Z}_{T_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 1) + \mathcal{Z}_{T_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 0) \right) \\ &= \lambda_r \cdot \prod_{i=1}^{d} \left(\beta \cdot \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 1) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}} + \mathcal{Z}_{G_i}^{\sigma_{\Lambda_i}}(v_i \leftarrow 0) \cdot P_{G_i,v_i}^{\sigma_{\Lambda_i}} \right) \\ &= \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 1) \cdot Q_{G,r}^{\sigma_\Lambda} \cdot \prod_{i=1}^{d} P_{G_i,v_i}^{\sigma_{\Lambda_i}} \\ &= \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 1) \cdot P_{G,r}^{\sigma_\Lambda}.\end{aligned}$$

The other equality $\mathcal{Z}_T^{\sigma_\Lambda}(r \leftarrow 0) = \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 0) \cdot P_{G,r}^{\sigma_\Lambda}$ is established in the same way. This completes the proof for the case that $(G \setminus \{r\}) \setminus \Lambda$ is connected.

If $(G \setminus \{r\}) \setminus \Lambda$ has two or more connected components, then we can construct $\mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ 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 $G' \setminus \Lambda$ 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 $T_{(j)} = \mathrm{T}_{\mathsf{SAW}}(G'_{(j)}, r_{(j)})$. Observe that once we contract the roots $r_{(1)}, \dots, r_{(k)}$ of $T_{(1)}, \dots, T_{(k)}$, the resulting tree is $\mathrm{T}_{\mathsf{SAW}}(G, r)$.

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 to $\Lambda_{(j)}$. Then $G_{(j)} \setminus \Lambda_{(j)}$ 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 $P_{G_{(i)},r_{(i)}}^{\sigma_{\Lambda_{(j)}}} = P_{G_{(i)},r_{(i)}}^{\sigma_{\Lambda_{(j)}}}(\boldsymbol{\lambda})$, which is independent of λ_r, such that

$$\begin{aligned}&\mathcal{Z}_{T_{(j)}}^{\sigma_{\Lambda_{(j)}}}(r_{(j)} \leftarrow 1) = \mathcal{Z}_{G_{(j)}}^{\sigma_{\Lambda_{(j)}}}(r_{(j)} \leftarrow 1) \cdot P_{G_{(j)},r_{(j)}}^{\sigma_{\Lambda_{(j)}}} \\ \text{and} \quad &\mathcal{Z}_{T_{(j)}}^{\sigma_{\Lambda_{(j)}}}(r_{(j)} \leftarrow 0) = \mathcal{Z}_{G_{(j)}}^{\sigma_{\Lambda_{(j)}}}(r_{(j)} \leftarrow 0) \cdot P_{G_{(j)},r_{(j)}}^{\sigma_{\Lambda_{(j)}}}.\end{aligned}$$

We define the polynomial $P_{G,r}^{\sigma_\Lambda} = P_{G,r}^{\sigma_\Lambda}(\boldsymbol{\lambda})$ to be

$$P_{G,r}^{\sigma_\Lambda} = \prod_{j=1}^{k} P_{G_{(j)},r_{(j)}}^{\sigma_{\Lambda_{(j)}}}.$$

It is then easy to check that

$$\mathcal{Z}_T^{\sigma_\Lambda}(r \leftarrow 1) = \lambda_r \cdot \prod_{j=1}^{k} \mathcal{Z}_{T_{(j)}}^{\sigma_{\Lambda_{(j)}}}(r_{(j)} \leftarrow 1) = \lambda_r \cdot \prod_{j=1}^{k} \left(\mathcal{Z}_{G_{(j)}}^{\sigma_{\Lambda_{(j)}}}(r_{(j)} \leftarrow 1) \cdot P_{G_{(j)}, r_{(j)}}^{\sigma_{\Lambda_{(j)}}} \right)$$

$$= \mathcal{Z}_G \sigma_\Lambda(r \leftarrow 1) \cdot \prod_{j=1}^{k} P_{G_{(j)}, r_{(j)}}^{\sigma_{\Lambda_{(j)}}} = \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 1) \cdot P_{G,r}^{\sigma_\Lambda},$$

and similarly $\mathcal{Z}_T^{\sigma_\Lambda}(r \leftarrow 0) = \mathcal{Z}_G^{\sigma_\Lambda}(r \leftarrow 0) \cdot P_{G,r}^{\sigma_\Lambda}$. The theorem then follows. □

7.3 Influence Bounds on Trees

In this section, we bound the total influence of the root node in an arbitrary tree of maximum degree Δ, assuming $\lambda < \lambda_c(\Delta)$. We prove the following.

Theorem 7.3.1. *Let $T = (V_T, E_T)$ be a tree of maximum degree Δ rooted at some vertex r, and suppose $\lambda \leq (1-\delta)\lambda_c(\Delta)$ for some $0 < \delta < 1$. Then for any boundary condition $\sigma_\Lambda : \Lambda \to \{0,1\}$ where $\Lambda \subseteq V_T$, we have the bound*

$$\sum_{v \in V_T : v \neq r} |\Psi_\mu(r \to v)| \leq \frac{12}{\delta}.$$

The proof utilizes two key ingredients which are specific to trees. The first says that influence factorizes along paths. This kind of chain rule only holds for trees because any pair of vertices are connected by a *unique path*.

Lemma 7.3.2 (Influence Factorization in Trees)**.** *Let μ denote the Gibbs distribution of a two-state spin system on an arbitrary tree $T = (V, E)$ with arbitrary parameters β, γ, λ and arbitrary boundary conditions $\sigma_\Lambda : \Lambda \to \{0,1\}$, where $\Lambda \subseteq V$. Let $u, v, w \in V \setminus \Lambda$ be distinct vertices such that v is on the unique path from u to w. Then $\Psi_\mu^{\sigma_\Lambda}(u \to w) = \Psi_\mu^{\sigma_\Lambda}(u \to v) \cdot \Psi_\mu^{\sigma_\Lambda}(v \to w)$.*

We prove this lemma in Section 7.4. The importance of this lemma is that it allows us to focus on the influence between *neighboring* vertices. This turns out to also have a simple form in terms of the marginal ratios used in the definition of the tree recursions. This is crucial because it will allow us to relate these influences to *derivatives* of the tree recursions.

Here, if $v \in V_T$ is a vertex, then we write R_v for $\mu_{T_v,v}^{\sigma_\Lambda}(1)/\mu_{T_v,v}^{\sigma_\Lambda}(0)$ w.r.t. the subtree rooted at v.

Lemma 7.3.3. *Let $u \in V_T$ and v be a child of u in the subtree T_u. Then*

$$\Psi_T^{\sigma_\Lambda}(u \to v) = -\frac{R_v}{R_v + 1}.$$

Proof. Let us explicitly compute the conditional marginal probabilities which constitute $\Psi_T^{\sigma_\Lambda}(u \to v)$. Since u is a neighbor of v, $\Pr_T[\sigma(v) = 1 \mid \sigma(u) = 1] = 0$ due to the hardcore constraint. On the other hand, conditioning on $\sigma(u) = 0$ is equivalent to deleting u from T since we're forcing u to not be in the independent set, in which case we are left with the subtree T_v of T rooted at v. Hence, by definition, $\Pr_T[\sigma(v) = 1 \mid \sigma(u) = 0] = \frac{R_v}{R_v+1}$. □

We are now ready to prove Theorem 7.3.1. Let us first sketch the argument. The statement of a much more general result showing that contraction implies spectral independence is given in Theorem A.1.1.

Just like in the first step of Eq. (7.2) above, where we did an informal analysis directly using spatial mixing, we can first split the sum over all vertices into a sum over vertices at each distance level away from the root.

$$\sum_{v \in V_T : v \neq r} |\Psi_T^{\sigma_\Lambda}(r \to v)| = \sum_{k=1}^{\infty} \sum_{v \in L_r(k)} |\Psi_T^{\sigma_\Lambda}(r \to v)|$$

If we can show that each level is upper bounded by $\lesssim (1-\Theta(\delta))^k$, then we would get a convergent series and the desired upper bound of $O(1/\delta)$, regardless of how large the tree T is. If for each $v \in L_r(k)$, we write $r = u_0, \ldots, u_k = v$ for the unique path from r to v, then

$$\begin{aligned}
&|\Psi_T^{\sigma_\Lambda}(r \to v)| \\
&= \prod_{i=1}^{k} |\Psi_T^{\sigma_\Lambda}(u_{i-1} \to u_i)| && \text{(Lemma 7.3.2)} \\
&= \prod_{i=1}^{k} \frac{R_{u_i}}{R_{u_i}+1} && \text{(Lemma 7.3.3)} \\
&= \frac{R_v}{R_r} \cdot \prod_{i=1}^{k} \frac{R_{u_{i-1}}}{R_{u_i}+1} && \text{(Telescoping Trick)} \\
&= \frac{R_v}{R_r} \cdot \prod_{i=1}^{k} (\partial_{u_i} F)(R_u : u \in L_{u_{i-1}}(1)) && \text{(Eq. (7.5) and } R_{u_{i-1}} = F(R_u : u \in L_{u_{i-1}}(1))\text{)}
\end{aligned}$$

so that

$$\begin{aligned}
\sum_{v \in L_r(k)} |\Psi_T^{\sigma_\Lambda}(r \to v)| &= \sum_{v \in L_r(k)} \frac{R_v}{R_r} \cdot \prod_{i=1}^{k} (\partial_{u_i} F)(R_u : u \in L_{u_{i-1}}(1)) \\
&\le \max_{v \in L_r(k)} \left\{ \frac{R_v}{R_r} \right\} \cdot \left(\max_{1 \le d < \Delta} \sup_{\boldsymbol{R} \in \mathbb{R}^{\Delta}_{\ge 0}} \|\nabla F_d(\boldsymbol{R})\|_1 \right)^k . && \text{(Induction)}
\end{aligned}$$

This is great because on the right-hand side, we have derivatives of the tree recursion F, which we expect to be less than 1 since F should be a contraction whenever $\lambda < \lambda_c(\Delta)$.

This is almost true. Unfortunately, $\sup_{\boldsymbol{R} \in \mathbb{R}^{\Delta}_{\ge 0}} \|\nabla F(\boldsymbol{R})\|_1$ can be greater than 1 even if $\lambda < \lambda_c(\Delta)$. However, what saves us is the existence of a "good" *potential function* which "amortizes" the correlation decay; this is now a standard tool in the analysis of correlation decay. In particular, [LLY13] discovered an intriguing potential function $\Phi : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ defined as

$$\begin{aligned}
\Phi(R) &\stackrel{\text{def}}{=} 2 \log \left(\sqrt{R} + \sqrt{R+1} \right) \\
\Phi'(R) &= \frac{1}{\sqrt{R(R+1)}}
\end{aligned} \tag{7.10}$$

such that the modified tree recurrence $K_u = F_d^{\Phi}(K_v : v \in L_u(1))$ where

$$F_d^{\Phi}(\boldsymbol{K}) \stackrel{\text{def}}{=} (\Phi \circ F_d \circ \Phi^{-1})(\boldsymbol{K}) = \Phi(F_d(\Phi^{-1}(K_1), \ldots, \Phi^{-1}(K_d)))$$

satisfies the following contractive property.

Theorem 7.3.4 ([LLY13])**.** *Assume $\lambda \ge 0$ is such that the hardcore model with parameter λ is up-to-Δ unique with gap $0 < \delta < 1$. Then for every $1 \le d < \Delta$,*

$$\sup_{\boldsymbol{K} \in \mathrm{Im}(\Phi)^d} \left\| \nabla F_d^{\Phi}(\boldsymbol{K}) \right\|_1 \le \sqrt{1-\delta} < 1.$$

This contractive property is crucial, and implies strong spatial mixing; no such good potential function exists when $\lambda > \lambda_c(\Delta)$. With this final tool in hand, we now complete the proof of Theorem 7.3.1.

Proof of Theorem 7.3.1. It suffices to show that for every $k \ge 1$,

$$\sum_{v \in L_r(k)} |\Psi_T^{\sigma_\Lambda}(r \to v)| \le 6 \cdot (1-\delta)^{k/2}$$

since then,

$$\sum_{v \in V_T : v \ne r} |\Psi_T^{\sigma_\Lambda}(r \to v)| = \sum_{k=1}^{\infty} \sum_{v \in L_r(k)} |\Psi_T^{\sigma_\Lambda}(r \to v)| \le 6 \sum_{k=1}^{\infty} (1-\delta)^{k/2} \le 6 \sum_{k=0}^{\infty} \left(1 - \frac{\delta}{2}\right)^k \le \frac{12}{\delta}$$

as desired.

By the same analysis as above, using Lemmas 7.3.2 and 7.3.3 and the tree recursion,

$$\begin{aligned}
|\Psi_T^{\sigma_\Lambda}(r \to v)| &= \frac{R_v}{R_r} \cdot \prod_{i=1}^{k} (\partial_{u_i} F)(R_u : u \in L_{u_{i-1}}(1)) \\
&= \frac{R_v \cdot \Phi'(R_v)}{R_r \cdot \Phi'(R_r)} \cdot \prod_{i=1}^{k} \frac{\Phi'(R_{u_{i-1}})}{\Phi'(R_{u_i})} \cdot (\partial_{u_i} F)(R_u : u \in L_{u_{i-1}}(1)) \qquad \text{(Telescoping Trick)} \\
&= \frac{R_v \cdot \Phi'(R_v)}{R_r \cdot \Phi'(R_r)} \cdot \prod_{i=1}^{k} (\partial_{u_i} F^\Phi)(K_u : u \in L_{u_{i-1}}(1)). \\
&\qquad \text{($K_u \stackrel{\text{def}}{=} \Phi(R_u)$, Chain Rule, and Inverse Function Theorem)}
\end{aligned}$$

By the same inductive argument, it follows that

$$\begin{aligned}
\sum_{v \in L_r(k)} |\Psi_T^{\sigma_\Lambda}(r \to v)| &= \sum_{v \in L_r(k)} \frac{R_v \cdot \Phi'(R_v)}{R_r \cdot \Phi'(R_r)} \cdot \prod_{i=1}^{k} (\partial_{u_i} F^\Phi)(K_u : u \in L_{u_{i-1}}(1)) \\
&\le \max_{v \in L_r(k)} \left\{ \frac{R_v \cdot \Phi'(R_v)}{R_r \cdot \Phi'(R_r)} \right\} \cdot \left(\max_{1 \le d < \Delta} \sup_{\boldsymbol{K} \in \mathbb{R}^\Delta_{\ge 0}} \left\| \nabla F_d^\Phi(\boldsymbol{K}) \right\|_1 \right)^k \\
&\le \underbrace{\max_{v \in L_r(k)} \left\{ \frac{R_v \cdot \sqrt{R_r + 1}}{R_r \cdot \sqrt{R_v + 1}} \right\}}_{(*)} \cdot \sqrt{1-\delta}^k. \qquad \text{(Theorem 7.3.4 and Eq. (7.10))}
\end{aligned}$$

All that remains is to upper bound $(*)$. By Fact 7.1.5, we have $\lambda \left(\frac{1}{1+\lambda}\right)^\Delta \le R_v, R_r \le \lambda$, so that

$$\frac{R_v \cdot \sqrt{R_r + 1}}{R_r \cdot \sqrt{R_v + 1}} \le \sqrt{\frac{\lambda}{1+\lambda} + (1+\lambda)^{\Delta - 1}}.$$

Since $\lambda < \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$, it follows that $(*) \le \sqrt{26} < 6$. □

Remark 38. The careful reader will notice that the root vertex $\hat{r}$ of $\mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)$ can have degree Δ, instead of degree d for $1 \le d < \Delta$. In this case, we would get $\left\|\nabla F_\Delta^\Phi(\boldsymbol{R})\right\|_1$ instead of $\max_{1 \le d < \Delta} \sup_{\boldsymbol{R}} \left\|\nabla F_\Delta^\Phi(\boldsymbol{R})\right\|_1$. Very roughly speaking, even though λ is only up-to-Δ unique (rather than up-to-$(\Delta+1)$ unique) so that $\left\|\nabla F_\Delta^\Phi(\boldsymbol{R})\right\|_1$ possibly can exceed 1, the potential function Φ satisfies nice boundedness assumptions which ensure that

$$\max_{v \ne r} \left\{ \frac{R_v \cdot \Phi'(R_v)}{R_r \cdot \Phi'(R_r)} \right\} \cdot \left\|\nabla F_\Delta^\Phi(\boldsymbol{R})\right\|_1 \le O(1)$$

independent of Δ. We refer the interested reader to Lemma A.1.2 and its proof for a more detailed discussion of how to remedy this.

7.4 Wrapping Up Unfinished Proofs

Proof of Theorem 7.1.2. Fix $0 \le k \le n-2$ and some boundary condition $\sigma_\Lambda : \Lambda \to \{0,1\}$ where $\Lambda \subseteq V$ with $|\Lambda| = k$. Since

$$\boldsymbol{\lambda}_{\max}\left(\Psi_\mu^{\sigma_\Lambda}\right) \le \left\|\Psi_\mu^{\sigma_\Lambda}\right\|_\infty = \max_{r \in V \setminus \Lambda} \sum_{v \in V \setminus \Lambda : v \ne r} \left|\Psi_\mu^{\sigma_\Lambda}(r \to v)\right|,$$

it suffices to show that for every vertex $r \in V \setminus \Lambda$, the total influence of r is upper bounded by

$$\sum_{v \in V \setminus \Lambda : v \ne r} \left|\Psi_\mu^{\sigma_\Lambda}(r \to v)\right| \le \min\left\{ \frac{32}{\delta}, \frac{\lambda}{1+\lambda}(n-k-1) \right\}.$$

The first bound is the nontrivial one. By Corollary 7.2.3, if $T = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda) = (V_T, E_T)$ is the corresponding self-avoiding walk tree rooted at $\hat{r}$,

$$\begin{aligned}\sum_{v \in V \setminus \Lambda : v \neq r} |\Psi_G^{\sigma_\Lambda}(r \to v)| &\leq \sum_{\hat{v} \in V_T \setminus \Lambda : \hat{v} \neq \hat{r}} |\Psi_T^{\sigma_\Lambda}(\hat{r} \to \hat{v})| \\ &\leq \frac{32}{\delta}. && \text{(Theorem 7.3.1)}\end{aligned}$$

This concludes the first bound. For the second bound, simply observe that by Fact 7.1.5, the conditional marginal probabilities $\mu_{G,v}^{\sigma_\Lambda, r \leftarrow 1}(1), \mu_{G,v}^{\sigma_\Lambda, r \leftarrow 1}(1), \mu_{G,v}^{\sigma_\Lambda, r \leftarrow 0}(1)$ lie in the interval $\left[0, \frac{\lambda}{1+\lambda}\right]$. It immediately follows that $|\Psi_G^{\sigma_\Lambda}(r \to v)| \leq \frac{\lambda}{1+\lambda}$ for every $v \in V \setminus \Lambda$ with $v \neq r$. The second bound immediately follows. □

Proof of Lemma 7.3.2. This lemma follows in a straightforward manner from conditional independence of u and w when the spin of v is fixed, and the Law of Total Probability. We formalize this now. For convenience, for the purposes of this proof, we write v for the event that $\sigma(v) = 1$, and $\overline{v}$ for the event that $\sigma(v) = 0$; we do the same for the events $u, \overline{u}$ and $w, \overline{w}$. All probabilities are with respect to $\sigma \sim \mu_T^{\sigma_\Lambda}$. Thus, for instance, $\Pr[v] = \Pr_{\sigma \sim \mu^{\sigma_\Lambda}}[\sigma(v) = 1]$ and $\Pr[v \mid \overline{u}] = \Pr_{\sigma \sim \mu^{\sigma_\Lambda}}[\sigma(v) = 1 \mid \sigma(u) = 0]$. We this notation in hand, we may expand the conditional probability $\Pr[w \mid u]$ as

$$\begin{aligned}\Pr[w \mid u] &= \Pr[v, w \mid u] + \Pr[\overline{v}, w \mid u] && \text{(Law of Total Probability)} \\ &= \Pr[w \mid u, v] \cdot \Pr[v \mid u] + \Pr[w \mid u, \overline{v}] \cdot \Pr[\overline{v} \mid u] \\ &= \Pr[w \mid v] \cdot \Pr[v \mid u] + \Pr[w \mid \overline{v}] \cdot \Pr[\overline{v} \mid u] && \text{(Conditional Independence)}\end{aligned}$$

Similarly,

$$\Pr[w \mid \overline{u}] = \Pr[w \mid v] \cdot \Pr[v \mid \overline{u}] + \Pr[w \mid \overline{v}] \cdot \Pr[\overline{v} \mid \overline{u}]$$

Combining the above two displays yields

$$\begin{aligned}\Psi_\mu^{\sigma_\Lambda}(u \to w) &= \Pr[w \mid u] - \Pr[w \mid \overline{u}] \\ &= \Pr[w \mid v] \cdot \left(\underbrace{\Pr[v \mid u] - \Pr[v \mid \overline{u}]}_{= \Psi_\mu^{\sigma_\Lambda}(u \to v)} \right) + \Pr[w \mid \overline{v}] \cdot \left(\underbrace{\Pr[\overline{v} \mid u] - \Pr[\overline{v} \mid \overline{u}]}_{= -\Psi_\mu^{\sigma_\Lambda}(u \to v)} \right) \\ &= \Psi_\mu^{\sigma_\Lambda}(u \to v) \cdot \Psi_\mu^{\sigma_\Lambda}(v \to w)\end{aligned}$$

as desired. □

Proof of Lemma 7.1.4. Let $1 < d < \Delta$. First, we calculate that

$$f_d'(R) = -d\lambda \cdot \left(\frac{1}{R+1} \right)^{d+1} = -d \cdot \frac{f_d(R)}{R+1}$$

In particular, at the unique fixed point of f_d, we have

$$\left| f_d'(\hat{R}_d) \right| = d \cdot \left(1 - \frac{1}{\hat{R}_d + 1} \right)$$

Up-to-Δ uniqueness holds only if $\left| f_d'(\hat{R}_d) \right| \leq 1 - \delta$. In terms of $\hat{R}_d$, this holds if and only if $\hat{R}_d \leq \frac{1-\delta}{d-1+\delta}$. Observe that since f_d is monotone decreasing and $\hat{R}_d$ is the unique fixed point of f_d, we have $f_d(R) < R$ for all $R > \hat{R}_d$ and $f_d(R) > R$ for all $R < \hat{R}_d$. Hence, $\frac{1-\delta}{d-1+\delta} \geq \hat{R}_d$ holds if and only if

$$\begin{aligned}&\lambda \left(\frac{d-1+\delta}{d} \right)^d = f_d\left(\frac{1-\delta}{d-1+\delta} \right) \leq \frac{1-\delta}{d-1+\delta} \\ &\iff \lambda \leq \left(\frac{d}{d-1+\delta} \right)^d \cdot \frac{1-\delta}{d-1+\delta} \stackrel{\mathsf{def}}{=} \lambda_c(\delta, d+1)\end{aligned}$$

Now, let us compare this with $\lambda_c(d+1) = \left(\frac{d}{d-1}\right)^d \cdot \frac{1}{d-1}$. Define $c(\delta)$ such that $(1-c(\delta))\lambda_c(d+1) = \lambda_c(\delta, d+1)$. Since $\lambda_c(d+1)$ is monotone decreasing in d and $d < \Delta$, we have $\lambda_c(d+1) \geq \lambda_c(\Delta)$. Thus, we have shown that λ is up-to-Δ unique with gap $0 < \delta < 1$ if and only if $\lambda < (1-c(\delta))\lambda_c(\Delta)$. All that remains is to show $c(\delta) = \Theta(\delta)$.

For this, we first calculate that

$$1 - c(\delta) = (1-\delta)\left(\frac{d-1}{d-1+\delta}\right)^{d+1}$$

Clearly, $\left(\frac{d-1}{d-1+\delta}\right)^{d+1} \leq 1$, which implies $c(\delta) \geq \delta$. On the other hand, by Bernoulli's Inequality, we also have that

$$\left(\frac{d-1}{d-1+\delta}\right)^{d+1} = \left(1 - \frac{\delta}{d-1+\delta}\right)^{d+1} \geq 1 - \frac{d+1}{d-1+\delta} \cdot \delta$$

which implies $c(\delta) \leq O(\delta)$ as well. This concludes the proof. □

Chapter 8

Coupling, Ollivier-Ricci Curvature, and Stein's Method for Markov Chains

In this chapter, we show that the existence of a "good" coupling w.r.t. Hamming distance for any local Markov chain sampling from a high-dimensional discrete distribution μ implies strong bounds on the spectral independence parameter of μ. This result yields rapid mixing for the simple down-up walk in a completely blackbox fashion, and can be viewed as a Markov chain *comparison* result.

Our primary application is to sampling proper list-colorings on bounded-degree graphs. In particular, combining the coupling for the flip dynamics given by [Vig00; Che+19] with our techniques, we show optimal $O(n \log n)$ mixing[1] for the Glauber dynamics for sampling proper list-colorings on any bounded-degree graph with maximum degree Δ whenever the size of the color lists are at least $\left(\frac{11}{6} - \epsilon\right)\Delta$, where $\epsilon \approx 10^{-5}$ is small constant. This approach is markedly different from prior works establishing spectral independence for spin systems using spatial mixing [ALO21; CLV20; Che+21d; Fen+21] (see Chapter 7), which crucially is still open in this regime for proper list-colorings as of this writing.

Along the way, we show how curvature conditions [Oll09] and bounds on the Dobrushin influence matrix imply bounds on the spectral independence of the distribution. However, as mentioned earlier, the coupling we allow will be much weaker; for instance, we can accommodate variable-length path couplings. In particular, for a coupling to be "good", we only require the expected distance between successive iterates under the coupling to be summable, as opposed to being one-step contractive in the worst case. The main technique we use to achieve these results is known as Stein's method for Markov chains [BN19; RR19].

This chapter is based on [Liu21]. We note that these results were independently discovered by [Bla+22].

8.1 Spectral Independence via Coupling and a Blackbox Comparison Result

We now state the main results in this chapter. To do this, we first precisely define our notion of a "good" coupling and locality.

Definition 36 (Amortized Convergent Coupling)**.** *Let* M *be the transition matrix of a reversible, irreducible Markov chain on a finite metric space* (Ω, d) *with stationary distribution* μ*. For* $C > 0$*, we say a coupling of two faithful copies of the chain* $(X^{(t)})_{t\geq 0}, (Y^{(t)})_{t\geq 0}$ *is* C***-amortized convergent (w.r.t. the metric*** $d(\cdot,\cdot)$***)*** *if the following holds for all* $x, y \in \Omega$*:*

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

[1] $O(n^2)$ mixing was already previously known.

Definition 37 (Locality of Dynamics)**.** *Let* M *be the transition matrix of a reversible, irreducible Markov chain on a finite metric space* (Ω, d) *with stationary distribution* μ*. For a positive real number* $\ell > 0$*, we say the dynamics* M *is* ℓ***-local (w.r.t. the metric*** $d(\cdot,\cdot)$***)*** *if*

$$\max_{x,y\in\Omega:\mathsf{M}(x,y)>0} d(x,y) \le \ell.$$

Throughout this chapter, unless stated otherwise, we endow $\binom{\mathscr{U}}{n}$ with Hamming distance. With these notions in hand, we now state our blackbox comparison result.

Theorem 8.1.1 (Blackbox Comparison with Down-Up Walk)**.** *Let* μ *be a distribution on* $\binom{\mathscr{U}}{n}$*, where* $\mathscr{U}$ *is a finite universe and* $n \ge 1$ *is a positive integer. For each feasible* $\sigma \subseteq \mathscr{U}$ *with* $|\sigma| \le n-2$*, let* M_{μ^σ} *be a Markov chain which is reversible w.r.t. the conditional distribution* μ^σ*. Assume the family of Markov chains* $\{\mathsf{M}_{\mu^\sigma}\}_\sigma$ *satisfy the following:*

1. ***Locality:*** *For some* $\ell \ge 0$*,* M_{μ^σ} *is* ℓ*-local w.r.t. Hamming distance for all* σ*.*
2. ***Good Coupling:*** *For some* $C_{n-k} > 0$*,* M_{μ^σ} *admits a* C_{n-k}*-amortized convergent coupling w.r.t. Hamming distance for all* k *and* σ *with* $|\sigma| = k$*.*
3. ***Bounded Differences Between Chains:*** *For some* $\widetilde{C}_{n-k} > 0$*, we have the following bound for every* $0 \le k \le n-2$*,* $i \in \mathscr{U}$ *and feasible* $\sigma \subseteq \mathscr{U}$ *with* $|\sigma| = k$*:*

$$\max_{\tau\in\mathrm{supp}(\mu^{\sigma+i})} \left\{ \sum_{\tau'\neq\tau} \left|\mathsf{M}_{\mu^{\sigma+i}}(\tau\to\tau') - \mathsf{M}_{\mu^\sigma}(\tau\to\tau')\right| \right\} \le \widetilde{C}_{n-k}.$$

If $\ell \cdot C_{n-k}\cdot\widetilde{C}_{n-k} \le O(1)$ *for all* k*, then the down-up walk* $\mathcal{P}_\mu$ *has spectral gap at least* $n^{-O(1)}$*. If, in addition,* μ *is the Gibbs distribution of a spin system on a bounded-degree graph, then the spectral gap, standard and modified log-Sobolev constants (see Definition 42) for the down-up walk are all* $\Omega(1/n)$*.*

Remark 39. While initially it may seem inconvenient to first build an entire family of Markov chains, one for each conditional distribution, this is very natural for many classes of distributions, in particular those which are closed under conditioning. As we will see, in practice, it is easy to obtain bounded differences between chains with $\widetilde{C}_{n-k} \lesssim \frac{1}{n-k}$ simply via brute force calculation. While $C_{n-k} \gtrsim n-k$ is often unavoidable, particularly for ℓ-local chains with $\ell \le O(1)$, we will see that in many settings, we have $C_{n-k} \lesssim n-k$ as well. If additionally our dynamics are ℓ-local with $\ell \le O(1)$, then the above yields a $n^{-O(1)}$ spectral gap for the down-up walk. It will turn out that our notion of ℓ-locality can also be relaxed; see Remark 42.

Our primary concrete application is to sampling proper list-colorings on graphs via the Glauber dynamics, which may be realized as a down-up walk. In this setting, we compare with another useful Markov chain known as the flip dynamics. The flip dynamics is ℓ-local w.r.t. unweighted Hamming distance with $\ell \le 12$, and was analyzed in [Vig00], who gave a greedy coupling which is one-step contractive whenever the number of available colors is at least $\frac{11}{6}\Delta$, implying it is C-amortized convergent with $C \le O(n)$. [Che+19] tweaked the parameters of the flip dynamics slightly while preserving locality, and further constructed a variable-length coupling which contracts by a constant factor every expected $O(n)$ steps whenever the number of available colors is at least $\left(\frac{11}{6}-\epsilon\right)\Delta$ for a small constant $\epsilon \approx 10^{-5}$. We will show this variable-length coupling is also C-amortized convergent with $C \le O(n)$, and deduce optimal mixing for list-colorings in this regime.

Theorem 8.1.2. *Let* $(G, \mathcal{L})$ *be a list-coloring instance where* $G = (V, E)$ *is a graph of maximum degree* $\Delta \le O(1)$ *and* $\mathcal{L} = (\mathcal{L}(v))_{v\in V}$ *is a collection of color lists. Then for some absolute constant* $\epsilon \approx 10^{-5}$*, if* $|\mathcal{L}(v)| \ge \left(\frac{11}{6}-\epsilon\right)\Delta$ *for all* $v \in V$*, then the uniform distribution over proper list-colorings for* $(G, \mathcal{L})$ *is* $(\eta_0, \ldots, \eta_{n-2})$*-spectrally independent where* $\eta_k \le O(1)$ *for all* k*. Furthermore, the spectral gap, standard and modified log-Sobolev constants (see Definition 42) for the Glauber dynamics are all* $\Omega(1/n)$*, and the mixing time is* $O(n\log n)$*.*

Remark 40. Our running time dependence on Δ is roughly Δ^{Δ^c} for a mild constant c, which is rather poor. The main bottleneck in improving this dependence lies in the local-to-global result of [CLV21a], although our spectral independence bound, which depends polynomially on Δ, can also be significantly improved.

To prove Theorem 8.1.1, we leverage recent local-to-global results [AL20; CLV21a] (see Theorems 2.3.1 and 10.0.1 for formal statements), which show that if one has sufficiently strong upper bounds on the total pairwise correlation $\sum_{j\in\mathscr{U}} |\Pr_{\sigma\sim\mu}[j\in\sigma \mid i\in\sigma] - \Pr_{\sigma\sim\mu}[j\in\sigma]|$, then one can deduce rapid mixing for the down-up walk [KM17; DK17; KO18; Opp18]. To upper bound these correlations, we considerably generalize a result simultaneously due to [BN19; RR19], which was discovered in the context of bounding the Wasserstein 1-distance between Ising models, or more generally, two measures on the discrete hypercube $\{-1,+1\}^n$. More specifically, we extend their results in several different directions:

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

Theorem 8.1.3. *Let μ be a distribution on $\binom{\mathscr{U}}{n}$, where $\mathscr{U}$ is some finite universe and $n\geq 1$ is a positive integer. Fix an arbitrary $i\in\mathscr{U}$. Let M_μ (resp. M_{μ^i}) be the transition kernel of any irreducible Markov chain on $\mathrm{supp}(\mu)$ (resp. $\mathrm{supp}(\mu^i)$) which is reversible w.r.t. μ (resp. $\mathrm{supp}(\mu^i)$). Suppose that M_μ is ℓ-local and admits a C-amortized convergent coupling, both w.r.t. the Hamming metric $d(\cdot,\cdot)$. Then we have the bound*

$$\sum_{j\in\mathscr{U}} \left| \Pr_{S\sim\mu}[j\in\sigma \mid i\in\sigma] - \Pr_{\sigma\sim\mu}[j\in\sigma] \right| \leq C\cdot\ell\cdot \max_{\tau\in\mathrm{supp}(\mu^i)} \left\{ \sum_{\tau'\neq\tau} \left|\mathsf{M}_{\mu^i}(\tau\to\tau') - \mathsf{M}_\mu(\tau\to\tau')\right| \right\}.$$

8.1.1 Bounding the Distance Between Two Distributions

We now state our main technical result, which provides the most general bound on the difference between marginals of two distributions μ,ν. We immediately use it to deduce Theorem 8.1.3.

Theorem 8.1.4. *Let μ,ν be any two distributions on $\binom{\mathscr{U}}{n}$ for a finite set $\mathscr{U}$ with $\mathrm{supp}(\nu)\subseteq\mathrm{supp}(\mu)$, where $\mathscr{U}$ is a finite universe and $n\geq 1$ is a positive integer. Further, let M_μ (resp. M_ν) be the transition kernel of any Markov chain on $\mathrm{supp}(\mu)$ (resp. $\mathrm{supp}(\nu)$) with stationary distribution μ (resp. ν). Assume M_μ is irreducible and reversible w.r.t. μ. Then we may bound both $\sum_{j\in\mathscr{U}} |\Pr_{\sigma\sim\mu}[j\in\sigma] - \Pr_{\sigma\sim\nu}[j\in\sigma]|$ and the 1-Wasserstein distance $\mathscr{W}_1(\mu,\nu)$ (see Eq. (1.1)) by the following quantity:*

$$\mathbb{E}_{\tau\sim\nu}\left[\sum_{\tau'\neq\tau} |\mathsf{M}_\mu(\tau\to\tau') - \mathsf{M}_\nu(\tau\to\tau')| \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)},Y^{(t)}}\left[d_H\left(X^{(t)},Y^{(t)}\right) \mid \substack{X^{(0)}=\tau\\ Y^{(0)}=\tau'}\right]\right],$$

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

Remark 41. The technical condition $\mathrm{supp}(\nu)\subseteq\mathrm{supp}(\mu)$ is just for convenience, as it ensures the transition probability $\mathsf{M}_\mu(\tau\to\tau')$ also makes sense when $\tau\sim\nu$. This assumption is certainly satisfied in our application where ν is a conditional distribution of μ.

Proof of Theorem 8.1.3. We use Theorem 8.1.4 with $\nu=\mu^i$ to obtain the upper bound

$$\begin{aligned}
&\mathbb{E}_{\tau\sim\mu^i}\left[\sum_{\tau'\neq\tau} \left|\mathsf{M}_\mu(\tau\to\tau') - \mathsf{M}_{\mu^i}(\tau\to\tau')\right| \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)},Y^{(t)}}\left[d_H\left(X^{(t)},Y^{(t)}\right) \mid \substack{X^{(0)}=\tau\\ Y^{(0)}=\tau'}\right]\right]\\
&\leq \max_{\tau\in\mathrm{supp}(\mu^i)} \left\{ \sum_{\tau'\neq\tau} \left|\mathsf{M}_\mu(\tau\to\tau') - \mathsf{M}_{\mu^i}(\tau\to\tau')\right| \right\}\\
&\qquad \cdot \underbrace{\mathbb{E}_{\tau\sim\mu^i}\left[\max_{\tau':\mathsf{M}_\mu(\tau\to\tau')>0} \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)},Y^{(t)}}\left[d_H\left(X^{(t)},Y^{(t)}\right) \mid \substack{X^{(0)}=\tau\\ Y^{(0)}=\tau'}\right]\right]}_{(*)}.
\end{aligned}$$

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

$$\sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)},Y^{(t)}}\left[d_H\left(X^{(t)},Y^{(t)}\right) \mid \substack{X^{(0)}=\tau\\ Y^{(0)}=\tau'}\right] \leq C\cdot d_H(\tau,\tau').$$

Hence,

$$(*) \leq C \cdot \mathbb{E}_{\tau \sim \mu^i} \left[\max_{\tau' : \mathsf{M}_\mu(\tau \to \tau') > 0} d_H(\tau, \tau') \right] \leq C \cdot \ell.$$

by ℓ-locality of M_μ. □

Remark 42. One can see from the proof that we only needed that

$$\mathbb{E}_{\tau \sim \mu^i} \left[\max_{\tau' : \mathsf{M}_\mu(\tau \to \tau') > 0} d_H(\tau, \tau') \right] \leq \ell,$$

as opposed to the stronger notion of ℓ-locality, where we have $\max_{\tau, \tau' : \mathsf{M}_\mu(\tau \to \tau') > 0} d_H(\tau, \tau') \leq \ell$. Thus, in some sense, we only need the dynamics to make local moves "on average". We leave it to future work to exploit this additional flexibility.

8.2 Stein's Method for Markov Chains

Our goal in this section is to prove Theorem 8.1.4. We follow [BN19; RR19], using what is known as *Stein's method for Markov chains*. Historically, Stein's method [Ste72] was developed as a method to bound distances between probability measures, with the primary motivation being to prove quantitative central limit theorems. [BN19; RR19] adapted this method to bound the distance between two probability measures μ, ν on the discrete hypercube $\{-1, +1\}^n$ assuming the Glauber dynamics of either measure admits a contractive coupling. Our main intuition lies in viewing spectral independence as a measure of distance between different conditionings of the same distribution. Thus, one can try to apply this method to bound the spectral independence of a distribution. Let us now elucidate this method.

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

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

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

Lemma 8.2.1 (see Lemma 2.1 [BN19], Lemma 2.3 [RR19])**.** *Let M be the transition matrix of a reversible, irreducible Markov chain on a finite state space Ω with stationary distribution μ. Let $(X^{(t)})_{t=0}^{\infty}$ be the Markov chain generated by M, and for a fixed function $f : \Omega \to \mathbb{R}$, define $h : \Omega \to \mathbb{R}$ by*

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

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

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

With this lemma in hand, we can immediately prove Theorem 8.1.4.

Proof of Theorem 8.1.4. Fix a function $f : \binom{\mathscr{U}}{n} \to \mathbb{R}$, and let h be the solution to the Poisson equation $h - \mathsf{M}_\mu h = f - \mathbb{E}_\mu f$ given in Lemma 8.2.1. Since ν is stationary w.r.t. M_ν, we have $\mathbb{E}_\nu \mathsf{M}_\nu h = \mathbb{E}_\nu h$, so that using the Poisson equation yields

$$\mathbb{E}_\nu(\mathsf{M}_\nu - \mathsf{M}_\mu)h = \mathbb{E}_\nu h - \mathbb{E}_\nu\left[h - f + \mathbb{E}_\mu f\right] = \mathbb{E}_\nu f - \mathbb{E}_\mu f.$$

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

Now, let us bound $|(\mathsf{M}_\nu - \mathsf{M}_\mu)h|$ entrywise. For each $\tau \in \mathrm{supp}(\nu)$, using the identity $\mathsf{M}_\mu(\tau \to \tau) = 1 - \sum_{\tau' \neq \tau} \mathsf{M}_\mu(\tau \to \tau')$ (and analogously for M_ν), we see that

$$\begin{aligned}(\mathsf{M}_\nu - \mathsf{M}_\mu)h(\tau) &= \sum_{\tau'} (\mathsf{M}_\nu(\tau \to \tau') - \mathsf{M}_\mu(\tau \to \tau')) \cdot h(\tau') \\ &= \sum_{\tau' \neq \tau} (\mathsf{M}_\nu(\tau \to \tau') - \mathsf{M}_\mu(\tau \to \tau')) \cdot (h(\tau') - h(\tau)) \\ &= \sum_{\tau' \neq \tau} (\mathsf{M}_\nu(\tau \to \tau') - \mathsf{M}_\mu(\tau \to \tau')) \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)},Y^{(t)}} \left[f(Y^{(t)}) - f(X^{(t)}) \mid \substack{X^{(0)}=\tau \\ Y^{(0)}=\tau'} \right].\end{aligned}$$

(Lemma 8.2.1)

It follows by the Triangle Inequality that

$$|(\mathsf{M}_\nu - \mathsf{M}_\mu)h(\tau)| \leq \sum_{\tau' \neq \tau} |\mathsf{M}_\mu(\tau \to \tau') - \mathsf{M}_\nu(\tau \to \tau')| \cdot \sum_{t=0}^{\infty} \mathbb{E}_{X^{(t)},Y^{(t)}} \left[\left| f(X^{(t)}) - f(Y^{(t)}) \right| \mid \substack{X^{(0)}=\tau \\ Y^{(0)}=\tau'} \right]. \tag{8.1}$$

Taking expectations w.r.t. ν finally yields a bound on $|\mathbb{E}_\mu f - \mathbb{E}_\nu f|$. The bound on the 1-Wasserstein distance follows immediately by taking f to be an arbitrary function which is 1-Lipschitz the metric $d_H(\cdot,\cdot)$.

To obtain the bound on the total difference between marginals $\sum_{j \in \mathscr{U}} |\Pr_{\sigma\sim\mu}[j \in \sigma] - \Pr_{S\sim\nu}[j \in \sigma]|$, we apply the above inequality to the indicator function $f = \mathbb{I}_j$ for each $j \in \mathscr{U}$ and sum over all $j \in \mathscr{U}$, noting that $d_H(\tau,\tau') = \sum_{j \in \mathscr{U}} |\mathbb{I}_j(S) - \mathbb{I}_j(T)|$ and $\mathbb{E}_\mu \mathbb{I}_j = \Pr_{S\sim\mu}[j \in S]$ (and analogously for ν). □

8.3 Ollivier-Ricci Curvature on Discrete Product Spaces

In this section, we discuss applications of our results to general distributions on discrete product spaces. We show that the existence of a contractive coupling w.r.t. Hamming distance for the Glauber dynamics implies $O(1)$-spectral independence. Such a condition is known as the Ollivier-Ricci curvature condition for the dynamics in the sense of [Oll09]. This also shows that the Dobrushin uniqueness condition implies $O(1)$-spectral independence. When combined with the local-to-global result of [CLV21a], we resolve an unpublished conjecture due Peres–Tetali in the special case of Glauber dynamics for spin systems on bounded-degree graphs; see [ELL17] and references therein for recent progress on this conjecture on general Markov chains. We also give an alternative proof of the $\Omega(1/n)$ lower bound on the standard and modified log-Sobolev constants of the Glauber dynamics in this setting when a Dobrushin-type condition is satisfied, essentially recovering a result of [Mar19].

Classical work on Dobrushin-type conditions [Dob70; DS85a; DS85b; DS87b; Hay06; DGJ09] yield relatively simple and direct criteria for rapid mixing of the Glauber dynamics [BD97a; BD97b]. The main idea here is intuitively similar to that of spectral independence (although the notion of Dobrushin influence here historically precedes spectral independence): so long as some measure of "total influence" is small, then μ is close in some sense to a product distribution, for which rapid mixing holds. However, prior to our work, the precise relationship between Dobrushin influence and our notion of influence used in spectral independence was unclear. This is an additional conceptual contribution of this chapter.

Definition 38 (Ollivier-Ricci Curvature [Oll09])**.** *Let M be the transition matrix of a reversible, irreducible Markov chain on a finite metric space (Ω, d) with stationary distribution μ. We define the **coarse Ricci curvature (or Ollivier-Ricci curvature)** of the Markov chain M w.r.t. the metric $d(\cdot,\cdot)$ by*

$$\alpha \stackrel{\mathsf{def}}{=} \inf_{x,y \in \Omega : x \neq y} \left\{ 1 - \frac{\mathscr{W}_1(\mathsf{M}(x \to \cdot), \mathsf{M}(y \to \cdot))}{d(x,y)} \right\},$$

*where $\mathscr{W}_1(\cdot,\cdot)$ is again the 1-Wasserstein distance w.r.t. $d(\cdot,\cdot)$. In other words, for every pair $x, y \in \Omega$, there is a coupling of the transitions $\mathsf{M}(x \to \cdot), \mathsf{M}(y \to \cdot)$ such that the expected distance $d(\cdot,\cdot)$ under the coupling contracts by a $(1-\alpha)$-multiplicative factor. In this case, we will say M admits a $(1-\alpha)$-**contractive coupling w.r.t.** $d(\cdot,\cdot)$.*

Fact 8.3.1. *Suppose* M *admits a* $(1-\alpha)$*-contractive coupling w.r.t.* $d(\cdot,\cdot)$*. Then this coupling is* C*-amortized convergent with* $C = \frac{1}{\alpha}$*.*

The following is an immediate application of Theorem 8.1.3, and yields a positive resolution to the Peres-Tetali conjecture in the special case of Glauber dynamics for spin systems on bounded-degree graphs.

Theorem 8.3.2 (Curvature Implies Spectral Independence on Product Spaces)**.** *Let* μ *be a probability measure on a discrete product space* $\Omega = \prod_{v\in V}\Sigma_v$*, where* V *is a finite index set with* $|V| = n$ *(e.g.* $[n]$*) and* Σ_v *is finite for all* $v \in V$*. Endow* Ω *with the Hamming metric* $d_H(\cdot,\cdot)$*, and let* α *be the Ollivier-Ricci curvature of the Glauber dynamics w.r.t.* (Ω, d_H)*. Then, the distribution is* $(\eta_0,\dots,\eta_{n-2})$*-spectrally independent where* $\eta_k \le \frac{4}{\alpha n} - 1$ *for all* k*. In particular, if* $\alpha \ge \Omega(1/n)$*, then the Glauber dynamics has spectral gap* $n^{-O(1)}$*. If in addition* μ *is the Gibbs distribution of a spin system on a bounded-degree graph, then the spectral gap, standard and modified log-Sobolev constants for the Glauber dynamics are all* $\Omega(1/n)$*.*

Note that since the Glauber dynamics only updates the assignment to a single $v \in V$ in each step, it must be that $\alpha \le O(1/n)$.

Proof. We show that $\eta_0 \le \frac{4}{\alpha n} - 1$. The bound $\eta_k \le \frac{4}{\alpha n} - 1$ follows by the same argument by instead considering the Glauber dynamics for the conditional distributions μ^σ of μ for each feasible σ. Because the Glauber dynamics only updates at most one coordinate in each step, it is 2-local w.r.t. $d_H(\cdot,\cdot)$. By Fact 8.3.1, we also have there is a C-amortized convergent coupling with $C = \frac{1}{\alpha}$. It follows from Theorem 8.1.3 that for every $u \in V$ and every $s_u \in \Sigma_u$,

$$\sum_{v\in V}\sum_{s_v\in\Sigma_v}\left|\Pr_{\sigma\sim\mu}[\sigma(v) = s_v \mid \sigma(u) = s_u] - \Pr_{\sigma\sim\mu}[\sigma(v) = s_v]\right| \le \frac{2}{\alpha}\cdot\max_{\sigma\in\mathrm{supp}(\mu^{u\leftarrow s_u})}\sum_{\tau\neq\sigma}\left|\mathsf{M}_{\mu^{u\leftarrow s_u}}(\sigma\to\tau) - \mathsf{M}_\mu(\sigma\to\tau)\right|.$$

Now, by the definition of the Glauber dynamics, for each $\sigma \in \mathrm{supp}(\mu^{u\leftarrow s_u})$, we have

$$\begin{aligned}
&\sum_{\tau\neq\sigma}\left|\mathsf{M}_{\mu^{u\leftarrow s_u}}(\sigma\to\tau) - \mathsf{M}_\mu(\sigma\to\tau)\right| \\
&= \sum_{v\in V}\sum_{s_v\in\Sigma_v : s_v\neq\sigma(v)}\left|\frac{1}{n}\cdot\mu_v^{\sigma_{-v}}(s_v) - \frac{1}{n-1}\cdot\mu_v^{\sigma_{-v},u\leftarrow s_u}(s_v)\right| \\
&= \sum_{v\in V : v\neq u}\sum_{s_v\in\Sigma_v : s_v\neq\sigma(v)}\left(\frac{1}{n-1} - \frac{1}{n}\right)\mu_v^{\sigma_{-v}}(s_v) + \sum_{s\in\Sigma_u : s\neq s_u}\frac{1}{n}\mu_u^{\sigma_{-u}}(s) \\
&\le \frac{2}{n}.
\end{aligned}$$

The claim for the spectral gap in the case $\alpha \ge \Omega(1/n)$ follows by combining with Theorem 2.3.1. The final claim for spin systems on bounded-degree graphs follows by combining with Theorem 10.0.1. □

8.3.1 Dobrushin Uniqueness and Spectral Independence

We now use Theorem 8.3.2 to show that Dobrushin's uniqueness condition implies spectral independence.

Definition 39 (Dobrushin Influence)**.** *Fix a probability measure* μ *on a discrete product space* $\prod_{v\in V}\Sigma_v$*, where* V *is a finite indexing set. For each* $u \in V$*, let* D_u *be the collection of pairs* $\tau,\sigma \in \prod_{v\in V}\Sigma_v$ *such that* $\tau_{-u} = \sigma_{-u}$ *while* $\tau(u) \neq \sigma(u)$*. For distinct* $u, v \in V$*, we may then define the* ***Dobrushin influence of*** u ***on*** v *by*

$$\mathscr{R}_\mu(u\to v) = \max_{(\tau,\sigma)\in D_u}\|\mu_v^{\tau_{-v}} - \mu_v^{\sigma_{-v}}\|_{\mathsf{TV}}.$$

We write $\mathscr{R}_\mu \in \mathbb{R}^{V\times V}$ *for the* ***Dobrushin influence matrix*** *whose entries are given* $\mathscr{R}_\mu(u,v) = \mathscr{R}_\mu(u\to v)$*. We say the distribution* μ *satisfies the*

- ***Dobrushin (uniqueness) condition*** *if* $\|\mathscr{R}_\mu\|_1 = \max_{v\in V}\sum_{u\in V}\mathscr{R}_\mu(u\to v) < 1$.
- ***Dobrushin-Shlosman condition*** *if* $\|\mathscr{R}_\mu\|_\infty = \max_{u\in V}\sum_{v\in V}\mathscr{R}_\mu(u\to v) < 1$.
- ℓ_2***-Dobrushin condition*** *if* $\|\mathscr{R}_\mu\|_2 < 1$.[2]

A straightforward application of the path coupling technique of [BD97a; BD97b] shows that if $\|\mathscr{R}_\mu\|_1 < 1$, then there is a coupling for the Glauber dynamics which is one-step contractive w.r.t. Hamming distance. We state this well-known implication formally here, and refer to [DGJ09] for the proof.

Fact 8.3.3. *Let μ be a distribution on some discrete product space $\prod_{v\in V}\Sigma_v$, where V is a finite index set. If $\|\mathscr{R}_\mu\|_1 \le \gamma < 1$, then the Glauber dynamics is $(1-\alpha)$-contractive w.r.t. Hamming distance with $\alpha = \frac{1}{n}(1-\gamma)$.*

In particular, combining Theorem 8.3.2 and Fact 8.3.3 immediately yields spectral independence under the Dobrushin uniqueness condition. Combined with Theorem 10.0.1, this additionally recovers a version of a result due to [Mar19], which says that a weaker ℓ_2-version of the Dobrushin uniqueness condition (see also [Hay06; DGJ09]) implies a $\Omega(1/n)$ log-Sobolev constant for the Glauber dynamics.

Corollary 8.3.4 (Dobrushin Uniqueness Implies Spectral Independence)**.** *Let μ be a distribution on some discrete product space $\prod_{v\in V}\Sigma_v$, where V is a finite index set with $|V| = n$ (e.g. $[n]$). If $\|\mathscr{R}_\mu\|_1 \le \gamma < 1$, then μ is $(\eta_0,\dots,\eta_{n-2})$-spectrally independent with $\eta_k \le \frac{4}{1-\gamma} - 1$ for all k. If in addition μ is the Gibbs distribution of a spin system on a bounded-degree graph, then the spectral gap, standard and modified log-Sobolev constants for the Glauber dynamics are all $\Omega(1/n)$.*

8.4 Spectral Independence for Proper List-Colorings

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

A well-known result due to [Jer95] using path coupling shows that if $|\mathcal{L}(v)| > 2\Delta$ for all $v \in V$, then there is a contractive one-step coupling for the Glauber dynamics which yields $O(n\log n)$ mixing. As noted in [CLV21a], one can adapt the argument of [GKM15] to obtain strong spatial mixing when $|\mathcal{L}(v)| > 2\Delta$, and use the arguments of [Che+21d; Fen+21] to deduce spectral independence in this regime. However, it is still open whether one can obtain strong spatial mixing below the 2Δ threshold; see [GKM15; Eft+19] for results going below 2Δ on special classes of graphs.

In the seminal work of Vigoda [Vig00], it was shown that there is a contractive one-step coupling for a different local Markov chain known as the flip dynamics whenever $|\mathcal{L}(v)| \ge \frac{11}{6}\Delta$. This threshold was further improved to $|\mathcal{L}(v)| \ge \left(\frac{11}{6} - \epsilon\right)\Delta$ in a recent breakthrough by [Che+19], this time using a more sophisticated variable-length coupling. Both works further showed that Glauber dynamics mixes in $O(n^2)$ time in this regime using a spectral gap comparison argument [DS93].

Our goal is to use these coupling results along with Theorem 8.1.3 to obtain spectral independence for the uniform distribution over proper list-colorings in the regime $|\mathcal{L}(v)| \ge \left(\frac{11}{6} - \epsilon\right)\Delta$. Combined with Theorem 10.0.1, we improve the previous $O(n^2)$ mixing time bound to the optimal $O(n\log n)$, as well as show Chernoff-type concentration bounds for Lipschitz functions, which were not known before.

8.4.1 The Flip Dynamics

We follow the presentation in [Che+19], which generalizes the flip dynamics analyzed in [Vig00] to list-colorings. Fix a list-coloring σ. We say a path $u = w_1,\dots,w_\ell = v$ in G is an *alternating path* from u to v using colors $\sigma(u), c$ if for all i, we have $\sigma_{w_i} \in \{\sigma(u), c\}$ and $\sigma_{w_i} \neq \sigma_{w_{i+1}}$. For a

[2] Other matrix norms for $\mathscr{R}_\mu$ were studied in [DGJ09].

fixed list-coloring σ, $v \in V$ and color c, we define the *Kempe component for* σ, u, c by the following subset of vertices.

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

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

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

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

We write $\mathsf{M}_\mu^{\mathsf{flip}}$ for the transition probability matrix of the flip dynamics. It is straightforward to verify that the stationary distribution of the flip dynamics is uniform over proper list-colorings, regardless of the choice of the flip parameters. One can recover the Wang-Swendsen-Kotecký Markov chain by setting $p_j = j$ for all $j \in \mathbb{N}$ [WSK89].

[Vig00] showed that with flip parameters

$$p_1 = 1 \quad p_2 = \frac{13}{42} \quad p_3 = \frac{1}{6} \quad p_4 = \frac{2}{21} \quad p_5 = \frac{1}{21} \quad p_6 = \frac{1}{84} \quad p_j = 0, \forall j \ge 7, \tag{8.2}$$

there is a one-step coupling which is contractive w.r.t. Hamming distance whenever $|\mathcal{L}(v)| \ge \frac{11}{6}\Delta$. [Che+19] showed using linear programming arguments that this is optimal in the sense that when $|\mathcal{L}(v)| < \frac{11}{6}$, there is no choice of the flip parameters which has a one-step contractive coupling w.r.t. Hamming distance. They additionally construct an explicit family of hard instances witnessing optimality.

One of the key insights of [Che+19] is that the optimal choice of flip parameters comes out of the solution to a linear program, with the objective value of the program governing the contraction properties of the coupling. By solving this linear program, they show that for the following choice of flip parameters

$$\begin{aligned} &\hat{p}_1 = 1 \quad \hat{p}_2 \approx 0.296706 \quad \hat{p}_3 \approx 0.166762 \quad \hat{p}_4 \approx 0.101790 \\ &\hat{p}_5 \approx 0.058475 \quad \hat{p}_6 = 0.025989 \quad \hat{p}_j = 0, \forall j \ge 7, \end{aligned} \tag{8.3}$$

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

We formalize their main coupling result in the following subsection. For the moment, we state two intermediate lemmas, prove one of them, and show how they imply Theorem 8.1.2.

Lemma 8.4.1. *Assume the input graph $G = (V, E)$ has maximum degree $\Delta \le O(1)$. Then, the flip dynamics with parameters given in Eq. (8.3) satisfy the following:*

$$\max_{\tau \in \mathrm{supp}(\mu|_{uc})} \left\{ \sum_{\sigma \ne \tau} \left| \mathsf{M}_\mu^{\mathsf{flip}}(\tau \to \sigma) - \mathsf{M}_{\mu|_{uc}}^{\mathsf{flip}}(\tau \to \sigma) \right| \right\} \le O(1/n).$$

Lemma 8.4.2. *Let $(G, \mathcal{L})$ be a list-coloring instance, where $\Delta \le O(1)$ and $|\mathcal{L}(v)| \ge \lambda^*\Delta$ for all $v \in V$, where $\lambda^* = \frac{11}{6} - \epsilon$ and $\epsilon \approx 10^{-5}$ is a small constant. Then the flip dynamics with parameters given in Eq. (8.3) admits a C-amortized convergent coupling w.r.t. Hamming distance where $C \le O(n)$.*

Proof of Theorem 8.1.2. The flip dynamics is clearly $O(1)$-local w.r.t. Hamming distance since only Kempe components of size at most 6 can be flipped. $(\eta_0, \dots, \eta_{n-2})$-spectral independence where $\eta_k \le O(1)$ for all k then follows immediately by combining Lemma 8.4.1 and Lemma 8.4.2 with Theorem 8.1.1. The lower bounds on the spectral gap, standard and modified log-Sobolev constants then follow from Theorem 10.0.1. □

Proof of Lemma 8.4.1. The main detail one must be careful of is that the flip dynamics for sampling from $\mu \mid uc$ always leaves the color for u fixed to c. Hence, flipping any Kempe component containing u leads to potentially different list-colorings under $\mathsf{M}_{\mu}^{\mathsf{flip}}$ versus $\mathsf{M}_{\mu|uc}^{\mathsf{flip}}$. However, since we only flip components of $O(1)$-size, this isn't an issue for us.

Fix a τ with $\tau(u) = c$, and let $B(u, 6)$ denote the set of vertices of shortest path distance at most 6 away from u in G. Since we only flip Kempe components of size at most 6, we have that for any $v \in V \setminus B(u, 6)$ and $c \in \mathcal{L}(u)$, the flippable Kempe component $S_\tau(v, c')$ does not contain u, and hence, flipping it leads to the same list-coloring under $\mathsf{M}_{\mu}^{\mathsf{flip}}$ and $\mathsf{M}_{\mu|vc}^{\mathsf{flip}}$. Hence, we have

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

□

Remark 43. As one can see in the proof from the factor of Δ^6, we have made no attempt to optimize constants.

At this point, all that remains is to prove Lemma 8.4.2, which we do using the variable-length path coupling constructed in [Che+19].

8.4.2 Variable-Length Path Coupling: Proof of Lemma 8.4.2

To begin, we first define the notion of variable-length coupling following [HV07; Che+19].

Definition 40 (Path-Generating Set)**.** *For a finite state space Ω, a **path generating set** is a subset $S \subseteq \binom{\Omega}{2}$ such that the undirected graph (Ω, S) is connected. We let $d_S(\cdot, \cdot)$ denote the induced shortest-path metric on Ω, and write $d(\cdot, \cdot)$ when the path generating set S is clear from context. We also write $x \sim y$ whenever $\{x, y\} \in S$.*

Definition 41 (Variable-Length Path Coupling [HV07])**.** *Fix an irreducible transition probability matrix M which is reversible w.r.t. a distribution μ on a finite state space Ω, and let $d(\cdot, \cdot)$ be a metric on Ω induced by a path generating set $S \subseteq \binom{\Omega}{2}$. For every pair of starting states $x^{(0)}, y^{(0)} \in \Omega$ with $x^{(0)} \sim y^{(0)}$, we let $(\overline{x}, \overline{y}, T) = (\overline{x}(x^{(0)}, y^{(0)}), \overline{y}(x^{(0)}, y^{(0)}), T(x^{(0)}, y^{(0)}))$ denote a random variable where T is a (potentially random) nonnegative integer and $\overline{x} = (x^{(0)}, x^{(1)}, \ldots, x^{(T)}), \overline{y} = (y^{(0)}, y^{(1)}, \ldots, y^{(T)})$ are length-T sequences of states in Ω.*

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

Remark 44. In our application to colorings, the random time T will be a stopping time in the sense that its value only depends on the past, i.e. $x^{(0)}, y^{(0)}, \ldots, x^{(t)}, y^{(t)}$ for $t \leq T$.

Given a variable-length path coupling, [HV07] showed one can construct a full coupling, generalizing the original path coupling theorem of [BD97a; BD97b]. Furthermore, the contraction properties of the full coupling are inherited from the path coupling. While the original statement in [HV07] merely states rapid mixing given a variable-length path coupling, its proof implies the following.

Theorem 8.4.3 (Proof of Corollary 4 from [HV07]). *Let $(\overline{x}, \overline{y}, T)$ be a variable-length path coupling w.r.t. a path generating set S for a reversible Markov chain M on a state space Ω with stationary distribution μ. Let*

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

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

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

Given this, all we need now is a good variable-length path coupling. This is given by the following result due to [Che+19].

Theorem 8.4.4 ([Che+19]). *Let $(G, \mathcal{L})$ be a list-coloring instance, where $G = (V, E)$ is a graph with maximum degree $\Delta \leq O(1)$, and $\mathcal{L} = (\mathcal{L}(v))_{v \in V}$ is a collection of color lists. Let the path generating set S be given by the set of pairs $\{\tau, \sigma\}$ such that τ, σ differ on the coloring of exactly one vertex. Assume $|\mathcal{L}(v)| \geq \lambda^* \Delta$ for all $v \in V$ where $\lambda^* = \frac{11}{6} - \epsilon$ for an absolute constant $\epsilon \approx 10^{-5}$. Then there exists a variable-length path coupling $(\overline{\tau}, \overline{\sigma}, T)$ for the flip dynamics w.r.t. S with flip parameters given in Eq. (8.3), where T is the first time such that the Hamming distance changes, such that $\alpha = \frac{q - \lambda^* \Delta}{q - \Delta - 2} = \Theta(1)$, $W = 13$ and $\beta \leq \frac{qn}{q - \Delta - 2} \leq O(n)$*

With these tools in hand, we may now finally prove Lemma 8.4.2 and complete the proof of Theorem 8.1.2.

Proof of Lemma 8.4.2. First, note that the path generating set S generates the Hamming metric $d_H(\cdot, \cdot)$ on proper list-colorings. Now, given the variable-length path coupling furnished by Theorem 8.4.4, we use Theorem 8.4.3 to construct an M-step coupling with $M = \left\lceil \frac{2\beta W}{\alpha} \right\rceil \leq O(n)$ which contracts with rate $1 - \alpha$ every M steps, where α is is a constant independent of n. Under this coupling, for every $k = 0, \ldots, M-1$ and every positive integer j, we have that

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

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

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

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

Part II

Fast Algorithms and Optimal Mixing Time Analyses

Chapter 9

Entropy Decay

In this chapter, we lay the foundations for optimal analysis of the down-up walk's mixing time, which is the focus of Part II of this thesis. We previously saw in Chapter 2 how to control the spectral gap of the down-up walk using spectral independence. However, for many high-dimensional discrete distributions arising in applications like statistical physics, the spectral gap by itself does *not* tightly capture the mixing time. As a simple example, it is well-known that the spectral gap of the simple random walk on the discrete hypercube $\{0,1\}^n$ (with edges corresponding to coordinate flips) has spectral gap exactly $1/n$, which leads to an $O(n^2)$ upper bound on the mixing time. On the other hand, a simple coupling argument demonstrates that the same random walk actually mixes in $O(n \log n)$ steps, which is tight via a reduction to the *Coupon Collector Problem*. So, even if we were to obtain the best possible bound on the spectral gap, there will still be a deficiency in the mixing time bounds we obtain.

Roughly speaking, this deficiency lies in the dependence of the mixing time on the stationary probability of the starting state. More specifically, the spectral gap controls the decay in (relative) variance of the current distribution w.r.t. the equilibrium distribution. However, variance is a very loose upper bound on the total variation distance between two probability distributions when they are far apart. For instance, while the total variation distance between two probability distributions is always at most 1, the (relative) variance between a Dirac mass at some state $x \in \Omega$ w.r.t. another distribution μ over Ω is $\frac{1}{\mu(x)} - 1$.

This motivates using better proxies for total variation distance such as (relative) entropy. For example, the (relative) entropy of a Dirac mass at some state $x \in \Omega$ w.r.t. another distribution μ over Ω is now $\log \frac{1}{\mu(x)}$, instead of $\frac{1}{\mu(x)} - 1$. Naturally, this now demands techniques for studying the decay of (relative) entropy over the evolution of a Markov chain. In complete analogy with the spectral gap (or Poincaré constant), this decay turns out to be controlled by fundamental quantities known as the standard and modified logarithmic Sobolev constants, which originated in functional analysis [Gro75; BT03]. Historically, these constants have proven difficult to control, especially in settings lacking product structure or special symmetries [DS81; DS87a; DS96; Sca97; LY98; DH02; ST10; FOW22].

The goal of this chapter is to develop new tools to study the decay of (relative) entropy using spectral independence and related ideas. We will consider both global and local forms of entropy decay, and prove a local-to-global theorem analogous to Theorem 2.3.1 from Chapter 2, where we replace (relative) variance by (relative) entropy. We then show how to deduce the local version of entropy decay using spectral independence. The ideas in this chapter are mostly based on [CLV21a], although some of the ideas were independently discovered in [CGM21; GM20; Ali+21].

9.1 Entropy Decay, Mixing, and Concentration

We begin with some preliminary definitions concerning the decay of (relative) entropy.

(Modified) Logarithmic Sobolev Inequalities Theorem 1.4.19 is useful but it own gives suboptimal mixing time upper bounds. The primary reason for this is that the spectral gap only captures the decay of (relative) variance, which is a loose upper bound on the total variation distance. To remedy this, one can instead study the decay of (relative) entropy, which is a much tighter upper bound on the total variation distance, and often leads to optimal mixing times. The

decay of relative entropy is captured by the following *logarithmic Sobolev constants*, which are reminiscent of Fact 1.4.21 in spirit.

Definition 42 (Standard and Modified Logarithmic Sobolev Constants)**.** *Let* P *be a Markov chain which is reversible w.r.t. a distribution* μ *on a domain* Ω*. We define the* ***modified log-Sobolev constant of*** P *as*

$$\varrho(\mathsf{P}) \stackrel{\mathsf{def}}{=} \inf_{f:\Omega\to\mathbb{R}_{\geq 0}} \left\{ \frac{\mathcal{E}_{\mathsf{P}}(f, \log f)}{\mathrm{Ent}_{\mu}(f)} : \mathrm{Ent}_{\mu}(f) \neq 0 \right\}.$$

We define the ***(standard) log-Sobolev constant of*** P *as*

$$\kappa(\mathsf{P}) \stackrel{\mathsf{def}}{=} \inf_{f:\Omega\to\mathbb{R}_{\geq 0}} \left\{ \frac{\mathcal{E}_{\mathsf{P}}\left(\sqrt{f}, \sqrt{f}\right)}{\mathrm{Ent}_{\mu}(f)} : \mathrm{Ent}_{\mu}(f) \neq 0 \right\}.$$

Inequalities of the form $\mathcal{E}_{\mathsf{P}}(f, \log f) \geq \varrho\, \mathrm{Ent}_{\mu}(f)$ are often called *modified log-Sobolev inequalities*. Similarly, inequalities of the form $\mathcal{E}_{\mathsf{P}}\left(\sqrt{f}, \sqrt{f}\right) \geq \kappa\, \mathrm{Ent}_{\mu}(f)$ are called *(standard) log-Sobolev inequalities*. The standard version was first proposed by Gross [Gro75] in the continuous space, where the two versions are essentially equivalent as observed by [ELL17]; see [Led99; GZ03; MT06] for more comprehensive material on these constants and inequalities.

Like the spectral gap, lower bounds on these constants yield upper bounds on the mixing time.

Theorem 9.1.1 ((Modified) Log-Sobolev Implies Rapid Mixing)**.** *Let* P *be a reversible ergodic Markov chain with stationary distribution* μ *on a domain* Ω*. Then for every* $\epsilon > 0$*, the following inequalities hold*

$$\mathrm{T}_{\mathsf{mix}}(\epsilon) \leq \frac{1}{\varrho(\mathsf{P})}\left(\log\log\frac{1}{\mu_{\min}} + \log\frac{1}{2\epsilon^2}\right) \qquad \text{[BT03]}$$

$$\mathrm{T}_{\mathsf{mix}}(\epsilon) \leq \frac{1}{4\kappa(\mathsf{P})}\left(\log\log\frac{1}{\mu_{\min}} + \log\frac{1}{2\epsilon^2}\right) \qquad \text{[DS96]}$$

where recall that $\mu_{\min} = \min_{x\in\Omega:\mu(x)>0}\mu(x)$.

Besides mixing, these constants turn out to also have incredibly useful consequences for concentration of measure phenomena.

Theorem 9.1.2 ((Modified) Log-Sobolev Implies Concentration; see e.g. [Goe04; Sam05; BLM16])**.** *Let* P *be a reversible ergodic Markov chain with stationary distribution* μ *on a domain* Ω*. Fix an arbitrary function* $f : \Omega \to \mathbb{R}$*, and define the* ***maximum one-step variance of*** f *by*

$$v(f) \stackrel{\mathsf{def}}{=} \max_{x\in\Omega}\left\{\sum_{y\in\Omega}\mathsf{P}(x\to y)\cdot(f(x)-f(y))^2\right\}. \tag{9.1}$$

Then for every $t \geq 0$*, we have the following sub-Gaussian concentration inequalities*

$$\Pr_{x\sim\mu}\left[f(x) \geq \mathbb{E}_{\mu}(f) + t\right] \leq \exp\left(-\frac{\varrho(\mathsf{P})t^2}{2v(f)}\right)$$

$$\Pr_{x\sim\mu}\left[f(x) \geq \mathbb{E}_{\mu}(f) + t\right] \leq \exp\left(-\frac{\kappa(\mathsf{P})t^2}{2v(f)}\right).$$

It is known that $4\kappa(\mathsf{P}) \leq \varrho(\mathsf{P}) \leq 2\gamma(\mathsf{P})$ [BT03], and so lower bounds on the standard and modified log-Sobolev constants are harder to obtain than lower bounds on the spectral gap. Historically, the standard and modified log-Sobolev constants are notoriously difficult to lower bound, especially in the absence of product structure or special symmetries [DS81; DS87a; DS96; Sca97; LY98; DH02; ST10; FOW22]. We will develop new techniques in this thesis which break these old barriers, and establish lower bounds on these constants in a variety of challenging settings.

We close this section with a brief remark on the standard log-Sobolev constant, which was previously observed in [HS20]. Unlike the modified log-Sobolev constant, the standard log-Sobolev

constant is sensitive to smallest probability of a state in the support of μ. In particular, if we take $f = \mathbb{I}_x$ for some state $x \in \Omega$, we see that

$$\mathcal{E}_{\mathsf{P}}(\mathbb{I}_x, \mathbb{I}_x) = \mu(x) \cdot (1 - \mathsf{P}(x \to x))$$
$$\mathrm{Ent}_\mu(\mathbb{I}_x) = \mu(x) \log \frac{1}{\mu(x)}$$

so that

$$\kappa(\mathsf{P}) \leq \min_{x \in \mathrm{supp}(\mu)} \frac{\mathcal{E}_{\mathsf{P}}(\mathbb{I}_x, \mathbb{I}_x)}{\mathrm{Ent}_\mu(\mathbb{I}_x)} = \min_{x \in \mathrm{supp}(\mu)} \frac{1 - \mathsf{P}(x \to x)}{\log \frac{1}{\mu(x)}} = \frac{1}{\log \frac{1}{\mu_{\min}}}.$$

Many of the distributions we have seen already have potentially arbitrarily small $\mu_{\min}$ (e.g. determinantal point processes, or Strongly Rayleigh and discrete log-concave distributions more broadly). Hence, for such distributions, we cannot hope to lower bound the standard log-Sobolev constant.

However, for many other distributions (e.g. Gibbs distributions of spin systems on bounded-degree graphs), $\mu_{\min}$ can be lower bounded by an exponentially small constant, which is in general tolerable. Hence, towards developing tools which not only control the modified log-Sobolev constant but the standard log-Sobolev constant as well, we introduce the following definition.

Definition 43 (Marginal Boundedness; [CLV21a])**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for a finite ground set $\mathscr{U}$ and integer $0 \leq n \leq |\mathscr{U}|$. For $0 < \mathfrak{b}_0, \ldots, \mathfrak{b}_{n-1} \leq 1$, we say μ is $(\mathfrak{b}_0, \ldots, \mathfrak{b}_{n-1})$-**marginally bounded** if for every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k \leq n-1$ and every $u \in \mathrm{supp}(\mu_1^\sigma)$, we have the lower bound $\mu_1^\sigma(u) \geq \frac{\mathfrak{b}_k}{n-k}$ on the marginal probability of u conditioned on σ. For brevity, we say μ is $\mathfrak{b}$-marginally bounded for some $0 < \mathfrak{b} \leq 1$ if μ is $(\mathfrak{b}_0, \ldots, \mathfrak{b}_{n-1})$-marginally bounded where $\mathfrak{b}_i = \mathfrak{b}$ for every $i = 0, \ldots, n-1$.*

Remark 45. This deviates slightly from the original definition given in [CLV21a], where the $\frac{1}{n-k}$ normalization wasn't included. Note that $\mathfrak{b} \leq 1$ follows automatically from Lemma 2.1.1 (or Eq. (2.5) more generally).

Note that $\mathfrak{b}$-marginal boundedness immediately implies that $\mu_{\min} \geq \mathfrak{b}^n$, which again is tolerable in general. We note that subsequent work on analyzing the mixing time of the Glauber dynamics for spin systems has shown that weaker control on the marginals is sufficient for obtaining optimal mixing times [Che+22a; Ana+21b; CE22].

9.2 Local-to-Global Entropy Contraction

Our goal here is to prove an analog of Theorem 2.3.1 for analyzing the rate of entropy decay. However, we deviate from Chapter 2 in that we aim for a greater level of generality. For instance, our analysis will extend to other high-order random walks such as $\mathcal{D}_\mu^{n \searrow k} \mathcal{U}_\mu^{k \nearrow n}$. This will be vital to establishing stronger mixing time bounds on the down-up walk. It will also give us new ways to bound the standard and modified log-Sobolev constants. The proof of our local-to-global entropy contraction theorem in this section essentially mirrors the second proof of Theorem 2.3.1 given in Section 2.4.2, except we replace variance by entropy everywhere. We emphasize that almost everything we say in this section can be extended to general *φ-entropies* (or *φ-divergences*) (see e.g. Section 1.4.3), including variance and entropy, in the obvious manner.

We begin by defining global entropy contraction in a similar way to Proposition 2.4.3 but for more general high-dimensional walks. Similar to our notation for variance, we will write $\mathrm{Ent}_k(\cdot)$ for $\mathrm{Ent}_{\mu_k}(\cdot)$ and $\mathrm{Ent}_k^\sigma(\cdot)$ for $\mathrm{Ent}_{\mu_k^\sigma}(\cdot)$.

Definition 44 (Global Entropy Contraction)**.** *For $0 \leq k \leq n$ and $0 \leq \beta_k \leq 1$, we say μ satisfies k-**step global entropy contraction with constant** β_k if for every global function $f_n = \mathrm{supp}(\mu) \to \mathbb{R}_{\geq 0}$, the induced projection $f_{n-k} = \mathcal{U}_\mu^{n-k \nearrow n} f_n : \mathrm{supp}(\mu_{n-k}) \to \mathbb{R}_{\geq 0}$ satisfies*

$$\mathrm{Ent}_{n-k}(f_{n-k}) \leq \beta_k \cdot \mathrm{Ent}_n(f_n)$$

Remark 46. In Chapter 2, more specifically Proposition 2.4.3, we only defined (1-step) global variance contraction since that was all we needed. Naturally, one can also define k-step global variance contraction paralleling the above definition.

We have the following direct analog of Proposition 2.4.3, which shows the connection between k-step global entropy contraction and fundamental functional analytic constants such as the standard and modified log-Sobolev constants.

Proposition 9.2.1 (Implications of Entropy Contraction)**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. Fix $0 \le k \le n$, and assume μ satisfies k-step global entropy contraction with constant $0 \le \beta_k \le 1$. Then the following hold.*

1. ***(Modified) Log-Sobolev Inequality:*** *The $n-k \leftrightarrow n$ down-up walk $\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$ has modified log-Sobolev constant lower bounded as*
$$\varrho\left(\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}\right) \ge 1 - \beta_k.$$
2. ***(Standard) Log-Sobolev Inequality:*** *If μ is $\mathfrak{b}$-marginally bounded for some $0 < b \le \frac{1}{2^{1/k}}$, then there exists a constant $C_{\mathfrak{b},k}$ depending only on $\mathfrak{b}$ and k such that the $n-k \leftrightarrow n$ down-up walk $\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$ has standard log-Sobolev constant lower bounded as*
$$\kappa\left(\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}\right) \ge (1-\beta_k) \cdot C_{\mathfrak{b},k}.$$
3. ***(Relative) Entropy Decay:*** *For every probability distribution ν on $\operatorname{supp}(\mu)$, we have that*
$$\mathscr{D}_{\mathrm{KL}}\left(\nu \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n \nearrow n-k} \| \mu\right) \le \beta_k \cdot \mathscr{D}_{\mathrm{KL}}(\nu \| \mu).$$
4. ***Mixing:*** *The $(n-k) \leftrightarrow n$ down-up walk $\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$ mixes in $O\left(C_k \log\log \frac{1}{\mu_{\min}}\right)$-steps.*
5. ***Concentration:*** *For every function $f : \operatorname{supp}(\mu) \to \mathbb{R}$ and every $t \ge 0$, we have the sub-Gaussian concentration inequality*
$$\Pr_{\sigma \sim \mu}\left[f(\sigma) \ge \mathbb{E}_\mu(f) + t\right] \le \exp\left(-\frac{(1-\beta_k)t^2}{2v(f)}\right),$$
where recall $v(f)$ is the maximum one-step variance of f w.r.t. the graph metric induced by $\mathcal{D}_\mu^{n \searrow k} \mathcal{U}_\mu^{k \nearrow n}$ (see Eq. (9.1)).

We defer the proof to Section 9.4 below, as the ideas are standard, and are not the main focus of this chapter.

Now that we have defined global entropy contraction, we now define local entropy contraction in complete analogy to Definition 16.

Definition 45 (Local Entropy Contraction; [CLV21a])**.** *For $0 \le \alpha \le 1$, we say μ satisfies α-**local entropy contraction** if for every global function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}_{\ge 0}$, we have that the induced projections f_1, f_2 satisfy*
$$\operatorname{Ent}_1(f_1) \le \alpha \cdot \operatorname{Ent}_2(f_2).$$
*Similarly, for $0 \le \alpha_0, \dots, \alpha_{n-2} \le 1$, we say μ satisfies $(\alpha_0, \dots, \alpha_{n-2})$-**local entropy contraction** if for every global function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}_{\ge 0}$, every $0 \le k \le n-2$, and every $\sigma \in \operatorname{supp}(\mu_k)$, the induced local functions f_1^σ, f_2^σ satisfy*
$$\operatorname{Ent}_1^\sigma(f_1^\sigma) \le \alpha_k \cdot \operatorname{Ent}_2^\sigma(f_2^\sigma).$$

Remark 47. In a similar manner, for each $0 \le k \le \ell \le n$, one can define $\ell \searrow k$ entropy contraction as an inequality of the form
$$\operatorname{Ent}_k(f_k) \le \beta_{\ell \searrow k} \cdot \operatorname{Ent}_\ell(f_\ell).$$
Local entropy contraction would then be equivalent to $2 \searrow 1$ entropy contraction, and k-step global entropy contraction would be equivalent to $n \searrow n-k$ entropy contraction. Such an inequality corresponds to analyzing mixing and entropy decay under the Markov chain $\mathcal{D}_\mu^{\ell \searrow k} \mathcal{U}_\mu^{k \nearrow \ell}$, although one would need the inequality for all level-ℓ functions $f_\ell : \operatorname{supp}(\mu_\ell) \to \mathbb{R}_{\ge 0}$, not just those induced by a global level-n function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}_{\ge 0}$.

Again, our goal in this section is to reduce global entropy contraction to local entropy contraction, in complete analogy with the local-to-global theorem(s) we saw in Chapter 2 (more precisely, Theorem 2.4.5).

Theorem 9.2.2 (Local-to-Global Entropy Contraction; [CLV21a])**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. If μ satisfies $(\alpha_0, \dots, \alpha_{n-2})$-local entropy contraction, then for every $0 \le k \le \ell \le n$ and every global function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}_{\ge 0}$, the induced projections $f_k = \mathcal{U}^{k \nearrow n} f_n : \operatorname{supp}(\mu_k) \to \mathbb{R}_{\ge 0}$ and $f_\ell = \mathcal{U}^{\ell \nearrow n} f_n : \operatorname{supp}(\mu_\ell) \to \mathbb{R}_{\ge 0}$ satisfy the inequality*

$$\frac{\operatorname{Ent}_k(f_k)}{\sum_{j=0}^{k-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)} \le \frac{\operatorname{Ent}_\ell(f_\ell)}{\sum_{j=0}^{\ell-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}. \tag{9.2}$$

In particular, for every $0 \le k \le n$, the distribution μ satisfies k-step global entropy contraction with constant

$$\beta_k \le \frac{\sum_{j=0}^{n-k-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}{\sum_{j=0}^{n-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}.$$

We follow the proof of Theorem 2.4.5. Much like Lemma 2.4.2, we have the following entropy decomposition lemma, which is essentially the Law of Total Entropy.

Lemma 9.2.3 (Entropy Decomposition)**.** *Fix a finite ground set $\mathscr{U}$ and an integer $0 \le n \le |\mathscr{U}|$, and let μ be a probability distribution over $\binom{\mathscr{U}}{n}$. Then for every $0 \le k \le \ell \le n$ and every function $f_\ell : \operatorname{supp}(\mu_\ell) \to \mathbb{R}$, we have that*

$$\operatorname{Ent}_\ell(f_\ell) = \operatorname{Ent}_k(f_k) + \mathbb{E}_{\sigma \sim \mu_k}\left[\operatorname{Ent}^\sigma_{\ell-k}\left(f^\sigma_{\ell-k}\right)\right]$$

where f_k and $f^\sigma_{\ell-k}$ are local functions induced by f_ℓ.

Proof. Using the Law of Total Expectation and the fact that $f_\ell(\tau) = f^\sigma_{\ell-k}(\tau \setminus \sigma)$ whenever $\tau \supseteq \sigma$, we have that

$$\begin{aligned}
&\mathbb{E}_{\tau \sim \mu_\ell}\left[f_\ell(\tau) \log f_\ell(\tau)\right] \\
&= \mathbb{E}_{\sigma \sim \mu_k}\left[\mathbb{E}_{\tau \sim \mu_\ell}\left[f_\ell(\tau) \log f_\ell(\tau) \mid \tau \supseteq \sigma\right]\right] \\
&= \mathbb{E}_{\sigma \sim \mu_k}\left[\mathbb{E}_{\sigma' \sim \mu^\sigma_{\ell-k}}\left[f^\sigma_{\ell-k}(\sigma')\right] \log \mathbb{E}_{\sigma' \sim \mu^\sigma_{\ell-k}}\left[f^\sigma_{\ell-k}(\sigma')\right]\right] + \mathbb{E}_{\sigma \sim \mu_k}\left[\operatorname{Ent}^\sigma_{\ell-k}\left(f^\sigma_{\ell-k}\right)\right] \\
&= \mathbb{E}_{\sigma \sim \mu_k}\left[f_k(\sigma) \log f_k(\sigma)\right] + \mathbb{E}_{\sigma \sim \mu_k}\left[\operatorname{Ent}^\sigma_{\ell-k}\left(f^\sigma_{\ell-k}\right)\right].
\end{aligned}$$

Since $\mathbb{E}_{\mu_\ell}[f_\ell] = \mathbb{E}_{\mu_k}[f_k]$, subtracting $x \mapsto x \log x$ applied to each expectation from both sides gives the desired result. □

Proof of Theorem 9.2.2. It suffices to show that for every $0 \le k < n$, we have the inequality

$$\frac{\operatorname{Ent}_k(f_k)}{\sum_{j=0}^{k-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)} \le \frac{\operatorname{Ent}_{k+1}(f_{k+1})}{\sum_{j=0}^{k} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)} \tag{9.3}$$

since the general case Eq. (9.2) follows by chaining these together. We prove Eq. (9.3) by induction. The case where $k = 0$ is trivial since the left-hand side is 0, and the first nontrivial base case $k = 1$ follows immediately by the definition of local entropy contraction. For the inductive step, we have that

$$\begin{aligned}
\operatorname{Ent}_{k+1}(f_{k+1}) - \operatorname{Ent}_{k-1}(f_{k-1}) &= \mathbb{E}_{\sigma \sim \mu_{k-1}}\left[\operatorname{Ent}^\sigma_2\left(f^\sigma_2\right)\right] && \text{(Lemma 9.2.3)} \\
&\ge \frac{1}{\alpha_{k-1}} \cdot \mathbb{E}_{\sigma \sim \mu_{k-1}}\left[\operatorname{Ent}^\sigma_1\left(f^\sigma_1\right)\right] && \text{(Definition 45)} \\
&= \frac{1}{\alpha_{k-1}} \cdot \left(\operatorname{Ent}_k(f_k) - \operatorname{Ent}_{k-1}(f_{k-1})\right) && \text{(Lemma 9.2.3)}
\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{Ent}_{k+1}(f_{k+1}) &\geq \frac{1}{\alpha_{k-1}} \cdot \mathrm{Ent}_k(f_k) - \left(\frac{1}{\alpha_{k-1}} - 1\right) \cdot \mathrm{Ent}_{k-1}(f_{k-1}) \\ &\geq \left(\frac{1}{\alpha_{k-1}} - \left(\frac{1}{\alpha_{k-1}} - 1\right) \cdot \frac{\sum_{j=0}^{k-2} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}{\sum_{j=0}^{k-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}\right) \cdot \mathrm{Ent}_k(f_k) \\ &\qquad \text{(Inductive Hypothesis)} \\ &= \frac{\sum_{j=0}^{k} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)}{\sum_{j=0}^{k-1} \prod_{i=0}^{j-1} \left(\frac{1}{\alpha_i} - 1\right)} \cdot \mathrm{Ent}_k(f_k)\end{aligned}$$

as desired. □

9.3 Local Entropy Contraction via Local Spectral Expansion

Theorem 9.2.2 shows us how to reduce k-step global entropy contraction to local entropy contraction. Our goal in this section is to show how to deduce bounds on the local entropy contraction constants using spectral independence, since we abundant tools for establishing spectral independence. This is akin to Lemma 2.4.4, where we showed that spectral independence implies local variance contraction. However, the implication for local entropy contraction is much less clean, since spectral gaps are inherently related to variance rather than entropy. For the rest of this chapter, it will be convenient to instead formulate everything using local spectral expansion, and then appeal to Lemma 2.3.3 to convert between spectral independence and local spectral expansion.

The crucial technical lemma we will need is the following, which relates the "local entropies" $\mathrm{Ent}_2(f_2)$ and $\mathrm{Ent}_1(f_1)$, which appear in the definition of local entropy contraction, to the spectral structure of the local walk $\mathcal{Q}_\mu$. The case when $\lambda_2(\mathcal{Q}_\mu) = 0$ was first proved in [CGM21]. We generalize this lemma in a straightforward manner.

Lemma 9.3.1 ([CLV21a])**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ where $\mathscr{U}$ is a finite ground set and $0 \leq n \leq |\mathscr{U}|$ is an integer. Let $f_2 : \mathrm{supp}(\mu_2) \to \mathbb{R}$ be an arbitrary function, and let $f_1 = \mathcal{U}_\mu^{1\nearrow 2} f_2 : \mathrm{supp}(\mu_1) \to \mathbb{R}$ be the level-1 projection of f_2. Then we have the inequality*

$$\mathrm{Ent}_2(f_2) - 2 \cdot \mathrm{Ent}_1(f_1) \geq -\lambda_2(\mathcal{Q}_\mu) \cdot \frac{\mathrm{Var}_1(f_1)}{\mathbb{E}_{\mu_1}(f_1)}.$$

In particular, if $\lambda_2(\mathcal{Q}_\mu) = 0$, then $\mathrm{Ent}_1(f_1) \leq \frac{1}{2} \cdot \mathrm{Ent}_2(f_2)$.

Proof. First, observe that the desired inequality is scale invariant, and hence we may assume without loss that $\mathbb{E}_{\mu_2}(f_2) = \mathbb{E}_{\mu_1}(f_1) = 1$. For convenience and simplicity, we shall write ij to represent $\{i, j\} \in \mathcal{X}_2 = \mathrm{supp}(\mu_2)$. Let us rewrite $2 \cdot \mathrm{Ent}_1(f_1)$ in a form which is more convenient to compare with $\mathrm{Ent}_2(f_2)$. Observe that

$$\begin{aligned}\mathrm{Ent}_1(f_1) &= \sum_{i \in \mathscr{U}} \mu_1(i) f_1(i) \log f_1(i) \\ &= \sum_{i \in \mathscr{U}} \mu_1(i) \left(\sum_{j \in \mathscr{U} : ij \in \mathcal{X}_2} \frac{\mu_2(ij)}{2\mu_1(i)} \cdot f_2(ij)\right) \log f_1(i) \\ &= \sum_{i \in \mathscr{U}} \sum_{j \in \mathscr{U} : ij \in \mathcal{X}_2} \frac{\mu_2(ij)}{2} \cdot f_2(ij) \log f_1(i) \\ &= \frac{1}{2} \sum_{ij \in \mathcal{X}_2} \mu_2(ij) \cdot f_2(ij) \log\big(f_1(i) f_1(j)\big).\end{aligned}$$

Using the inequality $a \log \frac{a}{b} \geq a - b$, which holds for every $a \geq 0$ and $b > 0$, see that

$$\begin{aligned}
\mathrm{Ent}_2(f_2) - 2 \cdot \mathrm{Ent}_1(f_1) &= \sum_{ij \in \mathcal{X}_2} \mu_2(ij) \cdot f_2(ij) \left(\log f_2(ij) - \log\left(f_1(i) f_1(j)\right)\right) \\
&\geq \sum_{ij \in \mathcal{X}_2} \mu_2(ij) \cdot \left(f_2(ij) - f_1(i) f_1(j)\right) \\
&= \sum_{ij \in \mathcal{X}_2} \mu_2(ij) f_2(ij) - \sum_{ij \in \mathcal{X}_2} \mu_2(ij) f_1(i) f_1(j) \\
&= 1 - \left\langle f_1, \mathcal{Q}_\mu f_1 \right\rangle_{\mu_1} && \left(\mathbb{E}_{\mu_2}(f_2) = 1 \text{ and Fact 2.3.2}\right) \\
&= \mathcal{E}_{\mathcal{Q}_\mu}(f_1, f_1) - \mathrm{Var}_1(f_1) && \left(\text{Adding and subtracting } \left\langle f_1, f_1 \right\rangle_{\mu_1}\right) \\
&\geq -\lambda_2(\mathcal{Q}_\mu) \cdot \mathrm{Var}_1(f_1). && \left(\text{Poincaré Inequality for } \mathcal{Q}_\mu\right)
\end{aligned}$$

and we are done. □

Now that we have this lemma, we show how this can be used to deduce local entropy contraction bounds given spectral independence bounds. We begin with the case of 0-spectral independence, which leads to simple and clean local entropy contraction bounds. This is result was proved in [CGM21], but we include it here for the sake of completeness.

Theorem 9.3.2 (Local Entropy Contraction for 0-Spectrally Independent Distributions; [CGM21])**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and integer $0 \leq n \leq |\mathscr{U}|$. Assume μ is $(0, \ldots, 0)$-spectrally independent, or equivalently, assume μ is a (one-sided) $(0, \ldots, 0)$-local spectral expander. Then μ satisfies $\left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$-local entropy contraction.*

Proof. This follows immediately by applying Lemma 9.3.1 to the conditional distribution μ^σ of μ for each feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| \leq n - 2$. □

Corollary 9.3.3 (Near-Optimal Mixing for Discrete Log-Concave Distributions)**.** *Let μ be a discrete log-concave distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and integer $0 \leq n \leq |\mathscr{U}|$. Then for every distribution ν on $\binom{\mathscr{U}}{n}$, we have the inequality*

$$\mathscr{D}_{\mathrm{KL}}\left(\nu \mathcal{P}_\mu \| \mu \mathcal{P}_\mu\right) \leq \mathscr{D}_{\mathrm{KL}}\left(\nu \mathcal{D}_\mu^{n \searrow n-1} \| \mu \mathcal{D}_\mu^{n \searrow n-1}\right) \leq \left(1 - \frac{1}{n}\right) \cdot \mathscr{D}_{\mathrm{KL}}(\nu \| \mu).$$

In particular, the down-up walk $\mathcal{P}_\mu$ has modified log-Sobolev constant $\varrho(\mathcal{P}_\mu) \geq 1/n$ and mixing time

$$\mathrm{T}_{\mathsf{mix}}(\epsilon; \mathcal{P}_\mu) \leq n \left(\log \log \frac{1}{\mu_{\min}} + \log \frac{1}{2\epsilon^2}\right).$$

Corollary 9.3.4 (Near-Optimal Mixing for Bases Exchange Walk on Matroid Bases)**.** *Let $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ be an rank-r matroid. Then the bases exchange walk $\mathcal{P}_\mu$ for sampling from the uniform distribution over bases of $\mathcal{M}$ mixes in $O\left(r\left(\log r + \log \log |\mathscr{U}|\right)\right)$-steps.*

Remark 48. As mentioned previously, since $\mu_{\min}$ can be arbitrarily small for general discrete log-concave distributions, we cannot hope for such as strong lower bound on the standard log-Sobolev constant. However, in Chapter 11, we will see how to further improve the mixing time for these Markov chains. For instance, for the bases exchange walk, we will be able to completely remove the dependence on the size of the ground set $|\mathscr{U}|$.

These results completely settle the case of 0-spectrally independent distributions. However, many of the distributions we've encountered in this thesis are not 0-spectrally independent, and so we need to dig deeper. The rest of this chapter is devoted to showing how η-spectral independence for positive η, combined with marginal boundedness in the sense of Definition 43, can still be used to deduce bounds on local entropy contraction.

9.3.1 Local Entropy Contraction via Spectral Independence and Marginal Boundedness

The key challenge in using Lemma 9.3.1 when $\lambda_2(\mathcal{Q}_\mu)$ is positive is showing that $\frac{\mathrm{Var}_1(f_1)}{\mathbb{E}_{\mu_1}(f_1)}$ is comparable to $\mathrm{Ent}_1(f_1)$. We overcome this obstacle in this subsection, and prove the following theorem.

Theorem 9.3.5. *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for some finite ground set $\mathscr{U}$ and integer $0 \le n \le |\mathscr{U}|$. Assume μ is a (one-sided) $(\lambda_0, \dots, \lambda_{n-2})$-local spectral expander which is also $(\mathfrak{b}_0, \dots, \mathfrak{b}_{n-1})$-marginally bounded. Then μ satisfies $(\alpha_0, \dots, \alpha_{n-2})$-local entropy contraction with*

$$\alpha_k = \max\left\{\frac{1}{2}\cdot\frac{1}{1-\frac{2\lambda_k}{\mathfrak{b}_k^2}}, 1 - \frac{1-\lambda_k}{4+2\log\left(\frac{(n-k)(n-k-1)}{2\mathfrak{b}_k\mathfrak{b}_{k+1}}\right)}\right\}, \quad \forall 0 \le k \le n-2. \tag{9.4}$$

The rest of this subsection aims to prove Theorem 9.3.5. We will show separately the two bounds in Eq. (9.4) on the rate of local entropy contraction, and we will refer to them as the *first* and *second* bounds, respectively. The first bound is the most interesting one; it is more subtle and indicates that $\alpha_k = \frac{1}{2} + \Theta(\lambda_k)$ as long as μ is $\mathfrak{b}$-marginally bounded with $\mathfrak{b} = \Theta(1)$. The second bound is crude but may still be helpful when the first bound is vacuous. We will first prove the first bound. Afterwards, we will appeal to existing results in the literature to prove the second bound.

Proof of the First Bound

Our strategy is to use marginal boundedness to show that $\frac{\mathrm{Var}_1(f_1)}{\mathbb{E}_{\mu_1}(f_1)}$ is comparable to $\mathrm{Ent}_1(f_1)$. This will then allow us to again use Lemma 9.3.1. We begin with the following lemma, which shows that for marginally bounded distributions, the local functions f_1^σ induced by a global function $f_n : \mathrm{supp}(\mu_n) \to \mathbb{R}_{\ge 0}$ are "balanced" in the sense that the values of f_1^σ cannot be too large compared to its expectation under μ_1^σ.

Lemma 9.3.6. *If μ is $(\mathfrak{b}_0, \dots, \mathfrak{b}_{n-1})$-marginally bounded, then for every global function $f_n : \mathcal{X}_n \to \mathbb{R}_{\ge 0}$, every feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k \le n-1$ and $f_k(\sigma) > 0$, and every $i \in \mathcal{X}_1^\sigma$, we have the inequality*

$$f_1^\sigma(i) \le \frac{1}{\mathfrak{b}_k}\cdot\mathbb{E}_{\mu_1^\sigma}(f_1^\sigma). \tag{9.5}$$

Next, we show that for such "balanced" functions, the entropy and variance differ only by a constant factor (after appropriate normalization).

Lemma 9.3.7. *Let π be a distribution over a finite set Ω, and let $f : \Omega \to \mathbb{R}_{\ge 0}$ such that $\mathbb{E}_\pi(f) > 0$. If there exists a constant $c \ge 1$ such that $f(x) \le c\cdot\pi(f)$ for all $x \in \Omega$, then*

$$\frac{\mathrm{Var}_\pi(f)}{\mathbb{E}_\pi(f)} \le 4c^2\cdot\mathrm{Ent}_\pi(f). \tag{9.6}$$

Note we always have the inequality $\mathrm{Ent}_\pi(f) \le \frac{\mathrm{Var}_\pi(f)}{\mathbb{E}_\pi(f)}$ (see e.g. [Sal97]) and so the above shows that for "balanced" functions, the left-hand and right-hand sides are the same up to constant factors. We then show how to use these lemmas to prove the first bound in Eq. (9.4).

Proof of the first bound in Eq. (9.4). Fix an arbitrary feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k \le n-2$, and let $f_n : \mathrm{supp}(\mu) \to \mathbb{R}_{\ge 0}$ be an arbitrary global function. Then applying, Lemma 9.3.1 to the condition distribution μ^σ and using $\lambda_2(\mathcal{Q}_{\mu^\sigma}) \le \lambda_k$, we have the inequality

$$\mathrm{Ent}_2^\sigma(f_2^\sigma) - 2\cdot\mathrm{Ent}_1^\sigma(f_1^\sigma) \ge -\lambda_k\cdot\frac{\mathrm{Var}_1^\sigma(f_1^\sigma)}{\mathbb{E}_{\mu_1^\sigma}(f_1^\sigma)} \tag{9.7}$$

By Lemma 9.3.6, we have $f_1^\sigma(i) \le \frac{1}{\mathfrak{b}_k}\cdot\mathbb{E}_{\mu_1^\sigma}(f_1^\sigma)$ for all $i \in \mathcal{X}_1^\sigma$. It follows by Lemma 9.3.7 that

$$\frac{\mathrm{Var}_1^\sigma(f_1^\sigma)}{\mathbb{E}_{\mu_1^\sigma}(f_1^\sigma)} \le \frac{4}{\mathfrak{b}_k^2}\cdot\mathrm{Ent}_1^\sigma(f_1^\sigma) \tag{9.8}$$

It follows that

$$\mathrm{Ent}_1(f_1) \le \frac{1}{2}\cdot\frac{1}{1-\frac{2\lambda_k}{\mathfrak{b}_k^2}}\cdot\mathrm{Ent}_2(f_2)$$

as desired. □

It remains to prove Lemmas 9.3.6 and 9.3.7. We note that these lemmas are logically independent of each other.

Proof of Lemma 9.3.6. Without loss of generality, we may assume $\mathbb{E}_\mu(f_n) = 1$, i.e. f_n is the density of some other probability distribution ν w.r.t. μ. Using Fact 2.4.1, we see that

$$\frac{f_1^\sigma(i)}{\mathbb{E}_{\mu_1^\sigma}(f_1^\sigma)} = \frac{f_{k+1}(\sigma+i)}{f_k(\sigma)} = \frac{\Pr_{\tau\sim\nu}[\tau \ni i \mid \tau \supseteq \sigma]}{\Pr_{\tau\sim\mu}[\tau \ni i \mid \tau \supseteq \sigma]} = \frac{\nu_1^\sigma(i)}{\mu_1^\sigma(i)}.$$

We have $\nu_1^\sigma(i) \le \frac{1}{n-k}$ using Lemma 2.1.1 (or Eq. (2.5) more generally). Furthermore, by assumption, we have $\mu_1^\sigma(i) \ge \frac{\mathtt{b}_k}{n-k}$. The claim follows. □

Proof of Lemma 9.3.7. Since the inequality is scale invariant, we may assume without loss of generality that $\mathbb{E}_\pi(f) = 1$. Then, f is a c-bounded function, i.e. $f(x) \le c$ for all $x \in \Omega$. Furthermore, f is the relative density of some other distribution ν with respect to π. Under this normalization, our goal is to show that $\mathrm{Var}_\pi(f) \le 4c^2 \cdot \mathrm{Ent}_\pi(f)$.

By the Donsker-Varadhan variational representation of entropy (see Proposition 1.4.11), for every function $g : \Omega \to \mathbb{R}$, we have

$$\mathbb{E}_\nu(g) \le \mathscr{D}_{\mathrm{KL}}(\nu\|\pi) + \log \mathbb{E}_\pi(e^g) = \mathrm{Ent}_\pi(f) + \log \mathbb{E}_\pi(e^g).$$

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

$$\mathbb{E}_\nu(g) = t \cdot (\mathbb{E}_\nu(f) - 1) = t \cdot \left(\mathbb{E}_\pi\left(f^2\right) - 1\right) = t \cdot \mathrm{Var}_\pi(f).$$

Hence, we obtain that

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

This is known as the *entropy inequality* [MSW03].

Notice that $c \ge 1$ always and $\mathrm{Ent}_\pi(f) \le \mathrm{Var}_\pi(f)$ when $\mathbb{E}_\pi(f) = 1$ (see e.g. [Sal97]). Consider first the case when $1 \le c \le 2$. We shall pick

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

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

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

It follows that

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

With our choice of t, we obtain

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

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

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

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

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

We then deduce that

$$\mathrm{Var}_\pi(f) \le \frac{1}{t} \cdot \mathrm{Ent}_\pi(f) + t\left(\frac{c}{2\log c}\right)^2 \cdot \mathrm{Var}_\pi(f)$$

With our choice of t, we obtain

$$\mathrm{Var}_\pi(f) \le \left(\frac{c}{\log c}\right)^2 \mathrm{Ent}_\pi(f) \le 4c^2 \cdot \mathrm{Ent}_\pi(f).$$

This establishes the lemma. □

Proof of the Second Bound

Here we prove the second bound in Eq. (9.4). We do this by reducing entropy contraction to bounding the standard log-Sobolev constant for the Markov chain $\mathcal{D}_\mu^{2\searrow 1}\mathcal{U}_\mu^{1\nearrow 2}$, which recall is intimately related to $\mathcal{Q}_\mu$ (see Definition 13). Since we assume μ is marginally bounded, a comparison inequality between the standard log-Sobolev constant and spectral gap then finishes the proof. Throughout this section, we will consider the special case $\sigma = \emptyset$ for simplicity, since one can extend the analysis mutatis mutandis to all feasible $\sigma \subseteq \mathscr{U}$ by considering the conditional distribution μ^σ.

Lemma 9.3.8 (Entropy Contraction via Standard Log-Sobolev)**.** *For every (local) function $f_2 : \mathrm{supp}(\mu_2) \to \mathbb{R}_{\ge 0}$, we have the inequality*

$$\mathrm{Ent}_1(f_1) \le \left(1 - \kappa\left(\mathcal{D}_\mu^{2\searrow 1}\mathcal{U}_\mu^{1\nearrow 2}\right)\right) \cdot \mathrm{Ent}_2(f_2).$$

In particular, we have the upper bound $\alpha_0 \le 1 - \kappa\left(\mathcal{D}_\mu^{2\searrow 1}\mathcal{U}_\mu^{1\nearrow 2}\right)$ on the local entropy contraction constant.

Remark 49. This lemma generalizes significantly. For instance, the same proof demonstrates that a lower bound of κ on the standard log-Sobolev constant of $\mathcal{D}_\mu^{n\searrow n-k}\mathcal{U}_\mu^{n-k\nearrow n}$ implies k-step global entropy contraction with constant $1-\kappa$ without assuming marginal boundedness. For context, recall that in Proposition 9.2.1, we showed the converse, namely that under additional marginal boundedness assumptions, k-step global entropy contraction implies lower bounds on the standard log-Sobolev constant of $\mathcal{D}_\mu^{n\searrow n-k}\mathcal{U}_\mu^{n-k\nearrow n}$.

The following lemma gives the final piece, which compares the standard log-Sobolev constant with the spectral gap.

Lemma 9.3.9 (Equations (3.9) and (3.10) from [DS96])**.** *For every reversible Markov chain P with stationary distribution μ on a finite state space Ω, we have the inequality*

$$\kappa(\mathsf{P}) \ge \frac{(1-2\mu_{\min})\cdot\gamma(\mathsf{P})}{\log\left(\frac{1}{\mu_{\min}}-1\right)} \ge \frac{\gamma(\mathsf{P})}{2+\log\left(\frac{1}{\mu_{\min}}\right)}$$

where recall $\mu_{\min} = \min_{x\in\mathrm{supp}(\mu)}\mu(x)$.

We are now ready to prove the second bound of Eq. (9.4).

Proof of the second bound in Eq. (9.4). Combining Lemmas 9.3.8 and 9.3.9, we see that

$$\mathrm{Ent}_1(f_1) \le \left(1-\kappa\left(\mathcal{D}_\mu^{2\searrow 1}\mathcal{U}_\mu^{1\nearrow 2}\right)\right)\cdot\mathrm{Ent}_2(f_2) \le \left(1 - \frac{\gamma\left(\mathcal{D}_\mu^{2\searrow 1}\mathcal{U}_\mu^{1\nearrow 2}\right)}{2+\log(1/\mu_2^*)}\right)\cdot\mathrm{Ent}_2(f_2)$$

where μ_2^* is the minimum probability of any state of μ_2. To finish the proof, observe that by Definition 13,

$$\gamma\left(\mathcal{D}_\mu^{2\searrow 1}\mathcal{U}_\mu^{1\nearrow 2}\right) = \frac{1}{2}\cdot\gamma(\mathcal{Q}_\mu) = \frac{1}{2}\cdot(1-\lambda_2(\mathcal{Q}_\mu)) \ge \frac{1}{2}\cdot(1-\lambda_0)$$

while for every $ij = \{i,j\} \in \mathrm{supp}(\mu_2)$, we have

$$\mu_2(ij) = 2\cdot\mu_1(i)\cdot\mu_1^i(j) \ge \frac{2\mathfrak{b}_0\mathfrak{b}_1}{n(n-1)}.$$

In particular, $\mu_2^* \geq \frac{2\mathfrak{b}_0\mathfrak{b}_1}{n(n-1)}$. Combining these bounds yields the claim for α_0. Applying a nearly identical analysis to the conditional distribution μ^σ for each feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = k \leq n-2$ finishes the proof. □

We refer the interested reader to [DS96] for the proof of Lemma 9.3.9. All that remains is to prove Lemma 9.3.8, which we do now. We follow the proof of [Mic97, Proposition 6], using the following technical lemma.

Lemma 9.3.10 ([Mic97, 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 9.3.8. For convenience, we write $\mathcal{U} = \mathcal{U}_\mu^{1\nearrow 2}$, $\mathcal{D} = \mathcal{D}_\mu^{2\searrow 1}$. Since $\mathbb{E}_{\mu_1}(f_1) = \mathbb{E}_{\mu_2}(f_2)$ 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 $i \in \mathscr{U}$,

$$(\mathcal{U}f_2 \log f_2)(i) \geq (\mathcal{U}f_2)(i) \cdot \log(\mathcal{U}f_2)(i) + (\mathcal{U}f_2)(i) - \left(\mathcal{U}\sqrt{f_2}\right)(i)^2. \tag{9.9}$$

Let us first see how to use this inequality to prove the desired contraction inequality. Observe that

$$\begin{aligned}
\mathrm{Ent}_1(f_1) &= \mathrm{Ent}_1(\mathcal{U}f_2) \\
&= \sum_{i\in\mathscr{U}} \mu_1(i) \cdot (\mathcal{U}f_2)(i) \cdot \log(\mathcal{U}f_2)(i) \\
&\leq \sum_{i\in\mathscr{U}} \mu_1(i)(\mathcal{U}f_2 \log f_2)(i) - \sum_{i\in\mathscr{U}} \mu_1(i)\left((\mathcal{U}f_2)(i) - \left(\mathcal{U}\sqrt{f_2}\right)(i)^2\right) && \text{(Eq. (9.9))} \\
&= \langle \mathbf{1}, \mathcal{U}f_2 \log f_2\rangle_{\mu_1} - \langle \mathbf{1}, \mathcal{U}f_2\rangle_{\mu_1} + \left\langle \mathcal{U}\sqrt{f_2}, \mathcal{U}\sqrt{f_2}\right\rangle_{\mu_1} \\
&= \langle \mathbf{1}, f_2 \log f_2\rangle_{\mu_2} - \langle \mathbf{1}, f_2\rangle_{\mu_2} + \left\langle \sqrt{f_2}, \mathcal{D}\mathcal{U}\sqrt{f_2}\right\rangle_{\mu_2} \\
&= \mathrm{Ent}_2(f_2) - \mathcal{E}_{\mathcal{D}\mathcal{U}}\left(\sqrt{f_2}, \sqrt{f_2}\right) \\
&\leq (1 - \rho(\mathcal{D}\mathcal{U}))\ \mathrm{Ent}_2(f_2).
\end{aligned}$$

All that remains is to prove Eq. (9.9). For every $i \in \mathscr{U}$, taking $t = (\mathcal{U}f_2)(i)$, we have that

$$\begin{aligned}
&(\mathcal{U}f_2 \log f_2)(i) \\
&= \sum_{\sigma\in\mathrm{supp}(\mu_2)} \mathcal{U}(i,\sigma) f_2(\sigma) \log f_2(\sigma) \\
&= \sum_{\sigma\in\mathrm{supp}(\mu_2)} \mathcal{U}(i,\sigma)(t + f_2(\sigma) - t)\log(t + f_2(\sigma) - t) \\
&\geq \sum_{\sigma\in\mathrm{supp}(\mu_2)} \mathcal{U}(i,\sigma)\left(t\log t + (f_2(\sigma)-t)(1+\log t) + \left(\sqrt{f_2(\sigma)} - \sqrt{t}\right)^2\right) && \text{(Lemma 9.3.10)} \\
&= (\mathcal{U}f_2)(i)\log(\mathcal{U}f_2)(i) + \sum_{\sigma\in\mathrm{supp}(\mu_2)} \mathcal{U}(i,\sigma)\underbrace{\left(\sqrt{f_2(\sigma)} - \sqrt{(\mathcal{U}f_2)(i)}\right)^2}_{\text{Expand}} \\
&= (\mathcal{U}f_2)(i)\log(\mathcal{U}f_2)(i) + \underbrace{2(\mathcal{U}f_2)(i) - 2\sqrt{(\mathcal{U}f_2)(i)} \cdot \left(\mathcal{U}\sqrt{f_2}\right)(i)}_{(*)}.
\end{aligned}$$

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

$$(*) - \left((\mathcal{U}f_2)(i) - \left(\mathcal{U}\sqrt{f_2}\right)(i)^2\right) = \left(\sqrt{(\mathcal{U}f_2)(i)} - \left(\mathcal{U}\sqrt{f_2}\right)(i)\right)^2 \geq 0.$$

Eq. (9.9) then follows and we are done. □

9.4 Entropy Contraction Implications: Proof of Proposition 9.2.1

We end this section by filling in the missing proof of Proposition 9.2.1, which recall shows the useful implications of global entropy contraction. We note that some of these results have been previously proved for special cases such as the Glauber dynamics for Gibbs distributions of spin systems on graphs. The arguments used to prove these previous results essentially extend in a straightforward manner to our setting, so we do not claim any novelty for the material presented here. For completeness, we provide full proofs.

Fix $f_n : \operatorname{supp}(\mu) \to \mathbb{R}_{\geq 0}$, and observe that to show Item 1, it suffices to show that

$$\left\langle f_n, \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n} \log f_n \right\rangle_\mu \leq \operatorname{Ent}_{n-k}(f_{n-k}) \tag{9.10}$$

since then, by rearranging the k-step global entropy contraction inequality, we would obtain

$$\begin{aligned}
(1-\beta_k) \cdot \operatorname{Ent}_n(f_n) &\leq \operatorname{Ent}_n(f_n) - \operatorname{Ent}_{n-k}(f_{n-k}) && \text{(Rearranging)} \\
&\leq \langle f_n, \log f_n \rangle_\mu - \left\langle f_n, \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n} \log f_n \right\rangle_\mu && \text{(Eq. (9.10))} \\
&= \mathcal{E}_{\mathcal{D}^{n \searrow n-k} \mathcal{U}^{n-k \nearrow n}}(f_n, \log f_n).
\end{aligned}$$

This is precisely the desired modified log-Sobolev inequality with constant $1-\beta_k$. It remains to justify Eq. (9.10), which can be done via a simple application of Jensen's Inequality. We have that

$$\begin{aligned}
\left\langle f_n, \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n} \log f_n \right\rangle_\mu &= \left\langle \mathcal{U}_\mu^{n-k \nearrow n} f_n, \mathcal{U}_\mu^{n-k \nearrow n} \log f_n \right\rangle_{\mu_{n-k}} && \text{(Lemma 2.1.2)} \\
&\leq \left\langle \mathcal{U}_\mu^{n-k \nearrow n} f_n, \log \mathcal{U}_\mu^{n-k \nearrow n} f_n \right\rangle_{\mu_{n-k}} && \text{(Concavity of } x \mapsto \log x\text{)} \\
&= \langle f_{n-k}, \log f_{n-k} \rangle_{\mu_{n-k}} && \text{(Fact 2.4.1)} \\
&= \operatorname{Ent}_{n-k}(f_{n-k})
\end{aligned}$$

as desired.

To show Item 2, we first observe that

$$\begin{aligned}
\frac{\mathcal{E}_{\mathcal{D}^{n \searrow n-k} \mathcal{U}^{n-k \nearrow n}}\left(\sqrt{f_n}, \sqrt{f_n}\right)}{\operatorname{Ent}_n(f_n)} &= \frac{\operatorname{Ent}_n(f_n) - \operatorname{Ent}_{n-k}(f_{n-k})}{\operatorname{Ent}_n(f_n)} \cdot \frac{\mathcal{E}_{\mathcal{D}^{n \searrow n-k} \mathcal{U}^{n-k \nearrow n}}\left(\sqrt{f_n}, \sqrt{f_n}\right)}{\operatorname{Ent}_n(f_n) - \operatorname{Ent}_{n-k}(f_{n-k})} \\
&\geq (1-\beta_k) \cdot \frac{\mathcal{E}_{\mathcal{D}^{n \searrow n-k} \mathcal{U}^{n-k \nearrow n}}\left(\sqrt{f_n}, \sqrt{f_n}\right)}{\operatorname{Ent}_n(f_n) - \operatorname{Ent}_{n-k}(f_{n-k})} \\
&\qquad \text{(k-Step Global Entropy Contraction)} \\
&= (1-\beta_k) \cdot \frac{\left\langle \sqrt{f_n}, \sqrt{f_n} \right\rangle_\mu - \left\langle \mathcal{U}^{n-k \nearrow n} \sqrt{f_n}, \mathcal{U}^{n-k \nearrow n} \sqrt{f_n} \right\rangle_{\mu_{n-k}}}{\mathbb{E}_{\sigma \sim \mu_{n-k}}\left[\operatorname{Ent}_k^\sigma(f_k^\sigma)\right]} \\
&\qquad \text{(Lemma 9.2.3)} \\
&= (1-\beta_k) \cdot \frac{\mathbb{E}_{\sigma \sim \mu_{n-k}}\left[\left\langle \sqrt{f_k^\sigma}, \sqrt{f_k^\sigma} \right\rangle_{\mu_k^\sigma} - \left(\mathcal{U}^{n-k \nearrow n} \sqrt{f_n}\right)(\sigma)^2\right]}{\mathbb{E}_{\sigma \sim \mu_{n-k}}\left[\operatorname{Ent}_k^\sigma(f_k^\sigma)\right]} \\
&= (1-\beta_k) \cdot \frac{\mathbb{E}_{\sigma \sim \mu_{n-k}}\left[\mathcal{E}_{\mathbf{1}_k^\sigma \otimes \mu_k^\sigma}\left(\sqrt{f_k^\sigma}, \sqrt{f_k^\sigma}\right)\right]}{\mathbb{E}_{\sigma \sim \mu_{n-k}}\left[\operatorname{Ent}_k^\sigma(f_k^\sigma)\right]}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\kappa\left(\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}\right) &\geq (1-\beta_k) \cdot \min_{\sigma \in \operatorname{supp}(\mu_{n-k})} \left\{ \frac{\mathcal{E}_{\mathbf{1}_k^\sigma \otimes \mu_k^\sigma}\left(\sqrt{f_k^\sigma}, \sqrt{f_k^\sigma}\right)}{\operatorname{Ent}_k^\sigma(f_k^\sigma)} \right\} \\
&\geq (1-\beta_k) \cdot \min_{\sigma \in \operatorname{supp}(\mu_{n-k})} \left\{ \kappa\left(\mathbf{1}_k^\sigma \otimes \mu_k^\sigma\right) \right\}.
\end{aligned}$$

Thus, we have reduced lower bounding the standard log-Sobolev constant of $\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$ to lower bounding the standard log-Sobolev constant of each of the "trivial" walks $\mathbf{1}_k^\sigma \otimes \mu_k^\sigma$ which directly sample from μ_k^σ in a single step. For this, we can appeal to existing bounds in the literature. Since μ is $\mathfrak{b}$-marginally bounded, we have that the minimum probability under μ_k^σ of any state in the support of μ_k^σ is at least $0 < \mathfrak{b}^k \leq 1/2$. Using Theorem A.1 from [DS96], it follows that

$$\kappa\left(\mathbf{1}_k^\sigma \otimes \mu_k^\sigma\right) \geq \frac{1 - 2 \cdot \mathfrak{b}^k}{\log\left(\frac{1}{\mathfrak{b}^k} - 1\right)}$$

independent of $\sigma \in \operatorname{supp}(\mu_{n-k})$. Taking the right-hand side to be $C_{\mathfrak{b},k}$ yields the claim.

For Item 3, write $f = \frac{d\nu}{d\mu} : \Omega \to \mathbb{R}_{\geq 0}$ and observe that

$$\begin{aligned}
\mathscr{D}_{\mathrm{KL}}\left(\nu \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n} \| \mu\right) &= \mathscr{D}_{\mathrm{KL}}\left(\nu \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n} \| \mu \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}\right) \\
&\leq \mathscr{D}_{\mathrm{KL}}\left(\nu \mathcal{D}_\mu^{n \searrow n-k} \| \mu \mathcal{D}_\mu^{n \searrow n-k}\right) \\
&\qquad \text{(Data Processing Inequality; see Theorem 1.4.12)} \\
&= \mathrm{Ent}_{n-k}(f_{n-k}) \qquad \text{(Fact 2.4.1)} \\
&\leq \beta_k \cdot \mathrm{Ent}_n(f) \qquad (k\text{-Step Global Entropy Contraction}) \\
&= \beta_k \cdot \mathscr{D}_{\mathrm{KL}}(\nu \| \mu).
\end{aligned}$$

Item 4 follows by combining Item 1 with Theorem 9.1.1, while Item 5 follows by combining Item 1 with Theorem 9.1.2 (or directly appealing to Item 3 and using Proposition 1.4.15 to compare total variation distance with relative entropy).

Chapter 10

Optimal Mixing of Glauber Dynamics for Sparse Graphical Models

In this chapter, we study the Glauber dynamics or Gibbs sampler in the context of *(probabilistic) graphical models*. These distributions are well-studied in Bayesian machine learning and statistical physics, and encompass many of the distributions we have seen in previous chapters. We'll see how the tools we developed in Chapter 9 can be used to establish *optimal* mixing times for these problems. This chapter is based on [CLV21a].

We begin with a definition of a (discrete, undirected) graphical model.

Definition 46 ((Discrete, Undirected) Graphical Model; see e.g. [KF09])**.** *Let μ be a probability distribution over a discrete product space $\Omega = \prod_{v \in V} \Sigma_v$, where Σ_v is a nonempty finite set for each $v \in V$, and V is a finite index set. We say μ is (the Gibbs distribution of) a* ***(discrete, undirected) graphical model*** *w.r.t. an undirected graph $G = (V, E)$ with vertex set V if it satisfies the following* ***(global) Markov property w.r.t.*** *G: For $\sigma \sim \mu$ and every triple of mutually disjoint nonempty subsets of vertices $A, B, S \subseteq V$ such that every path from any vertex $a \in A$ to any vertex $b \in B$ must contain a vertex in S, the distributions μ_A, μ_B of σ_A, σ_B respectively are mutually independent conditioned on a fixed configuration σ_S on S.*

In other words, graphical models are distribution satisfying *conditional independence* relations w.r.t. some underlying undirected graph. Any two subsets of variables σ_A, σ_B are independent once conditioned on a *separator*. Such distributions are also sometimes called *(discrete) Markov random fields* [KF09]. Notable examples of such distributions which we have seen in previous chapters include Gibbs distributions of spin systems such as graph colorings (Chapters 4 and 8) and independent sets (Chapter 7), Holant-type problems such as even subgraphs and edge covers (Chapter 6), tensor network contractions (Chapter 6), and more.

We prove the following theorem on the mixing time of the Glauber dynamics for sampling from the Gibbs distribution of a bounded-degree graphical model.

Theorem 10.0.1 (Optimal Mixing for Sparse Spectrally Independent Graphical Models; [CLV21a])**.** *Let μ be the Gibbs distribution of a discrete graphical model w.r.t. an underlying graph $G = (V, E)$. Suppose G has maximum degree $\leq \Delta$, and μ is both $\mathfrak{b}$-marginally bounded and η-spectrally independent for constants $0 < \mathfrak{b} \leq 1$ and $\eta \geq 0$. Then μ satisfies 1-global entropy contraction with constant*

$$\beta_1 \leq 1 - \frac{1}{C_1} \qquad \text{where} \qquad C_1 \leq \left(\frac{\Delta}{\mathfrak{b}}\right)^{O\left(1+\frac{\eta}{\mathfrak{b}^2}\right)} \cdot n.$$

In particular, if $\Delta \leq O(1)$, $\mathfrak{b} \geq \Omega(1)$ and $\eta \leq O(1)$ are absolute constants, then μ satisfies 1-step global entropy contraction with constant $1 - O(1/n)$, and the Glauber dynamics $\mathcal{P}_\mu$ satisfies

$$\varrho\left(\mathcal{P}_\mu\right), \kappa\left(\mathcal{P}_\mu\right) \geq \Omega(1/n) \qquad \text{(Log-Sobolev Constants)}$$

$$\mathrm{T}_{\mathsf{mix}}\left(\epsilon; \mathcal{P}_\mu\right) \leq O\left(n\left(\log n + \log \frac{1}{\epsilon}\right)\right). \qquad \text{(Mixing Time)}$$

Remark 50. More specifically, when $n \geq \frac{24\Delta}{\mathfrak{b}^2}\left(\frac{4\eta}{\mathfrak{b}^2}+1\right)$, we can choose

$$C_1 \leq \frac{18\log(1/\mathfrak{b})}{\mathfrak{b}^4}\cdot\left(\frac{24\Delta}{\mathfrak{b}^2}\right)^{1+\frac{4\eta}{\mathfrak{b}^2}}\cdot n$$

and the mixing time of the Glauber dynamics is bounded by

$$\mathrm{T}_{\mathsf{mix}}(\epsilon;\mathcal{P}_\mu) \leq \left\lceil \frac{18\log(1/\mathfrak{b})}{\mathfrak{b}^4}\cdot\left(\frac{24\Delta}{\mathfrak{b}^2}\right)^{1+\frac{4\eta}{\mathfrak{b}^2}}\cdot n\cdot\left(\log n+\log\log\frac{1}{\mathfrak{b}}+\log\frac{1}{2\epsilon^2}\right)\right\rceil.$$

Remark 51. In general, the marginal boundedness $\mathfrak{b}$ of μ will depend (perhaps very poorly) on Δ. However, if $\mathfrak{b} \geq \Omega(1)$ independent of Δ, then Theorem 10.0.1 yields nearly optimal mixing even if $\Delta \leq \mathsf{polylog}(n)$, and already improves upon the mixing time furnished by Theorem 2.3.1 if $\Delta \leq O(n^\delta)$ where $\delta = \delta(\eta,\mathfrak{b})$ is a small constant.

We note that subsequent work of [Bla+22] has reduced the exponential dependence on $\mathfrak{b}$ slightly, yielding $C_1 \leq \left(\frac{\Delta}{\mathfrak{b}}\right)^{O(1+\frac{\eta}{\mathfrak{b}})}$. In the case of 1-global variance contraction, which recall corresponds to lower bounding the spectral gap of the Glauber dynamics, [JPV22] removed the dependence on $\mathfrak{b}$ and further improved the dependence on Δ to polynomial. This allows one to smoothly interpolate with Theorem 2.3.1 as one varies Δ.

[Che+21b] then completely removed the dependence of the spectral gap on $\mathfrak{b},\Delta$ for the special case of two-state spin systems in the uniqueness regime. [Ana+21a; Ana+21b; Ana+22c] established a slightly restricted form of 1-step global entropy contraction, and used it to deduce $O_\delta(n\log n)$-mixing for a variant of the Glauber dynamics they call the *balanced Glauber dynamics*. Independently, [Che+21a; Che+22a] establish the full 1-step global entropy contraction, again for the special case of two-spin systems. [CE22] also independently achieves this entropy contraction result for hardcore model.

Remark 52. This mixing time is *optimal* due to a $\Omega(n\log n)$ lower bound of [HS07]. Intuitively, one must resample each of the n vertices at least once over the evolution of the Markov chain, since otherwise, the resulting sample still have a strong correlation with the starting configuration. From this (very informal) heuristic reasoning, a Coupon Collector argument then demonstrates that $\Theta(n\log n)$ is the correct bound.

We already saw in Chapter 9 how to derive bounds on the rate of entropy decay for the down-up walk (in this case, the Glauber dynamics) given bounds on the local entropy contraction constants; this was the local-to-global entropy contraction result stated in Theorem 9.2.2. We also saw how to deduce bounds on the local entropy contraction constants given spectral independence and marginal boundedness. However, while Theorem 9.2.2 is useful in that it can lead to entropy decay rates of $1-O\left(\frac{1}{n^c}\right)$ for a constant $c \geq 1$, it isn't strong enough to yield entropy decay rates of $1-O(1/n)$, which is necessary to establish optimal mixing times and in particular, Theorem 10.0.1. In a sense, Theorem 9.2.2 is the main bottleneck, and so our goal in this chapter is to tighten it in the setting of bounded-degree graphical models. Before we give the proof, let us first try to understand what exactly makes Theorem 9.2.2 is insufficient.

The first key observation is that while the original local-to-global entropy contraction theorem (see Theorem 9.2.2) by itself isn't strong enough to establish optimal global entropy contraction rates for the Glauber dynamics (i.e. 1-step global entropy contraction with constant $C_1 = O(n)$), it is strong enough to obtain k-step global entropy contraction with constant $C_k = O(1)$ for $k = \Theta(n)$. In a sense, Theorem 9.2.2 only fails to be tight for small k.

Now, suppose we only wanted fast sampling algorithm, setting aside the Glauber dynamics for the moment. Since μ satisfies $\Theta(n)$-step global entropy contraction with constant $O(1)$, we could try to run $\mathcal{D}_\mu^{n\searrow n-k}\mathcal{U}_\mu^{n-k\nearrow n}$ with $k=\Theta(n)$ instead, since it mixes in $O(\log\log(1/\mu_{\min}))$-steps, which is very fast. This dynamics randomly updates the configuration on a random subset $S \subseteq V$ of vertices with $|S| = \Theta(n)$ in each step.

The issue of course is that in general, this dynamics is not implementable, since one would need to be able to sample perfectly from the marginal distribution μ_S^σ over configurations on vertices in S conditioned on a fixed configuration σ on vertices in $V\setminus S$. However, the second key observation which saves us is that when μ comes from a bounded-degree graphical model, sampling from this conditional marginal distribution over S *is possible "most of the time"*. By fixing a configuration σ over the vertices of $V\setminus S$, one *shatters* the graph G into many *small* connected components which are mutually disconnected. More specifically, with high probability, the induced subgraph

$G[S]$ consists of connected components $G[S_1], \dots, G[S_\ell]$ where $S = S_1 \sqcup \cdots \sqcup S_\ell$, there are no edges between S_i, S_j for $i \neq j$, and $S_i \leq O(\log n)$ for every $i = 1, \dots, \ell$. If this happens, then conditional independence tells us that μ_S^σ factorizes as $\mu_S^\sigma = \mu_{S_1}^\sigma \otimes \cdots \otimes \mu_{S_\ell}^\sigma$ and so to sample from μ_S^σ, it suffices to independently sample a configuration over each component from $\mu_{S_i}^\sigma$. This can be done perfectly via brute force since each component S_i has size $|S_i| \leq O(\log n)$.

Of course, this is just informal intuition, and there are many pieces missing. For one, the above shattering phenomenon only happens with high probability, rather than with probability 1, and so at best, one can only hope to obtain a running time bound which holds with high probability. The other glaring issue is that the above does not suggest any way to analyze the simple Glauber dynamics $\mathcal{P}_\mu = \mathcal{D}_\mu^{n \searrow n-1} \mathcal{U}_\mu^{n-1 \nearrow n}$, which was our original goal. However, it turns out we can combine an alternative version of entropy contraction called *entropy factorization*, which enables a more refined analysis, with the shattering phenomenon to optimally *compare* the much more complicated dynamics $\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$ with the simple Glauber dynamics $\mathcal{P}_\mu = \mathcal{D}_\mu^{n \searrow n-1} \mathcal{U}_\mu^{n-1 \nearrow n}$ when $k = \Theta(n)$. We already saw this "comparison intuition" more explicitly in Remark 9.

10.1 Tensorization and Block Factorization of Entropy

To build up to the proof of Theorem 10.0.1, we first switch perspectives, and consider an equivalent but more convenient version of global entropy contraction which is known as entropy factorization. This will be useful as per the discussion in the previous section.

Definition 47 (Uniform Block Factorization of Entropy)**.** *For $0 \leq k \leq n$ and $C_k \geq 0$, we say μ satisfies k-**uniform block factorization of entropy with constant** C_k if for every global function $f_n : \operatorname{supp}(\mu) \to \mathbb{R}_{\geq 0}$, we have the inequality*

$$\operatorname{Ent}_n(f_n) \leq C_k \cdot \mathbb{E}_{\sigma \sim \mu_{n-k}} \left[\operatorname{Ent}_k^\sigma (f_k^\sigma) \right] \tag{10.1}$$

*where f_k^σ are local functions induced by f_n. When $k = 1$, we also say μ satisfies **approximate tensorization of entropy with constant** C_1.*

Remark 53. One can also define a *k-uniform block factorization of variance* (and *approximate tensorization of variance* for the case $k = 1$) by replacing all occurrences of entropy with variance in Definition 47. In this case, the constant C_k becomes the inverse of the spectral gap of $\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$. Again, these notions can all be readily extended to φ-entropies, although we will not need these in this thesis.

Remark 54 (Factorization Formulation of Standard Log-Sobolev). Recall that the standard log-Sobolev constant of a reversible Markov chain P with stationary distribution μ on a finite state space Ω is defined as the smallest constant κ such that for every function $f : \Omega \to \mathbb{R}_{\geq 0}$, we have the standard log-Sobolev inequality

$$\kappa \cdot \operatorname{Ent}_\mu(f) \leq \mathcal{E}_{\mathsf{P}} \left(\sqrt{f}, \sqrt{f} \right).$$

If $\mathsf{P} = \mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$, then recall that

$$\mathcal{E}_{\mathsf{P}}(f, f) = \operatorname{Var}_n(f) - \operatorname{Var}_{n-k}(f_{n-k}) = \mathbb{E}_{\sigma \sim \mu_{n-k}} \left[\operatorname{Var}_k^\sigma (f_k^\sigma) \right]$$

where in the last step we use variance decomposition (see Lemma 2.4.2). Hence, the standard log-Sobolev inequality for $\mathcal{D}_\mu^{n \searrow n-k} \mathcal{U}_\mu^{n-k \nearrow n}$ can be equivalently formulated as

$$\operatorname{Ent}_n(f) \leq \frac{1}{\kappa} \cdot \mathbb{E}_{\sigma \sim \mu_{n-k}} \left[\operatorname{Var}_k^\sigma \left(\sqrt{f_k^\sigma} \right) \right]$$

in a manner analogous to entropy factorization.

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 $C_1 = 1$ [Led99; GZ03; Ces01; CMT15] (see Lemma 10.1.2 below for a more general statement). 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.

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; see e.g. [Mar99; GZ03; Ces01; CP21a]. Despite successes on $\mathbb{Z}^d$, there has not been much study for spin systems on general bounded-degree graphs prior to our work. The works of [CMT15; Mar19] 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.

As previously mentioned, uniform block factorization of entropy is equivalent to global entropy contraction. We state and prove this now.

Lemma 10.1.1 (Contraction $\iff$ Factorization)**.** *Let μ be a probability distribution over $\binom{\mathscr{U}}{n}$ for a finite ground set $\mathscr{U}$ and integer $0 \leq n \leq |\mathscr{U}|$, and fix some integer $0 \leq k \leq n$. Then μ satisfies k-uniform block factorization of entropy with constant C_k if and only if μ satisfies k-step global entropy contraction with constant $1 - \frac{1}{C_k}$.*

Proof. By Lemma 9.2.3, we have the equality

$$\mathbb{E}_{\sigma \sim \mu_{n-k}} \left[\mathrm{Ent}_k^{\sigma} \left(f_k^{\sigma} \right) \right] = \mathrm{Ent}_n \left(f_n \right) - \mathrm{Ent}_{n-k} \left(f_{n-k} \right).$$

Hence, by rearranging Eq. (10.1), we see that uniform block factorization of entropy with constant C_k is equivalent to

$$\mathrm{Ent}_{n-k} \left(f_{n-k} \right) \leq \left(1 - \frac{1}{C_k} \right) \cdot \mathrm{Ent}_n \left(f_n \right)$$

which is exactly k-step global entropy contraction with constant $1 - \frac{1}{C_k}$. □

Remark 55 (Non-Uniform Block Factorization). We have just seen that k-uniform block factorization is equivalent to k-step global entropy contraction, which is a mixing condition for the k-uniform block dynamics. In the context of spin systems, one can generalize this considerably by fixing some arbitrary distribution over all subsets of sites (i.e. "blocks"), and define a block dynamics w.r.t. this distribution over blocks. Analyzing the rate of entropy decay for these more general dynamics can then be done with a generalized version of block factorization of entropy, which was first proposed in [CP21a]. [Bla+22] showed how to deduce these more general block factorization inequalities from spectral independence and marginal boundedness.

Since we will also be considering graphical models where the underlying graph shatters into small connected components, it will be useful to under how entropy factorizes under independent products of probability distributions. The following is also a special case of Shearer's Inequalities in information theory. We refer the interested reader to [Led99; GZ03; Ces01; CMT15] for a more general statement (and its proof).

Lemma 10.1.2 (Entropy Factorization for Product Measures)**.** *Let μ, ν be probability distributions over state spaces Ω_μ, Ω_ν, respectively. Let $\mu \otimes \nu$ be the **product distribution** over $\Omega_\mu \times \Omega_\nu$ defined by $(\mu \otimes \nu)(\tau, \sigma) = \mu(\tau) \cdot \nu(\sigma)$ for all $\tau \in \Omega_\mu, \sigma \in \Omega_\nu$. Let $f : \Omega_\mu \times \Omega_\nu \to \mathbb{R}_{\geq 0}$ be an arbitrary global function. Then we have the inequality*

$$\mathrm{Ent}_{\mu \otimes \nu} \left(f \right) \leq \mathbb{E}_{\sigma \sim \nu} \left[\mathrm{Ent}_{\mu} \left(f^{\sigma} \right) \right] + \mathbb{E}_{\tau \sim \mu} \left[\mathrm{Ent}_{\nu} \left(f^{\tau} \right) \right],$$

where for $\sigma \in \mathrm{supp}\,(\nu)$, $f^{\sigma} : \Omega_\mu \to \mathbb{R}_{\geq 0}$ is defined by $f^{\sigma}(\tau) \stackrel{\mathsf{def}}{=} f(\tau, \sigma)$ (and $f^{\tau} : \Omega_\nu \to \mathbb{R}_{\geq 0}$ is defined analogously).

Definition 47 defines entropy factorization at a very general level which goes beyond discrete graphical models or even distributions over discrete product spaces. Since we are focusing on discrete graphical models in this chapter, it will be convenient to slightly modify our notation so as to make certain salient aspects of the analysis more transparent (e.g. the different components of the graph after deleting the subset of pinned vertices).

For a subset $S \subseteq V$ of vertices, we write μ_S for the *marginal distribution* over feasible configurations over S. As the name suggests, the marginal distributions μ_S are intimately related to the level-k marginal distributions μ_k we have considered throughout this thesis. For instance, we may

write μ_k as a mixture $\mu_k = \frac{1}{\binom{n}{k}} \sum_{S \in \binom{V}{k}} \mu_S$. Similarly, we write μ_S^σ for the marginal distribution over S conditioned on a feasible configuration σ over a subset of vertices of $V \setminus S$. We typically call such σ *pinnings* or *boundary conditions*. In a similar manner, we may then write $\mathrm{Ent}_S^\sigma(\cdot)$ for $\mathrm{Ent}_{\mu_S^\sigma}(\cdot)$, etc. Most of the time, σ be a configuration on all of $V \setminus S$, in which case we can view μ_S^σ not just a distribution over configurations ξ on S, but also as a distribution over configurations τ on all of V, where τ agrees with ξ on S and with σ on $V \setminus S$. In this case, even if f_n is a global function on configurations over V, the quantities $\mathbb{E}_{\mu_S^\sigma}(f_n)$ and $\mathrm{Ent}_S^\sigma(f_n)$ make sense.

With this notation in hand, we may more convenient rewrite k-uniform block factorization of entropy (see Eq. (10.1)) for discrete graphical models as follows. For every global function f on configurations over V, we have the inequality

$$\mathrm{Ent}_n(f) \le C_k \cdot \frac{1}{\binom{n}{k}} \sum_{S \in \binom{V}{k}} \mathbb{E}_{\sigma \sim \mu_{V \setminus S}} \left[\mathrm{Ent}_S^\sigma(f) \right]. \tag{10.2}$$

Similarly, for the case $k = 1$, approximate tensorization of entropy can be rewritten as

$$\mathrm{Ent}_n(f) \le C_1 \cdot \frac{1}{n} \sum_{v \in V} \mathbb{E}_{\sigma \sim \mu_{V-v}} \left[\mathrm{Ent}_v^\sigma(f) \right]. \tag{10.3}$$

Note Eqs. (10.2) and (10.3) deviate slightly from the original definitions of uniform block factorization and approximate tensorization that we used in [CLV21a], where there is an additional multiplicative factor of $\frac{k}{n}$ on the left-hand side.

Corollary 10.1.3. *For every subset $S \subseteq V$, every boundary condition σ on $V \setminus S$, and every global function $f : \Omega \to \mathbb{R}_{\ge 0}$, we have*

$$\mathrm{Ent}_S^\sigma(f) \le \sum_{U \in \mathcal{C}(S)} \mathbb{E}_{\xi \sim \mu_{S \setminus U}^\sigma} \left[\mathrm{Ent}_U^{\sigma \sqcup \xi}(f) \right].$$

Proof. This is an immediate consequence of Lemma 10.1.2 and conditional independence. □

Lemma 10.1.4 (Crude Entropy Factorization)**.** *If μ is $\mathfrak{b}$-marginally bounded, then for every subset $S \subseteq V$, every boundary condition $\sigma \in \Omega_{V \setminus S}$, and every function $f : \Omega_S^\sigma \to \mathbb{R}_{\ge 0}$, we have*

$$\mathrm{Ent}_S^\sigma(f) \le \frac{3 |S|^2 \log(1/\mathfrak{b})}{2\mathfrak{b}^{2|S|+2}} \sum_{v \in S} \mathbb{E}_{\xi \sim \mu_{S-v}^\sigma} \left[\mathrm{Ent}_v^{\sigma \sqcup \xi}(f) \right].$$

This lemma is purely technical, and we defer its proof to Section 10.3.

10.2 Entropy Factorization in Sparse Graphical Models

In the previous section, we saw that entropy factorization is equivalent global entropy contraction. Hence, to prove Theorem 10.0.1, it suffices to show that if μ is the Gibbs distribution of a bounded-degree graphical model satisfying $\mathfrak{b}$-marginal boundedness and η-spectral independence, then it satisfies approximate tensorization of entropy with constant $C_1 \le O(n)$. As discussed earlier, our strategy is to compare the approximate tensorization constant (or, equivalently, the 1-step global entropy contraction constant) with the k-uniform block factorization constant (or, equivalently, the k-step global entropy contraction constant) with $k = \Theta(n)$. In particular, we prove the following.

Proposition 10.2.1 (Entropy Factorization Comparison)**.** *Let $\Delta \ge 3$ be an integer and $0 < \mathfrak{b} \le 1$. Let μ be the Gibbs distribution of a discrete graphical model w.r.t. an n-vertex graph G of maximum degree $\le \Delta$, and assume that μ is $\mathfrak{b}$-marginally bounded. Suppose there exist constants $0 < \theta \le \frac{\mathfrak{b}^2}{12\Delta}$ and $C = C_{\lceil \theta n \rceil} > 0$ 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*

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

Proof of Theorem 10.0.1. By Proposition 10.2.1, using Lemma 10.1.1 to convert between approximate tensorization of entropy and 1-step global entropy contraction, it suffices to show that for some constant $0 < \theta < \frac{\mathfrak{b}^2}{12\Delta}$, the distribution μ satisfies $\lceil \theta n \rceil$-uniform block factorization of entropy with constant

$$C_{\lceil \theta n \rceil} \leq \left(\frac{24\Delta}{\mathfrak{b}^2}\right)^{1+\frac{4\eta}{\mathfrak{b}^2}}.$$

For this, we use Theorem 9.2.2 to reduce this to bounds on the local entropy contraction constants of μ, which we bound by appealing to Theorem 9.3.5 and using both $\mathfrak{b}$-marginal boundedness and η-spectral independence. □

Our goal now is to prove Proposition 10.2.1. As previously mentioned, use the following shattering property of sparse graphs.

Lemma 10.2.2 (Shattering Lemma for Sparse Graphs)**.** *Let $G = (V, E)$ be an n-vertex graph of maximum degree $\leq \Delta$. Then for every positive integer $\ell > 0$, we have*

$$\Pr_S\left[|S_v| = \ell\right] \leq \frac{k}{n} \cdot (2e\Delta\theta)^{\ell-1},$$

where S is a uniformly random subset of V of size $k = \lceil \theta n \rceil$, and S_v is the unique maximal connected component of $G[S]$ containing v.

Again, we defer its proof to Section 10.3. We close this section with a proof of Proposition 10.2.1, which recall compares the $\lceil \theta n \rceil$-uniform block factorization constant with the approximate tensorization constant.

Proof of Proposition 10.2.1. Set $k = \lceil \theta n \rceil$. Combining all of the lemmas we've seen in this chapter, we deduce that

$$\begin{aligned}
\mathrm{Ent}_n(f) &\leq C_k \cdot \frac{1}{\binom{n}{k}} \sum_{S \in \binom{V}{k}} \mathbb{E}_{\sigma \sim \mu_{V \setminus S}}\left[\mathrm{Ent}_S^\sigma(f)\right] && (k\text{-uniform block factorization}) \\
&\leq C_k \cdot \frac{1}{\binom{n}{k}} \sum_{S \in \binom{V}{k}} \sum_{U \in \mathcal{C}(S)} \mathbb{E}_{\sigma \sim \mu_{V \setminus U}}\left[\mathrm{Ent}_U^\sigma(f)\right] && (\text{Corollary 10.1.3}) \\
&\leq C_k \cdot \frac{1}{\binom{n}{k}} \sum_{S \in \binom{V}{k}} \sum_{U \in \mathcal{C}(S)} \frac{3\,|U|^2 \log(1/\mathfrak{b})}{2\mathfrak{b}^{2|U|+2}} \sum_{v \in U} \mathbb{E}_{\sigma \sim \mu_{V-v}}\left[\mathrm{Ent}_v^\sigma(f)\right] && (\text{Lemma 10.1.4}) \\
&= \frac{3C_k \log(1/\mathfrak{b})}{2\mathfrak{b}^4} \sum_{v \in V} \mathbb{E}_{\sigma \sim \mu_{V-v}}\left[\mathrm{Ent}_v^\sigma(f)\right] \sum_{\ell=1}^{k} \Pr_S\left[|S_v| = \ell\right] \cdot \frac{\ell^2}{\mathfrak{b}^{2(\ell-1)}} && (\text{Rearranging}) \\
&\leq \frac{3C_k \log(1/\mathfrak{b})}{2\mathfrak{b}^4} \sum_{v \in V} \mathbb{E}_{\sigma \sim \mu_{V-v}}\left[\mathrm{Ent}_v^\sigma(f)\right] \sum_{\ell=1}^{k} \ell^2 \left(\frac{2e\Delta\theta}{\mathfrak{b}^2}\right)^{k-1} && (\text{Lemma 10.2.2}) \\
&\leq \frac{3C_k \log(1/\mathfrak{b})}{2\mathfrak{b}^4} \left(\sum_{\ell=1}^{k} \frac{\ell^2}{2^{\ell-1}}\right) \sum_{v \in V} \mathbb{E}_{\sigma \sim \mu_{V-v}}\left[\mathrm{Ent}_v^\sigma(f)\right] && (\theta \leq \tfrac{\mathfrak{b}^2}{12\Delta}) \\
&\leq \frac{18C_k n \log(1/\mathfrak{b})}{\mathfrak{b}^4} \cdot \frac{1}{n} \sum_{v \in V} \mathbb{E}_{\sigma \sim \mu_{V-v}}\left[\mathrm{Ent}_v^\sigma(f)\right]. && (\textstyle\sum_{k=1}^{\infty} \frac{k^2}{2^{k-1}} = 12)
\end{aligned}$$

This is precisely yields the desired approximate tensorization constant. □

10.3 Proofs of Technical Lemmas 10.1.4 and 10.2.2

We first prove Lemma 10.1.4 which gives a crude bound on the approximate tensorization constant for any subset and boundary condition.

Proof of Lemma 10.1.4. Fix a subset $S \subseteq V$ of size $k \geq 1$ and some boundary condition $\sigma \in \Omega_{V \setminus S}$. Let $C_1 = C_1(S, \sigma)$ be the optimal approximate tensorization constant for μ_S^σ; hence, for every function $f : \Omega_S^\sigma \to \mathbb{R}_{\geq 0}$ one has

$$\mathrm{Ent}_S^\sigma(f) \leq C_1 \cdot \frac{1}{k} \sum_{v \in S} \mu_S^\sigma[\mathrm{Ent}_v(f)].$$

Let $\gamma = \gamma(U, \xi)$ be the spectral gap of the Glauber dynamics for μ_S^σ, and let $\kappa = \kappa(U, \xi)$ be the standard log-Sobolev constant. Thus, for every function $f : \Omega_S^\sigma \to \mathbb{R}_{\geq 0}$ it holds that

$$\gamma \cdot \mathrm{Var}_S^\sigma(f) \leq \frac{1}{k} \sum_{v \in S} \mathbb{E}_{\xi \sim \mu_{S-v}^\sigma} \left[\mathrm{Var}_v^{\sigma \sqcup \xi}(f) \right] \qquad \text{(Remark 53)}$$

$$\kappa \cdot \mathrm{Ent}_S^\sigma(f) \leq \frac{1}{k} \sum_{v \in S} \mathbb{E}_{\xi \sim \mu_{S-v}^\sigma} \left[\mathrm{Var}_v^{\sigma \sqcup \xi} \left(\sqrt{f} \right) \right]. \qquad \text{(Remark 54)}$$

Since $\mathrm{Var}_v^{\sigma \sqcup \xi} \left(\sqrt{f} \right) \leq \mathrm{Ent}_v^{\sigma \sqcup \xi}(f)$, we have

$$C_1 \leq \frac{1}{\kappa}. \tag{10.4}$$

This is essentially the content of Remark 49 and Lemma 9.3.8; see also [CMT15, Proposition 1.1]. By comparing the standard log-Sobolev constant with the spectral gap (see Lemma 9.3.9), we obtain

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

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

$$\kappa \geq \frac{\gamma}{3k \log(1/\mathfrak{b})}. \tag{10.5}$$

We will lower bound γ by lower bounding the conductance $\Phi = \Phi\left(\mathcal{P}_{\mu_S^\sigma}\right)$ of the Glauber dynamics $\mathcal{P}_{\mu_S^\sigma}$ and appealing to Cheeger's Inequality $\gamma \geq \frac{\Phi^2}{2}$ (see Theorem 3.3.2). Since we assume μ is totally connected, for any subset $\Omega_0 \subseteq \Omega_S^\sigma$ of configurations, we have that $\Phi(\Omega_0) > 0$, i.e. there exists an edge crossing the cut $(\Omega_0, \Omega_S^\sigma \setminus \Omega_0)$ with nonzero weight. If $\xi \in \Omega_0, \tau \in \Omega_S^\sigma \setminus \Omega_0$ are the endpoints of any such edge, then

$$\mu_S^\sigma(\xi) \cdot \mathcal{P}_{\mu_S^\sigma}(\xi \to \tau) \geq \mathfrak{b}^k \cdot \frac{\mathfrak{b}}{k} = \frac{\mathfrak{b}^{k+1}}{k}$$

from which it follows that

$$\Phi \geq \frac{2\mathfrak{b}^{k+1}}{k}. \tag{10.6}$$

Combining Eqs. (10.4) to (10.6) and Cheeger's Inequality, we finally conclude that

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

as claimed. □

Next we establish Lemma 10.2.2. We use the following lemma concerning the number of connected induced subgraphs in a bounded degree graph.

Lemma 10.3.1 (Lemma 2.1 from [Bor+13])**.** *Let $G = (V, E)$ be a graph with maximum degree $\leq \Delta$, and fix a vertex $v \in V$. Then for every positive integer ℓ, the number of connected induced subgraphs of G containing v with ℓ vertices is at most $(e\Delta)^{\ell-1}$.*

We then prove Lemma 10.2.2.

Proof of Lemma 10.2.2. Let $\mathcal{A}_v(\ell)$ denote the collection of subsets of vertices $U \subseteq V$ such that $|U| = \ell$, $v \in U$, and $G[U]$ is connected. By the Union Bound, we have

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

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

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

The lemma then follows immediately from Lemma 10.3.1 since $|\mathcal{A}_v(\ell)| \leq (e\Delta)^{\ell-1}$. □

Chapter 11

Tight Mixing for Discrete Log-Concave Distributions

In this chapter, which is based on [Ana+21c], we prove *tight* mixing time bounds for the down-up walk on bases of matroids, determinantal distributions, and more generally distributions associated with log-concave polynomials. Such discrete log-concave distributions were originally introduced in [AOV21], and we discussed them in Chapter 5. Previously, [Ana+19] proved that if μ is a discrete log-concave distribution over $\binom{[n]}{k}$, the spectral gap of the random walk P is at least $1/k$; this was one of the main results of Chapter 5. This implies that

$$\mathrm{T}_{\mathsf{mix}}(\epsilon; S_0, \mathsf{P}) \leq O\left(k \cdot \left(\log \frac{1}{\mu(S_0)} + \log \frac{1}{\epsilon}\right)\right).$$

Later, [CGM21] proved proved that the modified log-Sobolev constant for the same random walk is at least $1/k$; we saw this in Chapter 9. This resulted in a tighter mixing time bound of

$$\mathrm{T}_{\mathsf{mix}}(\epsilon; S_0, \mathsf{P}) \leq O\left(k \cdot \left(\log \log \frac{1}{\mu(S_0)} + \log \frac{1}{\epsilon}\right)\right).$$

These results lead to efficient algorithms assuming that the mass of the starting set, $\mu(S_0)$, is not terribly small; this can often be achieved in practice. For example, for matroids, any starting basis S_0 will satisfy $\mu(S_0) \geq 1/\binom{n}{k} \geq n^{-k}$, because the number of bases is at most $\binom{n}{k}$. Consequently the above bounds turn into $\mathrm{T}_{\mathsf{mix}}(\epsilon; S_0, \mathsf{P}) \leq O(k(k \log(n) + \log(1/\epsilon)))$ and $\mathrm{T}_{\mathsf{mix}}(\epsilon; S_0, \mathsf{P}) \leq O(k(\log k + \log\log n + \log(1/\epsilon)))$ respectively. However, for other distributions μ with a log-concave generating polynomial, even in the very special case of determinantal point processes, there is no control on $\min\{\mu(S_0) \mid S_0 \in \mathrm{supp}(\mu)\}$, so one has to rely on clever tricks to find a good starting set S_0; even then, the best hope is to find a set S_0 with $\mu(S_0) \gtrsim 1/\binom{n}{k}$, which results in a mixing time mildly depending on n.

Here, we show that for a discrete log-concave distribution over $\binom{[n]}{k}$, the down-up random walk started from an arbitrary point in the support mixes $O(k \log k)$-steps *independent* of n and the mass of the starting point $\mu(S_0)$. This improves significantly upon the previous analyses [Ana+19; CGM21] which were based purely on the spectral gap and the modified log-Sobolev constant.

Theorem 11.0.1. *For any distribution defined by $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ with a log-concave generating polynomial g_μ, the mixing time of the down-up random walk P, starting from any S_0 in the support of μ is*

$$\mathrm{T}_{\mathsf{mix}}(\epsilon; S_0, \mathsf{P}) \leq O(k \log(k/\epsilon)).$$

Our $O(k \log k)$ mixing time upper bound is tight up to constant factors, since generally we cannot hope for a better mixing time than $k \log k$. Indeed, each step of the random walk P replaces one element of the current set, and by a Coupon Collector argument, at least $\Omega(k \log k)$ steps are needed to replace every element of the starting set S_0. As long as k is not too close to n, say $k < 0.99n$, replacing every starting element is needed for sufficient mixing, even for the simple distribution μ which is uniform over $\binom{[n]}{k}$.

The main new ingredient which allows us to remove the dependence on n and the starting state is a property we call *approximate exchange*, a generalization of well-studied exchange properties

for matroids and valuated matroids. Our mixing time bound is an asymptotic improvement over prior work for $k = O(1)$, or more generally when k is smaller than $\log(n)^\epsilon$ for all $\epsilon > 0$. Another consequence of the new mixing time bound is that it enables the analysis of the down-up random walk when n is infinitely large; for example, this is the case for continuous determinantal point processes [see, e.g., OR19].[1] To avoid complicating the notation, we do not consider infinitely large ground sets, but note that the results do generalize to such cases.

Historically, earlier works on a subclass of matroids, called *balanced matroids*, followed a similar development, where initially a spectral gap lower bound was proved, resulting in a running time[2] of $O(nk(k\log n + \log(1/\epsilon)))$ followed by a modified log-Sobolev inequality which resulted in a mixing time of $O(k(\log k + \log\log n + \log(1/\epsilon)))$ [see MT06, for a survey]. Noting that the term $\log\log n$ seems unnecessary, [MT06] raised the question of proving a better inequality that would result in a running time of $O(nk\log(k/\epsilon))$. They specifically hoped for the possibility of proving a *Nash Inequality*, an advanced type of functional inequality used to derive very tight mixing times for some Markov chains [MT06]. We believe there are barriers to using functional inequalities in general to prove $O(k\log(k/\epsilon))$ mixing time for the down-up random walk; we defer an explanation of this to a future version of [Ana+21c]. However, without proving new functional inequalities, we manage to sidestep this barrier and improve the running time to the conjectured $O(nk\log(k/\epsilon))$ for not just balanced matroids, but the class of all matroids.

11.1 Tight Mixing via Approximate Exchange

In order to prove Theorem 11.0.1, we combine a new analysis of the initial steps of the down-up random walk with the previously established modified log-Sobolev inequality proved by [CGM21]; see Corollary 9.3.3 from Chapter 9. Our high-level strategy is to prove that conditioned on having replaced every element of the starting set S_0 at least once by time t, the set at time t can be used as a *"warm start"* for the rest of the steps. Again, by a straightforward Coupon Collector argument, this event happens in $O(k\log k)$ steps with good probability. Furthermore, after this occurs, the warm start distribution quickly converges to μ in an additional $O(k\log k)$ steps. To be more specific, by "warm start", we mean that the density of the set at time t w.r.t. μ, conditioned on this event, is upper-bounded by only a function of k.

In order to prove this, we introduce a new property of functions $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ that we call α-approximate exchange.

Definition 48 (Approximate Exchange Property)**.** *For $\alpha \geq 0$, we say $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ satisfies the α-approximate exchange property if for every $S, T \in \binom{[n]}{k}$, and $i \in S$, there exists $j \in T$ such that*

$$\mu(S)\mu(T) \leq \alpha \cdot \mu(S - i + j)\mu(T + i - j).$$

For brevity, we sometimes simply say μ has α-exchange.

Note that when μ takes values in $\{0, 1\}$ and $\alpha \geq 1$, this property becomes equivalent to the famous *strong basis exchange axiom of matroids* [Oxl11]; indeed, this property says that for every two bases S, T of a matroid $\mathcal{M}$ and every $i \in S$, there exists $j \in T$ such that $S - i + j$ and $T + i - j$ are also bases in $\mathcal{M}$. Definition 48 can be seen as a *quantitative variant* of strong basis exchange. Alternatively, it can be viewed as an approximate and multiplicative form of $M^\natural$*-concavity*, a cornerstone of discrete convex analysis [MS99].

We prove that every μ with a log-concave generating polynomial satisfies α-approximate exchange with $\alpha = 2^{O(k)}$.

Lemma 11.1.1. *Any $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ with a log-concave generating polynomial satisfies a $2^{O(k)}$-approximate exchange property. That is, for every $S, T \in \binom{[n]}{k}$ and $i \in S$ there exists $j \in T$ such that*

$$\mu(S)\mu(T) \leq 2^{O(k)}\mu(S - i + j)\mu(T + i - j).$$

Crucially, our α does not depend on n. We remark that [BH18] showed a result that can be thought of as a converse to this. They proved that $M^\natural$-concavity of $\log\mu$, which is equivalent to the

[1] We note however that one still needs to be able to implement each step of the random walk efficiently when n is infinitely large. For examples where this is possible see [OR19].

[2] Note that the running time is n times the mixing time for the down-up walk.

1-approximate exchange property, implies that the generating polynomial of μ is log-concave. We establish Lemma 11.1.1, as well as improved exchange inequalities in certain special settings (e.g. determinantal point processes), in Section 11.2. In Appendix F Appendix F.4, we show how these approximate exchange properties can also be used to give multiplicative approximation guarantees for simple local search algorithms in discrete optimization.

11.1.1 From Approximate Exchange to a "Warm Start"

Let us first see how to implement our warm start idea. Let τ be the first time such that every element in our initial set has been replaced at least once. In other words think of initial elements as unmarked, and every time we replace an element we mark the new element brought in. Then τ is the first time that every element is marked.

We will prove the following.

Lemma 11.1.2 (Warm Start)**.** *Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be a probability distribution with the α-approximate exchange property. Let S_t be the set at time t in the down-up random walk. Then for any $X \in \binom{[n]}{k}$ and any time t,*

$$\Pr[S_t = X \mid \tau \leq t] \leq \alpha^k \cdot k! \cdot \mu(X).$$

Note that without $\alpha^k \cdot k!$, the right-hand side is simply the stationary distribution. So, this statement can be understood to say that as long as we have replaced each element at least once, we cannot be too far off from the stationary distribution.

Before proving Lemma 11.1.2, let us finish the proof of Theorem 11.0.1 assuming it and Lemma 11.1.1,

Proof of Theorem 11.0.1 assuming Lemmas 11.1.1 and 11.1.2. Note that for any fixed time t, we can simply bound $\Pr[\tau > t]$ by $k(1-1/k)^t \leq ke^{-t/k}$. In particular this probability rapidly converges to 0 after about $k \log k$ steps.

Now let $t_1 < t_2$ be two time indices. Let ν_t denote the distribution of the state S_t of random walk at time t. Our goal is to bound $\|\nu_t - \mu\|_{\mathsf{TV}}$, where for simplicity of notation, we assume μ is properly normalized to be a probability distribution. Let ν'_t be the distribution of S_t conditioned on $\tau \leq t$, and let ν''_t be the distribution of S_t conditioned on $\tau \geq t$. Then we can write

$$\nu_{t_1} = \Pr[\tau \leq t_1] \cdot \nu'_{t_1} + \Pr[\tau > t_1] \cdot \nu''_{t_1}.$$

If P denotes the random walk operator, then note that $\nu_{t_2} = \nu_{t_1}\mathsf{P}^{t_2-t_1}$. So we get

$$\nu_{t_2} = \Pr\left[\tau \leq t_1\right] \cdot \nu'_{t_1}\mathsf{P}^{t_2-t_1} + \Pr\left[\tau > t_1\right] \cdot \nu''_{t_1}\mathsf{P}^{t_2-t_1}.$$

Using the Triangle Inequality, we can bound

$$\|\nu_{t_2} - \mu\|_{\mathsf{TV}} \leq \left\|\nu'_{t_1}\mathsf{P}^{t_2-t_1} - \mu\right\|_{\mathsf{TV}} + \Pr[\tau > t_1].$$

Here we used the fact that $\Pr[\tau \leq t_1] \leq 1$, and $\left\|\nu''_{t_1}\mathsf{P}^{t_2-t_1} - \mu\right\|_{\mathsf{TV}} \leq 1$; the latter inequality is because $\|\cdot\|_{\mathsf{TV}}$ is always upper bounded by 1.

We can bound the second term in the above inequality by $ke^{-t_1/k}$ as stated before. For the first term, note that the KL-divergence between ν'_{t_1} and μ is at most $O(k^2)$ by Lemmas 11.1.1 and 11.1.2. This is because $k! = 2^{O(k \log k)}$ can be absorbed into the $2^{O(k^2)}$ without loss, and so

$$\mathscr{D}_{\mathrm{KL}}\left(\nu'_{t_1} \| \mu\right) = \mathbb{E}_{S \sim \nu'_{t_1}}\left[\log \frac{\nu'_{t_1}(S)}{\mu(S)}\right] \leq \log\left(2^{O(k^2)}\right) = O(k^2).$$

So by Corollary 9.3.3 in $t_2 - t_1$ steps this KL-divergence decreases to $(1-1/k)^{t_2-t_1} \cdot O(k^2) = O(k^2 e^{-(t_2-t_1)/k})$. By Pinsker's Inequality (see Proposition 1.4.15), we get that

$$\left\|\nu'_{t_1}\mathsf{P}^{t_2-t_1} - \mu\right\|_{\mathsf{TV}} \leq O\left(ke^{-(t_2-t_1)/2k}\right).$$

So in the end we get the following bound

$$\|\nu_{t_2} - \mu\|_{\mathsf{TV}} \leq O\left(ke^{-(t_2-t_1)/2k} + ke^{-t_1/k}\right).$$

In order for this to be at most ϵ, it is enough to make sure that $\min\{t_1, t_2 - t_1\} = \Omega(k \log k + k \log \frac{1}{\epsilon})$. So we can simply let $t_1 = t_2/2$, and then make sure that $t_2 = \Omega(k \log(k/\epsilon))$. □

Let us now prove Lemma 11.1.2 using Lemma 11.1.1.

Proof of Lemma 11.1.2. Let's look at the down-up walk process *with orders*. This means that we start with some elements $e_1, \dots, e_k$ that together form the starting set. In each time step we replace one of the e_i. But, we keep track of the ordering and do not convert these to sets. So, we can talk about e_i^t as element i at time t. In particular S_t is simply the *unordered* collection $\{e_1^t, \dots, e_k^t\}$. Let's say that $X = \{f_1, \dots, f_k\}$. Then to have $S_t = X$, there must be some permutation of $f_1, \dots, f_k$ that equals $e_1^t, \dots, e_k^t$. We will show that for any such permutation σ, the following bound holds

$$\Pr[e_1^t = f_{\sigma(1)}, \dots, e_k^t = f_{\sigma(k)} \mid \tau \le t] \le \alpha^k \cdot \mu(X). \tag{11.1}$$

Since there are $k!$ many permutations, the proving this is sufficient. So we fix an arbitrary permutation; without loss of generality, we may take the identity permutation. We try to bound the following

$$\Pr\left[e_1^t = f_1, \dots, e_k^t = f_k \mid \tau \le t\right].$$

Since we are conditioning on $\tau \le t$, note that there must be some time $\tau_i \le t$, which is the last time before t where element i gets replaced by the down-up random walk. We will bound the above probability, even conditioned on $\tau_1, \dots, \tau_k$ having any set of fixed values up to t. Note that the index of the element that gets replaced in every step is uniformly random and independent of everything else that happens in the random walk, in particular the identity of the elements that come in as replacements. In the rest of the proof, we condition on the indices of the elements that get replaced at every step up to time t; note that this also uniquely determines $\tau_1, \dots, \tau_k$, so we assume $\tau_1, \dots, \tau_k$ are some fixed time indices. Without loss of generality, assume that $\tau_1 < \dots < \tau_k$. We will use induction to prove the following statement for $i = 0, \dots, k$:

$$\Pr\left[e_1^{\tau_1} = f_1, \dots, e_i^{\tau_i} = f_i \mid \text{replacement indices}\right] \le \alpha^i \Pr_{U \sim \mu}\left[f_1, \dots, f_i \in U\right].$$

Notice that for $i = 0$, both sides are trivially equal to 1, and for $i = k$, this inequality is exactly Eq. (11.1).

It remains to show the inductive step. We will show that going from $i-1$ to i, the left-hand side gets multiplied by a smaller quantity compared to the right-hand side. If we have the inequality below in hand, then it is not hard to see that we can complete the induction, since the factors that get multiplied on each side are the two sides of this inequality.

$$\begin{aligned}&\Pr[e_i^{\tau_i} = f_i \mid e_1^{\tau_1} = f_1, \dots, e_{i-1}^{\tau_{i-1}} = f_{i-1} \text{ and replacement indices}]\\ &\le \alpha \cdot \Pr_{U \sim \mu}\left[f_i \in U \mid f_1, \dots, f_{i-1} \in U\right].\end{aligned}$$

Instead of conditioning only on $f_1, \dots, f_{i-1}$ being chosen at the appropriate times on the left-hand side, we will refine the conditioning and condition on the *history of the random walk* up to time $\tau_i - 1$. This means we can in particular assume that the elements $e_{i+1}^{\tau_i}, \dots, e_k^{\tau_i}$ are fixed, that $e_1^{\tau_i} = f_1, \dots, e_{i-1}^{\tau_i} = f_{i-1}$, and the only uncertain thing is what element i is being replaced by at time τ_i.

Let $S = \left\{f_1, \dots, f_i, e_{i+1}^{\tau_i}, \dots, e_k^{\tau_i}\right\}$. Then the conditional probability of choosing f_i at time τ_i is

$$\Pr[e_i^{\tau_i} = f_i \mid e_1^{\tau_1} = f_1, \dots, e_{i-1}^{\tau_{i-1}} = f_{i-1} \text{ and replacement indices}] = \frac{\mu(S)}{\sum_{V \supset S - f_i} \mu(V)}.$$

On the other hand,

$$\Pr_{U \sim \mu}\left[f_i \in U \mid f_1, \dots, f_{i-1} \in U\right] = \frac{\sum_{U \ni f_1, \dots, f_i} \mu(U)}{\sum_{T \ni f_1, \dots, f_{i-1}} \mu(T)}.$$

So we have to show the following:

$$\mu(S)\left(\sum_{T \ni f_1, \dots, f_{i-1}} \mu(T)\right) \le \alpha\left(\sum_{V \supseteq S - f_i} \mu(V)\right)\left(\sum_{U \ni f_1, \dots, f_i} \mu(U)\right). \tag{11.2}$$

We will give an injection from the terms on the left-hand side to the terms in the expanded form of the right-hand side. Choose some set $T \ni f_1, \ldots, f_{i-1}$. Apply the α-approximate exchange property to S and T with the element $f_i \in S$. We get that there must be some element $e \in T$ such that

$$\mu(S)\mu(T) \leq \alpha \cdot \mu(S - f_i + e)\mu(T + f_i - e).$$

Note that $V \stackrel{\text{def}}{=} S - f_i + e$ contains $S - f_i$, and $U \stackrel{\text{def}}{=} T + f_i - e$ contains $\{f_1, \ldots, f_i\}$. So $\mu(U)\mu(V)$ appears on the right-hand side of the desired inequality. So for each T appearing on the left-hand side of the desired inequality we produced a pair of U and V. Note that this mapping from T to (U, V) is injective. This is because given (U, V), we can recover T as the symmetric difference of the other three sets, that is $T = S \Delta U \Delta V$. □

11.2 Establishing Approximate Exchange Properties

In this section, we prove approximate exchange properties in a variety of settings. We also prove Lemma 11.1.1, namely that if $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ has a log-concave generating polynomial g_μ, then μ has a $2^{O(k)}$-approximate exchange property. We conjecture that a $k^{O(1)}$-approximate exchange property holds, but even if true, this will not improve the mixing time results in this chapter beyond constants hidden in the $O(\cdot)$ notation.

Although we do not directly need it, we first give an example where approximate exchange can be proven by elementary means. This is the class of k-determinantal point processes [BBL09; KT12].

Proposition 11.2.1. *Suppose that $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is defined as*

$$\mu(S) = \det\left([v_i]_{i \in S}\right)^2,$$

for some vectors $v_1, \ldots, v_n \in \mathbb{R}^k$. Then μ has a k^2-approximate exchange property.

Proof. It is enough to consider the case where S and T are disjoint; otherwise, the problem can be reduced to lower values of k by taking out the intersection, and projecting all vectors on the orthogonal complement of the space spanned by the intersection.

Define the number β_j as $\sqrt{\mu(S - i + j)\mu(T + i - j)}$ and let α be $\sqrt{\mu(S)\mu(T)}$. The *Plücker Relations for the Grassmanian* [see, e.g., Abe80] say that a signed sum of α and β_j is zero:

$$\alpha + \sum_{j \in T} \pm \beta_j = 0.$$

This means that there is at least one j such that $|\beta_j| \geq \frac{1}{k}\alpha$, and this concludes the proof. □

We now proceed to prove Lemma 11.1.1. Our strategy is to first prove the case $k = 2$ by using log-concavity of g_μ; note that $k = 1$ is trivial. We will then use induction to prove the general case. We remark that the type of induction we use is a standard procedure used in many other places, such as in the context of proving the Plücker Relations and $M^\natural$-concavity [MS18].

Before delving into the proof, note that we can always assume $S \cap T = \emptyset$. This is because we can always condition the distribution μ on having any set of elements, and then throwing out those elements; this operation corresponds to taking partial derivatives of g_μ which results in a log-concave polynomial by Lemma 5.3.4. In particular, we can condition μ on having $S \cap T$, and then throwing out $S \cap T$ from the ground set.

Proof of Lemma 11.1.1 for $k = 2$. When $k = 2$, we might as well assume that $n = 4$, because no element outside of $S \cup T$ is important, and we can condition the distribution μ on not having those elements. This corresponds to substituting 0 for variables outside $S \cup T$ in g_μ which preserves log-concavity.

So our goal now is to show that for a log-concave quadratic polynomial in four variables

$$g_\mu = \sum_{\{i,j\} \in \binom{[4]}{2}} \mu(\{i, j\}) \cdot z_i z_j,$$

we have an $O(1)$-exchange property. Without loss of generality assume that $S = \{1, 2\}$ and $T = \{3, 4\}$.

Let us consider $\nabla^2 g_\mu$. This is a constant matrix, which has at most one positive eigenvalue by Proposition 5.0.3. On the other hand it is a matrix with nonnegative entries, so it must have at least one nonnegative eigenvalue as well. Analyzing the possible signs of the eigenvalues, we see that their product, i.e. the determinant is nonpositive:

$$\det(\nabla^2 g_\mu) \leq 0.$$

This determinant can be written in a special way. Let us define:

$$A \stackrel{\text{def}}{=} \mu(\{1,2\})\mu(\{3,4\}),$$

$$B \stackrel{\text{def}}{=} \mu(\{1,3\})\mu(\{2,4\}),$$

$$C \stackrel{\text{def}}{=} \mu(\{1,4\})\mu(\{2,3\}).$$

Notice that approximate exchange for S, T any any $i \in S$ is equivalent to saying that $A \leq O(1) \cdot \max\{B, C\}$. We can write $\det(\nabla^2 g_\mu) = A^2 + B^2 + C^2 - 2(AB + AC + BC)$. So we get the inequality

$$A^2 + B^2 + C^2 \leq 2(AB + AC + BC).$$

This is the same as

$$(A - B - C)^2 \leq 4BC.$$

Taking square-roots we get

$$A - B - C \leq 2\sqrt{BC},$$

which is the same as saying

$$A \leq (\sqrt{B} + \sqrt{C})^2.$$

Taking square-roots again we get

$$\sqrt{A} \leq \sqrt{B} + \sqrt{C}.$$

In particular one of $\sqrt{B}$ and $\sqrt{C}$ must be at least $\frac{1}{2}\sqrt{A}$. This proves that μ satisfies a $2^2 = 4$-approximate exchange property for $S = \{1,2\}$ and $T = \{3,4\}$. □

We now complete the proof by inducting on k.

Proof of Lemma 11.1.1 for General k. We can assume that for any S, T such that $|S \cap T| \geq 1$, we have a $2^{O(k-|S\cap T|)}$-approximate exchange property. This is because by the arguments we had, such nonempty intersections can be reduced to smaller values of k by conditioning and throwing out $S \cap T$.

Now let $S \cap T = \emptyset$ and let $i \in S$ be given. Our goal is to find j such that

$$\mu(S)\mu(T) \leq 2^{O(k)}\mu(S - i + j)\mu(T + i - j).$$

Let $i' \neq i$ be another, arbitrary, element of S. We will exchange i' with an element $j' \in T$ and use induction on $S - i' + j'$ and T. We need to be careful how we choose j' though. Let us choose j' to be the element of T that maximizes the expression $\mu(T + i - j')\mu(S - i' + j')$. The reason for this choice will become apparent in the rest of he proof.

Then the sets $S - i' + j'$ and T have an intersection of one element, so by induction we know an approximate exchange property for them. Therefore, there must be a $j \in T$ such that

$$\mu(S - i' + j')\mu(T) \leq 2^{O(k-1)}\mu(S - i - i' + j + j')\mu(T + i - j). \tag{11.3}$$

We will apply approximate exchange a second time. The sets S and $S - i - i' + j + j'$ have a very large intersection. In particular their exchange property reduces to the case of $k = 2$ of Lemma 11.1.1, which we have already proven. By this exchange property, we have

$$\begin{aligned} &\mu(S)\mu(S - i - i' + j + j') \\ &\leq 2^{O(1)} \max\{\mu(S - i + j)\mu(S - i' + j'), \mu(S - i + j')\mu(S - i' + j)\}. \end{aligned} \tag{11.4}$$

If the first term in Eq. (11.4) achieves the maximum, then we are done, because multiplying Eqs. (11.3) and (11.4) yields

$$\begin{aligned} &\mu(S - i' + j')\mu(T)\mu(S)\mu(S - i - i' + j + j') \\ &\leq 2^{O(k)} \cdot \mu(S - i - i' + j + j')\mu(T + i - j)\mu(S - i + j)\mu(S - i' + j'), \end{aligned}$$

which simplifies to

$$\mu(S)\mu(T) \le 2^{O(k)}\mu(S-i+j)\mu(T+i-j),$$

showing that i can be exchange for j.

So assume that the second term in Eq. (11.4) achieves the maximum. We will show that in this case i can be exchanged for j'. Multiplying Eqs. (11.3) and (11.4) yields

$$\begin{aligned}&\mu(S-i'+j')\mu(T)\mu(S)\mu(S-i-i'+j+j')\\ &\le 2^{O(k)}\mu(S-i-i'+j+j')\mu(T+i-j)\mu(S-i+j')\mu(S-i'+j),\end{aligned}$$

which simplifies to

$$\mu(S)\mu(T) \le 2^{O(k)}\cdot\mu(S-i+j')\mu(T+i-j')\frac{\mu(T+i-j)\mu(S-i'+j)}{\mu(T+i-j')\mu(S-i'+j')}$$

Notice that by our choice of j', the fraction appearing on the right-hand side is ≤ 1. So we can conclude that

$$\mu(S)\mu(T) \le 2^{O(k)}\mu(S-i+j')\mu(T+i-j').$$

□

If we require the stricter assumption that g_μ is real stable, then we obtain an exponential improvement. This is a generalization of Proposition 11.2.1.

Lemma 11.2.2. *Fix $\mu : \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ and suppose its multivariate generating polynomial g_μ is real stable. Then for every $S,T \in \binom{[n]}{k}$ and $i \in S \setminus T$*

$$\sqrt{\mu(S)\mu(T)} \le \sum_{j\in T\setminus S}\sqrt{\mu(S-i+j)\mu(T+i-j)}. \tag{11.5}$$

Consequently, there exists $j \in T \setminus S$ such that

$$\mu(S)\mu(T) \le k^2\mu(S-i+j)\mu(T+i-j). \tag{11.6}$$

Thus μ satisfies a k^2-approximate exchange property. Moreover, for $S \in \binom{[n]}{k}$ and $j \notin S$,

$$\mu(S)\mu(j) \le k\sum_{e\in S}\mu(S+j-e)\mu(e) \tag{11.7}$$

where $\mu(t) = \sum_{T\in\binom{[n]}{k}:t\in T}\mu(T)$ for $t \in \{j,e\}$.

To prove it, we need the following theorem about univariate Hurwitz stable polynomials due to [Asn70].

Theorem 11.2.3. *Consider a Hurwitz stable polynomial $f(z) = \sum_{i=0}^n a_i z^i$ with nonnegative coefficients. Define its Hurwitz matrix $H = (h_{ij}) \in \mathbb{R}^{n\times n}$ by $h_{ij} = a_{2j-i}$ for $0 \le 2j-i \le n$, and $h_{ij} = 0$ otherwise. Then H is totally nonnegative, in the sense that all its minors are nonnegative.*

As an immediate consequence, we obtain the following lemma about coefficients of univariate Hurwitz stable polynomials.

Lemma 11.2.4. *Consider a Hurwitz stable polynomial $f(z) = \sum_{i=0}^{2t-1} a_i z^i$ with nonnegative coefficients. Then $a_{2t-1}a_0 \le a_{2t-2}a_1$.*

Proof. By total-nonnegativity of the Hurwitz matrix H, we have

$$\det\begin{bmatrix}h_{1,1} & h_{1,t}\\ h_{2,1} & h_{2,t}\end{bmatrix} = \det\begin{bmatrix}a_1 & a_{2t-1}\\ a_0 & a_{2t-2}\end{bmatrix} = a_1a_{2t-2} - a_0a_{2t-1} \ge 0.$$

□

We are ready to prove Lemma 11.2.2. The idea is to construct a Hurwitz stable polynomial whose coefficients correspond to the left-hand side and right-hand side of (11.5), then use Lemma 11.2.4 to derive (11.5).

Proof of Lemma 11.2.2. We first show that (11.5) implies (11.6) and (11.7). Indeed,

$$\sqrt{\mu(S)\mu(T)} \le \sum_{j\in T\setminus S} \sqrt{\mu(S-i+j)\mu(T+i-j)} \le k \max_{j\in T\setminus S} \sqrt{\mu(S-i+j)\mu(T+i-j)}.$$

For $j \notin S$ and $T \in \binom{[n]}{k}$ containing j, using (11.5) and Hölder's Inequality, we get

$$\mu(S)\mu(T) \le \left(\sum_{e\in S\setminus T} \sqrt{\mu(S-e+j)\mu(T+e-j)}\right)^2 \le k \sum_{e\in S\setminus T} \mu(S-e+j)\mu(T+e-j).$$

Summing over all such T, while observing that $\sum_{T\in\binom{[n]}{k}:j\in T} \mu(T+e-j) \le \mu(e)$, gives

$$\mu(S)\mu(j) = \sum_{T\in\binom{[n]}{k}:j\in T} \mu(S)\mu(T) \le k \sum_{T\in\binom{[n]}{k}:j\in T} \sum_{e\in S\setminus T} \mu(S-e+j)\mu(T+e-j) \le k\sum_{e\in S} \mu(S-e+j)\mu(e).$$

Any $\mu : \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ with a real stable generating polynomial can be approximated by a strictly real stable $\tilde{\mu} : \binom{[n]}{k} \to \mathbb{R}_{>0}$, in the sense that $|\tilde{\mu}(S) - \mu(S)| < \epsilon$ where ϵ can be made arbitrarily small. This statement appears in [NUI69] and [BH20, Prop. 2.2]. We can prove the lemma for $\tilde{\mu}$ then send $\epsilon \to 0$ to get the corresponding inequality for μ. Thus, we can assume $\mu(S) > 0$ for all $S \in \binom{[n]}{k}$.

We deal with the case when $S \cap T = \emptyset$ and $[n] = S \cup T$. Other cases can be reduced to this scenario by setting z_i to 0 for $i \notin S \cup T$, and taking derivatives with respect to $i \in S \cap T$. Let $t \stackrel{\text{def}}{=} |S| = |T|$; then $n = 2t$. We can rewrite f as

$$f(z_1, \dots, z_{2t}) = \sum_{W\in\binom{[2t]}{t}} \mu(W) z_W$$

with $\mu(W) > 0$ for all W.

For $W \in \binom{[2t]}{t}$ let $\Delta(W) \stackrel{\text{def}}{=} \sqrt{\mu(W) \times \mu([2t] \setminus W)}$. Fix $e \in T$. We want to show

$$\sum_{i\in S} \Delta(S+e-i) \ge \Delta(S).$$

Since f is homogeneous and real stable, it is also Hurwitz stable. We set the variables of f as follows.

1. Set $z_e = 1$.
2. For all $i \in S$, set $z_i = z^{-1}\delta_i$ where $\delta_i = \sqrt{\frac{\mu(S+e-i)}{\mu(T-e+i)}}$.
3. Let $B = \prod_{i\in S} \delta_i > 0$, and for all $j \in T - e$, set $z_j = zB^{\frac{1}{t-1}}$.

Finally, after multiplying f by $B^{-1}z^t$, we obtain a Hurwitz stable $\tilde{f}(z)$ with positive coefficients and degree $2t-1$. We rewrite $\tilde{f}(z) = b_0 z^{2t-1} + b_1 z^{2t-2} + \cdots + b_{2t-2}z_1 + b_{2t-1}$.

Note that the monomial $z^W\mu(W)$ in f contributes to b_{2t-2} if and only if $|S\Delta W| = 2$ and $e \in W$ i.e. $W = S - i + e$ for some $i \in S$. Similarly, $z^W\mu(W)$ contributes to b_1 if and only if $|S\Delta W| = 2t-2$ and $e \notin W$ i.e. $W = T - e + i$ for some $i \in S$.

A routine calculation gives

$$b_{2t-1} = \mu(S)B^{-1}\prod_{i\in S}\delta_i = \mu(S)$$

$$b_0 = \mu(T)\prod_{j\in T-e} B^{\frac{1}{t-1}} = \mu(T)$$

$$b_{2t-2} = B^{-1}\sum_{i\in S}\left(\mu(S-i+e)\prod_{j\in S-i}\delta_j\right) = \sum_{i\in S}\mu(S-i+e)\delta_i^{-1} = \sum_{i\in S}\Delta(S+e-i)$$

$$b_1 = B^{-1}\sum_{i\in S}\left(\mu(T-e+i)\delta_i\prod_{j\in T\setminus e} B^{1/(t-1)}\right) = \sum_{i\in S}\Delta(S+e-i)$$

Since $\tilde{f}$ is Hurwitz stable with nonnegative coefficients, Lemma 11.2.4 implies $b_1 b_{2t-2} \geq b_0 b_{2t-1}$ i.e. $\left(\sum_{i \in S} \Delta(S + e - i)\right)^2 \geq \Delta(S)^2$ as required. □

Chapter 12

Fast Algorithms for Sampling Forests and Spanning Trees

In this chapter, we show how to leverage down-up random walks to approximately sample random forests or random spanning trees in a graph with n edges in time $O(n \log^2 n)$; for applications of sampling random forests, see e.g. [Goe+15]. The best previous results [Ana+19; CGM21] for sampling from all forests employ approximate sampling to approximate counting reductions [JVV86; AD20], which introduce a large polynomial blow-up in running time.

In addition, we show a similar algorithm that also runs in quasi-linear time and samples from the uniform distribution over spanning trees of G, improving upon the almost-linear time algorithm by [Sch18]. Much attention has been paid to the problem of sampling random spanning trees over the years, starting from the seminal works of [Ald90; Bro89] who proposed a simple routine to extract a random spanning tree from the trace of a random walk on G itself. Subsequent works introduced improved algorithms [Wil96; CMN96; KM09; MST15; Dur+17a; Dur+17b] until finally [Sch18] managed to obtain an almost-linear time algorithm, running in time $n^{1+o(1)}$ on graphs with n edges. This algorithm and that of several prior works were all based on the original approach of [Ald90; Bro89]; they achieved an improved running time by employing several clever, but complicated, tricks to shortcut the trace of a random walk over G. Our algorithms for sampling a random spanning tree or a random forest are wholly different, being based on the down-up random walk. They achieve a nearly-linear running time of $O(n \log^2 n)$, while being arguably much simpler to describe and implement.

Our analysis works on *weighted* graphs too, and is the first to achieve nearly-linear running time for these problems. Our algorithms can be naturally extended to support approximate sampling from random forests of size between k_1 and k_2 in time $O(n \log^2 n)$, for fixed parameters k_1, k_2, as well as approximate sampling from random independent sets of a matroid $\mathcal{M}$ of rank k on a ground set of n elements using $O(kn \log k)$ calls to an independence oracle of $\mathcal{M}$.

Theorem 12.0.1. *There is a randomized algorithm that takes a weighted graph $G = (V, E)$ on n edges with weight function $w : E \to \mathbb{R}_{\geq 0}$, parameters $q \geq 0$ and $\epsilon > 0$ as input, and outputs a (random) forest $F \subseteq E$ in time $O(n \log(n) \log(n/\epsilon))$. The distribution of F is guaranteed to be ϵ-close in total variation distance to the distribution μ over forests of G defined by*

$$\mu(F) \propto q^{k-|F|} w^F$$

where k is the rank of the graphic matroid of G, and $|F|$ denotes the number of edges in F. In particular, when $w(e) = 1$ for all e, μ is the uniform distribution on forests of G if $q = 1$, and is the uniform distribution on spanning trees of G if G is connected and $q = 0$.

In fact, we can extend Theorem 12.0.1 to allow sampling from the uniform distribution over forests of size between k_1 and k_2, for any parameters k_1, k_2, in quasi-linear time.

Theorem 12.0.2. *There is a randomized algorithm that takes a weighted graph $G = (V, E)$ on n edges with weight function $w : E \to \mathbb{R}_{\geq 0}$, parameters $q \geq 0$, $k_1, k_2 \in \mathbb{N}$ and $\epsilon > 0$ as input, and outputs a (random) forest $F \subseteq E$ in time $O(n \log(n) \log(n/\epsilon))$. The distribution of F is guaranteed to be ϵ-close in total variation distance to the distribution $\mu^{(k_1,k_2)}$ over forests of G defined by*

$$\mu^{(k_1,k_2)}(F) \propto q^{k_2-|F|} w^F$$

restricted to forests of size $|F| \in [k_1, k_2]$.

Since our algorithms are based on Markov chains, they can only approximately sample from the forest or spanning tree distribution. In contrast, some of the prior works, including [Sch18], can sample *exactly* from the spanning tree distribution. This is mostly an inconsequential difference in practice, as no polynomial-time user of the algorithm can detect a difference between exact sampling and approximate sampling; one simply needs to set ϵ to be inverse-polynomially small.

We remark that our technique also leads to algorithm(s) that perform the more general task of approximately sampling from the uniform distribution over the family of independent sets of an arbitrary matroid, given access to suitable oracles. Specifically, for a matroid $\mathcal{M} = ([n], \mathcal{X})$ of rank-k, an algorithm similar to the one from Theorem 12.0.1 samples from a distribution that is ϵ-close to the uniform distribution over the independent set $\mathcal{X}$ using $O(n \log(n/\epsilon))$ calls to a data structure $\mathcal{O}'$ that maintains a set $S \subseteq [n]$, guaranteed to contain at most one circuit and supports the following operations:

- Addition and removal of an element from S, provided we maintain the property that S contains at most one circuit.
- Outputting a uniformly random element from the unique circuit in S if such a circuit exists.

For graphic matroids, we can implement $\mathcal{O}'$ with amortized quasi-constant query time using *link-cut trees* [ST83; RTF18]. In general, since the input S is guaranteed to have size at most $k+1$, we can implement each call to $\mathcal{O}'$ using $O(k)$ calls to the more familiar *independence oracle* $\mathcal{O}_I$ for $\mathcal{M}$, resulting in a $O(kn \log(n/\epsilon))$-time algorithm.

Overview of the Approach Let us first discuss the high-level ideas for proving Theorem 12.0.1. For simplicity sake, we consider the unweighted case i.e. $w(e) = 1$ for all $e \in E$. It will also be helpful to first discuss the special case $q = 0$, G is connected, and μ is the uniform distribution over spanning trees of G. We would like to use the down-up random walk from Theorem 11.0.1 to sample from μ.

Though the down-up walk on the support of μ mixes in nearly-linear time, we do not see a way to implement each step of it in $\mathsf{polylog}(n)$ time. Fortunately, the down-up random walk on an equivalent family of sets, the *dual* of the graphic matroid of G, which consists of the *complements* of spanning trees, also mixes in nearly-linear time, and we can implement each step in amortized $O(\log n)$-time using *link-cut trees* [ST83; RTF18].

For $q \neq 0$, the distribution μ over forests of G is not *homogeneous* i.e. the support of μ contains different-size subsets of E, so we cannot immediately apply Theorem 11.0.1. Let $\overline{\mu}$ be the *complement* distribution of μ i.e. $\overline{\mu}(E \setminus F) = \mu(F)$ if F is a forest. Then sampling from μ and from $\overline{\mu}$ are equivalent. We add auxiliary elements to each $\overline{F} \in \mathrm{supp}(\overline{\mu})$ to obtain a homogeneous distribution. More precisely, we design a homogeneous distribution $\mu^{\uparrow} : \binom{E \cup Y}{n} \to \mathbb{R}_{\geq 0}$ whose projection to E is $\overline{\mu}$, i.e. $\Pr_{T \sim \mu^{\uparrow}}[T \cap E] = \overline{\mu}(T \cap E)$ where Y is the set of auxiliary elements, such that the generating polynomial of $\mu^{\uparrow}$ is log-concave. Specifically, in Lemma 12.1.1, we prove that for any matroid $\mathcal{M}$ of rank-r over ground set $[n]$, the polynomial $f_{\mathcal{M}}(y, z_1, \ldots, z_n) = \sum_{S \in \mathcal{X}} y^{|S|} z^{[n] \setminus S}$ is log-concave, then use *polarization* (see Proposition 5.3.3) to transform $f_{\mathcal{M}}$ into a multiaffine homogeneous log-concave polynomial

$$f^{\uparrow}_{\mathcal{M}}(y_1, \ldots, y_r, z_1, \ldots, z_n) = \sum_{S \in \mathcal{X}, T \in \binom{[r]}{|S|}} \frac{1}{\binom{r}{|S|}} y^T z^{[n] \setminus S}.$$

The distribution $\mu^{\uparrow}$ has generating polynomial $f^{\uparrow}_{\mathcal{M}}$. Our algorithm runs the down-up random walk on $\mu^{\uparrow}$, which mixes fast by Theorem 11.0.1, then outputs $E \setminus (T_t \cap E)$ where $T_t \in \mathrm{supp}(\mu^{\uparrow})$ is the random set we obtained after $t = O(n \log(n/\epsilon))$ down-up steps. Each step of the walk, even in the weighted case, can again be implemented in amortized $O(\log n)$-time using link-cut trees [ST83; RTF18].

If we only consider the effect of the down-up walk on $\overline{T}_{t,E} \stackrel{\mathsf{def}}{=} E \setminus (T_t \cap E)$, then each step of the down-up walk can be viewed as follows:

- With probability $1 - \frac{|\overline{T}_{t,E}|}{n}$, sample an edge $e \notin \overline{T}_{t,E}$ uniformly at random and add e to $\overline{T}_{t,E}$

- If there is a cycle formed in $\overline{T}_{t,E}$ by the previous operation, remove an edge uniformly at random from the cycle. Else, with probability $\frac{q}{1+q}$, remove an edge uniformly at random from $\overline{T}_{t,E}$. Note that this has no effects if $\overline{T}_{t,E}$ is already empty.

Observe that if $q = 0$, we never remove an edge from $\overline{T}_{t,E}$ unless $\overline{T}_{t,E}$ contains a cycle, thus if $\overline{T}_{0,E}$ is a spanning tree then so is $\overline{T}_{t,E}$ for all t. For $q = 0$, our algorithm (to sample random spanning tree) is the same as the one proposed by [RTF18]. Despite not having the tight mixing time analysis, they empirically observed fast mixing times for the proposed algorithm, and additionally showed how link-cut trees can be used to implement each step.

To prove Theorem 12.0.2, we only need to show that the following polynomial is completely log-concave

$$f^{k_1}_{\mathcal{M}}(z_0, z_1, \ldots, z_n) = \sum_{S\in\mathcal{X}, |S|\geq k_1} z_0^{|S|} z^{[n]\setminus S}$$

and then set $\mathcal{M}$ to be the matroid whose bases are the size-k_2 forests of graph G. We then employ the same *polarization* trick and run the down-up walk framework just like how we prove Theorem 12.0.1.

12.1 Some New Log-Concave Polynomials: Duality, Polarization, and Other Tricks

This section is devoted to establishing log-concavity for the polynomials we need to prove Theorems 12.0.1 and 12.0.2.

Lemma 12.1.1. *Let $\mathcal{M} = ([n], \mathcal{X})$ be a matroid of rank-r over ground set $[n]$, where $\mathcal{X}$ is its family of independent sets. Then the following polynomials are completely log-concave.*

(a)

$$f_{\mathcal{M}}(y, z_1, \ldots, z_n) = \sum_{S\in\mathcal{X}} y^{|S|} z^{[n]\setminus S}$$

(b)

$$f_{\mathcal{M},q,w}(y, z_1, \ldots, z_n) = \sum_{S\in\mathcal{X}} q^{r-|S|} w^S y^{|S|} z^{[n]\setminus S}$$

where $q \geq 0$ and $w_1, \ldots, w_n > 0$.

(c)

$$f^{\uparrow}_{\mathcal{M},q,w}(y_1, \ldots, y_r, z_1, \ldots, z_n) = \sum_{S\in\mathcal{X}, T\in\binom{[r]}{|S|}} \frac{1}{\binom{r}{|S|}} q^{r-|S|} w^S y^T z^{[n]\setminus S}$$

where $q \geq 0$ and $w_1, \ldots, w_n > 0$.

Proof. We first show that Item (a) implies Item (b), which then implies Item (c). Since

$$f^{\uparrow}_{\mathcal{M},q,w}(y_1, \ldots, y_r, z_1, \ldots, z_n) = \prod_{\kappa}^{\uparrow}(f_{\mathcal{M},q,w})$$

with $\kappa_0 = r$ and $\kappa_i = 1$ for all $i \in [n]$, Item (b) implies Item (c) by Proposition 5.3.3.

We next show that Item (a) implies Item (b). If $q > 0$ then

$$f_{\mathcal{M},q,w}(y, z_1, \ldots, z_n) \propto f_{\mathcal{M}}\left(\frac{y}{q}, \frac{z_1}{w_1}, \ldots, \frac{z_n}{w_n}\right)$$

is completely log-concave since composing with a linear map preserves complete log-concavity. If $q = 0$ then

$$f_{\mathcal{M},q,w}(y, z_1, \ldots, z_n) = \sum_{S\in B(\mathcal{M})} w^S z^{[n]\setminus S} \propto \frac{\partial^r f_{\mathcal{M}}}{\partial y^r}\left(\frac{z_1}{w_1}, \ldots, \frac{z_n}{w_n}\right)$$

is completely log-concave since taking derivative preserves complete log-concavity.

Now we show Item (a). For convenience, let $f \stackrel{\text{def}}{=} f_{\mathcal{M}}$. We first show that $\text{supp}(f)$ is M-convex (see Definition 30). First, the support of $g(y, z_1, \ldots, z_n) = \sum_{S \in \mathcal{X}} y^{n-|S|} w^S$ is M-convex. Note that

$$\text{supp}(f) = \{\vec{v} - \tilde{w} \mid \tilde{w} \in \text{supp}(g)\}$$

where $v_0 = n$ and $v_i = 1$ for all $i \in [n]$. Thus, $\text{supp}(f)$ is also M-convex. Indeed, consider any $\alpha, \beta \in \text{supp}(f)$ and $i \in \{0, \ldots, n\}$ s.t. $\alpha_i < \beta_i$. Then $\vec{v} - \alpha, \vec{v} - \beta \in \text{supp}(g)$ and $(\vec{v} - \alpha)_i > (\vec{v} - \beta)_i$ so there exists j s.t. $(\vec{v} - \alpha)_j < (\vec{v} - \beta)_j$, and $(\vec{v} - \alpha) - \mathbf{1}_i + \mathbf{1}_j$ and $(\vec{v} - \beta) - \mathbf{1}_j + \mathbf{1}_i$ are in $\text{supp}(g)$. This implies $\alpha_j > \beta_j$ and $\alpha - \mathbf{1}_j + \mathbf{1}_i$ and $\beta - \mathbf{1}_i + \mathbf{1}_j$ are in $\text{supp}(f)$.

We now proceed using induction on n. Obviously, for $n = 1$, f is a linear function in y, z_1 with positive coefficients, which is trivially completely log-concave. Suppose the statement is true for all matroids $\mathcal{M}'$ on ground set $[n-1]$ with $n \geq 2$. We only need to verify $\partial^\alpha f$ is log-concave for all α with $|\alpha| = n - 2$. Note that for $i \in [n]$,

$$\partial_i f = \sum_{S \in \mathcal{X} : i \notin S} y^{|S|} z^{([n] \setminus i) \setminus S} = f_{\mathcal{M} \setminus i}$$

is completely log-concave by applying induction hypothesis to $\mathcal{M} \setminus i$. We only need to show $\partial_y^{n-2} f$ is log-concave. Note that $\partial_y^{n-2} f \neq 0$ only if $r = \text{rank}(\mathcal{M}) \geq n - 2$. Also, for $n = 2$, $\partial_y^{n-2} f$ is exactly f.

For $r = n - 2$,

$$\partial_y^{n-2} f = (n-2)! \sum_{S \in \mathcal{B}(\mathcal{M})} z^{[n] \setminus S} = (n-2)! \sum_{S' \in \mathcal{B}(\mathcal{M}^*)} z^{S'}$$

is strongly log-concave (and of degree-2, thus real stable), since it is the sum over the bases of the dual matroid $\mathcal{M}^*$ of $\mathcal{M}$ (see e.g. Example 10).

For $r = n$, then $\mathcal{B}(\mathcal{M}) = \{[n]\}$ and $f = \prod_{i \in [n]} (y + z_i)$ is real stable as product of real stable polynomials $y + z_i$ for $i \in [n]$, thus so is $\partial_y^{n-2} f$.

For $r = n - 1$,

$$\partial_y^{n-2} f = (n-1)! \cdot y \cdot e_1(z_T) + (n-2)! \cdot \left(e_2(z_T) + e_1(z_T) \cdot e_1(z_{[n] \setminus T})\right)$$

where $T \subseteq [n]$ is such that $\mathcal{B}(\mathcal{M}) = \{[n] \setminus i : i \in T\}$ are the bases of $\mathcal{M}$, e_k is elementary symmetric polynomial of degree-k, and z_S is shorthand for $\{z_i : i \in S\}$. This is because $\{S \in \mathcal{X} : |S| = n - 2\}$ is exactly

$$\{[n] - t - t' : t, t' \in T\} \cup \left\{[n] - t - \bar{t} : t \in T, \bar{t} \notin T\right\}.$$

Set $u \stackrel{\text{def}}{=} (n-1)y + e_1(z_{[n] \setminus T})$ then

$$\partial_y^{n-2} f = (n-2)! \left(e_1(z_T) u + e_2(z_T)\right) = (n-2)! \, e_2(u, z_T)$$

Note that e_2 is real stable (Theorem 5.7.1), and that $u \in \mathcal{H}$ whenever $(y, z_{[n] \setminus T}) \in \mathcal{H}^{n+1-|T|}$, where $\mathcal{H}$ is the upper half-plane. Thus $\partial_y^{n-2} f$ is nonzero for any $(y, z_{[n]}) \in \mathcal{H}^{n+1}$ i.e. $\partial_y^{n-2} f$ is real stable. □

Remark 56. Observe that $f_{\mathcal{M}}$ is the *dual* or *complement* of the strongly log-concave polynomial

$$g_{\mathcal{M}}(y, z_1, \ldots, z_n) = \sum_{S \in \mathcal{X}} y^{n-|S|} z^S$$

we saw in Chapter 5, in the sense that

$$f_{\mathcal{M}}(y, z_1, \ldots, z_n) = y^r z^{[n]} g_{\mathcal{M}}\left(y^{-1}, z_1^{-1}, \ldots, z_n^{-1}\right)$$

We remark that while the dual of a real stable polynomial is real stable, the dual of a strongly log-concave polynomials is not in general necessarily strongly log-concave.

Lemma 12.1.2. *For any matroid $\mathcal{M}$ of rank-r over ground set $[n]$ and parameter $h \in \mathbb{N}$, the following polynomial is completely log-concave*

$$f_{\mathcal{M}}^h(y, z_1, \cdots, z_n) = \sum_{S \in \mathcal{X} : |S| \geq h} y^{|S|} z^{[n] \setminus S}.$$

Proof. Note that if $h > r$ then $f^h_{\mathcal{M}} \equiv 0$, and if $h = r$ then $f^h_{\mathcal{M}}$ is completely log-concave since it is the bases generating polynomial of the dual matroid of $\mathcal{M}$. Below, assume $h < r$.

Lemma 12.1.1 implies $f^0_{\mathcal{M}}$ is completely log-concave. Now, $f^h_{\mathcal{M}}$ and $y^h \partial^h_y f^0_{\mathcal{M}}$ have the same support, and this support is M-convex since $y^h \partial^h_y f^0_{\mathcal{M}}$ is Lorentzian (see Lemma 5.3.4 and Proposition 5.3.2 Item 3).

We show $f^h_{\mathcal{M}}$ is Lorentzian by inducting on n as in Lemma 12.1.1. We only need to check $\partial^{n-2}_y f^h_{\mathcal{M}}$ is Lorentzian. If $h \le n-2$, then $\partial^{n-2}_y f^h_{\mathcal{M}}$ is exactly $\partial^{n-2}_y f^0_{\mathcal{M}}$, thus is Lorentzian because $f^0_{\mathcal{M}}$ is Lorentzian. The only remaining case is $h = n-1$ and $r = n$. But $\partial^{n-2}_y f^h_{\mathcal{M}} = (n-1)! y \sum_{i=1}^n z_i + \frac{n!}{2} y^2 \propto y(\sum_{i=1}^n z_i + \frac{n}{2} y)$ is real stable. □

12.2 Nearly-Linear Time Samplers

In this section we prove Theorems 12.0.1 and 12.0.2. We remark that for these random spanning tree and random forest results, Corollary 9.3.4 from Chapter 9 was already sufficient, and we do not truly need Theorem 11.0.1. However, for fast sampling from all independent sets of a general matroid which is not necessarily graphic, Theorem 11.0.1 is needed to obtain the fastest possible mixing times.

Let μ be the distribution over forests of G defined in Theorem 12.0.1. In Lemma 12.1.1, we showed the homogeneous multiaffine log-concave polynomial $f^{\uparrow}_{\mathcal{M},q,w}$ that generates a homogeneous distribution $\mu^{\uparrow} : \binom{E \cup Y}{n} \to \mathbb{R}_{\ge 0}$ whose projection to E is the *complement* distribution of μ, i.e., $\Pr_{T \sim \mu^{\uparrow}}[T \cap E] = \mu(E \setminus (T \cap E))$.

We then run the down-up random walk on the distribution $\mu^{\uparrow}$ for some t steps, and obtain a random set $T_t \in \mathrm{supp}(\mu^{\uparrow})$. We argue that the distribution of $E \setminus (T_t \cap E)$ is ϵ-close to μ for some $t = O\left(|E| \log \frac{|E|}{\epsilon}\right)$ using the mixing time bound proved in Theorem 11.0.1. For completeness, we briefly discuss how to implement each step of the random walk in $O(\log |E|)$ time.

We are ready to prove Theorem 12.0.1.

Proof of Theorem 12.0.1. Let $\mathcal{M}$ be the graphic matroid on graph $G = (V, E)$ with n edges. Let $k \stackrel{\mathsf{def}}{=} \mathrm{rank}(\mathcal{M})$. Without loss of generality, we can label the edges by $1, 2 \ldots, n$ and assume $E = [n]$.

Let μ be the distribution over independent sets of $\mathcal{M}$ (i.e. forest of G), where $\mu(F) \propto q^{k-|F|} w^F$ for $F \in \mathcal{X}$. Note that we can remove all edges of weight 0 from E without changing μ. Without loss of generality, we assume this is already done; thus $w(e) > 0$ for all $e \in E$.

Let $Y \stackrel{\mathsf{def}}{=} \{y_1, \ldots, y_k\}, Z \stackrel{\mathsf{def}}{=} \{z_1, \ldots, z_n\}$. We identify the variable z_i with the edge labeled by i. Let $f \stackrel{\mathsf{def}}{=} f^{\uparrow}_{\mathcal{M},q,w}$ then f is multiaffine, homogeneous, and completely log-concave by Lemma 12.1.1. Observe that f is the generating polynomial for the distribution $\mu^{\uparrow} : \binom{Y \cup Z}{n} \to \mathbb{R}_{\ge 0}$ defined by $\mu^{\uparrow}(T) \propto \frac{q^{k-|T \cap Y|}}{\binom{k}{|T \cap Y|}} w^{Z \setminus T}$ if $Z \setminus T \in \mathcal{X}$, and 0 otherwise.

We run the down-up walk starting from $T_0 \in \mathrm{supp}(\mu^{\uparrow})$ (e.g. $T_0 = (Z \setminus \mathcal{F}) \cup Y$ for some spanning forest $\mathcal{F}$). Let $\nu^{\uparrow}$ be the distribution of the set $T \in \mathrm{supp}(\mu^{\uparrow})$ we obtained after $O(n \log(n/\epsilon))$ steps; Theorem 11.0.1 implies $\left\|\mu^{\uparrow} - \nu^{\uparrow}\right\|_{\mathsf{TV}} \le \epsilon$. We then remove all y_j from S i.e. collapse $T \subseteq Y \cup Z$ to $T_Z \stackrel{\mathsf{def}}{=} Z \setminus (T \cap Z)$. Let ν be the distribution of T_Z. Clearly, $\mathrm{supp}(\nu) = \mathcal{X}$, and if $\nu^{\uparrow}$ is the same distribution as $\mu^{\uparrow}$, then ν is the same as μ. By the Data Processing Inequality (see Theorem 1.4.12), the total variation distance between ν and μ is at most ϵ since $\|\mu - \nu\|_{\mathsf{TV}} \le \left\|\mu^{\uparrow} - \nu^{\uparrow}\right\|_{\mathsf{TV}} \le \epsilon$.

Now, we show each step of the random walk can be implemented in $O(\log n)$ time. In the down step, we keep track of whether the level-$(n-1)$ set S is still such that $S_Z \stackrel{\mathsf{def}}{=} Z \setminus S$ is a forest. Note that if we dropped a y_j in the down step to arrive at S, then S_Z is always a forest; if, instead, we dropped a z_i (equivalently, added z_i to S_Z), then we can check whether S_Z stays a forest in $O(\log n)$-amortized time using *link-cut trees* [ST83; RTF18]. If S_Z is not a forest, then let $\mathcal{C}_S$ be the unique cycle in S_Z which contains the edge z_i that was added to S_Z. When we perform an up-step from S, if S_Z is not a forest, select z_f among the edges f in $\mathcal{C}_S$ with probability $\propto 1/w_f$ and add it to S (equivalently, remove z_f from S_Z). This can again be done in $O(\log n)$-amortized time (see [RTF18]). If S_Z is empty, then we can only add y_j which is not already in $S_Y \stackrel{\mathsf{def}}{=} S \cap Y$ with uniform probability. If S_Z is a nonempty forest, then we can add any variable $y_j \notin S_Y$ or $z_i \in S_Z$. In this case, the probability of adding variable can be explicitly computed i.e. uniform

among $y_j \notin S_Y$, and $\Pr[z_i] / \Pr[y_j] = \frac{q^{k-(\ell-1)} w_i^{-1} / \binom{k}{\ell-1}}{q^{k-\ell}/\binom{k}{\ell}}$ where $\ell \stackrel{\text{def}}{=} |S_Z| = |S_Y|+1$. We can perform these operations in $O(\log n)$ time by:

- With probability $1/(1+\tau)$ where

$$\tau \stackrel{\text{def}}{=} \frac{q^{k-(\ell-1)} \sum_{i \in S_Z} w_i^{-1} / \binom{k}{\ell-1}}{q^{k-\ell}(k - |S_Y|)/\binom{k}{\ell}} = \frac{q \sum_{i \in S_Z} w_i^{-1}}{\ell},$$

  sample y_j uniformly at random from $Y \setminus S_Y$ and add y_j to S_Y. Note that this action will always be performed if $q = 0$.

- Maintain an array of cumulative sums $s_t \stackrel{\text{def}}{=} \sum_{h=1}^{t} w_{i_h}^{-1}$ for $t \in [\ell]$ where $w_{i_1}, w_{i_2}, \ldots, w_{i_\ell}$ are the weights corresponding to the edges in S_Z; this data structure supports amortized $O(\log n)$-time insertion and deletion from S_Z and binary search in the sorted array $[s_t]_{t=1}^{\ell}$. This data structure can be implemented using a *splay tree* where each node stores the sum of all leaves in its rooted subtree. With probability $\tau/(1+\tau)$, sample z_f from S_Z with probability $\propto 1/w_f$ by sampling a uniformly random $p \in [0, s_\ell]$, finding the minimum $t \in [\ell]$ where $p \le s_t$, and removing z_{i_t} from S_Z. This removal will split a tree in the forest S_Z, and we update the link-cut tree representation of S_Z accordingly in $O(\log n)$-time.

For completeness, we briefly summarize how to handle sampling and removing an edge from $\mathcal{C}_S$, which was described in [RTF18]. We represent S_Z as a forest of link-cut trees. When we add an edge $e = (u, v)$ that forms a cycle, splay u to be the root of its tree $\mathcal{T}_u$, then access v (which is also in T_u) so that the entire path $\mathcal{P}_{u,v}$ from u to v in T_u is stored in one auxiliary tree. This auxiliary splay tree can be augmented to support (weighted) sampling an edge f from $\mathcal{P}_{u,v}$ as described above with S_Z. Remove f (which disconnects $\mathcal{T}_u$ into two trees) then add e. Link-cut trees support these operations in amortized $O(\log n)$ time, and the augmentation increases the running time by only a constant factor. □

Proof Sketch of Theorem 12.0.2. Without loss of generality, we may assume $0 \le k_1 \le k_2 \le \operatorname{rank}(\mathcal{M})$ where $\mathcal{M}$ is the graphic matroid of G. For any $\ell \le \operatorname{rank}(\mathcal{M})$, size-$\ell$ forests of G form the bases of a matroid $\mathcal{M}^{(\ell)}$ called the *truncation*. For an arbitrary matroid $\mathcal{M}$ and parameters $k_1 \in \mathbb{N}$, Lemma 12.1.2 tells us that $f_{\mathcal{M}}^{k_1}(y, z_1, \cdots, z_n) = \sum_{S \in \mathcal{X}: |S| \ge k_1} y^{|S|} z^{[n]\setminus S}$ is completely log-concave. From here, we may proceed as in the proof of Theorem 12.0.1, while setting $\mathcal{M}$ to be the truncation $\mathcal{M}^{(k_2)}$. □

Finally, we briefly discuss how to sample from the family of independent sets of a matroid $\mathcal{M} = ([n], \mathcal{X})$ of rank-k using the framework developed here. Suppose we are given access to a data structure $\mathcal{O}'$ that maintains a set $S \subseteq [n]$ guaranteed to contain at most one circuit, and which supports the following operations:

- Addition and removal of an element from S, provided we maintain the property that S contains at most one circuit.
- Outputting a uniformly random element from the unique circuit in S if such a circuit exists.

Then each step of the down-up walk described in Theorem 12.0.1 can be implemented with $O(1)$ calls to $\mathcal{O}'$, resulting in a $O(n \log n)$-time algorithm to sample uniformly from the family of independent sets of $\mathcal{M}$. We remark that Theorem 5.5.2 and the polarization trick employed in proof of Theorem 12.0.1 already gives a $O(n^2 \log(n/\epsilon))$-time algorithm, given access to an independent set oracle $\mathcal{O}_I$ for $\mathcal{M}$. Indeed, the down-up walk on the distribution defined by the polarization of the strongly log-concave polynomial $g_{\mathcal{M}}(y, z_1, \ldots, z_n) = \sum_{S \in \mathcal{X}} y^{n-|S|} z^S$ (see Theorem 5.5.2) mixes in $O(n \log(n/\epsilon))$ steps, and each step can be implemented using $O(n)$ calls to $\mathcal{O}_I$.

Chapter 13

Conclusion and Future Directions

In this thesis, we saw a new technique called *spectral independence* for bounding mixing times of Markov chains, which then yield efficient algorithms for sampling from high-dimensional probability distributions supported over exponentially large domains. We developed a powerful new toolbox for bounding the spectral independence of many probability distributions arising in theory and in practice. However, many open questions remain, and many new applications are waiting to be discovered. This chapter is devoted to discussing recent applications of spectral independence which have emerged, as well as lines of inquiry to be explored in the future which I personally find interesting.

13.1 Subsequent Works

We start a few interesting lines of subsequent work which have extended or strengthened this theory. We also mention recent, newer applications of spectral independence not previously mentioned.

Entropic Independence We saw in Chapter 2 that spectral independence is intimately related to the decay of relative variance under the local down operator $\mathcal{D}_\mu^{2\searrow 1}$, which can then be used to establish a spectral gap lower bound for the down-up walk. We then saw in Chapter 9 how to go from spectral independence to decay of relative entropy under $\mathcal{D}_\mu^{2\searrow 1}$, with the motivation of establishing a lower bound on the modified log-Sobolev constant of the down-up walk. However, this implication required *marginal boundedness* of the distribution. While this assumption holds for many statistical physics distributions on sparse graphs, it fails for many other classes of distributions of interest (e.g. Gibbs distributions when the underlying graph does not have bounded degree, determinantal point processes, etc.).

A beautiful line of work recently showed that under a different but natural technical assumption, one can immediately obtain contraction of relative entropy under $\mathcal{D}_\mu^{n\searrow 1}$ in an extremely clean and elegant manner, which also yields lower bounds on the modified log-Sobolev constant of the down-up walk [Ana+21a; Ana+21b; Ana+22c]. This $n \leftrightarrow 1$ entropy contraction is called *entropic independence*, which is a strengthening of spectral independence. The technical assumption they make is that the distribution remains spectrally independent under *all external fields*, not just all conditionings. In the language of polynomials, this assumption is called *fractional log-concavity* of the associated multivariate generating polynomial [Ali+21], and it holds for instance for determinantal point processes, and discrete log-concave distributions more generally. It is further shown in [Ana+21a] that entropic independence for all external fields is equivalent to fractional log-concavity of the generating polynomial.

Entropic independence has since played an important role in many further algorithmic developments, particularly in techniques for *accelerated sampling*. We point interested readers to [ALV22; Ana+22a; Ana+22b] (see also [AD20]).

Beyond Discrete Spaces [CE22] has developed a new framework called *localization schemes* significantly generalizes the spectral independence framework. It also encompasses Eldan's groundbreaking *stochastic localization* method [Eld13], which has had huge impact in the analysis of probability distributions and high-dimensional geometry. We refer interested readers to [Eld13; LV17; Che21; KL22; JLV22] for applications to Bourgain's Slicing Conjecture [Bou86; Bal88], the

Thin-Shell Conjecture [ABP03], and the Kannan–Lovász–Simonovits Conjecture [KLS95] in convex geometry; see also [LV18; Eld20; ES22; EKZ22] for further applications of related ideas.

Finally, we also mention a recent work of [QW22] which extends the notion of spectral independence to arbitrary product spaces, including $\mathbb{R}^n$. They use their framework to analyze the Gibbs sampler, which is essentially the same as the Glauber dynamics but for products of more general spaces which need not be finite. Their extension follows a very similar analysis to our first proof of Theorem 2.3.1.

Refined Boosting Theorems for Modified Log-Sobolev Constants Based on the ideas and intuitions from Chapter 10, [Che+21b] developed an incredibly novel Markov chain called the *field dynamics*. Most notably, for the hardcore model and many two-spin systems more generally, the field dynamics can be used to *boost* lower bounds on the spectral gap of the Glauber dynamics when the parameters are far from criticality to when they are near criticality [Che+21b]. These boosting results were subsequently extended to the rate of entropy decay and the modified log-Sobolev constants independently in [Che+21a; Che+22a; Ana+22c; CE22]. These boosting theorems lie at the heart of recent optimal mixing results for spin systems, whose running times do not depend on the maximum degree of the underlying graph.

At a very high-level, the intuition behind these improvements compared to Chapter 10, say for the hardcore model, is the following. In Chapter 10, we used that after pinning linearly many vertices (say, in a single down step from the θn-uniform block dynamics), the remaining graph G' shatters into small $O(\log n)$-size components for which sufficiently good entropy factorization holds just from crude bounds.

However, one can do something better for the hardcore model. Namely, instead of using the shattering property, observe that one expects the maximum degree $\Delta' = \Delta(G')$ of the remaining graph to be smaller than $\Delta = \Delta(G)$ by a constant factor depending on the density of the pinned vertices. Since the distribution conditioned on the pinned vertices is the Gibbs distribution of the hardcore model on the remaining graph G' *with the same fugacity* λ, we expect that λ is much farther away from the critical threshold $\lambda_c(\Delta')$ for G' than $\lambda_c(\Delta)$. In particular, by taking the density of the pinned vertices to be large enough, one can ensure that λ is deep within the uniqueness regime, from which one can deduce strong modified log-Sobolev inequalities via other techniques (e.g. entropic Ricci curvature [Erb+17]).

The main issue then becomes getting enough concentration to ensure that indeed, with high probability, the maximum degree does sufficiently decrease after pinning linearly many vertices. To obtain the required level of concentration, [Che+21b] introduced another trick they call the "k-transformation" (also called a "blow-up" of the original distribution [Ana+21b], or a "subdivision" of the original distribution [AD20]). Roughly speaking, it turns out that the θn-uniform block dynamics on the k-transformation in the large k limit recovers the field dynamics after applying a natural projection operation.

Beyond Glauber Dynamics Spectral independence has recently also been used to bound mixing times of other more complex Markov chains, as well as devise *new* algorithms for sampling. For instance, in the context of spin systems, [Bla+22] established optimal mixing and optimal entropy decay for *arbitrary block dynamics*. Furthermore, in the so-called "high-temperature regime", they gave optimal mixing time bounds for the *Swendsen–Wang dynamics*, a well-studied but nonlocal Markov chain used to sample from the ferromagnetic Ising and Potts models. Historically, these more complex dynamics have proven challenging to understand theoretically. These results were achieved via spectral independence.

[Ana+21b] developed a modified version of the Glauber dynamics they dubbed the "balanced Glauber dynamics", and proved that it mixes in $O(n \log n)$ steps for sampling from the hardcore model in the tree uniqueness regime. Recently, [CZ22] simulated the field dynamics invented in [Che+21b] to give nearly-linear time sampling algorithms for the ferromagnetic Ising model at all temperatures with nonzero but consistent external fields. Besides the development of new Markov chains, simple and well-studied Markov chains have also been incorporated as part of more sophisticated overarching algorithms. We now highlight a couple of these applications.

[Bez+22] used spectral independence to build fast algorithms for sampling from the hardcore model on the Erdös–Rényi random graph $G(n, d/n)$ when $\lambda < \lambda_c(d)$, even though the maximum degree of such random graphs is much larger (in fact, growing in n with high probability). They further extended these results to the ferromagnetic Ising model and the monomer-dimer model.

They established spectral independence using correlation decay like in Chapter 7, but employ a much more refined analysis previously done in [SSY13; Sin+15] for graphs with bounded *connective constant*, a notion of "average degree". Their algorithm is based on similar intuitions to those in Chapter 10, but instead they simulate the Glauber dynamics on the marginal distribution of a carefully chosen set of low-degree vertices. They then show that the resulting random partial configuration can be efficiently extended to a random full configuration.

Building on this idea, recently [CMM22; Gal+22] independently used spectral independence to develop polynomial-time sampling algorithms for solutions of random k-SAT instances when the clause density is at most $2^{O(k)}$.[1] The significance of this density is that it is (very roughly) the threshold which delineates between whether or not a random k-SAT instance is satisfiable (with high probability) [DSS22a]. We refer interested readers to those papers for the history and importance of this problem. Like in [Bez+22], they again first run a Markov chain on a carefully chosen subset of variables (i.e. "vertices"), and then extend the resulting random partial configuration to a random full configuration. However, rather than running the Glauber dynamics, they simulate a linear-sized block dynamics, which they prove can be implemented efficiently with high probability.

Additional Applications [KKS21] have shown that for distributions over $\{0,1\}^n$, the condition $\|\Psi_\mu\|_\infty \leq O(1)$ known as (two-sided) ℓ_∞-independence for all conditional distributions implies Chernoff bounds for linear matrix and scalar functions. They achieve this by directly controlling the *moment generating function*, instead of appealing to modified log-Sobolev inequalities. Thus, they are able to establish much stronger concentration bounds, albeit only for linear functions as opposed to all Lipschitz functions. An interesting application of their methodology is proving that the union of $O(\log n)$ uniformly random spanning trees is a spectral sparsifier of the underlying graph with high probability.

[Fri+21] applied spectral independence to obtain algorithms for sampling from the *hard spheres model*, an important continuous model of a gas of which the hardcore gas model studied in Chapter 7 can be viewed as a discretization.

13.2 Some Major Unresolved Problems and Conjectures

We now provide a collection of specific counting and sampling problems. Most of these problems are folklore, and have been open for decades. For problems which do not include a citation, we could not find a reference which formally states them; regardless, we certainly do not claim credit for these problems. Our rough format here is to first state the problem/conjecture, and then briefly mention some known results. For some, we informally state some related conjectures which pertain to spectral independence of the distribution.

Problem 1 (Matroid Intersection)**.** *Given independence oracles for two matroids $\mathcal{M}_1, \mathcal{M}_2$ of the same rank r on a common ground set $\mathscr{U}$, is there an* FPRAS *for estimating the number of common bases in both $\mathcal{M}_1, \mathcal{M}_2$?*

The seminal work of [JSV04] resolved this question for intersections of *partition* matroids, which can alternatively be framed as counting perfect matchings in bipartite graphs. [AL20] also obtain rapid mixing for a down-up walk for sampling common size-k independent sets of two partition matroids, provided $k \leq r/3$. [Ali+21] obtain an FPRAS for the intersection of a *real stable* matroid and a partition matroid with $O(1)$-many parts. This was also studied previously in [KD16; AO17; Cel+17; SV17]. [AOV21] obtained an efficient deterministic simply-exponential multiplicative approximation algorithm for general matroids.

Problem 2 (Perfect Matchings in General Graphs)**.** *Given an undirected graph $G = (V, E)$ with an even number of vertices, is there an* FPRAS *for estimating the number of perfect matchings on G?*

Again, the seminal work of [JSV04] resolved this question for *bipartite* graphs. [JS89] obtained an FPRAS when the ratio of the number of near-perfect matchings to the number of perfect matchings is polynomial in the size of the graph. This includes dense graphs, or more specifically, graphs for which the minimum degree of a vertex is at least $n/2$. [ENO22] recently showed the

[1] We note that [HWY22] independently also achieved the same algorithmic result, although using a completely different algorithm whose analysis does not rely on spectral independence.

algorithm of [JS89] works for d-regular expander graphs. This encompasses almost all d-regular graphs [Fri08; Bor19].

Problem 3 (#BIS)**.** *Let* #BIS *denote the problem of approximately counting (unweighted) independent sets given a bipartite graph. Which of the following is true?*

1. *There is an* FPRAS *for* #BIS*.*
2. *There is no* FPRAS *for* #BIS *(unless* NP = RP*).*
3. *Neither of the above is true, in which case* #BIS *genuinely has* ***"intermediate"*** *complexity.*

#BIS is an important problem in the complexity theory of approximate counting and sampling. Many natural counting problems have been proved to reduce to #BIS, e.g. counting downsets in a poset [Dye+04a], counting stable matchings in the stable marriage problem [CGM12], computing the partition function of the ferromagnetic Potts model/random cluster models [GJ08; GJ12a; GJ12b; GJ13; GJ14], and computing the partition function of the ferromagnetic Ising model with inconsistent external fields [GJ07] (see also [LLZ14b]). #BIS is also *complete* for a certain logically-defined complexity class known as $\#\mathsf{RH\Pi}_1$ under approximation-preserving reductions [Dye+04a].

#BIS is sometimes colloquially referred to as the approximate counting analog of the *Unique Games Problem* UG in combinatorial optimization and hardness of approximation [Kho02]. This is because like UG, recent works have shown that many #BIS-Hard problems are tractable on many interesting classes of graphs [BR19; HPR20; Bor+20; CP20], including random graphs and expanders [Lia+19; Bla+20; JKP20; GGS21; Che+21c; HJP22; JPP22; Che+22b; Car+22; KLR22].

Conjecture 6 (Proper Colorings)**.** *Let G be an undirected n-vertex graph with maximum degree Δ, and let $q \geq \Delta+2$. Then the Glauber dynamics on proper vertex q-colorings on G mixes rapidly. Furthermore, if $q \geq \Delta+1$, there exists an* FPRAS *for estimating the total number of proper vertex q-colorings on G.*[2]

As previously mentioned, when combined with the hardness result of [GŠV15], this conjecture would provide another example of a *computational phase transition*. The threshold $q \geq \Delta + 1$ would precisely delineate between the computationally tractable and intractable regimes.[3] The current best is $q \geq \left(\frac{11}{6} - \epsilon_0\right)\Delta$ where $\epsilon_0 \approx 10^{-5}$ is a small universal constant. This is due to [Che+19], building on a seminal result of Vigoda [Vig00]. Establishing rapid mixing even when $q \geq C\Delta$ for any constant C "substantially" below $\frac{11}{6}$ would be a breakthrough. Better thresholds are known for special classes of graphs, including triangle-free graphs [Che+21d; Fen+21], locally sparse graphs more broadly [HV03; Mol04; HV05; FV06; FV07; Dye+13], planar graphs [HVV15], chordal graphs [Hei20], line graphs [ALO22; DHP20], random graphs [Dye+06; ES08; MS10; Eft14; Eft+18], graphs with bounded treewidth [Hei20], graphs with logarithmically bounded pathwidth [Var18], lattices [GMP05; Gol+06; Jal12], hyperbolic graphs [Ber+05], and trees [MSW07; Ber+05; LMP09; GJK10].

We firmly believe that the uniform distribution over proper q-colorings is $O(1)$-spectrally independent when $q \geq \Delta + 2$, although Theorem E.3.1 suggests that when q is in the $\Delta + O(1)$, the spectral independence of the Gibbs distribution should depend at least linearly on Δ. All of the known tools currently break down when going ("substantially") below the $\frac{11}{6}\Delta$ threshold. Thus, it seems such a result would require fundamentally new tools for establishing spectral independence.

Conjecture 7 (Random Cluster Model on Graphs when $1 \leq q \leq 2$)**.** *Let $G = (V, E)$ be an undirected (multi)graph with n vertices and m edges, and let $0 \leq p \leq 1, q \geq 0$ be parameters. Recall the* ***random cluster measure*** $\mu^{\mathsf{RC}}_{G,p,q}$ *over $\{0,1\}^E$ (or subsets of edges) is defined by*

$$\mu^{\mathsf{RC}}_{G,p,q}(F) \propto q^{k(F)} p^{|F|} (1-p)^{|E \setminus F|}, \quad \forall F \subseteq E, \tag{13.1}$$

where $k(F)$ is the number of connected components of the subgraph (V, F). Then the Glauber dynamics for sampling from $\mu^{\mathsf{RC}}_{G,p,q}$ mixes rapidly whenever $1 \leq q \leq 2$, irrespective of the choice of $0 \leq p \leq 1$.

[2] The slight discrepancy between the threshold $q \geq \Delta + 2$ for rapid mixing and $q \geq \Delta + 1$ for existence of an FPRAS comes from the fact that when $q = \Delta + 1$, there are examples of graphs on which the Glauber dynamics fails to be connected. A classic example is the 3-vertex cycle graph, on which the goal is to approximately count proper 3-colorings.

[3] Technically, the hardness results of [GŠV15] only hold for even $q \leq \Delta$. However, we fully expect hardness to hold for all $q \leq \Delta$.

[GJ21] have shown this is true for $q = 2$ by a comparison argument with the even subgraphs process studied in [JS93] (more precisely, the "worm process" [PS01]). The case $q = 1$ is trivial. [Ana+19] gave a polynomial-time sampler based for $0 < q < 1$ based on down-up walks, which can actually sample even when conditioning $\mu^{\mathsf{RC}}_{G,p,q}$ on a fixed number of edges; this was discussed in Section 5.4 (see also Appendix F.3 and [Mou22]). There are also well-known hardness results for the random cluster model. For instance, there is no FPRAS for estimating the partition function of the random cluster model on graphs when $q > 2$ unless there is an FPRAS for counting independent sets in bipartite graphs, i.e. the problem is #BIS-Hard [GJ12a]. Furthermore, it was shown in [GJ13] that for general binary matroids, it is also #BIS-Hard to sample from the random cluster model even when $q = 2$. See [GJ12b; GJ14] for further hardness results.

For this model, we believe that $\mu^{\mathsf{RC}}_{G,p,q}$ is $O(1)$-spectrally independent whenever $1 \leq q \leq 2$ and $0 \leq p \leq 1$. Here, the Fortuin–Kastelyn–Ginibre (FKG) Inequalities [FKG71] show that the entries of the influence matrix $\mathcal{I}_\mu$ are nonnegative. Furthermore, random cluster models behave nicely under conditioning [Gri09]; fixing an edge e to be in the sample F yields $\mu^{\mathsf{RC}}_{G/e,p,q}$, where G/e is the graph obtained by contracting the vertices of the edge e (and deleting the resulting loops), while fixing an edge e to not be in the sample F yields $\mu^{\mathsf{RC}}_{G\setminus e,p,q}$, where $G \setminus e$ is the graph obtained by simply deleting the edge e. One potential challenge of this is reconciling with known tight mixing results for the case of $\mathbb{Z}^2$ [BS17b],[4] which suggest that the scaling of the mixing time in n depends on p in a delicate manner.

Along this line, one can directly combine the spectral independence result for even subgraphs Theorem 6.1.2 from [CLV21b] with the Grimmett–Janson coupling [GJ09] to obtain spectral independence for a slightly reweighted version of the $q = 2$ random cluster model; for experts, this reweighted version corresponds to the ferromagnetic Ising model with consistent external fields. This was also observed recently in [FGW22]. However, the spectral independence obtained through this method depends on the maximum degree of the input graph. Even more recently, [CZ22] established spectral independence for the same reweighted $q = 2$ random cluster model but with no dependence on the maximum degree. Their method is completely different, relying on a sophisticated coupling procedure which doesn't use zero-freeness at all.

Conjecture 8 (Nonbroken Bases of Matroids)**.** *Let $\mathcal{M} = (\mathscr{U}, \mathscr{X})$ be a rank-r matroid, and fix a total order $<$ on the ground elements $\mathscr{U}$. A* ***broken circuit*** *is a set of the form $C - \min_< C$, where $C \subseteq \mathscr{U}$ is a circuit (i.e. a minimal dependent set). Let μ be the uniform distribution over bases of $\mathcal{M}$ which do not contain a broken circuit. Then the down-up walk for sampling from μ mixes rapidly.*

The collection of all subsets which do not contain a broken circuit forms a pure abstract simplicial complex called the *broken circuit complex* $\mathsf{BC}_<(\mathcal{M})$, originally introduced in [Wil76]. Such sets are also called *nonbroken*. It is easy to see that this simplicial complex is contained within the matroid complex $\mathcal{X}$, i.e. all sets in $\mathsf{BC}_<(\mathcal{M})$ are independent. The distribution μ described above is then the uniform distribution over the *nonbroken bases*. A remarkable fact is that while different orderings $<$ induce genuinely different simplicial complexes which are not isomorphic to each other, their *face numbers*

$$\mathfrak{f}_k(\mathsf{BC}_<(\mathcal{M})) \stackrel{\mathsf{def}}{=} \#\{S \in \mathsf{BC}_<(\mathcal{M}) : |S| = k\}, \quad \forall 0 \leq k \leq r$$

are invariant under the ordering $<$ [Bry77].[5] It is also well-known that the $\mathfrak{f}_k(\mathsf{BC}_<(\mathcal{M}))$ form a log-concave (and hence, unimodal) sequence. This is the content of the various conjectures of Read, Heron, Rota and Welsh [Rea68; Rot71; Her72; Hog74; Wel76], which were famously resolved first for graphic matroids in a breakthrough of Huh [Huh12], then for realizable matroids [HK12], and finally in full generality by [AHK18]; see [BES21; BL21] for recently simplified proofs.

These numbers $\mathfrak{f}_k(\mathsf{BC}_<(\mathcal{M}))$ have concrete combinatorial significance. For instance, consider the univariate polynomial

$$\chi_{\mathcal{M}}(t) \stackrel{\mathsf{def}}{=} \sum_{k=0}^{r} (-1)^k \cdot \mathfrak{f}_k(\mathsf{BC}_<(\mathcal{M})) \cdot t^{r-k}.$$

[4] There are also tight mixing results for the complete graph K_n, although these are mainly for other nonlocal Markov chains; see e.g. [BGJ96; BS15b; Bla16].

[5] In algebraic and topological combinatorics, $\mathfrak{f}_k$ is typically defined to be the number of *dimension-k* faces, i.e. cardinality-$(k+1)$ sets, as opposed to the number of cardinality-k sets.

This is known as the *characteristic polynomial* of the matroid $\mathcal{M}$, and it is a specialization of the Tutte polynomial discussed in Section 5.4.[6] When $\mathcal{M}$ is a graphic matroid w.r.t. some graph $G = (V, E)$, then $\chi_{\mathcal{M}}$ becomes the George Birkhoff's famous *chromatic polynomial* χ_G of the graph G, where $\chi_G(q) = \#\{\text{proper } q\text{-colorings of } G\}$ for all positive integers $q \geq 1$. In this case, it is well-known that $(-1)^r \chi_{\mathcal{M}}(-1) = \sum_{k=0}^{r} \mathfrak{f}_k(\mathsf{BC}_<(\mathcal{M}))$ counts the number of *acyclic orientations of* G. More generally, when $\mathcal{M}$ is a linear matroid w.r.t. some vectors in $\mathbb{R}^n$, then $(-1)^r \chi_{\mathcal{M}}(-1) = \sum_{k=0}^{r} \mathfrak{f}_k(\mathsf{BC}_<(\mathcal{M}))$ counts the number of *regions* in the corresponding *hyperplane arrangement*. Surveys on these correspondences can be found in [CR70; Bry72; BK80; Sta07]. Finally, the number of nonbroken bases $\mathfrak{f}_r(\mathsf{BC}_<(\mathcal{M}))$ is equal to the *Möbius number* of the *lattice of flats* of $\mathcal{M}$, a fundamental combinatorial parameter of the matroid with topological significance [Bry77]. For further discussion of all of these connections, we refer interested readers to [Bry77; BO80; BO81; Bjö92] and references therein.

The point of this discussion is that resolving Conjecture 8 would have a number of major consequences to approximate counting and sampling. For graphic matroids, it would allow us to approximately count acyclic orientations in graphs. For $\mathbb{R}$-linear matroids, it would allow us to approximately count regions in hyperplane arrangements, or equivalently, count vertices in *zonotopes*.[7] Some applications of counting and sampling vertices in zonotopes to numerical linear algebra can be found in [VWG17]. For the special case of rank-3 matroids, Conjecture 8 has been verified by Guo–Mousa; see [Mou22]. Based on preliminary numerical experiments, we believe the distribution described in Conjecture 8 to be $O(1)$-spectrally independent, but this remains wide open, even for special classes of matroids.

13.3 Some Additional Directions to Explore

We also briefly mention a few other open problems which we have encountered in our journey. Some of these are not really well-posed problems per se, but rather are directions for future inquiry and research that the author personally finds intriguing (although not necessarily deep or profound).

Problem 4 (Refined Trickle-Down Methods)**.** *How "far" can we push trickle-down type methods (see Chapters 3 and 4) for establishing spectral independence? For instance, are there principled or even "natural" ways of constructing "good" bounding matrices in the matrix trickle-down method discussed in Chapter 4?*

See [AO22] for some recent developments in this direction.

Problem 5 (Weaker Zero-Freeness Assumptions)**.** *We saw in Chapter 6 how multivariate zero-freeness (i.e. $\prod_{i \in \mathscr{U}} \Gamma_i$-stability) can be used to deduce spectral independence. How much can we weaken such an assumption? For instance, what if the multivariate generating polynomial g_μ is nonzero whenever $z \in \Gamma \subseteq \mathbb{C}^{\mathscr{U}}$ where Γ is not of the form $\prod_{i \in \mathscr{U}} \Gamma_i$?*

An example of a weaker notion of stability is *same-phase stability*, which was introduced in [LR19], and proven to hold for multivariate independence polynomials of line graphs, and *claw-free graphs* more generally.

Problem 6 (Reverse Implications of Spectral Independence)**.** *We previously saw that correlation decay (see Chapter 7) and the presence of large zero-free regions for the generating polynomial (see Chapter 6) imply strong spectral independence bounds. However, can either of these implications be* ***reversed****?*

Clearly, correlation decay only makes sense in a graph or metric space context. However, even for zero-freeness, one must be careful about the context in which to study this question. At the highest level of generality, spectral independence has no implications for stability. Indeed, strongly log-concave polynomials need not have any nontrivial zero-free region. For Gibbs distributions of spin systems in high temperature, the recent developments and intuitions in the field suggest that all three of these notions coincide.

[6] The characteristic polynomial is traditionally defined recursively using deletion-contraction identities, and is intimately related to the *Möbius function of the lattice of flats* of $\mathcal{M}$; see e.g. [Whi32; Tut54; Cra69]. It is a theorem that these two definitions coincide [Rot64]. From this, one can immediately deduce the invariance of the numbers $\mathfrak{f}_k(\mathsf{BC}_<(\mathcal{M}))$ under the ordering $<$.

[7] A *zonotope* is a Minkowski sum of centrally symmetric segments $[-v_i, v_i] \subseteq \mathbb{R}^n$.

One interesting implication of such a reverse implication would be spatial mixing and stability for the uniform distribution over proper q-colorings of a graph in the regime $\left(\frac{11}{6} - \epsilon_0\right)\Delta \leq q \leq 2\Delta$, which would pave the way for deterministic FPTAS for approximately counting proper q-colorings in this regime. Indeed, we proved $O(1)$-spectral independence for this model in Chapter 8 using coupling arguments. However, we only know correlation decay and zero-freeness when $q \geq 2\Delta$; see [GK07; GK12; GKM15] for the former, and [LSS20] for the latter.

Problem 7 (Correct Spectral Independence Bounds for Monomer-Dimer)**.** *We showed in Theorem B.0.2 that for graphs of maximum degree Δ, the Gibbs distribution of the monomer-dimer model with fugacity λ is $O(\sqrt{\lambda\Delta})$-spectrally independent. However, calculations on the infinite tree (see Theorem E.2.1) suggest that the correct spectral independence bound should be $O(\sqrt{\lambda})$ instead, independent of Δ, even though $O(\sqrt{\lambda\Delta})$ should be the correct bound if one were to consider ℓ_∞-independence. Also, generalizing beyond the monomer-dimer model, i.e. the hardcore model on line graphs, what is the correct spectral independence parameter for the hardcore model on claw-free graphs?*

Problem 8 (Formal Spectral Independence Lower Bounds)**.** *The calculations in Appendix E suggest lower bounds for the spectral independence of various graphical models on the class of bounded-degree graphs. However, those calculations were not formally justified, since they involved infinite matrices. So, are those predicted lower bounds correct? Ideally, to genuinely demonstrate a lower bound, we'd like an infinite family of finite graphs of maximum degree Δ, with number of vertices growing to ∞, on which the spectral independence of the associated Gibbs distribution is lower bounded.*

Problem 9 (Hardcore Model on Bounded-Degree Graphs at Criticality)**.** *Let G be an undirected graph with maximum degree Δ, and let $\lambda = \lambda_c(\Delta)$. Then what is the mixing time of the Glauber dynamics for sampling from the Gibbs distribution of the hardcore model on G with fugacity λ? What about other spin systems at criticality?*

Problem 10 (Log-Concavity of Stanley's Symmetric Chromatic Polynomial)**.** *For a graph $G = (V, E)$ and a positive integer $q \geq 1$, define the **Stanley's symmetric chromatic polynomial** in q variables by*

$$X_{G,q}(t_1, \ldots, t_q) = \sum_{\kappa} \prod_{v \in V} t_{\kappa(v)} = \sum_{\kappa} \prod_{c=1}^{q} t_c^{\left|\kappa^{-1}(c)\right|},$$

where the sum ranges over all proper q-colorings $\kappa : V \to [q]$ [Sta95].[8] For what classes of graphs is this polynomial strongly log-concave? Is it strongly log-concave for all claw-free graphs?[9]

Note that if one were to replace each by $t_{\kappa(v)}$ by $t_{v,\kappa(v)}$ (equivalently, expand each $t_c^{\left|\kappa^{-1}(c)\right|}$ as $\prod_{v \in V : \kappa(v) = c} t_{v,c}$), then we would recover the homogeneous multivariate generating polynomial of the uniform distribution over all proper q-colorings. For the claw graph, i.e. the star graph with 3 leaves, $X_{G,q}$ is not strongly log-concave (unless $q = 1$, in which case the polynomial is identically zero). Indeed, if $q \geq 2$, then setting $t_c = 0$ for all $c \geq 3$, we obtain the bivariate polynomial $t_1^3 t_2 + t_1 t_2^3 = t_1 t_2 (t_1^2 + t_2^2)$, which isn't log-concave. This is why we have posed the log-concavity question for claw-free graphs.

[8] The original function Stanley considered had countably infinitely many variables, and the sum is over all proper colorings using any number of colors. Here, we have specialized to a bounded number of colors q.

[9] Recently, we discovered that strong log-concavity of $X_{G,q}$ for claw-free graphs was independently conjectured by Matherne–Morales–Selover. They managed to prove some partial results in this direction; see [MMS22] for details.

Bibliography

[Abe80] Silvana Abeasis. "On the Plücker relations for the Grassmann varieties". In: *Advances in Mathematics* 36.3 (1980), pp. 277–282 (cit. on p. 184).

[ABF94] Y. Azar, A.Z. Broder, and A.M. Frieze. "On the problem of approximating the number of bases of a matroid". In: *Information Processing Letters* 50.1 (1994), pp. 9–11. ISSN: 0020-0190 (cit. on p. 59).

[ABP03] Milla Anttila, Keith Ball, and Irini Perissinaki. "The Central Limit Problem for Convex Bodies". In: *Transactions of the American Mathematical Society* 355.12 (2003), pp. 4723–4735 (cit. on p. 196).

[AD20] Nima Anari and Michał Dereziński. "Isotropy and Log-Concave Polynomials: Accelerated Sampling and High-Precision Counting of Matroid Bases". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. Los Alamitos, CA, USA: IEEE Computer Society, Nov. 2020, pp. 1331–1344 (cit. on pp. 189, 195, 196).

[AD78] Rudolf Ahlswede and David E. Daykin. "An inequality for the weights of two families of sets, their unions and intersections". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 43 (1978), pp. 183–185 (cit. on p. 18).

[ADH21] Federico Ardila, Graham Denham, and June Huh. "Lagrangian combinatorics of matroids". In: *arXiv preprint arXiv:2109.11565* (2021) (cit. on pp. 90, 110).

[ADH22] Federico Ardila, Graham Denham, and June Huh. "Lagrangian geometry of matroids". In: *Journal of the American Mathematical Society* (2022). DOI: https://doi.org/10.1090/jams/1009 (cit. on pp. 90, 110).

[AGT19a] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani. "Approximating Constraint Satisfaction Problems on High-Dimensional Expanders". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. 2019, pp. 180–201. DOI: 10.1109/FOCS.2019.00021 (cit. on p. 40).

[AGT19b] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani. "Approximating Constraint Satisfaction Problems on High-Dimensional Expanders". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. 2019, pp. 180–201. DOI: 10.1109/FOCS.2019.00021 (cit. on p. 42).

[AHK18] Karim Adiprasito, June Huh, and Eric Katz. "Hodge theory for combinatorial geometries". In: *Annals of Mathematics* 188.2 (2018), pp. 381–452 (cit. on pp. 12, 59, 60, 64, 90, 104, 105, 110, 199).

[AK96] Ron Aharoni and Uri Keich. "A generalization of the Ahlswede-Daykin inequality". In: *Discrete Mathematics* 152 (1–3 1996), pp. 1–12 (cit. on p. 18).

[AL10] Itai Arad and Zeph Landau. "Quantum Computation and the Evaluation of Tensor Networks". In: *SIAM Journal on Computing* 39.7 (2010), pp. 3089–3121 (cit. on p. 280).

[AL20] Vedat Levi Alev and Lap Chi Lau. "Improved Analysis of Higher Order Random Walks and Applications". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2020. Chicago, IL, USA: Association for Computing Machinery, 2020, pp. 1198–1211 (cit. on pp. 33, 40, 43, 45, 58, 150, 197).

[Ald90] David J. Aldous. "The random walk construction of uniform spanning trees and uniform labelled trees". In: *SIAM Journal on Discrete Mathematics* 3.4 (1990), pp. 450–465 (cit. on p. 189).

[Ale+20] Vedat Levi Alev, Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. "List Decoding of Direct Sum Codes". In: *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2020, pp. 1412–1425. DOI: 10.1137/1.9781611975994.85. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611975994.85 (cit. on p. 40).

[Ali+21] Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. "Fractionally Log-Concave and Sector-Stable Polynomials: Counting Planar Matchings and More". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2021. Virtual, Italy: Association for Computing Machinery, 2021, pp. 433–446 (cit. on pp. 17, 18, 37, 41, 90, 114, 116, 117, 120, 125, 159, 195, 197, 258, 303, 305).

[ALO21] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model". In: *SIAM Journal on Computing* 0.0 (2021). Special Section FOCS 2020, FOCS20-1-FOCS20–37. DOI: 10.1137/20M1367696. eprint: https://doi.org/10.1137/20M1367696 (cit. on pp. 10, 14, 17, 19, 36, 37, 40, 43, 51, 53, 67, 132, 148, 242, 259, 263, 267).

[ALO22] Dorna Abdolazimi, Kuikui Liu, and Shayan Oveis Gharan. "A Matrix Trickle-Down Theorem on Simplicial Complexes and Applications to Sampling Colorings". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. 2022, pp. 161–172 (cit. on pp. 16, 19, 65, 67, 198).

[Alo86] Noga Alon. "Eigenvalues and expanders". In: *Combinatorica* 6 (2 Jan. 1986), pp. 83–96. ISSN: 0209-9683 (cit. on p. 63).

[ALV22] Nima Anari, Yang P. Liu, and Thuy-Duong Vuong. "Optimal Sublinear Sampling of Spanning Trees and Determinantal Point Processes via Average-Case Entropic Independence". In: *arXiv preprint arXiv:2204.02570* (2022). To Appear in FOCS 2022 (cit. on pp. 41, 195).

[AM85] N. Alon and V. D. Milman. "λ_1, Isoperimetric inequalities for graphs, and superconcentrators". In: *Journal of Combinatorial Theory, Series B* 38.1 (Feb. 1985), pp. 73–88 (cit. on p. 63).

[Ana+17] Nima Anari, Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. "Nash Social Welfare, Matrix Permanent, and Stable Polynomials". In: *ITCS*. 2017, 36:1–36:12 (cit. on pp. 18, 107).

[Ana+18a] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. "Log-Concave Polynomials III: Mason's Ultra-Log-Concavity Conjecture for Independent Sets of Matroids". In: *arXiv preprint arXiv:1811.01600* (2018) (cit. on pp. 12, 19, 88, 91, 95, 101, 102, 104, 105, 109).

[Ana+18b] Nima Anari, Tung Mai, Shayan Oveis Gharan, and Vijay V. Vazirani. "Nash Social Welfare for Indivisible Items under Separable, Piecewise-Linear Concave Utilities". In: *SODA*. 2018, pp. 2274–2290 (cit. on pp. 18, 107).

[Ana+19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. "Log-Concave Polynomials II: High-Dimensional Walks and an FPRAS for Counting Bases of a Matroid". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2019. Pheonix, Arizona, USA: Association for Computing Machinery, 2019, pp. 1–12 (cit. on pp. 11, 12, 19, 58, 63, 88, 94, 97, 100, 101, 180, 189, 199).

[Ana+21a] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Entropic Independence I: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Distributions and High-Temperature Ising Models". In: *arXiv preprint arXiv:2106.04105* (2021) (cit. on pp. 15, 17, 173, 195, 305).

[Ana+21b] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Entropic Independence II: Optimal Sampling and Concentration via Restricted Modified Log-Sobolev Inequalities". In: *arXiv preprint arXiv:2111.03247* (2021) (cit. on pp. 15, 17, 132, 161, 173, 195, 196).

[Ana+21c] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, Cynthia Vinzant, and Thuy-Duong Vuong. "Log-Concave Polynomials IV: Approximate Exchange, Tight Mixing Times, and near-Optimal Sampling of Forests". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2021. Virtual, Italy: Association for Computing Machinery, 2021, pp. 408–420 (cit. on pp. 11, 12, 16, 19, 180, 181).

[Ana+22a] Nima Anari, Callum Burgess, Kevin Tian, and Thuy-Duong Vuong. "Improved Sampling-to-Counting Reductions in High-Dimensional Expanders and Faster Parallel Determinantal Sampling". In: *arXiv preprint arXiv:2203.11190* (2022) (cit. on p. 195).

[Ana+22b] Nima Anari, Michał Dereziński, Thuy-Duong Vuong, and Elizabeth Yang. "Domain Sparsification of Discrete Distributions Using Entropic Independence". In: *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Vol. 215. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 5:1–5:23 (cit. on p. 195).

[Ana+22c] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Entropic Independence: Optimal Mixing of down-up Random Walks". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 1418–1430 (cit. on pp. 15, 17, 37, 101, 173, 195, 196, 301, 303).

[AO15] Nima Anari and Shayan Oveis Gharan. "The Kadison-Singer Problem for Strongly Rayleigh Measures and Applications to Asymmetric TSP". In: *arXiv preprint arXiv:1412.1143* (2015) (cit. on p. 18).

[AO17] Nima Anari and Shayan Oveis Gharan. "A generalization of permanent inequalities and applications in counting and optimization". In: *STOC*. 2017, pp. 384–396 (cit. on pp. 18, 107, 197).

[AO22] Dorna Abdolazimi and Shayan Oveis Gharan. "An Improved Trickle-Down Theorem for Partite Complexes". In: *arXiv preprint arXiv:2208.04486* (2022) (cit. on pp. 65, 200).

[AOR16] Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. "Monte Carlo Markov Chain Algorithms for Sampling Strongly Rayleigh Distributions and Determinantal Point Processes". In: *COLT*. Vol. 49. Proceedings of Machine Learning Research. Columbia University, New York, New York, USA: PMLR, June 2016, pp. 103–115 (cit. on pp. 11, 17, 18, 38, 59, 98, 114).

[AOV21] Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. "Log-concave polynomials, I: Entropy and a deterministic approximation algorithm for counting bases of matroids". In: *Duke Mathematical Journal* (2021), pp. 1–46 (cit. on pp. 11, 12, 18, 19, 59, 60, 64, 89–91, 96, 104, 180, 197).

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. "Expander Flows, Geometric Embeddings and Graph Partitioning". In: *J. ACM* 56.2 (Apr. 2009). ISSN: 0004-5411 (cit. on p. 9).

[AS04] A.A. Ageev and M.I. Sviridenko. "Pipage Rounding: A New Method of Constructing Algorithms with Proven Performance Guarantee". In: *Journal of Combinatorial Optimization* 8 (2004), pp. 307–328 (cit. on p. 9).

[AS16] Karim A. Adiprasito and Raman Sanyal. "Whitney numbers of arrangements via measure concentration of intrinsic volumes". In: *arXiv preprint arXiv:1606.09412* (2016) (cit. on pp. 104, 110).

[Asa+10] Arash Asadpour, Michel X. Goemans, Aleksander Mądry, Shayan Oveis Gharan, and Amin Saberi. "An $O(\log n/\log\log n)$-Approximation Algorithm for the Asymmetric Traveling Salesman Problem". In: *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '10. Austin, Texas: Society for Industrial and Applied Mathematics, 2010, pp. 379–389. ISBN: 9780898716986 (cit. on pp. 9, 16, 18, 38).

[Asa70] 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 (cit. on pp. 114, 127).

[Asn70] Bernard A. Asner. "On the Total Nonnegativity of the Hurwitz Matrix". In: *SIAM Journal on Applied Mathematics* 18.2 (1970), pp. 407–414. ISSN: 00361399 (cit. on p. 186).

[AV22] Nima Anari and Thuy-Duong Vuong. "From Sampling to Optimization on Discrete Domains with Applications to Determinant Maximization". In: *Proceedings of Thirty Fifth Conference on Learning Theory*. Ed. by Po-Ling Loh and Maxim Raginsky. Vol. 178. Proceedings of Machine Learning Research. PMLR, July 2022, pp. 5596–5618 (cit. on p. 17).

[Baf+22a] Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. "High Dimensional Expanders: Eigenstripping, Pseudorandomness, and Unique Games". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2022, pp. 1069–1128. DOI: 10.1137/1.9781611977073.47. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611977073.47 (cit. on pp. 40, 42).

[Baf+22b] Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. "Hypercontractivity on High Dimensional Expanders". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 185–194. ISBN: 9781450392648. DOI: 10.1145/3519935.3520040 (cit. on pp. 40, 42).

[Bak18] Matthew Baker. "Hodge theory in combinatorics". In: *Bulletin of the American Mathematical Society* 55.1 (2018), pp. 57–80 (cit. on pp. 90, 110).

[Bal00] Cristina M. Ballantine. "Ramanujan Type Buildings". In: *Canadian Journal of Mathematics* 52 (6 2000), pp. 1121–1148 (cit. on p. 40).

[Bal88] Keith Ball. "Logarithmically concave functions and sections of convex sets in R^n". In: *Studia Mathematica* 88.1 (1988), pp. 69–84 (cit. on p. 195).

[Bar15] Alexander Barvinok. "Computing the Partition Function for Cliques in a Graph". In: *Theory of Computing* 11.13 (2015), pp. 339–355. DOI: 10.4086/toc.2015.v011a013 (cit. on p. 283).

[Bar16a] Alexander Barvinok. *Combinatorics and Complexity of Partition Functions*. Vol. 30. Springer Algorithms and Combinatorics, 2016 (cit. on pp. 17, 18, 111, 112, 115, 136, 279, 283).

[Bar16b] Alexander Barvinok. "Computing the Permanent of (Some) Complex Matrices". In: *Found. Comput. Math.* 16.2 (Apr. 2016), pp. 329–342. ISSN: 1615-3375 (cit. on pp. 18, 111, 115, 283).

[Bar17a] Alexander Barvinok. "Approximating permanents and hafnians". In: *Discrete Analysis* 2 (2017) (cit. on pp. 18, 111, 115, 283).

[Bar17b] Alexander Barvinok. "Computing the Partition Function of a Polynomial on the Boolean Cube". In: *A Journey Through Discrete Mathematics* (2017), pp. 135–164 (cit. on pp. 115, 282).

[Bay+07] Mohsen Bayati, David Gamarnik, Dimitriy Katz, Chandra Nair, and Prasad Tetali. "Simple Deterministic Approximation Algorithms for Counting Matchings". In: *STOC*. San Diego, California, USA, 2007, pp. 122–127. ISBN: 978-1-59593-631-8 (cit. on pp. 18, 131, 133, 136, 262, 265).

[BB08] Julius Borcea and Petter Brändén. "Applications of stable polynomials to mixed determinants: Johnson's conjectures, unimodality, and symmetrized Fischer products". In: *Duke Mathematical Journal* 143.2 (2008), pp. 205–223 (cit. on p. 108).

[BB09a] Julius Borcea and Petter Brändén. "Pólya-Schur master theorems for circular domains and their boundaries". In: *Annals of Mathematics* 170 (1 2009), pp. 465–492 (cit. on p. 108).

[BB09b] Julius Borcea and Petter Brändén. "The Lee-Yang and Pólya-Schur programs. I. Linear operators preserving stability". In: *Inventiones mathematicae* 177.541 (2009) (cit. on pp. 96, 108, 127).

[BB09c] Julius Borcea and Petter Brändén. "The Lee-Yang and Pólya-Schur programs. II. Theory of stable polynomials and applications". In: *Communications on Pure and Applied Mathematics* 62 (12 2009), pp. 1595–1631 (cit. on p. 108).

[BB10] Julius Borcea and Petter Brändén. "Multivariate Pólya–Schur classification problems in the Weyl algebra". In: *Proceedings of the London Mathematical Society* 101 (1 2010), pp. 73–104 (cit. on p. 108).

[BB21] Alexander Barvinok and Nicholas Barvinok. "More on zeros and approximation of the Ising partition function". In: *arXiv preprint arXiv:2005.11232* (2021) (cit. on p. 283).

[BBL09] Julius Borcea, Petter Brändén, and Thomas M. Liggett. "Negative dependence and the geometry of polynomials". In: *Journal of the American Mathematical Society* 22.2 (2009), pp. 521–567 (cit. on pp. 38, 97, 104, 108, 184).

[BCR20] Ferenc Bencs, Péter Csikvári, and Guus Regts. "Some applications of Wagner's weighted subgraph counting polynomial". In: *arXiv preprint arXiv:2012.00806* (2020) (cit. on pp. 18, 112, 113, 115, 127–129).

[BCV20] Antonio Blanca, Zongchen Chen, and Eric Vigoda. "Swendsen-Wang dynamics for general graphs in the tree uniqueness region". In: *Random Structures and Algorithms* 56 (2020), pp. 373–400 (cit. on pp. 267, 269).

[BD20] 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 (cit. on p. 283).

[BD97a] R. Bubley and M. Dyer. "Path Coupling: A Technique for Proving Rapid Mixing in Markov Chains". In: *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*. FOCS '97. Washington, DC, USA: IEEE Computer Society, 1997, pp. 223–. ISBN: 0-8186-8197-7 (cit. on pp. 40, 152, 154, 156).

[BD97b] R. Bubley and M. E. Dyer. *Path Coupling, Dobrushin Uniqueness, and Approximate Counting*. Tech. rep. University of Leeds, 1997 (cit. on pp. 152, 154, 156).

[Bea+22] Daniel Beaglehole, Max Hopkins, Daniel Kane, Sihan Liu, and Shachar Lovett. "Sampling Equilibria: Fast No-Regret Learning in Structured Games". In: *arXiv preprint arXiv:2201.10758* (2022) (cit. on pp. 11, 18).

[Ben+21] Ferenc Bencs, Ewan Davies, Viresh Patel, and Guus Regts. "On zero-free regions for the anti-ferromagnetic Potts model on bounded-degree graphs". In: *arXiv preprint arXiv:1812.07532* (2021) (cit. on p. 66).

[Ben18] Ferenc Bencs. "On trees with real-rooted independence polynomial". In: *Discrete Mathematics* 341.12 (2018), pp. 3321–3330 (cit. on p. 139).

[Ber+05] Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. "Glauber dynamics on trees and hyperbolic graphs". In: *Probability Theory and Related Fields* 131 (2005), pp. 311–340 (cit. on p. 198).

[Ber+22] Andrew Berget, Christopher Eur, Hunter Spink, and Dennis Tseng. "Tautological classes of matroids". In: *arXiv preprint arXiv:2103.08021* (2022) (cit. on pp. 90, 110).

[BES21] Spencer Backman, Christopher Eur, and Connor Simpson. "Simplicial generation of Chow rings of matroids". In: *Journal of the European Mathematical Society* (2021). To Appear (cit. on pp. 90, 104, 110, 199).

[Bet18] Michael Betancourt. "A Conceptual Introduction to Hamiltonian Monte Carlo". In: *arXiv preprint arXiv:1701.02434* (2018) (cit. on p. 9).

[Bez+08] Ivona Bezáková, Daniel Štefankovič, Vijay V. Vazirani, and Eric Vigoda. "Accelerating Simulated Annealing for the Permanent and Combinatorial Counting Problems". In: *SIAM Journal on Computing* 37 (5 2008) (cit. on p. 16).

[Bez+22] Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, and Daniel Štefankovič. "Fast Sampling via Spectral Independence Beyond Bounded-Degree Graphs". In: *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Ed. by Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff. Vol. 229. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 21:1–21:16. ISBN: 978-3-95977-235-8. DOI: 10.4230/LIPIcs.ICALP.2022.21 (cit. on pp. 17, 196, 197).

[BG08] Antar Bandyopadhyay and David Gamarnik. "Counting Without Sampling: Asymptotics of the Log-partition Function for Certain Statistical Physics Models". In: *Random Struct. Algorithms* 33.4 (Dec. 2008), pp. 452–479. ISSN: 1042-9832 (cit. on pp. 18, 131, 136).

[BGJ96] B. Bollobás, G. Grimmett, and S. Janson. "The random-cluster model on the complete graph". In: *Probability Theory and Related Fields* 104 (1996), pp. 283–317 (cit. on p. 199).

[BH18] Petter Brändén and June Huh. "Hodge-Riemann relations for Potts model partition functions". In: *arXiv preprint arXiv:1811.01696* (2018) (cit. on pp. 12, 19, 101, 102, 104, 105, 109, 181).

[BH20] Petter Brändén and June Huh. "Lorentzian polynomials". In: *Annals of Mathematics* 192 (3 2020), pp. 821–891 (cit. on pp. 11, 19, 96, 97, 100, 102, 104, 106, 107, 109, 110, 187).

[Bjö92] Anders Björner. "Homology and Shellability of Matroids and Geometric Lattices". In: *Matroid Applications*. Ed. by Neil White. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1992. Chap. 7, pp. 226–283 (cit. on p. 200).

[BK80] T.H. Brylawski and D. Kelly. *Matroids and Combinatorial Geometries*. University of North Carolina Press, 1980 (cit. on p. 200).

[BK99] Piotr Berman and Marek Karpinski. "On Some Tighter Inapproximability Results (Extended Abstract)". In: *Proceedings of the 26th International Colloquium on Automata, Languages, and Programming*. ICALP '99. Berlin, Heidelberg: Springer-Verlag, 1999, pp. 200–209. ISBN: 3540662243 (cit. on p. 14).

[BL21] Petter Brändén and Jonathan Leake. "Lorentzian polynomials on cones and the Heron-Rota-Welsh conjecture". In: *arXiv preprint arXiv:2110.00487* (2021) (cit. on pp. 105, 109, 110, 199).

[Bla+13] Antonio Blanca, David Galvin, Dana Randall, and Prasad Tetali. "Phase Coexistence and Slow Mixing for the Hard-Core Model on $\mathbb{Z}^2$". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Ed. by Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and José D. P. Rolim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 379–394 (cit. on p. 136).

[Bla+19] Antonio Blanca, Yuxuan Chen, David Galvin, Dana Randall, and Prasad Tetali. "Phase Coexistence for the Hard-Core Model on $\mathbb{Z}^2$". In: *Combinatorics, Probability & Computing* 28 (2019), pp. 1–22 (cit. on p. 136).

[Bla+20] Antonio Blanca, Andreas Galanis, Leslie Ann Goldberg, Daniel Štefankovič, Eric Vigoda, and Kuan Yang. "Sampling in Uniqueness from the Potts and Random-Cluster Models on Random Regular Graphs". In: *SIAM Journal on Discrete Mathematics* 34 (1 2020) (cit. on p. 198).

[Bla+22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. "On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2022, pp. 3670–3692 (cit. on pp. 15, 17, 132, 148, 173, 175, 196).

[Bla16] Antonio Blanca-Pimentel. "Random-cluster dynamics". PhD thesis. EECS Department, University of California, Berkeley, Aug. 2016 (cit. on p. 199).

[BLM16] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. 1st ed. Oxford University Press, 2016 (cit. on p. 160).

[BLP06] R.B. Bapat, A.K. Lal, and Sukanta Pati. "A q-analogue of the distance matrix of a tree". In: *Linear Algebra and its Applications* 416.2 (2006), pp. 799–814. ISSN: 0024-3795 (cit. on p. 290).

[BLP22] Petter Brändén, Jonathan Leake, and Igor Pak. "Lower bounds for contingency tables via Lorentzian polynomials". In: *Israel Journal of Mathematics* (2022). To Appear (cit. on p. 109).

[Blu+05] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. "Practical Privacy: The SuLQ Framework". In: PODS '05. Baltimore, Maryland: Association for Computing Machinery, 2005, pp. 128–138. ISBN: 1595930620 (cit. on p. 8).

[BN19] 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) (cit. on pp. 148, 150, 151).

[BNH80] Tomáš Boublik, Ivo Nezbeda, and Karel Hlavaty. *Statistical thermodynamics of simple liquids and their mixtures*. Fundamental Studies in Engineering. Elsevier, 1980 (cit. on p. 13).

[BO80] Tom Brylawski and James G. Oxley. "Several identities for the characteristic polynomial of a combinatorial geometry". In: *Discrete Mathematics* 31 (2 1980), pp. 161–170 (cit. on p. 200).

[BO81] Tom Brylawski and James G. Oxley. "The Broken-circuit Complex: Its Structure and Factorizations". In: *Eur. J. Comb.* 2 (1981), pp. 107–121 (cit. on p. 200).

[Bor+13] 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 (cit. on p. 178).

[Bor+20] Christian Borgs, Jennifer Chayes, Tyler Helmuth, Will Perkins, and Prasad Tetali. "Efficient Sampling and Counting Algorithms for the Potts Model on $\mathbb{Z}^d$ at All Temperatures". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2020. Chicago, IL, USA: Association for Computing Machinery, 2020, pp. 738–751. ISBN: 9781450369794. DOI: 10.1145/3357713.3384271 (cit. on p. 198).

[Bor19] Charles Bordenave. "A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts". In: *Annales scientifiques de l'École normale supérieure* (2019) (cit. on p. 198).

[Bou86] Jean Bourgain. "On High Dimensional Maximal Functions Associated to Convex Bodies". In: *American Journal of Mathematics* 108.6 (1986), pp. 1467–1476 (cit. on p. 195).

[Bov06] Anton Bovier. *Statistical Mechanics of Disordered Systems. A Mathematical Perspective*. Cambridge Series in Statistical and Probabilistic Mathematics 18. Cambridge University Press, 2006 (cit. on pp. 7, 18).

[BR09] 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 (cit. on p. 112).

[BR19] Alexander Barvinok and Guus Regts. "Weighted counting of solutions to sparse systems of equations". In: *Combinatorics, Probability and Computing* 28 (5 2019), pp. 696–719 (cit. on p. 198).

[BR94] Mihir Bellare and John Rompel. "Randomness-efficient oblivious sampling". In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. 1994, pp. 276–287. DOI: 10.1109/SFCS.1994.365687 (cit. on p. 17).

[Bra+20a] Tom Braden, June Huh, Jacob P. Matherne, Nicholas Proudfoot, and Botong Wang. "A semi-small decomposition of the Chow ring of a matroid". In: *arXiv preprint arXiv:2002.03341* (2020) (cit. on pp. 90, 110).

[Bra+20b] Tom Braden, June Huh, Jacob P. Matherne, Nicholas Proudfoot, and Botong Wang. "Singular Hodge theory for combinatorial geometries". In: *arXiv preprint arXiv:2010.06088* (2020) (cit. on pp. 90, 110).

[Brä07] Petter Brändén. "Polynomials with the half-plane property and matroid theory". In: *Advances in Mathematics* 216.1 (2007), pp. 302–320 (cit. on p. 109).

[Brä10] Petter Brändén. "Discrete concavity and the half-plane property". In: *SIAM J. Discrete Math.* 24.3 (2010), pp. 921–933. ISSN: 0895-4801. DOI: 10.1137/090758738 (cit. on p. 107).

[Bro+11] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press LLC, 2011 (cit. on pp. 7, 9).

[Bro89] Andrei Z. Broder. "Generating random spanning trees". In: *30th Annual Symposium on Foundations of Computer Science*. Vol. 89. 1989, pp. 442–447. DOI: 10.1109/SFCS.1989.63516 (cit. on p. 189).

[Bry72] Thomas H. Brylawski. "A Decomposition for Combinatorial Geometries". In: *Transactions of the American Mathematical Society* 171 (1972), pp. 235–282 (cit. on p. 200).

[Bry77] Tom Brylawski. "The Broken-Circuit Complex". In: *Transactions of the American Mathematical Society* 234.2 (1977) (cit. on pp. 199, 200).

[BS07] Alexander Barvinok and Alex Samorodnitsky. "Random weighting, asymptotic counting, and inverse isoperimetry". In: *Israel Journal of Mathematics* 158.1 (Mar. 2007), pp. 159–191. ISSN: 1565-8511 (cit. on p. 59).

[BS13] R. B. Bapat and S. Sivasubramanian. "Product Distance Matrix of a Graph and Squared Distance Matrix of a Tree". In: *Applicable Analysis and Discrete Mathematics* 7.2 (Oct. 2013), pp. 285–301 (cit. on p. 293).

[BS15a] R.B. Bapat and Sivaramakrishnan Sivasubramanian. "Product distance matrix of a tree with matrix weights". In: *Linear Algebra and its Applications* 468 (2015). 18th ILAS Conference, pp. 145–153. ISSN: 0024-3795. DOI: https://doi.org/10.1016/j.laa.2014.03.034 (cit. on p. 297).

[BS15b] Antonio Blanca and Alistair Sinclair. "Dynamics for the Mean-field Random-cluster Model". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*. Ed. by Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim. Vol. 40. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015, pp. 528–543. ISBN: 978-3-939897-89-7. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2015.528 (cit. on p. 199).

[BS16] 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 (cit. on pp. 115, 279).

[BS17a] Alexander Barvinok and Pablo Soberón. "Computing the partition function for graph homomorphisms". In: *Combinatorica* 37 (2017), pp. 633–650 (cit. on pp. 115, 279, 280, 284).

[BS17b] Antonio Blanca and Alistair Sinclair. "Random-cluster dynamics in $\mathbb{Z}^2$". In: *Probability Theory and Related Fields* 168 (3 2017), pp. 821–847 (cit. on p. 199).

[BS94] J. van den Berg and J. E. Steif. "Percolation and the hard-core lattice gas model". In: *Stochastic Processes and their Applications* 49 (2 1994), pp. 179–197 (cit. on p. 13).

[BŚ97] W. Ballmann and J. Świątkowski. "On L2-cohomology and property (T) for automorphism groups of polyhedral cell complexes". In: *Geom. Funct. Anal.* 7.4 (1997), pp. 615–645 (cit. on p. 46).

[BT03] Sergey Bobkov and Prasad Tetali. "Modified Log-Sobolev Inequalities, Mixing and Hypercontractivity". In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '03. San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 287–296 (cit. on pp. 159, 160).

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar. 2004. DOI: 10.1017/CBO9780511804441 (cit. on p. 10).

[BZ20] Adrian Barbu and Song-Chun Zhu. "Hamiltonian and Langevin Monte Carlo". In: *Monte Carlo Methods*. Singapore: Springer Singapore, 2020, pp. 281–325. ISBN: 978-981-13-2971-5. DOI: 10.1007/978-981-13-2971-5_9 (cit. on p. 9).

[Car+22] Charlie Carlson, Ewan Davies, Nicolas Fraiman, Alexandra Kolla, Aditya Potukuchi, and Corrine Yap. "Algorithms for the ferromagnetic Potts model on expanders". In: *arXiv preprint arXiv:2204.01923* (2022) (cit. on p. 198).

[Car72] Pierre Cartier. "Fonctions harmoniques sur un arbre". In: *Symposia Mathematica* IX (Convegno di Calcolo delle Probabilità, INDAM, Rome, 1971) (1972). MR0353467 (50:5950), pp. 203–270 (cit. on p. 289).

[CC09] Eric A. Carlen and Dario Cordero–Erausquin. "Subadditivity of The Entropy and its Relation to Brascamp–Lieb Type Inequalities". In: *Geometric and Functional Analysis* 19 (2009), pp. 373–405 (cit. on p. 301).

[CE22] Yuansi Chen and Ronen Eldan. "Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains". In: *arXiv preprint arXiv:2203.04163* (2022) (cit. on pp. 15, 17, 132, 161, 173, 195, 196).

[Cel+17] L. Elisa Celis, Amit Deshpande, Tarun Kathuria, Damian Straszak, and Nisheeth K. Vishnoi. "On the Complexity of Constrained Determinantal Point Processes". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017)*. Ed. by Klaus Jansen, José D. P. Rolim, David Williamson, and Santosh S. Vempala. Vol. 81. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017, 36:1–36:22. ISBN: 978-3-95977-044-6. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2017.36 (cit. on p. 197).

[Ces01] 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 (cit. on pp. 174, 175).

[CGM12] Prasad Chebolu, Leslie Ann Goldberg, and Russell Martin. "The Complexity of Approximately Counting Stable Matchings". In: *Theoretical Computer Science* Volume 437 (2012), pp. 35–68 (cit. on p. 198).

[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. "Modified log-Sobolev inequalities for strongly log-concave distributions". In: *Annals of Probability* 49.1 (2021). A preliminary version of this paper appeared in Proceedings of the 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)., pp. 506–525 (cit. on pp. 11, 12, 16, 159, 164, 165, 180, 181, 189).

[CH14] Thomas M. Carsey and Jeffrey J. Harden. *Monte Carlo Simulation and Resampling Methods for Social Science*. SAGE Publications, 2014 (cit. on p. 8).

[Cha+11] Venkat Chandrasekaran, Misha Chertkov, David Gamarnik, Devavrat Shah, and Jinwoo Shin. "Counting Independent Sets Using the Bethe Approximation". In: *SIAM Journal of Discrete Mathematics* 25.2 (2011) (cit. on p. 136).

[Che+19] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. "Improved Bounds for Randomly Sampling Colorings via Linear Programming". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '19. San Diego, California: Society for Industrial and Applied Mathematics, 2019, pp. 2216–2234 (cit. on pp. 15, 65, 66, 148, 149, 154–157, 198).

[Che+21a] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. "Optimal Mixing Time for the Ising Model in the Uniqueness Regime". In: *arXiv preprint arXiv:2111.03034* (2021) (cit. on pp. 132, 173, 196).

[Che+21b] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. "Rapid mixing of Glauber dynamics via spectral independence for all degrees". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. 2021, pp. 137–148 (cit. on pp. 15, 17, 132, 173, 196).

[Che+21c] Zongchen Chen, Andreas Galanis, Leslie Ann Goldberg, Will Perkins, James Stewart, and Eric Vigoda. "Fast algorithms at low temperatures via Markov chains". In: *Random Structures & Algorithms* 58 (2 2021), pp. 294–321 (cit. on p. 198).

[Che+21d] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Rapid Mixing for Colorings via Spectral Independence". In: *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '21. Virtual Event, Virginia: Society for Industrial and Applied Mathematics, 2021, pp. 1548–1557 (cit. on pp. 15, 53, 66, 67, 132, 134, 148, 154, 198, 259).

[Che+22a] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. "Optimal mixing for two-state anti-ferromagnetic spin systems". In: *arXiv preprint arXiv:2203.07771* (2022) (cit. on pp. 15, 161, 173, 196).

[Che+22b] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Sampling Colorings and Independent Sets of Random Regular Bipartite Graphs in the Non-Uniqueness Region". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '22. USA: Society for Industrial and Applied Mathematics, 2022, pp. 2198–2207 (cit. on p. 198).

[Che21] Yuansi Chen. "An Almost Constant Lower Bound of the Isoperimetric Coefficient in the KLS Conjecture". In: *Geometric and Functional Analysis* 31 (2021), pp. 34–61 (cit. on p. 195).

[Che70] Jeff Cheeger. "A lower bound for the smallest eigenvalue of the Laplacian". In: *Problems in Analysis: A Symposium in Honor of Salomon Bochner (PMS-31)* (1970) (cit. on p. 63).

[CHL12] 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 (cit. on p. 280).

[Cho+04] Young-Bin Choe, James G. Oxley, Alan D. Sokal, and David G. Wagner. "Homogeneous multivariate polynomials with the half-plane property". In: *Adv. in Appl. Math.* 32.1-2 (2004). Special issue on the Tutte polynomial, pp. 88–187. ISSN: 0196-8858. DOI: 10.1016/S0196-8858(03)00078-2 (cit. on pp. 106, 107, 109).

[Clo10] Brian D. Cloteaux. "Approximating the Number of Bases for Almost All Matroids". In: *Congressus Numerantium* 202 (2010), pp. 149–154 (cit. on p. 59).

[CLV20] Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Rapid Mixing of Glauber Dynamics up to Uniqueness via Contraction". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. FOCS 2020. 2020, pp. 1307–1318 (cit. on pp. 14, 17, 19, 67, 132, 148, 239, 240, 242, 259, 263).

[CLV21a] Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Optimal Mixing of Glauber Dynamics: Entropy Factorization via High-Dimensional Expansion". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2021. Virtual, Italy: Association for Computing Machinery, 2021, pp. 1537–1550 (cit. on pp. 14, 16, 19, 49, 127, 132, 149, 150, 152, 154, 159, 161–164, 172, 176, 258, 259, 266).

[CLV21b] Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Spectral Independence via Stability and Applications to Holant-Type Problems". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. FOCS 2021. 2021, pp. 149–160 (cit. on pp. 17–19, 112, 121, 125, 199, 279).

[CLX09] 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 (cit. on p. 280).

[CLX11] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. "Computational Complexity of Holant Problems". In: *SIAM Journal on Computing* 40 (4 2011), pp. 1101–1132 (cit. on p. 280).

[ÇM10] Ali Çivril and Malik Magdon-Ismail. "Exponential Inapproximability of Selecting a Maximum Volume Sub-matrix". In: *CoRR* abs/1006.4349 (2010). arXiv: 1006.4349 (cit. on p. 305).

[CMM22] Zongchen Chen, Nitya Mani, and Ankur Moitra. "From algorithms to connectivity and back: finding a giant component in random k-SAT". In: *arXiv preprint arXiv:2207.02841* (2022) (cit. on pp. 17, 197).

[CMN96] Charles J Colbourn, Wendy J Myrvold, and Eugene Neufeld. "Two algorithms for unranking arborescences". In: *Journal of Algorithms* 20.2 (1996), pp. 268–281 (cit. on p. 189).

[CMT15] Pietro Caputo, Georg Menz, and Prasad Tetali. "Approximate tensorization of entropy at high temperature". en. In: *Annals of the Faculty of Sciences of Toulouse: Math 'ematique* Ser. 6, 24.4 (2015), pp. 691–716. DOI: 10.5802/afst.1460 (cit. on pp. 174, 175, 178).

[Coh16] Michael B. Cohen. "Ramanujan Graphs in Polynomial Time". In: *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. 2016, pp. 276–281. DOI: 10.1109/FOCS.2016.37 (cit. on pp. 18, 107).

[CP20] Sarah Cannon and Will Perkins. "Counting Independent Sets in Unbalanced Bipartite Graphs". In: *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '20. Salt Lake City, Utah: Society for Industrial and Applied Mathematics, 2020, pp. 1456–1466 (cit. on p. 198).

[CP21a] Pietro Caputo and Daniel Parisi. "Block Factorization of the Relative Entropy via Spatial Mixing". In: *Communications in Mathematical Physics* 388 (2021), pp. 793–818 (cit. on pp. 18, 131, 175, 269).

[CP21b] Swee Hong Chan and Igor Pak. "Log-concave poset inequalities". In: *arXiv preprint arXiv:2110.10740* (2021) (cit. on p. 105).

[CP22] Swee Hong Chan and Igor Pak. "Introduction to the combinatorial atlas". In: *arXiv preprint arXiv:2203.01533* (2022) (cit. on p. 105).

[CR70] Henry H. Crapo and Gian-Carlo Rota. *On The Foundations of Combinatorial Theory: Combinatorial Geometries*. Preliminary. The MIT Press, 1970, p. 293 (cit. on p. 200).

[Cra69] Henry H. Crapo. "The Tutte polynomial". In: *aequationes mathematicae* 3 (1969), pp. 211–229 (cit. on p. 200).

[CS22] P. Csikvári and Á. Schweitzer. "A short survey on stable polynomials, orientations and matchings". In: *Acta Mathematica Hungarica* 166 (2022), pp. 1–16 (cit. on p. 107).

[CTY15] Emma Cohen, Prasad Tetali, and Damir Yeliussizov. "Lattice path matroids: negative correlation and fast mixing". In: *arXiv preprint arXiv:1505.06710* (2015) (cit. on p. 59).

[CVZ10] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. "Dependent Randomized Rounding via Exchange Properties of Combinatorial Structures". In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2010, pp. 575–584 (cit. on pp. 9, 38).

[CZ22] Xiaoyu Chen and Xinyuan Zhang. "A Near-Linear Time Sampler for the Ising Model". In: *arXiv preprint arXiv:2207.09391* (2022) (cit. on pp. 17, 196, 199).

[DD19] Yotam Dikstein and Irit Dinur. "Agreement Testing Theorems on Layered Set Systems". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. 2019, pp. 1495–1524. DOI: 10.1109/FOCS.2019.00088 (cit. on pp. 40, 42).

[Dey22] Papri Dey. "Polynomials with Lorentzian Signature and computing Permanents". In: *arXiv preprint arXiv:2206.02759* (2022) (cit. on p. 110).

[DFJ02] Martin Dyer, Alan Frieze, and Mark Jerrum. "On Counting Independent Sets in Sparse Graphs". In: *SIAM J. Comput.* 31.5 (May 2002), pp. 1527–1541. ISSN: 0097-5397 (cit. on p. 137).

[DG00] Martin Dyer and Catherine Greenhill. "On Markov Chains for Independent Sets". In: *Journal of Algorithms* 35.1 (2000), pp. 17–49 (cit. on pp. 14, 136).

[DGJ09] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. "Matrix Norms and Rapid Mixing for Spin Systems". In: *The Annals of Applied Probability* 19.1 (Feb. 2009), pp. 71–107 (cit. on pp. 40, 152, 154).

[DH02] Persi Diaconis and Susan Holmes. "Random Walks on Trees and Matchings". In: *Electronic Journal of Probability* 7.6 (2002), pp. 1–17 (cit. on pp. 159, 160).

[DH21] Claire Donnat and Susan Holmes. "Modeling the heterogeneity in COVID-19's reproductive number and its impact on predictive scenarios". In: *Journal of Applied Statistics* (2021). Special Issue: Statistical Perspectives on Analytics for COVID-19 Data, pp. 1–29 (cit. on p. 8).

[DHP20] Michelle Delcourt, Marc Heinrich, and Guillem Perarnau. "The Glauber dynamics for edge-colorings of trees". In: *Random Structures & Algorithms* 57 (4 Dec. 2020), pp. 1050–1076 (cit. on pp. 66, 198).

[Di +14] Marco Di Summa, Friedrich Eisenbrand, Yuri Faenza, and Carsten Moldenhauer. "On largest volume simplices and sub-determinants". In: *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. SIAM. 2014, pp. 315–323 (cit. on p. 305).

[Dia09] Persi Diaconis. "The Markov Chain Monte Carlo Revolution". In: *Bull. Amer. Math. Soc.* 46.2 (2009) (cit. on p. 9).

[Dik+18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. "Boolean Function Analysis on High-Dimensional Expanders". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Ed. by Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer. Vol. 116. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, 38:1–38:20. ISBN: 978-3-95977-085-9. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2018.38 (cit. on pp. 40, 42).

[Din+21a] Irit Dinur, Yuval Filmus, Prahladh Harsha, and Madhur Tulsiani. "Explicit SoS Lower Bounds from High-Dimensional Expanders". In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Ed. by James R. Lee. Vol. 185. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 38:1–38:16. ISBN: 978-3-95977-177-1. DOI: 10.4230/LIPIcs.ITCS.2021.38 (cit. on p. 40).

[Din+21b] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. "List-Decoding with Double Samplers". In: *SIAM Journal on Computing* 50.2 (2021), pp. 301–349. DOI: 10.1137/19M1276650. eprint: https://doi.org/10.1137/19M1276650 (cit. on p. 40).

[Din+22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. "Locally Testable Codes with Constant Rate, Distance, and Locality". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 357–374 (cit. on pp. 18, 40).

[Din07] Irit Dinur. "The PCP Theorem by Gap Amplification". In: *J. ACM* 54.3 (June 2007), 12–es. ISSN: 0004-5411. DOI: 10.1145/1236457.1236459 (cit. on p. 18).

[DK17] Irit Dinur and Tali Kaufman. "High Dimensional Expanders Imply Agreement Expanders". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. 2017, pp. 974–985. DOI: 10.1109/FOCS.2017.94 (cit. on pp. 18, 33, 40–43, 150).

[DN03] Irit Dinur and Kobbi Nissim. "Revealing Information While Preserving Privacy". In: *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. PODS '03. San Diego, California: Association for Computing Machinery, 2003, pp. 202–210. ISBN: 1581136706 (cit. on p. 8).

[DN04] Cynthia Dwork and Kobbi Nissim. "Privacy-Preserving Datamining on Vertically Partitioned Databases". In: *Advances in Cryptology - CRYPTO 2004, 24th Annual International Cryptology Conference*. Vol. 3152. Lecture Notes in Computer Science. Santa Barbara, California, USA: Springer, 2004, pp. 528–544 (cit. on p. 8).

[DNS21] Anindya De, Shivam Nadimpalli, and Rocco A. Servedio. "Quantitative Correlation Inequalities via Semigroup Interpolation". In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Ed. by James R. Lee. Vol. 185. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 69:1–69:20. ISBN: 978-3-95977-177-1. DOI: 10.4230/LIPIcs.ITCS.2021.69 (cit. on p. 18).

[Dob70] Roland Lvovich Dobrushin. "Prescribing a System of Random Variables by Conditional Distributions". In: *Theory of Probability and its Applications* 15.3 (1970), pp. 458–486 (cit. on pp. 40, 152).

[Dow80] Thomas A Dowling. "On the independent set numbers of a finite matroid". In: *Annals of Discrete Mathematics*. Vol. 8. Elsevier, 1980, pp. 21–28 (cit. on p. 104).

[DS81] Persi Diaconis and Mehrdad Shahshahani. "Generating a random permutation with random transpositions". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57 (1981), pp. 159–179 (cit. on pp. 159, 160).

[DS85a] R. L. Dobrushin and S. B. Shlosman. "Completely Analytical Gibbs Fields". In: *Statistical Physics and Dynamical Systems: Rigorous Results*. Boston, MA: Birkhäuser Boston, 1985, pp. 371–403. ISBN: 978-1-4899-6653-7 (cit. on pp. 40, 152).

[DS85b] R. L. Dobrushin and S. B. Shlosman. "Constructive Criterion for the Uniqueness of Gibbs Field". In: *Statistical Physics and Dynamical Systems: Rigorous Results*. Boston, MA: Birkhäuser Boston, 1985, pp. 347–370. ISBN: 978-1-4899-6653-7 (cit. on pp. 40, 152).

[DS87a] Persi Diaconis and Mehrdad Shahshahani. "Time to Reach Stationarity in the Bernoulli–Laplace Diffusion Model". In: *SIAM Journal on Mathematical Analysis* 18 (1 1987) (cit. on pp. 159, 160).

[DS87b] R. L. Dobrushin and S. B. Shlosman. "Completely analytical interactions: Constructive description". In: *Journal of Statistical Physics* 46.5 (Mar. 1987), pp. 983–1014 (cit. on pp. 40, 152).

[DS93] Persi Diaconis and Laurent Saloff-Coste. "Comparison Theorems for Reversible Markov Chains". In: *The Annals of Applied Probability* 3.3 (1993), pp. 696–730 (cit. on p. 154).

[DS96] Persi Diaconis and Laurent Saloff-Coste. "Logarithmic Sobolev inequalities for finite Markov chains". In: *The Annals of Applied Probability* 6.3 (Aug. 1996) (cit. on pp. 159, 160, 168–170).

[DSS22a] Jian Ding, Allan Sly, and Nike Sun. "Proof of the satisfiability conjecture for large k". In: *The Annals of Mathematics* 196 (1 2022), pp. 1–388 (cit. on p. 197).

[DSS22b] Jian Ding, Jian Song, and Rongfeng Sun. "A new correlation inequality for Ising models with external fields". In: *Probability Theory and Related Fields* (2022) (cit. on p. 18).

[Dua+87] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. "Hybrid Monte Carlo". In: *Physics Letters B* 195 (2 1987), pp. 216–222 (cit. on p. 9).

[Dur+17a] David Durfee, Rasmus Kyng, John Peebles, Anup B. Rao, and Sushant Sachdeva. "Sampling Random Spanning Trees Faster than Matrix Multiplication". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2017. Montreal, Canada: Association for Computing Machinery, 2017, pp. 730–742. ISBN: 9781450345286. DOI: 10.1145/3055399.3055499 (cit. on pp. 16, 189).

[Dur+17b] David Durfee, John Peebles, Richard Peng, and Anup B Rao. "Determinant-preserving sparsification of SDDM matrices with applications to counting and sampling spanning trees". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2017, pp. 926–937 (cit. on p. 189).

[DV76] M. D. Donsker and S. R. S. Varadhan. "Asymptotic evaluation of certain Markov process expectations for large time. III". In: *Communications on Pure and Applied Mathematics* 29 (4 1976), pp. 389–461 (cit. on p. 23).

[DV83] M. D. Donsker and S. R. S. Varadhan. "Asymptotic evaluation of certain markov process expectations for large time. IV". In: *Communications on Pure and Applied Mathematics* 36 (2 1983), pp. 183–212 (cit. on p. 23).

[Dwo+06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Proceedings of the Third Conference on Theory of Cryptography*. TCC'06. New York, NY: Springer-Verlag, 2006, pp. 265–284. ISBN: 3540327312 (cit. on p. 8).

[Dye+04a] Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, and Mark Jerrum. "The Relative Complexity of Approximate Counting Problems". In: *Algorithmica* 38 (2004), pp. 471–500 (cit. on p. 198).

[Dye+04b] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. "Mixing in time and space for lattice spin systems: A combinatorial view". In: *Random Structures & Algorithms* 24 (4 2004), pp. 461–479 (cit. on pp. 17, 18, 131, 136, 269).

[Dye+06] Martin Dyer, Abraham D. Flaxman, Alan M. Frieze, and Eric Vigoda. "Randomly coloring sparse random graphs with fewer colors than the maximum degree". In: *Random Structures & Algorithms* 29 (4 2006), pp. 450–465 (cit. on p. 198).

[Dye+13] Martin Dyer, Alan Frieze, Thomas P. Hayes, and Eric Vigoda. "Randomly coloring constant degree graphs". In: *Random Structures & Algorithms* 43.2 (2013). A preliminary version of this paper appeared in Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science., pp. 181–200 (cit. on pp. 15, 66, 198).

[Dye+21] 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 (cit. on pp. 113, 127, 279).

[Edm71] Jack Edmonds. "Matroids and the greedy algorithm". In: *Mathematical Programming* (1971), pp. 127–136 (cit. on p. 11).

[Eft+16] Charilaos Efthymiou, Thomas P. Hayes, Daniel Štefankovič, Eric Vigoda, and Yitong Yin. "Convergence of MCMC and Loopy BP in the Tree Uniqueness Region for the Hard-Core Model". In: *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. 2016, pp. 704–713 (cit. on pp. 14, 132, 136).

[Eft+18] Charilaos Efthymiou, Thomas P. Hayes, Daniel Štefankovič, and Eric Vigoda. "Sampling Random Colorings of Sparse Random Graphs". In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '18. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2018, pp. 1759–1771. ISBN: 9781611975031 (cit. on p. 198).

[Eft+19] Charilaos Efthymiou, Andreas Galanis, Thomas P. Hayes, Daniel Stefankovic, and Eric Vigoda. "Improved Strong Spatial Mixing for Colorings on Trees". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Ed. by Dimitris Achlioptas and László A. Végh. Vol. 145. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, 48:1–48:16. ISBN: 978-3-95977-125-2. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2019.48 (cit. on pp. 66, 154).

[Eft14] Charilaos Efthymiou. "MCMC Sampling Colourings and Independent Sets of $G(n, d/n)$ Near Uniqueness Threshold". In: *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '14. Portland, Oregon: Society for Industrial and Applied Mathematics, 2014, pp. 305–316. ISBN: 978-1-611973-38-9 (cit. on p. 198).

[EH20] Christopher Eur and June Huh. "Logarithmic concavity for morphisms of matroids". In: *Advances in Mathematics* 367 (2020), p. 107094. ISSN: 0001-8708. DOI: https://doi.org/10.1016/j.aim.2020.107094 (cit. on p. 109).

[EHL22] Christopher Eur, June Huh, and Matt Larson. "Stellahedral geometry of matroids". In: *arXiv preprint arXiv:2207.10605* (2022) (cit. on pp. 90, 110).

[EK10] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010 (cit. on p. 8).

[EKZ21] Shai Evra, Tali Kaufman, and Gilles Zémor. "Decodable Quantum LDPC Codes beyond the $\sqrt{n}$-Distance Barrier Using High-Dimensional Expanders". In: *SIAM Journal on Computing* 0.0 (2021). Special Section FOCS 2020. DOI: 10.1137/20M1383689. eprint: https://doi.org/10.1137/20M1383689 (cit. on p. 40).

[EKZ22] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. "A spectral condition for spectral gap: fast mixing in high-temperature Ising models". In: *Probability Theory and Related Fields* 182 (2022), pp. 1035–1051 (cit. on p. 196).

[Eld13] Ronen Eldan. "Thin Shell Implies Spectral Gap Up to Polylog via a Stochastic Localization Scheme". In: *Geometric and Functional Analysis* 23 (2013), pp. 532–569 (cit. on pp. 60, 195).

[Eld20] Ronen Eldan. "Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation". In: *Probability Theory and Related Fields* 176 (2020), pp. 737–755 (cit. on p. 196).

[ELL17] 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 (cit. on pp. 152, 160).

[ENO22] Farzam Ebrahimnejad, Ansh Nagda, and Shayan Oveis Gharan. "Counting and Sampling Perfect Matchings in Regular Expanding Non-Bipartite Graphs". In: *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Ed. by Mark Braverman. Vol. 215. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 61:1–61:12. ISBN: 978-3-95977-217-4. DOI: 10.4230/LIPIcs.ITCS.2022.61 (cit. on p. 197).

[Erb+17] Matthias Erbar, Christopher Henderson, Georg Menz, and Prasad Tetali. "Ricci curvature bounds for weakly interacting Markov chains". In: *Electronic Journal of Probability* 22 (2017), pp. 1–23 (cit. on p. 196).

[ES08] Charilaos Efthymiou and Paul G. Spirakis. "Random sampling of colourings of sparse random graphs with a constant number of colours". In: *Theoretical Computer Science* 407 (1–3 2008), pp. 134–154 (cit. on p. 198).

[ES22] Ronen Eldan and Omer Shamir. "Log concavity and concentration of Lipschitz functions on the Boolean hypercube". In: *Journal of Functional Analysis* 282 (8 2022), p. 109392 (cit. on p. 196).

[Eur20] Christopher Eur. "Divisors on matroids and their volumes". In: *Journal of Combinatorial Theory, Series A* 169 (2020) (cit. on pp. 90, 110).

[Fed72] Valerii Vadimovich Fedorov. *Theory of Optimal Experiments*. 1st ed. Academic Press, Inc., 1972 (cit. on p. 305).

[Fen+21] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. "Rapid Mixing from Spectral Independence beyond the Boolean Domain". In: *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '21. Virtual Event, Virginia: Society for Industrial and Applied Mathematics, 2021, pp. 1558–1577 (cit. on pp. 15, 54, 66, 67, 132, 134, 148, 154, 198, 259).

[FG22] Neil Jiuyu Fan and Peter Long Guo. "Upper bounds of Schubert polynomials". In: *Science China Mathematics* 65 (2022), pp. 1319–1330 (cit. on p. 109).

[FGW22] Weiming Feng, Heng Guo, and Jiaheng Wang. "Swendsen-Wang dynamics for the ferromagnetic Ising model with external fields". In: *arXiv preprint arXiv:2205.01985* (2022) (cit. on p. 199).

[FH10] Wai Shing Fung and Nicholas J. A. Harvey. "Graph Sparsification by Edge-Connectivity and Random Spanning Trees". In: *arXiv preprint arXiv:1005.0265* (2010) (cit. on p. 16).

[Fis84] Peter C. Fishburn. "A correlational inequality for linear extensions of a poset". In: *Order* 1 (1984), pp. 127–137 (cit. on p. 18).

[FK72] Cornelis Marius Fortuin and Pieter Willem Kasteleyn. "On the random-cluster model: I. Introduction and relation to other models". In: *Physica* 57 (4 1972), pp. 536–564 (cit. on p. 100).

[FKG71] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. "Correlation inequalities on some partially ordered sets". In: *Communications in Mathematical Physics* 22.2 (1971), pp. 89–103 (cit. on pp. 18, 199).

[FM92] Tomás Feder and Milena Mihail. "Balanced Matroids". In: *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*. STOC '92. Victoria, British Columbia, Canada: Association for Computing Machinery, 1992, pp. 26–38 (cit. on pp. 17, 38, 59, 114).

[For71] Cornelis Marius Fortuin. "On the random-cluster model". PhD thesis. Leiden University, 1971 (cit. on p. 100).

[For72a] Cornelis Marius Fortuin. "On the random-cluster model: II. The percolation model". In: *Physica* 58 (3 1972), pp. 393–418 (cit. on p. 100).

[For72b] Cornelis Marius Fortuin. "On the random-cluster model: III. The simple random-cluster model". In: *Physica* 59 (4 1972), pp. 545–570 (cit. on p. 100).

[FOW22] Yuval Filmus, Ryan O'Donnell, and Xinyu Wu. "Log-Sobolev inequality for the multislice, with applications". In: *Electronic Journal of Probability* 27 (2022), pp. 1–30 (cit. on pp. 159, 160).

[Fox+12] Jacob Fox, Mikhail Gromov, Vincent Lafforgue, Assaf Naor, and János Pach. "Overlap properties of geometric expanders". In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 2012 (671 2012), pp. 49–83 (cit. on p. 40).

[Fri+21] Tobias Friedrich, Andreas Göbel, Martin S. Krejca, and Marcus Pappik. "A Spectral Independence View on Hard Spheres via Block Dynamics". In: *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*. Vol. 198. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 66:1–66:15 (cit. on p. 197).

[Fri08] Joel Friedman. "A proof of Alon's second eigenvalue conjecture and related problems". In: *Memoirs of the American Mathematical Society* (2008) (cit. on p. 198).

[Fri91] Joel Friedman. "The Spectra of Infinite Hypertrees". In: *SIAM Journal on Computing* 20.5 (Oct. 1991), pp. 951–961. ISSN: 0097-5397. DOI: 10.1137/0220058 (cit. on p. 289).

[FV06] Alan Frieze and Juan Vera. "On randomly colouring locally sparse graphs". In: *Discrete Mathematics and Theoretical Computer Science* 8 (2006) (cit. on pp. 66, 198).

[FV07] Alan Frieze and Eric Vigoda. "A survey on the use of Markov chains to randomly sample colorings". In: *Combinatorica* (Jan. 2007). DOI: 10.1093/acprof:oso/9780198571278.003.0004 (cit. on pp. 15, 66, 198).

[FV17] Sacha Friedli and Yvan Velenik. "Statistical Mechanics of Lattice Systems. A Concrete Mathematical Introduction". In: (2017) (cit. on pp. 7, 18).

[Gal+14] Andreas Galanis, Qi Ge, Daniel Štefankovič, Eric Vigoda, and Linji Yang. "Improved Inapproximability Results for Counting Independent Sets in the Hard-core Model". In: *Random Struct. Algorithms* 45.1 (Aug. 2014), pp. 78–110. ISSN: 1042-9832 (cit. on pp. 14, 131, 137).

[Gal+22] Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Andrés Herrera-Poyatos. "Fast sampling of satisfying assignments from random k-SAT". In: *arXiv preprint arXiv:2206.15308* (2022) (cit. on pp. 17, 197).

[Gal71] Giovanni Gallavotti. "A Proof of the Griffiths Inequalities for the X-Y Model". In: *Studies in Applied Mathematics* 50 (1971), pp. 89–92 (cit. on p. 18).

[Gam99] Anna Gambin. "On approximating the number of bases of exchange preserving matroids". In: *International Symposium on Mathematical Foundations of Computer Science*. Springer. 1999, pp. 332–342 (cit. on p. 59).

[Gan+06] Rajiv Gandhi, Samir Khuller, Srinivasan Parthasarathy, and Aravind Srinivasan. "Dependent Rounding and Its Applications to Approximation Algorithms". In: *J. ACM* 53.3 (May 2006), pp. 324–360. ISSN: 0004-5411. DOI: 10.1145/1147954.1147956 (cit. on p. 38).

[Går59] Lars Gårding. "An Inequality for Hyperbolic Polynomials". In: *Journal of Mathematics and Mechanics* 8.6 (1959), pp. 957–965 (cit. on p. 108).

[Gar73] H. Garland. "p-adic curvature and the cohomology of discrete subgroups of p-adic groups". In: *Annals of Mathematics* 97.3 (1973), pp. 375–423 (cit. on pp. 43, 46).

[Gel+87] I.M Gelfand, R.M Goresky, R.D MacPherson, and V.V Serganova. "Combinatorial geometries, convex polyhedra, and schubert cells". In: *Advances in Mathematics* 63.3 (1987), pp. 301–316. ISSN: 0001-8708. DOI: https://doi.org/10.1016/0001-8708(87)90059-4 (cit. on p. 29).

[GG06] B. T. Graham and G. R. Grimmett. "Influence and sharp-threshold theorems for monotonic measures". In: *The Annals of Probability* 34.5 (2006), pp. 1726–1745 (cit. on p. 40).

[GG84] Stuart Geman and Donald Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6 (1984), pp. 721–741 (cit. on p. 9).

[GGS21] Andreas Galanis, Leslie Ann Goldberg, and James Stewart. "Fast Algorithms for General Spin Systems on Bipartite Expanders". In: *ACM Trans. Comput. Theory* 13.4 (Aug. 2021). ISSN: 1942-3454. DOI: 10.1145/3470865 (cit. on p. 198).

[GHS70] Robert B. Griffiths, Charles Angas Hurst, and Seymour Sherman. "Concavity of Magnetization of an Ising Ferromagnet in a Positive External Field". In: *Journal of Mathematical Physics* 11 (1970), pp. 790–795 (cit. on p. 18).

[Gil14] Jeff Gill. *Bayesian Methods A Social and Behavioral Sciences Approach*. 3rd ed. Chapman and Hall/CRC, 2014 (cit. on p. 8).

[Gin70] J. Ginibre. "General formulation of Griffiths' inequalities". In: *Communications in Mathematical Physics* 16 (1970), pp. 310–328 (cit. on p. 18).

[GJ07] Leslie Ann Goldberg and Mark Jerrum. "The Complexity of Ferromagnetic Ising with Local Fields". In: *Combinatorics, Probability and Computing* 16 (1 2007), pp. 43–61 (cit. on p. 198).

[GJ08] Leslie Ann Goldberg and Mark Jerrum. "Inapproximability of the Tutte polynomial". In: *Inform. and Comput* 206.7 (2008), pp. 908–929 (cit. on pp. 100, 198).

[GJ09] Geoffrey Grimmett and Svante Janson. "Random even graphs". In: *The Electronic Journal of Combinatorics* 16.1 (2009), R46 (cit. on pp. 112, 199).

[GJ12a] Leslie Ann Goldberg and Mark Jerrum. "Approximating the partition function of the ferromagnetic Potts model". In: *J. ACM* 59.5 (2012), pp. 1–31 (cit. on pp. 100, 198, 199).

[GJ12b] Leslie Ann Goldberg and Mark Jerrum. "Inapproximability of the Tutte polynomial of a planar graph". In: *Computational Complexity* 21 (4 2012), pp. 605–642 (cit. on pp. 100, 198, 199).

[GJ13] Leslie Ann Goldberg and Mark Jerrum. "Approximating the Tutte polynomial of a binary matroid and other related combinatorial polynomials". In: *Journal of Computer and System Sciences* 79 (1 2013), pp. 68–78 (cit. on pp. 100, 198, 199).

[GJ14] Leslie Ann Goldberg and Mark Jerrum. "The complexity of computing the sign of the Tutte polynomial". In: *SIAM J. Comput.* 43 (2014), pp. 1921–1952 (cit. on pp. 100, 198, 199).

[GJ18] Heng Guo and Mark Jerrum. "Random cluster dynamics for the Ising model is rapidly mixing". In: *The Annals of Applied Probability* 28.2 (2018), pp. 1292–1313 (cit. on p. 100).

[GJ19] Heng Guo and Mark Jerrum. "A Polynomial-Time Approximation Algorithm for All-Terminal Network Reliability". In: *SIAM Journal on Computing* 48 (3 2019), pp. 964–978 (cit. on p. 58).

[GJ21] Heng Guo and Mark Jerrum. "Approximately counting bases of bicircular matroids". In: *Combinatorics, Probability and Computing* 30 (1 2021), pp. 124–135 (cit. on pp. 59, 199).

[GJK10] Leslie Ann Goldberg, Mark Jerrum, and Marek Karpinski. "The mixing time of Glauber dynamics for coloring regular trees". In: *Random Structures & Algorithms* 36 (4 2010), pp. 464–476 (cit. on p. 198).

[GJP03] Leslie Ann Goldberg, Mark Jerrum, and Mike Paterson. "The computational complexity of two-state spin systems". In: *Random Structures and Algorithms* 23.2 (2003), pp. 133–154 (cit. on p. 137).

[GK07] David Gamarnik and Dmitriy Katz. "Correlation Decay and Deterministic FPTAS for Counting List-colorings of a Graph". In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '07. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2007, pp. 1245–1254. ISBN: 978-0-898716-24-5 (cit. on pp. 18, 131, 133, 134, 201).

[GK12] David Gamarnik and Dmitriy Katz. "Correlation decay and deterministic FPTAS for counting colorings of a graph". In: *Journal of Discrete Algorithms* 12 (2012), pp. 29–47 (cit. on pp. 18, 66, 131, 134, 201).

[GKM15] David Gamarnik, Dmitriy Katz, and Sidhant Misra. "Strong Spatial Mixing of List Coloring of Graphs". In: *Random Struct. Algorithms* 46.4 (July 2015), pp. 599–613. ISSN: 1042-9832. DOI: 10.1002/rsa.20518 (cit. on pp. 66, 67, 154, 201).

[GL18] Heng Guo and Pinyan Lu. "Uniqueness, Spatial Mixing, and Approximation for Ferromagnetic 2-Spin Systems". In: *ACM Transactions on Computation Theory* 10.4 (2018) (cit. on pp. 247, 255).

[Gla63] Roy J. Glauber. "Time-Dependent Statistics of the Ising Model". In: *Journal of Mathematical Physics* 4 (2 1963) (cit. on p. 9).

[GLL20] Heng Guo, Jingcheng Liu, and Pinyan Lu. "Zeros of ferromagnetic 2-spin systems". In: *Proceedings of the 31st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2020, pp. 181–192 (cit. on p. 247).

[GLL22] Tom Gur, Noam Lifshitz, and Siqi Liu. "Hypercontractivity on High Dimensional Expanders". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 176–184. ISBN: 9781450392648. DOI: 10.1145/3519935.3520004 (cit. on pp. 40, 42).

[GM07] Antoine Gerschenfeld and Andrea Montanari. "Reconstruction for Models on Random Graphs". In: *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2007, pp. 194–204 (cit. on p. 124).

[GM20] Heng Guo and Giorgos Mousa. "Local-to-Global Contraction in Simplicial Complexes". In: *arXiv preprint arXiv:2012.14317* (2020) (cit. on pp. 49, 159).

[GM94] Ulf Grenander and Michael I. Miller. "Representations of Knowledge in Complex Systems". In: *Journal of the Royal Statistical Society Series B* 56.4 (1994), pp. 549–603 (cit. on p. 9).

[GMP05] L.A. Goldberg, R. Martin, and M. Paterson. "Strong spatial mixing with fewer colors for lattice graphs". In: *SIAM Journal on Computing* 35.2 (2005), pp. 486–517 (cit. on pp. 66, 198).

[God93] Christopher David Godsil. *Algebraic Combinatorics*. 1st ed. Chapman & Hall, Inc., 1993 (cit. on pp. 259, 260).

[Goe+15] Ashish Goel, Sanjeev Khanna, Sharath Raghvendra, and Hongyang Zhang. "Connectivity in Random Forests and Credit Networks". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '15. San Diego, California: Society for Industrial and Applied Mathematics, 2015, pp. 2037–2048 (cit. on pp. 11, 189).

[Goe04] Sharad Goel. "Modified logarithmic Sobolev inequalities for some models of random walk". In: *Stochastic Processes and their Applications* 114 (1 2004), pp. 51–79 (cit. on p. 160).

[Gol+06] Leslie Ann Goldberg, Markus Jalsenius, Russell Martin, and Mike Paterson. "Improved mixing bounds for the anti-ferromagnetic Potts model on $\mathbb{Z}^2$". In: *LMS Journal of Computation and Mathematics* 9 (2006), pp. 1–20 (cit. on p. 198).

[Gol21] Louis Golowich. "Improved Product-Based High-Dimensional Expanders". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Ed. by Mary Wootters and Laura Sanità. Vol. 207. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 38:1–38:17. ISBN: 978-3-95977-207-5. DOI: 10.4230/LIPIcs.APPROX/RANDOM.2021.38 (cit. on p. 40).

[Gre00] Catherine S. Greenhill. "The complexity of counting colourings and independent sets in sparse graphs and hypergraphs". In: *computational complexity* 9 (2000), pp. 52–72 (cit. on p. 137).

[Gri09] G. R Grimmett. *The Random-Cluster Model*. Berlin: Springer-Verlag, 2009 (cit. on pp. 100, 199, 304).

[Gri67a] Robert B. Griffiths. "Correlations in Ising Ferromagnets. I". In: *Journal of Mathematical Physics* 8 (3 1967), p. 478 (cit. on p. 18).

[Gri67b] Robert B. Griffiths. "Correlations in Ising Ferromagnets. II. External Magnetic Fields". In: *Journal of Mathematical Physics* 8 (3 1967), p. 484 (cit. on p. 18).

[Gri67c] Robert B. Griffiths. "Correlations in Ising ferromagnets. III. A mean-field bound for binary correlations". In: *Communications in Mathematical Physics* 6 (1967), pp. 121–127 (cit. on p. 18).

[Gri69] Robert B. Griffiths. "Rigorous Results for Ising Ferromagnets of Arbitrary Spin". In: *Journal of Mathematical Physics* 10 (9 1969), p. 1559 (cit. on p. 18).

[Gro75] Leonard Gross. "Logarithmic Sobolev inequalities". In: *American Journal of Mathematics* 97.4 (1975), pp. 1061–1083 (cit. on pp. 159, 160).

[GRV09] Navin Goyal, Luis Rademacher, and Santosh Vempala. "Expanders via Random Spanning Trees". In: *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '09. USA: Society for Industrial and Applied Mathematics, 2009 (cit. on p. 16).

[GŠV15] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Inapproximability for Antiferromagnetic Spin Systems in the Tree Nonuniqueness Region". In: *J. ACM* 62.6 (Dec. 2015). ISSN: 0004-5411. DOI: 10.1145/2785964 (cit. on pp. 14, 15, 65, 66, 131, 137, 198).

[GŠV16] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Inapproximability of the Partition Function for the Antiferromagnetic Ising and Hard-Core Models". In: *Combinatorics, Probability and Computing* 25.4 (2016), pp. 500–559 (cit. on pp. 14, 131, 137).

[Gül97] Osman Güler. "Hyperbolic Polynomials and Interior Point Methods for Convex Programming". In: *Mathematics of Operations Research* 22.2 (1997), pp. 350–377 (cit. on p. 108).

[Guo+21] 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 (cit. on pp. 18, 112, 114, 127, 130).

[Gur05] Leonid Gurvits. "On the Complexity of Mixed Discriminants and Related Problems". In: *Mathematical Foundations of Computer Science 2005*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 447–458. ISBN: 978-3-540-31867-5 (cit. on p. 19).

[Gur06a] Leonid Gurvits. "Hyperbolic polynomials approach to Van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications". In: *STOC*. 2006, pp. 417–426 (cit. on p. 107).

[Gur06b] Leonid Gurvits. "The Van der Waerden conjecture for mixed discriminants". In: *Advances in Mathematics* 200 (3 2006), pp. 435–454 (cit. on p. 107).

[Gur08] Leonid Gurvits. "Van der Waerden/Schrijver-Valiant like Conjectures and Stable (aka Hyperbolic) Homogeneous Polynomials: One Theorem for all". In: *The Electronic Journal of Combinatorics* 15 (2008) (cit. on pp. 105, 107).

[Gur09] Leonid Gurvits. "A polynomial-time algorithm to approximate the mixed volume within a simply exponential factor". In: *Discrete & Computational Geometry* 41.4 (2009), pp. 533–555 (cit. on pp. 11, 89, 105).

[Gur10] Leonid Gurvits. "On multivariate Newton-like inequalities". In: *Advances in Combinatorial Mathematics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 61–78 (cit. on pp. 11, 12, 89, 105).

[GW12] Anna Gundert and Uli Wagner. "On Laplacians of Random Complexes". In: *Proceedings of the Twenty-Eighth Annual Symposium on Computational Geometry*. SoCG '12. Chapel Hill, North Carolina, USA: Association for Computing Machinery, 2012, pp. 151–160. ISBN: 9781450312998. DOI: 10.1145/2261250.2261272 (cit. on p. 46).

[GW13] Jeff Gill and Christopher Witko. "Bayesian Analytical Methods: A Methodological Prescription for Public Administration". In: *Journal of Public Administration Research and Theory* 23 (2 2013), pp. 457–494 (cit. on p. 8).

[GW95] Michel X. Goemans and David P. Williamson. "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming". In: *J. ACM* 42.6 (Nov. 1995), pp. 1115–1145. ISSN: 0004-5411. DOI: 10.1145/227683.227684 (cit. on p. 9).

[GZ03] A. Guionnet and B. Zegarlinksi. "Lectures on Logarithmic Sobolev Inequalities". In: *Séminaire de Probabilités, XXXVI, volume 1801 of Lecture Notes in Math.* Springer, Berlin, 2003, pp. 1–134 (cit. on pp. 160, 174, 175).

[Har60] T. E. Harris. "A lower bound for the critical probability in a certain percolation process". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 56 (1 1960), pp. 13–20 (cit. on p. 18).

[Hås01] Johan Håstad. "Some Optimal Inapproximability Results". In: *J. ACM* 48.4 (July 2001), pp. 798–859. ISSN: 0004-5411 (cit. on p. 14).

[Has70] W. K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970), pp. 97–109 (cit. on pp. 9, 13).

[Hås99] Johan Håstad. "Clique is hard to approximate within $n^{1-\epsilon}$". In: *Acta Mathematica* 182 (1999), pp. 105–142 (cit. on p. 14).

[Hay06] Thomas P. Hayes. "A simple condition implying rapid mixing of single-site dynamics on spin systems". In: *FOCS* (2006), pp. 39–46 (cit. on pp. 40, 136, 152, 154).

[Hei20] Marc Heinrich. "Glauber dynamics for colourings of chordal graphs and graphs of bounded treewidth". In: *arXiv preprint arXiv:2010.16158* (2020) (cit. on p. 198).

[Her72] Andrew Heron. "Matroid polynomials". In: *Combinatorics (Proc. Conf. Combinatorial Math., Math. Inst.)* Oxford: Inst. of Math. and its Appl., Southend-on-Sea, 1972, pp. 164–202 (cit. on p. 199).

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. 2nd ed. Cambridge university press, 2013 (cit. on pp. 21, 22).

[HJP22] Tyler Helmuth, Matthew Jenssen, and Will Perkins. "Finite-size scaling, phase coexistence, and algorithms for the random cluster model on random graphs". In: *Annales de l'institut Henri Poincare (B) Probability and Statistics* (2022). To Appear (cit. on p. 198).

[HK12] June Huh and Eric Katz. "Log-concavity of characteristic polynomials and the Bergman fan of matroids". In: *Math. Ann.* 354 (2012), pp. 1103–1116 (cit. on pp. 104, 110, 199).

[HK21] David G. Harris and Vladimir Kolmogorov. "Parameter estimation for Gibbs distributions". In: *arXiv preprint arXiv:2007.10824* (2021) (cit. on p. 26).

[HL72] Ole J. Heilmann and Elliott H. Lieb. "Theory of monomer-dimer systems". In: *Communications in Mathematical Physics* 25 (1972), pp. 190–232 (cit. on pp. 18, 292).

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. "Expander graphs and their applications". In: *Bulletin of the American Mathematical Society* 43.4 (2006), pp. 439–561 (cit. on p. 289).

[HLZ16] 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 (cit. on pp. 112, 127, 279).

[HM13] Jean-Pierre Hansen and Ian R. McDonald. *Theory of Simple Liquids With Applications to Soft Matter*. 4th ed. Academic Press, 2013 (cit. on p. 13).

[HMC19] Robert V. Hogg, Joseph W. McKean, and Allen T. Craig. "Introduction to Mathematical Statistics". In: (2019), p. 768 (cit. on p. 53).

[HMR13] Ghassan Hamra, Richard MacLehose, and David Richardson. "Markov chain Monte Carlo: an introduction for epidemiologists". In: *International Journal of Epidemiology* 42 (2 2013), pp. 627–634 (cit. on p. 8).

[HO14] Nicholas J. A. Harvey and Neil Olver. "Pipage Rounding, Pessimistic Estimators and Matrix Concentration". In: *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '14. Portland, Oregon: Society for Industrial and Applied Mathematics, 2014, pp. 926–945. ISBN: 9781611973389 (cit. on p. 9).

[Hog74] Stuart G. Hoggar. "Chromatic polynomials and logarithmic concavity". In: *Journal of Combinatorial Theory, Series B* 16 (3 1974), pp. 248–254 (cit. on p. 199).

[Hol74] Richard Holley. "Remarks on the FKG inequalities". In: *Communications in Mathematical Physics* 36.3 (1974), pp. 227–231 (cit. on p. 18).

[Hou+06] J. Ben Hough, Manjunath Krishnapur, Yuval Peres, and Bálint Virág. "Determinantal Processes and Independence". In: *Probability Surveys* 3 (2006), pp. 206–229 (cit. on p. 8).

[HPR20] Tyler Helmuth, Will Perkins, and Guus Regts. "Algorithmic Pirogov-Sinai theory". In: *Probability Theory and Related Fields* 176 (2020), pp. 851–895 (cit. on p. 198).

[HPS18] Chris Hall, Doron Puder, and William F. Sawin. "Ramanujan coverings of graphs". In: *Advances in Mathematics* 323 (2018). A preliminary version of this paper appeared in Proceedings of the 48th Annual ACM Symposium on Theory of Computing., pp. 367–410 (cit. on pp. 18, 107).

[HS07] 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 (cit. on p. 173).

[HS20] Jonathan Hermon and Justin Salez. "Modified log-Sobolev inequalities for strong-Rayleigh measures". In: *arXiv preprint arXiv:1902.02775* (2020) (cit. on pp. 18, 160).

[HS89] Yahya Ould Hamidoune and Isabelle Salaün. "On the independence numbers of a matroid". In: *Journal of Combinatorial Theory Series B* 47.2 (1989), pp. 146–152 (cit. on p. 104).

[HSV18] Nicholas J. A. Harvey, Piyush Srivastava, and Jan Vondrák. "Computing the Independence Polynomial: From the Tree Threshold Down to the Roots". In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '18. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2018, pp. 1557–1576. ISBN: 978-1-6119-7503-1 (cit. on p. 136).

[HSW21] June Huh, Benjamin Schröter, and Botong Wang. "Correlation bounds for fields and matroids". In: *Journal of the European Mathematical Society* 24.4 (2021), pp. 1335–1351. DOI: 10.4171/JEMS/1119 (cit. on pp. 12, 59, 90, 104, 110).

[Hub15] Mark Huber. "Approximation algorithms for the normalizing constant of Gibbs distributions". In: *The Annals of Applied Probability* 25.2 (2015), pp. 974–985 (cit. on p. 26).

[Huh+22] June Huh, Jacob P. Matherne, Karola Mészáros, and Avery St. Dizier. "Logarithmic concavity of Schur and related polynomials". In: *Transactions of the American Mathematical Society* 375 (2022), pp. 4411–4427 (cit. on p. 109).

[Huh12] June Huh. "Milnor numbers of projective hypersurfaces and the chromatic polynomial of graphs". In: *Journal of the American Mathematical Society* 25.3 (2012), pp. 907–927 (cit. on pp. 104, 110, 199).

[Huh18] June Huh. "Combinatorial applications of the Hodge–Riemann relations". In: 3 (2018), pp. 3079–3098 (cit. on pp. 90, 110).

[Huh22] June Huh. "Combinatorics and Hodge theory". In: 1 (2022), pp. 2–29 (cit. on pp. 90, 110).

[HV03] Thomas P. Hayes and Eric Vigoda. "A non-Markovian coupling for randomly sampling colorings". In: *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* 2003, pp. 618–627 (cit. on pp. 15, 66, 198).

[HV05] Tom Hayes and Eric Vigoda. "Coupling with the Stationary Distribution and Improved Sampling for Colorings and Independent Sets". In: *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '05. Vancouver, British Columbia: Society for Industrial and Applied Mathematics, 2005, pp. 971–979. ISBN: 0-89871-585-7 (cit. on pp. 66, 136, 198).

[HV07] Thomas P. Hayes and Eric Vigoda. "Variable length path coupling". In: *Random Structures and Algorithms* 31 (3 2007), pp. 251–272 (cit. on pp. 156, 157).

[HVV15] Thomas P. Hayes, Juan C. Vera, and Eric Vigoda. "Randomly coloring planar graphs with fewer colors than the maximum degree". In: *Random Structures & Algorithms* 47 (4 2015). An extended abstract of this paper appeared in Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC), pp. 731–759 (cit. on pp. 66, 198).

[HW17] June Huh and Botong Wang. “Enumeration of points, lines, planes, etc”. In: *Acta Mathematica* 218.2 (2017), pp. 297–317 (cit. on pp. 59, 64, 90, 110).

[HWY22] Kun He, Kewen Wu, and Kuan Yang. “Improved Bounds for Sampling Solutions of Random CNF Formulas”. In: *arXiv preprint arXiv:2207.11892* (2022) (cit. on p. 197).

[Isi25] Ernst Ising. “Beitrag zur Theorie des Ferromagnetismus”. In: *Zeitschrift für Physik* 31 (1925), pp. 253–258 (cit. on pp. 7, 13).

[ISŻ03] Donald I.Cartwright, Patrick Solé, and Andrzej Żukc. “Ramanujan geometries of type $\tilde{A}_n$”. In: *Discrete Mathematics* 269 (1–3 2003), pp. 35–43 (cit. on p. 40).

[Jal12] Markus Jalsenius. “Sampling colourings of the triangular lattice”. In: *Random Structures & Algorithms* 40 (4 2012), pp. 501–533 (cit. on p. 198).

[Jan98] Svante Janson. “New versions of Suen’s correlation inequality”. In: *Random Structures & Algorithms* 13 (3-4 1998). Special Issue: Proceedings of the Eighth International Conference “Random Structures and Algorithms” held August 4-9, 1997, in Poznan, Poland, pp. 467–483 (cit. on p. 18).

[Jer+04] 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* (2004), pp. 1741–1765 (cit. on pp. 18, 59).

[Jer06] Mark Jerrum. “Two Remarks Concerning Balanced Matroids”. In: *Combinatorica* 26.6 (2006), pp. 733–742 (cit. on p. 59).

[Jer95] Mark Jerrum. “A Very Simple Algorithm for Estimating the Number of k-Colorings of a Low-Degree Graph”. In: *Random Structures & Algorithms* 7.2 (Sept. 1995), pp. 157–165. ISSN: 1042-9832 (cit. on pp. 15, 65, 66, 154).

[JKP20] Matthew Jenssen, Peter Keevash, and Will Perkins. “Algorithms for #BIS-hard problems on expander graphs”. In: *SIAM Journal on Computing* 49 (4 2020) (cit. on p. 198).

[JLV22] Arun Jambulapati, Yin Tat Lee, and Santosh S. Vempala. “A Slightly Improved Bound for the KLS Constant”. In: *arXiv preprint arXiv:2208.11644* (2022) (cit. on p. 195).

[Joh96] Anders Johansson. “The Choice Number of Sparse Graphs”. Unpublished manuscript. 1996 (cit. on p. 66).

[Jon02] Johan Jonasson. “Uniqueness of uniform random colorings of regular trees”. In: *Statistics & Probability Letters* 57 (3 2002), pp. 243–248 (cit. on p. 295).

[JPP22] Matthew Jenssen, Aditya Potukuchi, and Will Perkins. “Approximately counting independent sets in bipartite graphs via graph containers”. In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA ’22. USA: Society for Industrial and Applied Mathematics, 2022, pp. 499–516 (cit. on p. 198).

[JPV22] Vishesh Jain, Huy Tuan Pham, and Thuy-Duong Vuong. “Spectral independence, coupling, and the spectral gap of the Glauber dynamics”. In: *Information Processing Letters* 177 (2022) (cit. on pp. 132, 173).

[JS02] Mark Jerrum and Jung Bae Son. “Spectral gap and log-Sobolev constant for balanced matroids”. In: *FOCS*. 2002, pp. 721–729 (cit. on p. 59).

[JS89] M. Jerrum and Alistair Sinclair. “Approximating the Permanent”. In: *SIAM Journal on Computing* 18.6 (Dec. 1989), pp. 1149–1178 (cit. on pp. 113, 127, 136, 197, 198, 279).

[JS93] Mark Jerrum and Alistair Sinclair. “Polynomial-time approximation algorithms for the Ising model”. In: *SIAM Journal on Computing* 22.5 (1993), pp. 1087–1116 (cit. on pp. 100, 112, 113, 127, 131, 199, 279).

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. “A Polynomial-Time Approximation Algorithm for the Permanent of a Matrix with Nonnegative Entries”. In: *Journal of the ACM* 51.4 (July 2004), pp. 671–697 (cit. on pp. 16, 197).

[JVV86] Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. “Random generation of combinatorial structures from a uniform distribution”. In: *Theoretical Computer Science* 43 (1986), pp. 169–188 (cit. on pp. 25, 26, 189).

[JVW90] F. Jaeger, D. L. Vertigan, and D. J. A. Welsh. "On the computational complexity of the Jones and Tutte polynomials". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 108 (1 1990), pp. 35–53 (cit. on p. 100).

[Kai04] Volker Kaibel. "On the Expansion of Graphs of 0/1-Polytopes". In: *The Sharpest Cut*. MOS-SIAM Series on Optimization. 2004. Chap. 13, pp. 199–216. DOI: 10.1137/1.9780898718805.ch13 (cit. on pp. 12, 63).

[KD16] Tarun Kathuria and Amit Deshpande. "On Sampling and Greedy MAP Inference of Constrained Determinantal Point Processes". In: *arXiv preprint arXiv:1607.01551* (2016) (cit. on pp. 197, 305).

[Kel85] F. P. Kelly. "Stochastic Models of Computer Communication Systems". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 47.3 (1985), pp. 379–395 (cit. on p. 14).

[KF09] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009 (cit. on pp. 8, 172).

[KGV83] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi. "Optimization by Simulated Annealing". In: *Science* 220 (4598 1983), pp. 671–680 (cit. on p. 8).

[Kha95] Leonid Khachiyan. "On the complexity of approximating extremal determinants in matrices". In: *Journal of Complexity* 11.1 (1995), pp. 138–153 (cit. on p. 305).

[Kho02] Subhash Khot. "On the Power of Unique 2-Prover 1-Round Games". In: *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*. STOC '02. Montreal, Quebec, Canada: Association for Computing Machinery, 2002, pp. 767–775 (cit. on p. 198).

[Kir47] G. Kirchhoff. "Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird". In: *Annalen der Physik* 148 (12 1847), pp. 497–508 (cit. on p. 16).

[KKG20] Anna R. Karlin, Nathan Klein, and Shayan Oveis Gharan. "An Improved Approximation Algorithm for TSP in the Half Integral Case". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2020. Chicago, IL, USA: Association for Computing Machinery, 2020, pp. 28–39. ISBN: 9781450369794. DOI: 10.1145/3357713.3384273 (cit. on p. 18).

[KKO21] Anna R. Karlin, Nathan Klein, and Shayan Oveis Gharan. "A (Slightly) Improved Approximation Algorithm for Metric TSP". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2021. Virtual, Italy: Association for Computing Machinery, 2021, pp. 32–45. ISBN: 9781450380539. DOI: 10.1145/3406325.3451009 (cit. on pp. 9, 16, 18).

[KKS21] Tali Kaufman, Rasmus Kyng, and Federico Soldá. "Scalar and Matrix Chernoff Bounds from ℓ_∞-Independence". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2021, pp. 3732–3753 (cit. on pp. 16, 17, 37, 197).

[KL22] Bo'az Klartag and Joseph Lehec. "Bourgain's slicing problem and KLS isoperimetry up to polylog". In: *arXiv preprint arXiv:2203.15551* (2022) (cit. on p. 195).

[Kle21] Pieter Kleer. "Sampling from the Gibbs Distribution in Congestion Games". In: *Proceedings of the 22nd ACM Conference on Economics and Computation*. EC '21. Budapest, Hungary: Association for Computing Machinery, 2021, pp. 679–680. ISBN: 9781450385541. DOI: 10.1145/3465456.3467597 (cit. on pp. 11, 18).

[Kle66] Daniel J. Kleitman. "Families of Non-disjoint subsets". In: *Journal of Combinatorial Theory* 1 (1 1966), pp. 153–155 (cit. on p. 18).

[KLR22] Frederic Koehler, Holden Lee, and Andrej Risteski. "Sampling Approximately Low-Rank Ising Models: MCMC meets Variational Methods". In: *Proceedings of Thirty Fifth Conference on Learning Theory*. Ed. by Po-Ling Loh and Maxim Raginsky. Vol. 178. Proceedings of Machine Learning Research. PMLR, July 2022, pp. 4945–4988 (cit. on p. 198).

[KLS95] Ravi Kannan, László Lovász, and Miklós Simonovits. "Isoperimetric problems for convex bodies and a localization lemma". In: *Discrete & Computational Geometry* 13 (1995), pp. 541–559 (cit. on pp. 60, 196).

[KLT22] Tsz Chiu Kwok, Lap Chi Lau, and Kam Chuen Tung. "Cheeger Inequalities for Vertex Expansion and Reweighted Eigenvalues". In: *arXiv preprint arXiv:2203.06168* (2022) (cit. on pp. 12, 63).

[KM09] Jonathan A. Kelner and Aleksander Mądry. "Faster Generation of Random Spanning Trees". In: *2009 50th Annual IEEE Symposium on Foundations of Computer Science*. 2009, pp. 13–21. DOI: 10.1109/FOCS.2009.75 (cit. on pp. 16, 189).

[KM17] Tali Kaufman and David Mass. "High Dimensional Random Walks and Colorful Expansion". In: *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Ed. by Christos H. Papadimitriou. Vol. 67. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017, 4:1–4:27. ISBN: 978-3-95977-029-3. DOI: 10.4230/LIPIcs.ITCS.2017.4 (cit. on pp. 18, 33, 40–43, 150).

[KM20] Tali Kaufman and David Mass. "Local-To-Global Agreement Expansion via the Variance Method". In: *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*. Ed. by Thomas Vidick. Vol. 151. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020, 74:1–74:14. ISBN: 978-3-95977-134-4. DOI: 10.4230/LIPIcs.ITCS.2020.74 (cit. on pp. 40, 49).

[KN10] Jeff Kahn and Michael Neiman. "Negative correlation and log-concavity". In: *Random Structures & Algorithms* 37.3 (2010), pp. 367–388 (cit. on p. 104).

[KN11] Jeff Kahn and Michael Neiman. "A strong log-concavity property for measures on Boolean algebras". In: *Journal of Combinatorial Theory Series A* 118.6 (2011), pp. 1749–1760 (cit. on p. 104).

[KO18] Tali Kaufman and Izhar Oppenheim. "Construction of New Local Spectral High Dimensional Expanders". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 773–786. ISBN: 9781450355599. DOI: 10.1145/3188745.3188782 (cit. on pp. 40, 45, 58, 64, 150).

[KO20a] Tali Kaufman and Izhar Oppenheim. "High dimensional expanders and coset geometries". In: *arXiv preprint arXiv:1710.05304* (2020) (cit. on pp. 40, 58, 64).

[KO20b] Tali Kaufman and Izhar Oppenheim. "High Order Random Walks: Beyond Spectral Gap". In: *Combinatorica* 40 (2020), pp. 245–281 (cit. on pp. 18, 33, 40–43).

[Kol18] Vladimir Kolmogorov. "A Faster Approximation Algorithm for the Gibbs Partition Function". In: *Proceedings of the 31st Conference On Learning Theory*. Ed. by Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet. Vol. 75. Proceedings of Machine Learning Research. PMLR, July 2018, pp. 228–249 (cit. on p. 26).

[KS18] Rasmus Kyng and Zhao Song. "A Matrix Chernoff Bound for Strongly Rayleigh Distributions and Spectral Sparsifiers from a few Random Spanning Trees". In: *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. 2018, pp. 373–384. DOI: 10.1109/FOCS.2018.00043 (cit. on p. 16).

[KS68] D. G. Kelly and S. Sherman. "General Griffiths' Inequalities on Correlations in Ising Ferromagnets". In: *Journal of Mathematical Physics* 9 (3 1968), p. 466 (cit. on p. 18).

[KSV79] A. Khachaturyan, S. Semenovskaya, and Armen Vainshtein B. "Statistical-Thermodynamic Approach to Determination of Structure Amplitude Phases". In: *Soviet Physics, Crystallography* 24 (5 1979), pp. 519–524 (cit. on p. 8).

[KSV81] A. Khachaturyan, S. Semenovskaya, and Armen Vainshtein B. "The Thermodynamic Approach to the Structure Analysis of Crystals". In: *Acta Crystallographica* A37 (5 1981), pp. 742–754 (cit. on p. 8).

[KT12] Alex Kulesza and Ben Taskar. "Determinantal point processes for machine learning". In: *Foundations and Trends in Machine Learning* 5(2-3) (2012), pp. 123–286 (cit. on pp. 8, 97, 184).

[KT21] Tali Kaufman and Ran J. Tessler. "New Cosystolic Expanders from Tensors Imply Explicit Quantum LDPC Codes with $\Omega(\sqrt{n}\log^k n)$ Distance". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 1317–1329. ISBN: 9781450380539. DOI: 10.1145/3406325.3451029 (cit. on p. 40).

[KT22] Tali Kaufman and Ran J. Tessler. "Stronger Trickling Down in Posets Implies Expanding Sparsification of the Grassmanian Complex". In: *arXiv preprint arXiv:2101.12621* (2022) (cit. on p. 58).

[KW17] Jacob Korevaar and Jan Wiegerinck. *Several complex variables*. Korteweg-de Vries Institute for Mathematics, 2017 (cit. on p. 116).

[LA87] Peter J. M. van Laarhoven and Emile H. L. Aarts. *Simulated annealing: theory and applications*. Vol. 37. Mathematics and Its Applications. D. Reidel Publishing Company, 1987 (cit. on p. 8).

[LB12] Hui Lin and Jeff Bilmes. "Learning Mixtures of Submodular Shells with Application to Document Summarization". In: *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*. UAI'12. Catalina Island, CA: AUAI Press, 2012, pp. 479–490. ISBN: 9780974903989 (cit. on p. 8).

[Led99] Michel Ledoux. "Concentration of measure and logarithmic Sobolev inequalities". In: *Séminaire de Probabilités XXXIII (Strasbourg)* 33 (1999), pp. 120–216 (cit. on pp. 160, 174, 175).

[Len13] Matthias Lenz. "The f-vector of a representable-matroid complex is log-concave". In: *Advances in Applied Mathematics* 51.5 (2013), pp. 543–545 (cit. on p. 104).

[Len20] W. Lenz. "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern". In: *Physikalische Zeitschrift* 21 (1920), pp. 613–615 (cit. on pp. 7, 13).

[Li04] Wen-Ching Winnie Li. "Ramanujan hypergraphs". In: *Geometric & Functional Analysis GAFA* 14 (2004), pp. 380–399 (cit. on p. 40).

[Lia+19] Chao Liao, Jiabao Lin, Pinyan Lu, and Zhenyu Mao. "Counting Independent Sets and Colorings on Random Regular Bipartite Graphs". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Ed. by Dimitris Achlioptas and László A. Végh. Vol. 145. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019, 34:1–34:12. ISBN: 978-3-95977-125-2. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2019.34 (cit. on p. 198).

[Liu19] Jingcheng Liu. "Approximate counting, phase transitions and geometry of polynomials". PhD thesis. EECS Department, University of California, Berkeley, Aug. 2019 (cit. on p. 137).

[Liu21] Kuikui Liu. "From Coupling to Spectral Independence and Blackbox Comparison with the Down-Up Walk". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Vol. 207. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 32:1–32:21 (cit. on pp. 15, 19, 148).

[LKR12] Dean Lusher, Johan Koskinen, and Garry Robins. *Exponential Random Graph Models for Social Networks Theory, Methods, and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press, 2012 (cit. on p. 8).

[LL15a] Jingcheng Liu and Pinyan Lu. "FPTAS for #BIS with Degree Bounds on One Side". In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '15. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 549–556. ISBN: 9781450335362. DOI: 10.1145/2746539.2746598 (cit. on pp. 18, 131).

[LL15b] Jingcheng Liu and Pinyan Lu. "FPTAS for Counting Monotone CNF". In: *Proceedings of the 2015 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2015, pp. 1531–1548. DOI: 10.1137/1.9781611973730.101. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611973730.101 (cit. on pp. 18, 131).

[LLL14] 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 (cit. on pp. 18, 112, 131).

[LLP20] Eyal Lubetzky, Alexander Lubotzky, and Ori Parzanchevski. "Random walks on Ramanujan complexes and digraphs". In: *Journal of the European Mathematical Society* 22.11 (2020), pp. 3441–3466 (cit. on p. 40).

[LLY12] 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 (cit. on pp. 18, 131, 134, 137, 272).

[LLY13] Liang Li, Pinyan Lu, and Yitong Yin. "Correlation Decay Up to Uniqueness in Spin Systems". In: *Proceedings of the Twenty-fourth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '13. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2013, pp. 67–84. ISBN: 978-1-611972-51-1 (cit. on pp. 18, 131, 133–135, 137, 144, 240–242, 246, 248, 249, 272, 273, 275, 276, 278, 289).

[LLZ14a] 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 (cit. on pp. 18, 112, 131).

[LLZ14b] Jingcheng Liu, Pinyan Lu, and Chihao Zhang. "The Complexity of Ferromagnetic Two-spin Systems with External Fields". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Ed. by Klaus Jansen, José D. P. Rolim, Nikhil R. Devanur, and Cristopher Moore. Vol. 28. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014, pp. 843–856. ISBN: 978-3-939897-74-3. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2014.843 (cit. on p. 198).

[LM17] Rafał Latała and Dariusz Matlak. "Royen's Proof of the Gaussian Correlation Inequality". In: *Geometric Aspects of Functional Analysis. Lecture Notes in Mathematics, vol 2169* (2017), pp. 265–275 (cit. on p. 18).

[LMP09] Brendan Lucier, Michael Molloy, and Yuval Peres. "The Glauber Dynamics for Colourings of Bounded Degree Trees". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Ed. by Irit Dinur, Klaus Jansen, Joseph Naor, and José Rolim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 631–645. ISBN: 978-3-642-03685-9 (cit. on p. 198).

[LMY20] Siqi Liu, Sidhanth Mohanty, and Elizabeth Yang. "High-Dimensional Expanders from Expanders". In: *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*. Ed. by Thomas Vidick. Vol. 151. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020, 12:1–12:32. ISBN: 978-3-95977-134-4. DOI: 10.4230/LIPIcs.ITCS.2020.12 (cit. on p. 40).

[Löw34] Karl Löwner. "Über monotone Matrixfunktionen". In: *Mathematische Zeitschrift* 38.1 (1934), pp. 177–216 (cit. on p. 68).

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. "Ramanujan graphs". In: *Combinatorica* 8 (1988), pp. 261–277 (cit. on p. 40).

[LPW17] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. 2nd ed. American Mathematical Society, 2017 (cit. on pp. 9, 24, 26–28).

[LR19] Jonathan D. Leake and Nick R. Ryder. "Generalizations of the Matching Polynomial to the Multivariate Independence Polynomial". In: *Algebraic Combinatorics* 2.5 (2019), pp. 781–802 (cit. on p. 200).

[LS88] Gregory F. Lawler and Alan D. Sokal. "Bounds on the L^2 Spectrum for Markov Chains and Markov Processes: A Generalization of Cheeger's Inequality". In: *Transactions of the American Mathematical Society* 309.2 (1988), pp. 557–580 (cit. on p. 64).

[LSS19] 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 (cit. on pp. 66, 113, 137, 139).

[LSS20] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. "Correlation Decay and Partition Function Zeros: Algorithms and Phase Transitions". In: *SIAM Journal on Computing* 0.0 (2020). Special Section FOCS 2019, FOCS19-200-FOCS19–252. DOI: 10.1137/20M1317384. eprint: https://doi.org/10.1137/20M1317384 (cit. on pp. 18, 131, 201).

[LSV05a] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. "Explicit constructions of Ramanujan complexes of type $\tilde{A}_d$". In: *European Journal of Combinatorics* 26 (6 2005), pp. 965–993 (cit. on p. 40).

[LSV05b] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. "Ramanujan complexes of type $\tilde{A}_d$". In: *Israel Journal of Mathematics* 149 (2005), pp. 267–299 (cit. on p. 40).

[Lub19] Alexander Lubotzky. "High Dimensional Expanders". In: Proceedings of the International Congress of Mathematicians (ICM 2018). 2019, pp. 705–730 (cit. on p. 18).

[LV07] László Lovász and Santosh Vempala. "The Geometry of Logconcave Functions and Sampling Algorithms". In: *Random Structures and Algorithms* 30 (3 2007) (cit. on p. 10).

[LV17] Yin Tat Lee and Santosh Srinivas Vempala. "Eldan's Stochastic Localization and the KLS Hyperplane Conjecture: An Improved Lower Bound for Expansion". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. 2017, pp. 998–1007. DOI: 10.1109/FOCS.2017.96 (cit. on p. 195).

[LV18] Yin Tat Lee and Santosh S. Vempala. "Stochastic Localization + Stieltjes Barrier = Tight Bound for Log-Sobolev". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 1122–1129. ISBN: 9781450355599. DOI: 10.1145/3188745.3188866 (cit. on p. 196).

[LV97] Michael Luby and Eric Vigoda. "Approximately Counting Up to Four (Extended Abstract)". In: *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*. STOC '97. El Paso, Texas, USA: ACM, 1997, pp. 682–687. ISBN: 0-89791-888-6 (cit. on pp. 14, 136, 137).

[LV99] Michael Luby and Eric Vigoda. "Fast convergence of the Glauber dynamics for sampling independent sets". In: *Random Structures & Algorithms* 15 (1999), pp. 229–241 (cit. on pp. 14, 136).

[LW06] Michael Luby and Avi Wigderson. "Pairwise Independence and Derandomization". In: *Foundations and Trends in Theoretical Computer Science* 1.4 (2006), pp. 237–301 (cit. on p. 17).

[LWZ14] Pinyan Lu, Menghui Wang, and Chihao Zhang. "FPTAS for Weighted Fibonacci Gates and Its Applications". In: *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming*. ICALP '14. Part of the Lecture Notes in Computer Science, vol 8572. Berlin, Heidelberg: Springer-Verlag, 2014, pp. 787–799 (cit. on pp. 18, 131).

[LY13] Pinyan Lu and Yitong Yin. "Improved FPTAS for Multi-spin Systems". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Ed. by Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and José D. P. Rolim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 639–654 (cit. on pp. 18, 66, 131, 134).

[LY52] 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 (cit. on pp. 18, 111, 113, 124).

[LY98] Tzong-Yow Lee and Horng-Tzer Yau. "Logarithmic Sobolev inequality for some models of random walks". In: *The Annals of Probability* 26.4 (1998), pp. 1855–1873 (cit. on pp. 159, 160).

[LYZ16] Pinyan Lu, Kuan Yang, and Chihao Zhang. "FPTAS for Hardcore and Ising Models on Hypergraphs". In: *33rd Symposium on Theoretical Aspects of Computer Science (STACS 2016)*. Ed. by Nicolas Ollinger and Heribert Vollmer. Vol. 47. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, 51:1–51:14. ISBN: 978-3-95977-001-9. DOI: 10.4230/LIPIcs.STACS.2016.51 (cit. on pp. 18, 131).

[Mah85] Carolyn Mahoney. "On the unimodality of the independent set numbers of a class of matroids". In: *Journal of Combinatorial Theory Series B* 39.1 (1985), pp. 77–85 (cit. on p. 104).

[Mar19] Katalin Marton. "Logarithmic Sobolev inequalities in discrete product spaces". In: *Combinatorics, Probability and Computing* 28.6 (2019), pp. 919–935. DOI: 10.1017/S0963548319000099 (cit. on pp. 152, 154, 175).

[Mar99] Fabio Martinelli. "Lectures on Glauber Dynamics for Discrete Spin Models". In: *Lectures on probability theory and statistics, volume 1717 of Lecture Notes in Math.* Springer, Berlin, 1999, pp. 93–191 (cit. on p. 175).

[Mas72] John H. Mason. "Matroids: unimodal conjectures and Motzkin's theorem". In: *Combinatorics (Proc. Conf. Combinatorial Math., Math. Inst.)* Oxford: Inst. Math. Appl., Southend-on-Sea, 1972, pp. 207–220 (cit. on pp. 12, 102, 103).

[McQ13] Colin McQuillan. "Approximating Holant problems by winding". In: *arXiv preprint arXiv:1301.2880* (2013) (cit. on p. 127).

[Met+53] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21 (6 1953) (cit. on pp. 9, 13).

[Mic97] 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 (cit. on p. 169).

[MMS22] Jacob P. Matherne, Alejandro H. Morales, and Jesse Selover. "The Newton polytope and Lorentzian property of chromatic symmetric functions". In: *arXiv preprint arXiv:2201.07333* (2022) (cit. on p. 201).

[Mol04] Michael Molloy. "The Glauber dynamics on colorings of a graph with high girth and maximum degree". In: *SIAM Journal on Computing* 33.3 (2004), pp. 721–737 (cit. on pp. 66, 198).

[Mou22] Giorgos Mousa. "Local-to-Global Functional Inequalities in Simplicial Complexes". PhD thesis. School of Informatics, University of Edinburgh, 2022 (cit. on pp. 101, 199, 200, 303, 304).

[MR02] Neal Madras and Dana Randall. "Markov chain decomposition for convergence rate analysis". In: *The Annals of Applied Probability* 12.2 (2002) (cit. on p. 18).

[MR06] Russell Martin and Dana Randall. "Disjoint Decomposition of Markov Chains and Sampling Circuits in Cayley Graphs". In: *Comb. Probab. Comput.* 15.3 (May 2006), pp. 411–448. ISSN: 0963-5483. DOI: 10.1017/S0963548305007352 (cit. on p. 18).

[MS07] Andrea Montanari and Devavrat Shah. "Counting Good Truth Assignments of Random K-SAT Formulae". In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '07. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2007, pp. 1255–1264. ISBN: 9780898716245 (cit. on pp. 18, 131).

[MS08a] Igor L. Markov and Yaoyun Shi. "Simulating Quantum Computation by Contracting Tensor Networks". In: *SIAM Journal on Computing* 38 (3 2008), pp. 963–981 (cit. on p. 280).

[MS08b] Elchanan Mossel and Allan Sly. "Rapid Mixing of Gibbs Sampling on Graphs That Are Sparse on Average". In: *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '08. San Francisco, California: Society for Industrial and Applied Mathematics, 2008, pp. 238–247 (cit. on p. 136).

[MS10] Elchanan Mossel and Allan Sly. "Gibbs rapidly samples colorings of $G(n, d/n)$". In: *Probability Theory and Related Fields* 148 (2010), pp. 37–69 (cit. on p. 198).

[MS13] Elchanan Mossel and Allan Sly. "Exact thresholds for Ising–Gibbs samplers on general graphs". In: *Annals of Probability* 41.1 (2013), pp. 294–328 (cit. on pp. 136, 247, 267, 269).

[MS18] Kazuo Murota and Akiyoshi Shioura. "Simpler exchange axioms for M-concave functions on generalized polymatroids". In: *Japan Journal of Industrial and Applied Mathematics* 35.1 (2018), pp. 235–259 (cit. on p. 184).

[MS91] Milena Mihail and Madhu Sudan. *Connectivity Properties of Matroids*. Tech. rep. EECS Department, University of California, Berkeley, Dec. 1991 (cit. on p. 59).

[MS99] Kazuo Murota and Akiyoshi Shioura. "M-convex function on generalized polymatroid". In: *Mathematics of operations research* 24.1 (1999), pp. 95–105 (cit. on p. 181).

[MSJ18] Zelda Mariet, Suvrit Sra, and Stefanie Jegelka. "Exponentiated Strongly Rayleigh Distributions". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. NIPS'18. Montréal, Canada: Curran Associates, Inc., 2018, pp. 4464–4474 (cit. on pp. 8, 98).

[MSS14] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. "Ramanujan Graphs and the Solution of the Kadison-Singer Problem". In: Invited Lectures 3 (2014), pp. 363–386 (cit. on p. 18).

[MSS15a] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. "Interlacing families I: Bipartite Ramanujan graphs of all degrees". In: *Annals of Mathematics* 182 (1 2015), pp. 307–325 (cit. on pp. 18, 107).

[MSS15b] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. "Interlacing families II: Mixed characteristic polynomials and the Kadison–Singer problem". In: *Annals of Mathematics* 182 (1 2015), pp. 327–350 (cit. on pp. 18, 107).

[MSS15c] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. "Interlacing Families IV: Bipartite Ramanujan Graphs of All Sizes". In: FOCS '15. USA: IEEE Computer Society, 2015, pp. 1358–1377. ISBN: 9781467381918. DOI: 10.1109/FOCS.2015.87 (cit. on pp. 18, 107).

[MSS22] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. "Interlacing families III: Sharper restricted invertibility estimates". In: *Israel Journal of Mathematics* 247 (2022), pp. 519–546 (cit. on pp. 18, 107).

[MST15] Aleksander Mądry, Damian Straszak, and Jakub Tarnawski. "Fast Generation of Random Spanning Trees and the Effective Resistance Metric". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '15. San Diego, California: Society for Industrial and Applied Mathematics, 2015, pp. 2019–2036 (cit. on pp. 16, 189).

[MSW03] F. Martinelli, A. Sinclair, and D. Weitz. "The Ising model on trees: boundary conditions and mixing time". In: *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* Oct. 2003, pp. 628–639 (cit. on pp. 136, 167).

[MSW07] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. "Fast mixing for independent sets, colorings, and other models on trees". In: *Random Structures & Algorithms* 31 (2 2007), pp. 134–172 (cit. on pp. 66, 136, 198).

[MT06] 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. 237–354 (cit. on pp. 9, 160, 181).

[MT07] Frank McSherry and Kunal Talwar. "Mechanism Design via Differential Privacy". In: *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. 2007, pp. 94–103 (cit. on p. 8).

[MT93] Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Communications and Control Engineering. Springer-Verlag, 1993 (cit. on p. 9).

[Mur03] Kazuo Murota. *Discrete Convex Analysis*. Monographs on Discrete Mathematics and Applications. Society for Industrial and Applied Mathematics, 2003, p. 389 (cit. on pp. 11, 109).

[MV89] M. Mihail and U. Vazirani. "On the expansion of 0/1 polytopes". In: *Journal of Combinatorial Theory*. B (1989) (cit. on pp. 11, 12, 58, 63).

[MW89] Bojan Mohar and Wolfgang Woess. "A Survey on Spectra of infinite Graphs". In: *Bulletin of the London Mathematical Society* 21 (3 May 1989), pp. 209–234 (cit. on p. 289).

[MWW07] Elchanan Mossel, Dror Weitz, and Nicholas C. Wormald. "On the hardness of sampling independent sets beyond the tree threshold". In: *Probability Theory and Related Fields* 143 (2007), pp. 401–439 (cit. on p. 137).

[MY09] Neal Madras and Wai Kong Yuen. "Spectral gaps of random walk metropolis chains". In: *Far East Journal of Theoretical Statistics* 27.2 (2009) (cit. on p. 18).

[Nik15] Aleksandar Nikolov. "Randomized Rounding for the Largest Simplex Problem". In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '15. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 861–870. ISBN: 9781450335362. DOI: 10.1145/2746539.2746628 (cit. on p. 305).

[NUI69] WIM NUIJ. "A Note on Hyperbolic Polynomials". In: *Mathematica Scandinavica* 23.1 (1969), pp. 69–72. ISSN: 00255521, 19031807 (cit. on p. 187).

[NV20] James E. Newman and Moshe Y. Vardi. "FPRAS Approximation of the Matrix Permanent in Practice". In: *arXiv preprint arXiv:2012.03367* (2020) (cit. on p. 16).

[ODo14] Ryan O'Donnell. *Analysis of Boolean Functions*. USA: Cambridge University Press, 2014. ISBN: 1107038324 (cit. on pp. 40, 282).

[Oll09] Yann Ollivier. "Ricci curvature of Markov chains on metric spaces". In: *Journal of Functional Analysis* 256 (3 Feb. 2009), pp. 810–864 (cit. on pp. 148, 152).

[OP22] Ryan O'Donnell and Kevin Pratt. "High-Dimensional Expanders from Chevalley Groups". In: *37th Computational Complexity Conference (CCC 2022)*. Ed. by Shachar Lovett. Vol. 234. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 18:1–18:26. ISBN: 978-3-95977-241-9. DOI: 10.4230/LIPIcs.CCC.2022.18 (cit. on pp. 40, 58, 64).

[Opp18] 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 (cit. on pp. 19, 40, 46, 53, 58–60, 150).

[OR19] Shayan Oveis Gharan and Alireza Rezaei. "A Polynomial Time MCMC Method for Sampling from Continuous Determinantal Point Processes". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, June 2019, pp. 5438–5447 (cit. on p. 181).

[Orú14] 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 (cit. on p. 280).

[OSS11] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. "A Randomized Rounding Approach to the Traveling Salesman Problem". In: *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*. 2011, pp. 550–559. DOI: 10.1109/FOCS.2011.80 (cit. on pp. 9, 16, 18).

[Ove14] Shayan Oveis Gharan. "New Rounding Techniques for the Design and Analysis of Approximation Algorithms". PhD thesis. Department of Management Science and Engineering, Stanford University, 2014 (cit. on pp. 9, 18).

[Oxl11] James Oxley. *Matroid Theory*. 2nd ed. Oxford Graduate Texts in Mathematics. Oxford University Press, 2011 (cit. on pp. 11, 29, 60, 181).

[Pac04] Asa Packer. "Polynomial-time approximation of largest simplices in V-polytopes". In: *Discret. Appl. Math.* 134 (2004), pp. 213–237 (cit. on p. 305).

[Pem00] Robin Pemantle. "Towards a theory of negative dependence". In: *Journal of Mathematical Physics* 41 (3 2000), pp. 1371–1390 (cit. on pp. 38, 104).

[Pin64] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. 1st ed. Translated and edited by Amiel Feinstein. Holden-Day, Inc., 1964, p. 243 (cit. on p. 25).

[Pin70] Martin Pincus. "A Monte-Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems". In: *Journal of the Operations Research Society of America* 18 (6 1970), pp. 967–1235 (cit. on p. 8).

[PK22] Pavel Panteleev and Gleb Kalachev. "Asymptotically Good Quantum and Locally Testable Classical LDPC Codes". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 375–388. ISBN: 9781450392648. DOI: 10.1145/3519935.3520017 (cit. on pp. 18, 40).

[PP14] Robin Pemantle and Yuval Peres. "Concentration of Lipschitz Functionals of Determinantal and Other Strong Rayleigh Measures". In: *Combinatorics, Probability and Computing* 23 (1 2014), pp. 140–160 (cit. on pp. 38, 107).

[PR17] 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 (cit. on pp. 18, 111, 112, 115, 136, 279).

[PR19] Han Peters and Guus Regts. "On a conjecture of Sokal concerning roots of the independence polynomial". In: *The Michigan Mathematical Journal* 68.1 (2019), pp. 33–55 (cit. on p. 136).

[PS01] Nikolay Prokof'ev and Boris Svistunov. "Worm Algorithms for Classical Statistical Models". In: *Phys. Rev. Lett.* 87 (16 Sept. 2001), p. 160601. DOI: 10.1103/PhysRevLett.87.160601 (cit. on p. 199).

[PS17] Natesh S. Pillai and Aaron Smith. "Elementary bounds on mixing times for decomposable Markov chains". In: *Stochastic Processes and their Applications* 127 (9 2017), pp. 3068–3109 (cit. on p. 18).

[QW22] Qian Qin and Guanyang Wang. "Spectral Telescope: Convergence Rate Bounds for Random-Scan Gibbs Samplers Based on a Hierarchical Structure". In: *arXiv preprint arXiv:2208.11299* (2022) (cit. on pp. 17, 54, 196).

[Rag16] Maxim Raginsky. "Strong Data Processing Inequalities and Φ -Sobolev Inequalities for Discrete Channels". In: *IEEE Transactions on Information Theory* 62.6 (2016), pp. 3355–3389. DOI: 10.1109/TIT.2016.2549542 (cit. on p. 48).

[RDF78] P. J. Rossky, J. D. Doll, and H. L. Friedman. "Brownian dynamics as smart Monte Carlo simulation". In: *The Journal of Chemical Physics* 69 (10 1978) (cit. on p. 9).

[Rea68] Ronald C. Read. "An introduction to chromatic polynomials". In: *Journal of Combinatorial Theory* 4 (1 1968), pp. 52–71 (cit. on p. 199).

[Reg18] Guus Regts. "Zero-Free Regions of Partition Functions with Applications to Algorithms and Graph Limits". In: *Combinatorica* 38 (2018), pp. 987–1015 (cit. on pp. 18, 115, 279, 281, 285–287).

[Res+13] Ricardo Restrepo, Jinwoo Shin, Prasad Tetali, Eric Vigoda, and Linji Yang. "Improved mixing condition on the grid for counting and sampling independent sets". In: *Probability Theory and Related Fields* 156.1 (June 2013), pp. 75–99 (cit. on pp. 14, 18, 131–134, 136, 272).

[Rot12] Thomas Rothvoß. "The Entropy Rounding Method in Approximation Algorithms". In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '12. Kyoto, Japan: Society for Industrial and Applied Mathematics, 2012, pp. 356–372 (cit. on p. 9).

[Rot64] Gian-Carlo Rota. "On the foundations of combinatorial theory I. Theory of Möbius Functions". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 2 (1964), pp. 340–368 (cit. on p. 200).

[Rot71] Gian-Carlo Rota. "Combinatorial theory, old and new". In: (1971), pp. 229–233 (cit. on p. 199).

[Roy14] Thomas Royen. "A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions". In: *arXiv preprint arXiv:1408.1028* (2014) (cit. on p. 18).

[RR19] 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) (cit. on pp. 148, 150, 151).

[RR98] Gareth O. Roberts and Jeffrey S. Rosenthal. "Optimal scaling of discrete approximations to Langevin diffusions". In: *Journal of the Royal Statistical Society Series B* 60 (1 1998), pp. 255–268 (cit. on p. 9).

[RS85] Gian-Carlo Rota and David Sharp. "Mathematics, Philosophy, and Artificial Intelligence: a Dialogue with Gian-Carlo Rota and David Sharp". In: (1985) (cit. on p. 111).

[RS93] Yosef Rinott and Michael Saks. "Correlation inequalities and a conjecture for permanents". In: *Combinatorica* 13 (1993), pp. 269–277 (cit. on p. 18).

[RSJ19] Joshua Robinson, Suvrit Sra, and Stefanie Jegelka. "Flexible Modeling of Diversity with Strongly Log-Concave Distributions". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc., 2019, pp. 15225–15235 (cit. on p. 8).

[RST05] Jürgen Richter-Gebert, Bernd Sturmfels, and Thorsten Theobald. "First Steps in Tropical Geometry". In: *Contemporary Mathematics*. Idempotent Mathematics and Mathematical Physics 377 (2005) (cit. on p. 106).

[RT87] Prabhakar Raghavan and Clark D. Tompson. "Randomized rounding: A technique for provably good algorithms and algorithmic proofs". In: *Combinatorica* 7 (1987), pp. 365–374 (cit. on p. 9).

[RT96] Gareth O. Roberts and Richard L. Tweedie. "Exponential convergence of Langevin distributions and their discrete approximations". In: *Bernoulli* 2.4 (1996), pp. 341–363 (cit. on p. 9).

[RTF18] Luis Russo, Andreia Sofia Teixeira, and Alexandre P Francisco. "Linking and Cutting Spanning Trees". In: *Algorithms* 11.4 (2018), p. 53 (cit. on pp. 190, 191, 193, 194).

[Rue71] David Ruelle. "Extension of the Lee-Yang circle theorem". In: *Physical Review Letters* 26.6 (1971), pp. 303–304 (cit. on pp. 114, 127).

[Sal97] 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 (cit. on pp. 166, 167).

[Sam05] Marcus D. Sammer. "Aspects of Mass Transportation in Discrete Concentration Inequalities". PhD thesis. Georgia Institute of Technology, Apr. 2005 (cit. on p. 160).

[Sar04] Alireza Sarveniazi. "Ramanujan $(n_1, n_2, \ldots, n_{d-1})$-regular hypergraphs based on Bruhat-Tits Buildings of type $\tilde{A}_{d-1}$". PhD thesis. University of Göttingen, 2004 (cit. on p. 40).

[Sca97] Fabio Scarabotti. "Time to Reach Stationarity in the Bernoulli–Laplace Diffusion Model with Many Urns". In: *Advances in Applied Mathematics* 18 (3 1997), pp. 351–371 (cit. on pp. 159, 160).

[Sch18] Aaron Schild. "An Almost-Linear Time Algorithm for Uniform Random Spanning Tree Generation". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 214–227. ISBN: 9781450355599. DOI: 10.1145/3188745.3188852 (cit. on pp. 16, 189, 190).

[Sey75] Paul Seymour. "Matroids, hypergraphs, and the max-flow min-cut theorem". PhD thesis. University of Oxford, 1975 (cit. on p. 104).

[She60] Geoffrey Colin Shephard. "Inequalities between mixed volumes of convex sets". In: *Mathematika* 7 (2 1960), pp. 125–138 (cit. on p. 106).

[She82] L. A. Shepp. "The XYZ Conjecture and the FKG Inequality". In: *The Annals of Probability* 10.3 (1982), pp. 824–827 (cit. on p. 18).

[Sil00] John R. Silvester. "Determinants of Block Matrices". In: *The Mathematical Gazette* 84.501 (Nov. 2000), pp. 460–467 (cit. on p. 52).

[Sin+15] Alistair Sinclair, Piyush Srivastava, Daniel Štefankovič, and Yitong Yin. "Spatial Mixing and the Connective Constant: Optimal Bounds". In: *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '15. San Diego, California: Society for Industrial and Applied Mathematics, 2015, pp. 1549–1563 (cit. on pp. 18, 131, 134, 136, 197, 272).

[SJ89] Alistair Sinclair and Mark Jerrum. "Approximate counting, uniform generation and rapidly mixing Markov chains". In: *Information and Computation* 82.1 (1989), pp. 93–133. ISSN: 0890-5401 (cit. on p. 64).

[Sly10] Allan Sly. "Computational Transition at the Uniqueness Threshold". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. 2010, pp. 287–296. DOI: 10.1109/FOCS.2010.34 (cit. on pp. 14, 131, 137).

[Sok01] A. D. Sokal. "Bounds on the complex zeros of (di) chromatic polynomials and Potts-model partition functions. Combinatorics". In: *Probability and Computing* 10.1 (2001), pp. 41–77 (cit. on p. 66).

[Son04] Jung-Bae Son. "Rapid mixing through decomposition and induction". PhD thesis. University of Edinburgh, 2004 (cit. on p. 18).

[Sri01] A. Srinivasan. "Distributions on level-sets with applications to approximation algorithms". In: *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. 2001, pp. 588–597. DOI: 10.1109/SFCS.2001.959935 (cit. on p. 38).

[Sri14] Piyush Srivastava. "Counting and Correlation Decay in Spin Systems". PhD thesis. EECS Department, University of California, Berkeley, Aug. 2014 (cit. on pp. 18, 131, 134).

[Sri99] Aravind Srinivasan. "Approximation Algorithms Via Randomized Rounding: A Survey". In: *Series in Advanced Topics in Mathematics, Polish Scientific Publishers PWN*. Publishers, 1999, pp. 9–71 (cit. on p. 17).

[SS05] 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 (cit. on p. 138).

[SS14] 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 (cit. on pp. 14, 131, 137).

[SS19] Shuai Shao and Yuxin Sun. "Contraction: a Unified Perspective of Correlation Decay and Zero-Freeness of 2-Spin Systems". In: *arXiv preprint arXiv:1909.04244* (2019) (cit. on pp. 137, 242, 247).

[SS97] Jesús Salas and Alan D. Sokal. "Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem". In: *Journal of Statistical Physics* 86.3 (1997), pp. 551–579 (cit. on p. 65).

[SSS95] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. "Chernoff–Hoeffding Bounds for Applications with Limited Independence". In: *SIAM Journal on Discrete Mathematics* 8 (2 1995) (cit. on p. 17).

[SST14] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. "Approximation Algorithms for Two-State Anti-Ferromagnetic Spin Systems on Bounded Degree Graphs". In: *Journal of Statistical Physics* 155.4 (2014), pp. 666–686 (cit. on pp. 18, 131, 134, 137, 272).

[SSY13] Alistair Sinclair, Piyush Srivastava, and Yitong Yin. "Spatial Mixing and Approximation Algorithms for Graphs with Bounded Connective Constant". In: *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. FOCS '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 300–309. ISBN: 978-0-7695-5135-7 (cit. on pp. 18, 131, 134, 136, 197).

[ST10] Fabio Scarabotti and Filippo Tolli. In: *Forum Mathematicum* 22 (5 2010), pp. 879–911 (cit. on pp. 159, 160).

[ST83] Daniel D Sleator and Robert Endre Tarjan. "A data structure for dynamic trees". In: *Journal of computer and system sciences* 26.3 (1983), pp. 362–391 (cit. on pp. 190, 193).

[Sta07] Richard P. Stanley. "An Introduction to Hyperplane Arrangements". In: IAS/Park City Mathematics Series 13 (2007), pp. 389–496 (cit. on p. 200).

[Sta89] Richard P. Stanley. "Log-Concave and Unimodal Sequences in Algebra, Combinatorics, and Geometry". In: *Annals of the New York Academy of Sciences*. Graph Theory and Its Applications East and West: Proceedings of the First China-USA International Graph Theory Conference 576 (1 1989), pp. 500–535 (cit. on p. 18).

[Sta95] Richard P. Stanley. "A Symmetric Function Generalization of the Chromatic Polynomial of a Graph". In: *Advances in Mathematics* 111 (1 1995), pp. 166–194 (cit. on p. 201).

[Ste72] Charles Stein. "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables". In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*. 1972, pp. 583–602 (cit. on p. 151).

[Suz73] Masuo Suzuki. "Correlation inequalities and phase transition in the generalized X-Y model". In: *Journal of Mathematical Physics* 14 (7 1973), p. 837 (cit. on p. 18).

[SV14] Mohit Singh and Nisheeth K. Vishnoi. "Entropy, Optimization and Counting". In: STOC '14. New York, New York: Association for Computing Machinery, 2014, pp. 50–59. ISBN: 9781450327107 (cit. on p. 9).

[SV17] Damian Straszak and Nisheeth K. Vishnoi. "Real Stable Polynomials and Matroids: Optimization and Counting". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2017. Montreal, Canada: Association for Computing Machinery, 2017, pp. 370–383. ISBN: 9781450345286. DOI: 10.1145/3055399.3055457 (cit. on pp. 18, 197).

[ŠVV09] 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 (cit. on p. 26).

[SW75] Paul Seymour and Dominic Welsh. "Combinatorial applications of an inequality from statistical mechanics". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 77.3 (1975), pp. 485–495 (cit. on p. 104).

[SYZ19] Renjie Song, Yitong Yin, and Jinman Zhao. "Counting hypergraph matchings up to uniqueness threshold". In: *Information and Computation* 266 (2019), pp. 75–96 (cit. on pp. 18, 131).

[Tsy09] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. 1st ed. Springer Series in Statistics (SSS). Springer, 2009, p. 214 (cit. on p. 25).

[Tut54] William Thomas Tutte. "A Contribution to the Theory of Chromatic Polynomials". In: *Canadian Journal of Mathematics* 6 (1954), pp. 80–91 (cit. on p. 200).

[Vad02] Salil P. Vadhan. "The Complexity of Counting in Sparse, Regular, and Planar Graphs". In: *SIAM J. Comput.* 31.2 (Feb. 2002), pp. 398–427. ISSN: 0097-5397 (cit. on p. 137).

[Vad95] Salil P. Vadhan. *The Complexity of Counting*. Undergraduate thesis, Harvard University. Available from https://people.seas.harvard.edu/ salil/research/ugthesis-abs.html. 1995 (cit. on p. 137).

[Val79] Leslie G. Valiant. "The Complexity of Enumeration and Reliability Problems". In: *SIAM J. Comput.* 8 (1979), pp. 410–421 (cit. on p. 137).

[Var18] Shai Vardi. "Randomly Coloring Graphs of Logarithmically Bounded Pathwidth". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Ed. by Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer. Vol. 116. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, 57:1–57:19. ISBN: 978-3-95977-085-9. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2018.57 (cit. on p. 198).

[Vaz03] Vijay V. Vazirani. *Approximation Algorithms*. Springer-Verlag Berlin Heidelberg, 2003, p. 380 (cit. on p. 9).

[Ver91] D.L. Vertigan. "On the computational complexity of tutte, jones, homfly and kauffman invariants". PhD thesis. University of Oxford, 1991 (cit. on p. 100).

[Vig00] Eric Vigoda. "Improved bounds for sampling colorings". In: *Journal of Mathematical Physics* 41 (3 2000), pp. 1555–1569 (cit. on pp. 15, 65, 66, 148, 149, 154, 155, 198).

[Vig01] Eric Vigoda. "A Note on the Glauber Dynamics for Sampling Independent Sets". In: *The Electronic Journal of Combinatorics* 8 (1 2001) (cit. on pp. 14, 136).

[Vil09] Cédric Villani. *Optimal Transport: Old and New*. 1st ed. Vol. 338. Grundlehren der mathematischen Wissenschaften A Series of Comprehensive Studies in Mathematics. Springer-Verlag Berlin Heidelberg, 2009, p. 976 (cit. on p. 25).

[VVY13] Juan C. Vera, Eric Vigoda, and Linji Yang. "Improved Bounds on the Phase Transition for the Hard-Core Model in 2-Dimensions". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Ed. by Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and José D. P. Rolim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 699–713 (cit. on p. 136).

[VWG17] Nate Veldt, Anthony I. Wirth, and David F. Gleich. "Correlation Clustering with Low-Rank Matrices". In: *Proceedings of the 26th International Conference on World Wide Web*. WWW '17. Perth, Australia: International World Wide Web Conferences Steering Committee, 2017, pp. 1025–1034. ISBN: 9781450349130. DOI: 10.1145/3038912.3052586 (cit. on p. 200).

[Wag08] David G. Wagner. "Negative Correlated Random Variables and Mason's Conjecture for Independent Sets in Matroids". In: *Annals of Combinatorics* 12.2 (2008), pp. 211–239 (cit. on p. 104).

[Wag09] David G. Wagner. "Weighted enumeration of spanning subgraphs with degree constraints". In: *Journal of Combinatorial Theory, Series B* 99.2 (2009), pp. 347–357 (cit. on pp. 115, 127).

[Wag11] David G. Wagner. "Multivariate stable polynomials: theory and applications". In: *Bulletin of the American Mathematical Society* 48.1 (2011), pp. 53–84 (cit. on p. 108).

[Wei04] Dror Weitz. "Mixing in Time and Space for Discrete Spin Systems". AAI3147046. PhD thesis. University of California, Berkeley, 2004. ISBN: 0-496-05573-9 (cit. on pp. 14, 17, 18, 131, 136, 269).

[Wei06] Dror Weitz. "Counting Independent Sets Up to the Tree Threshold". In: *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*. Seattle, WA, USA, 2006, pp. 140–149 (cit. on pp. 14, 15, 17, 18, 66, 131, 133, 134, 136, 138–140, 259, 267–269, 273, 289).

[Wel71] D. J. A. Welsh. "Combinatorial problems in matroid theory". In: (1971), pp. 291–306 (cit. on p. 104).

[Wel76] D. Welsh. "Matroid Theory". In: vol. 8. London Mathematical Society Monographs. Academic Press, London-New York, 1976 (cit. on p. 199).

[Wel94] D. J. A. Welsh. "The computational complexity of knot and matroid polynomials". In: *Discrete Mathematics* 124 (1-3 1994), pp. 251–269 (cit. on p. 100).

[Whi32] Hassler Whitney. "A logical expansion in mathematics". In: *Bull. Amer. Math. Soc.* 38.8 (1932), pp. 572–579 (cit. on p. 200).

[Whi35] Hassler Whitney. "On the Abstract Properties of Linear Dependence". In: *American Journal of Mathematics* 57.3 (1935), pp. 509–533 (cit. on p. 11).

[Wil76] Herbert S. Wilf. "Which polynomials are chromatic?" In: *Proc. 1973 Rome International Colloq. Combinatorial Theory I, Accademia Nazionale dei Lincei, Rome* (1976), pp. 247–257 (cit. on p. 199).

[Wil96] David Bruce Wilson. "Generating random spanning trees more quickly than the cover time". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 1996, pp. 296–303 (cit. on p. 189).

[WJ08] Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*. Vol. 1. Foundations and Trends in Machine Learning 1–2. Now Publishers Inc, 2008, pp. 1–305 (cit. on p. 8).

[WS11] David P. Williamson and David B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511921735 (cit. on p. 9).

[WSK89] Jian-Sheng Wang, Robert H. Swendsen, and Roman Kotecký. "Antiferromagnetic Potts models". In: *Phys. Rev. Lett.* 63 (2 July 1989), pp. 109–112. DOI: 10.1103/PhysRevLett.63.109 (cit. on p. 155).

[Yin14] Yitong Yin. "Spatial Mixing of Coloring Random Graphs". In: *Automata, Languages, and Programming*. Ed. by Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias. Springer Berlin Heidelberg, 2014, pp. 1075–1086 (cit. on p. 66).

[Zha85] Cui Kui Zhao. "A conjecture on matroids". In: *Neimenggu Daxue Xuebao* 16.3 (1985), pp. 321–326 (cit. on p. 104).

[ZLB11] Jinshan Zhang, Heng Liang, and Fengshan Bai. "Approximating partition functions of the two-state spin system". In: *Information Processing Letters* 111.14 (2011), pp. 702–710. ISSN: 0020-0190 (cit. on pp. 137, 242).

Part III

Appendices

Appendix A

General Two-Spin Systems in the Correlation Decay Regime

In this chapter, we extend the analysis of the hardcore model from Chapter 7 to general two-state spin systems, including the ferromagnetic and antiferromagnetic Ising models. The content of this chapter is based primarily on [CLV20].

Recall the input is a graph $G = (V, E)$, parameters $0 \leq \beta \leq \gamma < \infty$, and external fields $\boldsymbol{\lambda} = \{\lambda_v : v \in V\} \in \mathbb{R}^V_{\geq 0}$, and the resulting distribution and multivariate partition function are given in Eqs. (7.7) and (7.8). We say the system is *antiferromagnetic* if $\beta\gamma < 1$ and *ferromagnetic* if $\beta\gamma > 1$; in the former case, the system "prefers" configurations with more disagreements, while in the latter case, the system "favors" configurations with fewer disagreements. The case $\beta\gamma = 1$ is considered *trivial* since

$$\begin{aligned}\beta^{m_0(\sigma)}\gamma^{m_1(\sigma)} &= \beta^{m_0(\sigma)}\gamma^{m_1(\sigma)} \cdot (\beta\gamma)^{\#\{uv\in E:\sigma(u)\neq\sigma(v)\}} \\ &= \beta^{\#\{uv\in E:\sigma(u)=0\}}\gamma^{\#\{uv\in E:\sigma(v)=1\}} \\ &= \prod_{u\in V:\sigma(u)=0} \beta^{\deg_G(u)} \cdot \prod_{v\in V:\sigma(v)=1} \gamma^{\deg_G(v)}\end{aligned}$$

means we can absorb the β, γ edge interaction terms into the external fields by rescaling λ_v by $\left(\frac{\beta}{\gamma}\right)^{\deg_G(v)}$ for all $v \in V$; in particular, the resulting Gibbs distribution is a product of independent Bernoulli measures when $\beta\gamma = 1$.

The main mixing result in this section are the following.

Theorem A.0.1 (Tree Unique Antiferromagnetic Two-Spin Models)**.** *Let $\Delta \geq 3$ be an integer and $0 < \delta < 1$. 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 δ. Then for every n-vertex graph G of maximum degree Δ, the Gibbs distribution $\mu = \mu_{G,\beta,\gamma,\lambda}$ of the antiferromagnetic 2-spin system on G with parameters (β, γ, λ) is $(\eta_0, \ldots, \eta_{n-2})$-spectrally independent with*

$$\eta_k \leq \min\left\{\frac{c}{\delta}, C \cdot (n-k-1)\right\}, \quad \forall 0 \leq k \leq n-2,$$

where c is a universal constant and C is a constant depending on $\beta, \gamma, \lambda, \Delta$. Furthermore, the mixing time of the Glauber dynamics for sampling from μ is at most $O_{\Delta,\beta,\gamma,\lambda,\delta}(n \log n)$ if $\Delta \leq O(1)$ and $O(n^{2+c/\delta})$ in general

Remark 57. If we restrict attention to the case $\beta = \gamma$, i.e. the antiferromagnetic Ising model, then we may take $c = 1.5$. In the general setting, we may take $c = 72$.

We achieve this by proving $O(1/\delta)$-spectral independence for the Gibbs distribution like we did for the hardcore model in Chapter 7. In Appendix E, we give evidence that this $O(1/\delta)$-upper bound on the spectral independence is tight; see Theorem E.1.1.

Recall from Chapter 7 that our strategy is to first reduce bounding the total influence of a vertex in G to bounding the total influence of the root vertex of an associated self-avoiding walk tree; this was Theorem 7.2.2. We then bound the total influence of the root of any tree by leveraging the contraction properties of multivariate tree recursions. The first step, i.e. Theorem 7.2.2, extends

for free as we discussed in Section 7.2. The chain rule for influences in trees (see Lemma 7.3.2) extends for free as well. What is left to do is to state the multivariate tree recursion for generate two-state spin systems like Eq. (7.4), prove an analog of Lemma 7.3.3 giving a formula for the influence across an edge in a tree, and find suitable potential functions à la Eq. (7.10) which yield contraction when composed with the multivariate tree recursion like in Theorem 7.3.4. We do each in turn.

A.1 Two-State Spin Systems on Trees

We first state the multivariate tree recursion in full generality. The special case of the hardcore model was given in Eq. (7.4). Again, to state it, we work with the marginal ratios like in Eq. (7.3).

Consider a tree rooted at r. Suppose that r has d children, denoted by $v_1, \dots, v_d$. For $1 \le i \le d$ we define T_{v_i} to be the subtree of T rooted at v_i that contains all descendant of v_i. Let $R_r = \mu_{T,r}(1)/\mu_{T,r}(0)$ denote the marginal ratio of the root, and $R_{v_i} = \mu_{T_{v_i},v_i}(1)/\mu_{T_{v_i},v_i}(0)$ for each subtree. The *multivariate tree recursion* is given by $R_r = F_d(R_{v_1}, \dots, R_{v_d})$ where $F_d : [0,+\infty]^d \to [0,+\infty]$ is the multivariate function

$$F_d(R_1, \dots, R_d) \stackrel{\text{def}}{=} \lambda \prod_{i=1}^{d} \frac{\beta R_i + 1}{R_i + \gamma}. \tag{A.1}$$

Again, when T is a complete d-ary tree, the R_{v_i} are all equal by symmetry, and so the multivariate tree recursion reduces to the *univariate tree recursion* given by

$$f_d(R) \stackrel{\text{def}}{=} \lambda \left(\frac{\beta R + 1}{R + \gamma} \right)^d.$$

In a similar manner, $f_d(R)$ admits a unique fixed point $\hat{R}_d = \hat{R}_{d,\beta,\gamma,\lambda}$ and the behavior of f_d' at $\hat{R}_d$ turns out to govern the behavior of the system. We have the following definition analogous to Definition 34.

Definition 49 (Up-to-Δ Uniqueness General; [LLY13])**.** *For an integer $\Delta \ge 3$, we say the parameters (β, γ, λ) are **up-to-Δ unique** if $\left| f_d'\left(\hat{R}_d\right) \right| < 1$ for every $1 \le d < \Delta$. We say the parameters (β, γ, λ) are **up-to-Δ unique with gap** $0 < \delta < 1$ if $\left| f_d'\left(\hat{R}_d\right) \right| \le 1 - \delta$ for every $1 \le d < \Delta$.*

One particularly clean observation we made in [CLV20] is that the influence $\Psi_G(r \to v)$ of r on v can be viewed as the derivative of the *log-ratio* $\log R_r$ with respect to the (log-)external field at v (see Lemma A.1.4). Thus, it is more convenient for us to work with these log-ratios. To this end, we rewrite the multivariate tree recursion from Eq. (A.1) as $\log R_v = H_d(\log R_{v_1}, \dots, \log R_{v_d})$ where $H_d : [-\infty,+\infty]^d \to [-\infty,+\infty]$ is the multivariate function

$$H_d(y_1, \dots, y_d) \stackrel{\text{def}}{=} \log \lambda + \sum_{i=1}^{d} \log \left(\frac{\beta e^{y_i} + 1}{e^{y_i} + \gamma} \right). \tag{A.2}$$

Observe that $H_d = \log \circ F_d \circ \exp$ for every d. Moreover, we define

$$h(y) \stackrel{\text{def}}{=} -\frac{(1 - \beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}$$

for $y \in [-\infty,+\infty]$, so that $\partial_{y_i} H_d(y_1, \dots, y_d) = h(y_i)$ for each i.

A.1.1 Reduction to Finding Good Potential Functions

To prove our main results, we again 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. Let us first specify our requirements on the potential. For every integer $d \ge 0$, we define a bounded interval J_d which contains all log ratios at a vertex of degree d. More specifically, we let $J_d = \left[\log(\lambda\beta^d), \log(\lambda/\gamma^d)\right]$ when $\beta\gamma < 1$, and $J_d = \left[\log(\lambda/\gamma^d), \log(\lambda\beta^d)\right]$ 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 50 ((Relaxed) (α, c)-Potential function)**.** *Let $\Delta \geq 3$ be an integer, and 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 strictly increasing function which is differentiable on $(-\infty, +\infty)$. For any $\alpha \in (0,1)$ and $c > 0$, we say Ξ is an (α, c)-**potential function w.r.t.** Δ **and** (β, γ, λ) if it satisfies the following conditions:*

- *(Contraction) For every integer d such that $1 \leq d < \Delta$, and every tuple $(\tilde{y}_1, \ldots, \tilde{y}_d) \in \mathrm{Im}(\Xi)^d$, we have*
$$\left\|\nabla H_d^{\Xi}(\tilde{y}_1, \ldots, \tilde{y}_d)\right\|_1 = \sum_{i=1}^{d} \frac{\Xi'(y)}{\Xi'(y_i)} \cdot |h(y_i)| \leq 1 - \alpha$$
where $H_d^{\Xi} \stackrel{\mathsf{def}}{=} \Xi \circ H_d \circ \Xi^{-1}$, $y_i = \Xi^{-1}(\tilde{y}_i)$ for $1 \leq i \leq d$, and $y = H_d(y_1, \ldots, y_d)$.
- *(Boundedness) For every $y_1, y_2 \in J$, we have*
$$\frac{\Xi'(y_2)}{\Xi'(y_1)} \cdot |h(y_1)| \leq \frac{c}{\Delta}.$$

*We say Ξ is an **relaxed** (α, c)-**potential function w.r.t.** Δ **and** (β, γ, λ) if it satisfies Contraction and instead of Boundedness, it satisfies the following weaker version of the boundedness condition.*

- *(Relaxed 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*
$$\frac{\Xi'(y_2)}{\Xi'(y_1)} \cdot |h(y_1)| \leq \frac{2c}{d_1 + d_2 + 2}.$$

In the definition of (α, c)-potential, one should think of y as the log-ratio R at a vertex and the potential function is of $\log R$. The following theorem establishes spectral independence for the Gibbs distribution of a two-spin system given an (α, c)-potential function.

Theorem A.1.1 (Contraction Implies Spectral Independence and Mixing)**.** *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 $\mu = \mu_{G,\beta,\gamma,\lambda}$ of the two-spin system on G with parameters (β, γ, λ) is $(\eta_0, \ldots, \eta_{n-2})$-spectrally independent where*
$$\eta_k \leq \min\left\{\frac{c}{\alpha}, C \cdot (n - k - 1)\right\}$$
and C is a constant depending only on $\beta, \gamma, \lambda, \Delta$ but not n; furthermore, C can be made independent of Δ if (β, γ, λ) is up-to-Δ unique.

In particular, the mixing time of the Glauber dynamics for sampling from μ mixes in $O_{\Delta,\beta,\gamma,\lambda}(n \log n)$-steps if $\Delta, c, \alpha \leq O(1)$ and $O(n^{2+c/\alpha})$-steps in general. If only a relaxed (α, c)-potential function w.r.t. Δ and (β, γ, λ) exists, then the same conclusions hold except all occurrences of c/α are replaced by $2c/\alpha$.

Note that in both Definition 50 and Theorem A.1.1, 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 [LLY13] would give $c = \Theta(\Delta)$ for the Boundedness condition, resulting in $\Theta(\Delta)$-spectral independence and potentially $n^{\Theta(\Delta)}$-mixing. 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). This allows us to establish $O(1/\delta)$-spectral independence independent of $\Delta, \beta, \gamma, \lambda$ when (β, γ, λ) is up-to-Δ unique with gap $0 < \delta < 1$.

The reason we also define a *relaxed* version of boundedness in Relaxed Boundedness is that this is necessary for our analysis of antiferromagnetic 2-spin systems with $0 \leq \beta < 1 < \gamma$. The reason is that the first version of Boundedness turns out to imply bounds on $\|\Psi_G\|_\infty$, similar to what we did for the hardcore mode (see Section 7.4 and the proof of Theorem 7.1.2). However, when $0 \leq \beta < 1 < \gamma$, there are graphs for which $\lambda_{\max}(\Psi_G) \leq O(1)$ independent of Δ, but $\|\Psi_G\|_\infty = \Theta(\Delta)$. The following gives one such example.

Example 12. Consider the antiferromagnetic two-spin system specified by parameters $\beta = 0$, $\gamma > 1$ and $\lambda > 0$ on the *star graph* centered at r with $\Delta < \infty$ leaves. A simple calculation reveals that $|\Psi_G(r \to v)| = \frac{\lambda}{\lambda+\gamma}$ for any leaf vertex $v \neq r$. Hence, $\sum_{v \neq r} |\Psi_G(r \to v)| = \Delta \cdot \frac{\lambda}{\lambda+\gamma}$. 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 $\sum_{v \neq r} |\Psi_G(r \to v)| = \Theta_\gamma(\Delta)$ even when $\lambda < \lambda_c$ lies in the uniqueness region. However, we still have $\boldsymbol{\lambda}_{\max}(\Psi_G) = O(1)$ since $\sum_{v \neq r} |\Psi_G(v \to r)| = O(1)$ and more broadly, $\|\Psi_G\|_1 \leq O(1)$ independent of Δ.

To circumvent this issue, one might want to consider the absolute column sum as in [ALO21], involving the sum of absolute influences on a fixed vertex. However, this will not allow us to use the beautiful connection between graphs and self-avoiding walk trees as showed in Theorem 7.2.2 and Corollary 7.2.3. So instead, we consider a vertex-weighted version of the absolute row sum of Ψ_G, which also upper bounds the maximum eigenvalue. This motivates Relaxed Boundedness.

Finally, 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 for the log-ratios R, and Φ, a potential function of R. In fact, our results can also be translated into this language. 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 50 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 paper 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 [LLY13], by the transformation $\Xi = \Phi \circ \exp$.[1] It is intriguing to notice that the derivative of this potential is simply $\Xi = \sqrt{|h|}$. Then the Contraction condition has a nice form: $\sum_{i=1}^d \sqrt{h(y) \cdot h(y_i)} \leq 1 - \alpha$; 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 [LLY13], and also indicates that H_d is a meaningful variant of the tree recursion to consider. To add one more piece of 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, we can take Ξ to be the identity function, in which case H_d itself is contracting without any nontrivial potential (see e.g. [ZLB11; SS19]).

Revision in July 2021 After the publication of [CLV20], a small error was found in [LLY13] regarding descriptions of the uniqueness region for antiferromagnetic 2-spin systems. The error was fixed in the latest version of [LLY13]. We updated corresponding results and proofs in Appendices A.2 and A.4 that are affected by the changes in [LLY13]; in particular, Lemma A.4.2 is adjusted in accordance with the current description of uniqueness regions. We remark that these changes are purely technical and do not affect the validity of our main results like Theorem A.1.1.

A.1.2 Contraction Implies Influence Bounds on Trees

Our goal in this subsection is to prove Theorem A.1.1. The remaining sections in this chapter are then devoted to showing that certain potential functions constructed in prior work are suitable in the sense of Definition 50.

Lemma A.1.2. *Let $T = (V_T, E_T)$ be a tree rooted at r, and let $\Xi : [-\infty, +\infty] \to (-\infty, +\infty)$ be an increasing potential function which is differentiable on $(-\infty, +\infty)$. Denote the degree of the root r by Δ_r. Then for every weight function $\rho : V_T \to \mathbb{R}_{\geq 0}$ and every integer $k \geq 1$,*

$$\sum_{v \in L_r(k)} \rho_v \cdot |\Psi_T^{\sigma_\Lambda}(r \to v)| \leq \Delta_r A_\Xi B_\Xi^\rho \cdot \left(\max_{1 \leq d < \Delta} \sup_{\tilde{\boldsymbol{y}} \in \mathrm{Im}(\Xi)^d} \left\| \nabla H_d^\Xi(\tilde{\boldsymbol{y}}) \right\|_1 \right)^{k-1}$$

where

$$A_\Xi \stackrel{\text{def}}{=} \max_{u \in L_r(1)} \left\{ \frac{|h(\log R_u)|}{\Xi'(\log R_u)} \right\} \quad \text{and} \quad B_\Xi^\rho \stackrel{\text{def}}{=} \max_{v \neq r} \left\{ \rho_v \cdot \Xi'(\log R_v) \right\}.$$

[1] To be more precise, we also multiply a constant factor which only simplifies our calculation and does not matter much. Also, note that [LLY13] instead denotes the potential function by φ and its derivative by $\Phi = \varphi'$.

Before we prove this lemma, let us see how this implies Theorem A.1.1.

Proof of Theorem A.1.1. Fix $0 \le k \le n-2$ and some boundary condition $\sigma_\Lambda : \Lambda \to \{0,1\}$ where $\Lambda \subseteq V$ with $|\Lambda| = k$. We primarily focus on the first bound, since it is the nontrivial. If Ξ is an (α, c)-potential function, then since

$$\begin{aligned}
\boldsymbol{\lambda}_{\max}\left(\Psi_G^{\sigma_\Lambda}\right) \le \left\|\Psi_G^{\sigma_\Lambda}\right\|_\infty &= \max_{r\in V\setminus\Lambda} \sum_{v\in V\setminus\Lambda: v\neq r} \left|\Psi_G^{\sigma_\Lambda}(r\to v)\right| \\
&\le \max_{r\in V\setminus\Lambda} \sum_{\hat{v}\in V_{T(r)}:\hat{v}\neq\hat{r}} \left|\Psi_{T(r)}^{\sigma_\Lambda}(\hat{r}\to\hat{v})\right| && (T(r) = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)) \\
&= \max_{r\in V\setminus\Lambda} \sum_{k=1}^{\infty} \sum_{\hat{v}\in L_{\hat{r}}(k)} \left|\Psi_{T(r)}^{\sigma_\Lambda}(\hat{r}\to\hat{v})\right| && (\text{Split sum by distance levels}) \\
&\le \Delta_r A_\Xi B_\Xi^{\mathbf{1}} \sum_{k=1}^{\infty} \left(\max_{1\le d<\Delta} \sup_{\tilde{\boldsymbol{y}}\in \mathrm{Im}(\Xi)^d} \left\|\nabla H_d^\Xi(\tilde{\boldsymbol{y}})\right\|_1 \right)^{k-1} && (\text{Lemma A.1.2}) \\
&\le \Delta_r A_\Xi B_\Xi^{\mathbf{1}} \sum_{k=1}^{\infty} (1-\alpha)^{k-1} && (\text{Contraction}) \\
&\le \frac{\Delta_r A_\Xi B_\Xi^{\mathbf{1}}}{\alpha}.
\end{aligned}$$

By Boundedness, $A_\Xi B_\Xi^{\mathbf{1}} \le \frac{c}{\Delta}$ so that $\Delta_r A_\Xi B_\Xi^{\mathbf{1}} \le c$. The $\frac{c}{\alpha}$ bound follows.

The proof for the case Ξ is a relaxed (α, c)-potential function is nearly identical, requiring only an additional step which conjugates the two-sided influence matrix $\Psi_G^{\sigma_\Lambda}$ by the diagonal degree matrix $D = \mathrm{diag}(\Delta_v : v \in V)$. Taking the weight function $\rho : V_T \to \mathbb{R}_{\ge 0}$ to be $\rho_v = \Delta_v$, we have

$$\begin{aligned}
\boldsymbol{\lambda}_{\max}\left(\Psi_G^{\sigma_\Lambda}\right) = \boldsymbol{\lambda}_{\max}\left(D^{-1}\Psi_G^{\sigma_\Lambda} D\right) &\le \left\|D^{-1}\Psi_G^{\sigma_\Lambda} D\right\|_\infty \\
&= \max_{r\in V\setminus\Lambda} \frac{1}{\Delta_r} \sum_{v\in V\setminus\Lambda: v\neq r} \rho_v \cdot \left|\Psi_G^{\sigma_\Lambda}(r\to v)\right| \\
&\le \max_{r\in V\setminus\Lambda} \frac{1}{\Delta_r} \sum_{\hat{v}\in V_{T(r)}:\hat{v}\neq\hat{r}} \rho_v \cdot \left|\Psi_{T(r)}^{\sigma_\Lambda}(\hat{r}\to\hat{v})\right| && (T(r) = \mathrm{T}_{\mathsf{SAW}}(G, r; \sigma_\Lambda)) \\
&= \max_{r\in V\setminus\Lambda} \frac{1}{\Delta_r} \sum_{k=1}^{\infty} \sum_{\hat{v}\in L_{\hat{r}}(k)} \rho_v \cdot \left|\Psi_{T(r)}^{\sigma_\Lambda}(\hat{r}\to\hat{v})\right| && (\text{Split sum by distance levels}) \\
&\le A_\Xi B_\Xi^{\rho} \sum_{k=1}^{\infty} \left(\max_{1\le d<\Delta} \sup_{\tilde{\boldsymbol{y}}\in \mathrm{Im}(\Xi)^d} \left\|\nabla H_d^\Xi(\tilde{\boldsymbol{y}})\right\|_1 \right)^{k-1} && (\text{Lemma A.1.2}) \\
&\le A_\Xi B_\Xi^{\rho} \sum_{k=1}^{\infty} (1-\alpha)^{k-1} && (\text{Contraction}) \\
&\le \frac{A_\Xi B_\Xi^{\rho}}{\alpha}.
\end{aligned}$$

By Relaxed Boundedness, $A_\Xi B_\Xi^{\rho} \le 2c$, from which the $\frac{2c}{\alpha}$ bound follows. We refer to Fact A.6.1 and Fact A.6.2 for the $C \cdot (n-k-1)$ bound. □

All that remains is to prove Lemma A.1.2. We use the following, which is a straightforward generalization of Lemma 7.3.3 from the analysis of the hardcore model.

Lemma A.1.3. *Let $u \in V_T$ and v be a child of u in the subtree T_u. Then*

$$\Psi_T^{\sigma_\Lambda}(u \to v) = -(1-\beta\gamma) \cdot \frac{R_v}{(\beta R_v + 1)(R_v + \gamma)} = h(\log R_v).$$

The lemma can be proved through an explicit computation of the influence. Indeed, in a manner similar to Lemma 7.3.3, one just views the tree as being rooted at v instead, and applies the tree recursions Eq. (A.1) while setting the log-ratio at u to $+\infty$ or $-\infty$ depending on whether or not one is conditioning on $\sigma(u) = 1$ or $\sigma(u) = 0$. Here we present another proof utilizing the following fact, which we believe gives some insights into the relation between influence and the function h.

Lemma A.1.4. *For every graph $G = (V, E)$, $\Lambda \subseteq V$ and $\sigma_\Lambda : \Lambda \to \{0, 1\}$, the following hold:*

1. *For all $v \in V$,*
$$\lambda_v \partial_{\lambda_v} \log \mathcal{Z}_G^{\sigma_\Lambda} = \mu_{G,v}^{\sigma_\Lambda}(1);$$
2. *For all $u, v \in V$,*
$$(\lambda_u \partial_{\lambda_u})(\lambda_v \partial_{\lambda_v}) \log \mathcal{Z}_G^{\sigma_\Lambda} = \lambda_v \partial_{\lambda_v} \mu_{G,v}^{\sigma_\Lambda}(1) = \mathrm{Cov}_G^{\sigma_\Lambda}(u, v);$$
3. *For all $u, v \in V$,*
$$\lambda_v \partial_{\lambda_v} \log R_{G,u}^{\sigma_\Lambda} = \Psi_G^{\sigma_\Lambda}(u \to v).$$

Proof. Items 1 and 2 are standard folklore. One incarnation was proved, for instance, in Lemma 5.1.1. Again, the proof follows by observing that the differential operator ∂_{λ_v} picks out terms consistent with $\sigma(v) = 1$ and annihilates others. Differentiating again yields covariances. The only "new" observation is Item 3. We deduce it from Item 1.

$$\begin{aligned}
\lambda_v \partial_{\lambda_v} \log R_{G,u}^{\sigma_\Lambda} &= \lambda_v \partial_{\lambda_v} \log \left(\frac{\mathcal{Z}_G^{\sigma_\Lambda, u \leftarrow 1}}{\mathcal{Z}_G^{\sigma_\Lambda, u \leftarrow 0}} \right) \\
&= \mu_{G,v}^{\sigma_\Lambda, u \leftarrow 1}(1) - \mu_{G,v}^{\sigma_\Lambda, u \leftarrow 0}(1) \\
&= \Psi_G^{\sigma_\Lambda}(u \to v).
\end{aligned}$$

□

Proof of Lemma A.1.3. We assume that u has d children in the subtree T_u, denoted by $v_1 = v$ and $v_2, \ldots, v_d$ 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

$$\begin{aligned}
\Psi_T^{\sigma_\Lambda}(u \to v) &= \Psi_{T_u}^{\sigma_\Lambda}(u \to v) = \lambda_v \partial_{\lambda_v} \log R_u && \text{(Lemma A.1.4)} \\
&= \lambda_v \partial_{\lambda_v} H_d(\log R_{v_1}, \ldots, \log R_{v_d}) && \text{(Tree Recursion applied to } u\text{)} \\
&= \sum_{i=1}^{d} \frac{\partial}{\partial \log R_{v_i}} H_d(\log R_{v_1}, \ldots, \log R_{v_d}) \cdot \lambda_v \partial_{\lambda_v} \log R_{v_i} && \text{(Chain Rule)} \\
&= \sum_{i=1}^{d} h(\log R_{v_i}) \cdot \Psi_{T_{v_i}}^{\sigma_\Lambda}(v_i \to v) && \text{(Lemma A.1.4)} \\
&= h(\log R_v),
\end{aligned}$$

where the last equality is because $\Psi_{T_{v_i}}^{\sigma_\Lambda}(v_i \to v) = 0$ for $v_i \neq v$ since $v \notin T_{v_i}$ and the Markov property applies, and $\Psi_{T_v}^{\sigma_\Lambda}(v \to v) = 1$. Note that the argument still holds even if some children v_i are fixed to certain spins. □

We are now ready to Lemma A.1.2. We follow the proof of Theorem 7.3.1, using Lemma A.1.3 and Lemma 7.3.2, but written in the language of (H_d, Ξ) instead of (F_d, Φ).

Proof of Lemma A.1.2. For convenience, we prove it in the case when $\rho_v = 1$ for all $v \in V_T$; the general case is proved via a nearly identical argument. For a vertex $v \in V_T$, denote the number of its children by d_v; note that $d_r = \Delta_r$. Let $u_1, \ldots, u_{\Delta_r}$ 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 Lemmas 7.3.2 and A.1.3, we get

$$\begin{aligned}
\sum_{v \in L_r(k)} |\Psi_T^{\sigma_\Lambda}(r \to v)| &= \sum_{i=1}^{\Delta_r} |\Psi_T^{\sigma_\Lambda}(r \to u_i)| \cdot \left(\sum_{v \in L_{u_i}(k-1)} |\Psi_T^{\sigma_\Lambda}(u_i \to v)| \right) \\
&= \sum_{i=1}^{\Delta_r} |h(\log R_{u_i})| \cdot \left(\sum_{v \in L_{u_i}(k-1)} |\Psi_T^{\sigma_\Lambda}(u_i \to v)| \right) \\
&= \sum_{i=1}^{\Delta_r} \frac{|h(\log R_{u_i})|}{\Xi'(\log R_{u_i})} \cdot \left(\sum_{v \in L_{u_i}(k-1)} \Xi'(\log R_{u_i}) \, |\Psi_T^{\sigma_\Lambda}(u_i \to v)| \right).
\end{aligned}$$

Hence, we obtain that

$$\begin{aligned}&\sum_{v\in L_r(k)} |\Psi_T^{\sigma_\Lambda}(r\to v)|\\ &\le \Delta_r \cdot \max_{1\le i\le \Delta_r}\left\{\frac{|h(\log R_{u_i})|}{\Xi'(\log R_{u_i})}\right\}\cdot \max_{1\le i\le \Delta_r}\left\{\sum_{v\in L_{u_i}(k-1)} \Xi'(\log R_{u_i})\cdot|\Psi_T^{\sigma_\Lambda}(u_i\to v)|\right\}.\end{aligned} \tag{A.3}$$

Next, we show by induction that for every vertex $u\in V_T\setminus\{r\}$ and every integer $k\ge 0$ we have

$$\sum_{v\in L_u(k)} \Xi'(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to v)| \le \max_{v\in L_u(k)}\{\Xi'(\log R_v)\}\cdot\left(\max_{w\in V_{T_u}}\sup_{\tilde{\boldsymbol{y}}\in \mathrm{Im}(\Xi)^{d_w}} \left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^k. \tag{A.4}$$

Observe that once we establish Eq. (A.4), the lemma follows immediately by plugging Eq. (A.4) into Eq. (A.3). We will use induction on k to prove Eq. (A.4). When $k=0$, if $u\in\Lambda$ is fixed then $L_u(0)=\emptyset$ and there is nothing to show; otherwise, Eq. (A.4) becomes

$$\Xi'(\log R_u)\cdot|\Psi_T^{\sigma_\Lambda}(u\to u)|\le \Xi'(\log R_u),$$

which holds with equality since $\Psi_T^{\sigma_\Lambda}(u\to u)=1$. Now suppose that Eq. (A.4) holds for some integer $k-1\ge 0$ (and for every vertex $u\in V_T\setminus\{r\}$). Let $u\in V_T\setminus\{r\}$ be arbitrary and denote the children of u by $w_1,\dots,w_d$, where $1\le d<\Delta$ (if $d=0$ then $L_u(k)=\emptyset$ and Eq. (A.4) holds trivially). Again by Lemmas 7.3.2 and A.1.3 we have

$$\begin{aligned}&\sum_{v\in L_u(k)} \Xi'(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to v)|\\ &=\sum_{i=1}^{d}\Xi'(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to w_i)|\sum_{v\in L_{w_i}(k-1)}|\Psi_T^{\sigma_\Lambda}(w_i\to v)|\\ &=\sum_{i=1}^{d}\frac{\Xi'(\log R_u)}{\Xi'(\log R_{w_i})}\,|h(\log R_{w_i})|\sum_{v\in L_{w_i}(k-1)}\Xi'(\log R_{w_i})\,|\Psi_T^{\sigma_\Lambda}(w_i\to v)|.\end{aligned}$$

Using the induction hypothesis, we get

$$\begin{aligned}&\sum_{v\in L_u(k)} \Xi'(\log R_u)\,|\Psi_T^{\sigma_\Lambda}(u\to v)|\\ &\le\sum_{i=1}^{d}\frac{\Xi'(\log R_u)}{\Xi'(\log R_{w_i})}\,|h(\log R_{w_i})|\cdot\max_{v\in L_{w_i}(k-1)}\{\Xi'(\log R_v)\}\cdot\left(\max_{w\in V_{T_{w_i}}}\sup_{\tilde{\boldsymbol{y}}\in\mathrm{Im}(\Xi)^{d_w}}\left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^{k-1}\\ &\le\max_{v\in L_u(k)}\{\Xi'(\log R_v)\}\cdot\left(\max_{w\in V_{T_u}\setminus\{u\}}\sup_{\tilde{\boldsymbol{y}}\in\mathrm{Im}(\Xi)^{d_w}}\left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^{k-1}\cdot\sum_{i=1}^{d}\frac{\Xi'(\log R_u)}{\Xi'(\log R_{w_i})}\,|h(\log R_{w_i})|\\ &\le\max_{v\in L_u(k)}\{\Xi'(\log R_v)\}\cdot\left(\max_{w\in V_{T_u}}\sup_{\tilde{\boldsymbol{y}}\in\mathrm{Im}(\Xi)^{d_w}}\left\|\nabla H_{d_w}^{\Xi}(\tilde{\boldsymbol{y}})\right\|_1\right)^{k},\end{aligned}$$

where the last inequality follows from the following calculation:

$$\begin{aligned}\sum_{i=1}^{d}\frac{\Xi'(\log R_u)}{\Xi'(\log R_{w_i})}\,|h(\log R_{w_i})| &= \sum_{i=1}^{d}\left|\frac{\partial}{\partial\Xi(\log R_{w_i})}H_d^{\Xi}\left(\Xi(\log R_{w_1}),\dots,\Xi(\log R_{w_d})\right)\right|\\ &= \left\|\nabla H_d^{\Xi}\left(\Xi(\log R_{w_1}),\dots,\Xi(\log R_{w_d})\right)\right\|_1.\end{aligned}$$

This establishes Eq. (A.4), and thus completes the proof of the lemma. □

A.2 Antiferromagnetic Two-Spin Systems

In this section, we give (relaxed) (α, c)-potential functions (in the sense of Definition 50) for good α, c when the parameters (β, γ, λ) are up-to-Δ unique with gap $0 < \delta < 1$ and the system is antiferromagnetic (i.e. $\beta\gamma < 1$). Consider the following function $\Phi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ discovered in [LLY13], which generalizes the one in Eq. (7.10) for the hardcore model; note that Φ is only implicitly defined through its derivative.

$$\Phi'(R) \stackrel{\text{def}}{=} \frac{1}{\sqrt{R(\beta R + 1)(R + \gamma)}}, \quad \Phi(1) = 0. \tag{A.5}$$

We are not aware of a closed-form expression for the general case, unlike the special case of the hardcore model. We refer the interested reader to [LLY13] for a discussion of how this function was discovered.

Combining the results of Lemmas 12, 13 and 14 from [LLY13], we get that the potential function Φ satisfies the following gradient bound when the parameters (β, γ, λ) are in the tree uniqueness region. Note that this can be regarded as the Contraction condition but for (F_d, Φ). This is a generalization of Theorem 7.3.4.

Theorem A.2.1 ([LLY13])**.** *Assume $\beta\gamma < 1$ and the parameters (β, γ, λ) are up-to-Δ unique with gap $0 < \delta < 1$. Then for every integer $1 \leq d < \Delta$,*

$$\sup_{\boldsymbol{K} \in \mathrm{Im}(\Phi)^d} \left\| \nabla F_d^{\Phi}(\boldsymbol{K}) \right\|_1 \leq \sqrt{1 - \delta} < 1,$$

where recall that $F_d^{\Phi} \stackrel{\text{def}}{=} \Phi \circ F_d \circ \Phi^{-1}$.

Since in this chapter, we are working with the log-ratios and the corresponding tree recursion H_d, we build our potential function Ξ based on Φ, which recall is defined in Eq. (A.5) above. Specifically, define $\Xi \stackrel{\text{def}}{=} \sqrt{1 - \beta\gamma} \cdot (\Phi \circ \exp)$, i.e.

$$\Xi(y) \stackrel{\text{def}}{=} \sqrt{1 - \beta\gamma} \cdot \Phi\left(e^y\right). \tag{A.6}$$

Note that a straightforward calculation using the Chain Rule yields that

$$\Xi'(y) = \sqrt{1 - \beta\gamma} \cdot e^y \cdot \Phi'\left(e^y\right) = \sqrt{\frac{(1 - \beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}}.$$

We show that Ξ is a (relaxed) (α, c)-potential function. We include a short proof in Appendix A.7 to show that Ξ is well-defined.

Proposition A.2.2 (Potential Function for Tree Unique Antiferromagnetic Two-Spin Models)**.** *Let $\Delta \geq 3$ be an integer, and 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 $0 < \delta < 1$, and let Ξ be the function defined in Eq. (A.6).*

1. *If $\sqrt{\beta\gamma} > \frac{\Delta - 2}{\Delta}$, then Ξ is an (α, c)-potential function with $\alpha \geq \delta/2$ and $c \leq 1.5$.*
2. *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 $\beta = 0$.*
3. *If $\sqrt{\beta\gamma} \leq \frac{\Delta - 2}{\Delta}$ and $\gamma > 1$, then Ξ is a relaxed (α, c)-potential with $\alpha \geq \delta/2$ and $c \leq 18$; we can further take $c \leq 4$ if $\beta = 0$.*

We first prove Contraction as it follows immediately from Theorem A.2.1. The contraction rate $\alpha \geq \delta/2$ is the same in all cases, so long as the parameters are up-to-Δ unique with gap $0 < \delta < 1$ and the system is antiferromagnetic.

Proof of Contraction. It suffices to show that for every $1 \leq d < \Delta$ and every $(\tilde{y}_1, \ldots, \tilde{y}_d) \in \mathrm{Im}(\Xi)^d$, the following inequality holds

$$\left\| \nabla H_d^{\Xi}(\tilde{y}_1, \ldots, \tilde{y}_d) \right\|_1 \leq \sqrt{1 - \delta}.$$

Note that $\sqrt{1-\delta} \le 1 - \frac{\delta}{2}$.

Define the linear function $a : \mathbb{R} \to \mathbb{R}$ by $a(x) = \sqrt{1-\beta\gamma} \cdot x$ for $x \in \mathbb{R}$. Then Eq. (A.6) says that $\Xi = a \circ \Phi \circ \exp$, and thereby $\Xi \circ \log = a \circ \Phi$. It follows that for every $1 \le d < \Delta$,

$$H_d^{\Xi} = \Xi \circ H_d \circ \Xi^{-1} = \Xi \circ \log \circ F_d \circ \exp \circ \Xi^{-1} = a \circ \Phi \circ F_d \circ \Phi^{-1} \circ a^{-1} = a \circ F_d^{\Phi} \circ a^{-1}.$$

That means, for every $(\tilde{y}_1, \dots, \tilde{y}_d) \in \mathrm{Im}(\Xi)^d$, we have

$$H_d^{\Xi}(\tilde{y}_1, \dots, \tilde{y}_d) = \sqrt{1-\beta\gamma} \cdot F_d^{\Phi}(\tilde{x}_1, \dots, \tilde{x}_d)$$

where $\tilde{x}_i = \tilde{y}_i / \sqrt{1-\beta\gamma}$ for $1 \le i \le d$. Then, for each i,

$$\frac{\partial}{\partial \tilde{y}_i} H_d^{\Xi}(\tilde{y}_1, \dots, \tilde{y}_d) = \sqrt{1-\beta\gamma} \cdot \frac{\partial}{\partial \tilde{x}_i} F_d^{\Phi}(\tilde{x}_1, \dots, \tilde{x}_d) \cdot \frac{\mathrm{d}\tilde{x}_i}{\mathrm{d}\tilde{y}_i} = \frac{\partial}{\partial \tilde{x}_i} F_d^{\Phi}(\tilde{x}_1, \dots, \tilde{x}_d).$$

This implies that $\nabla H_d^{\Xi}(\tilde{y}_1, \dots, \tilde{y}_d) = \nabla F_d^{\Phi}(\tilde{x}_1, \dots, \tilde{x}_d)$ for all $(\tilde{y}_1, \dots, \tilde{y}_d) \in \mathrm{Im}(\Xi)^d$, and the desired contraction then follows from Theorem A.2.1. □

Proving Boundedness for Items 1 and 2, and Relaxed Boundedness for Item 3, are relegated to Appendix A.4.

A.3 Ferromagnetic Two-Spin Systems

In the ferromagnetic case, the best known correlation decay results are given in [GL18; SS19]. Using the potential functions in [GL18] and [SS19], we show the following two results, which match the known correlation decay results. See [GL18; GLL20] for further discussion on the tightness of these results.

To establish our next result, we use the potential function from [SS19], which turns out to be an (α, c)-potential function for constants $\alpha = \Theta(\delta)$ and $c = O(1)$.

Theorem A.3.1. *Fix an integer $\Delta \ge 3$, real numbers $\beta, \gamma, \lambda > 0$, and $0 < \delta < 1$, and assume (β, γ, λ) satisfies one of the following three conditions:*

1. $\frac{\Delta-2+\delta}{\Delta-\delta} \le \sqrt{\beta\gamma} \le \frac{\Delta-\delta}{\Delta-2+\delta}$, *and $\lambda > 0$ is arbitrary;*
2. $\sqrt{\beta\gamma} \ge \frac{\Delta}{\Delta-2}$ *and* $0 < \lambda \le (1-\delta)\frac{\gamma}{\max\{1, \beta^{\Delta-1}\} \cdot ((\Delta-2)\beta\gamma - \Delta)}$;
3. $\sqrt{\beta\gamma} \ge \frac{\Delta}{\Delta-2}$ *and* $\lambda \ge \frac{1}{1-\delta} \cdot \frac{(\Delta-2)\beta\gamma - \Delta}{\beta \cdot \min\{1, 1/\gamma^{\Delta-1}\}}$.

Then the identity function $\Xi(y) = y$ (based on the potential given in [SS19]) is an (α, c)-potential function for $\alpha = \Theta(\delta)$ and $c \le O(1)$. Furthermore, for every n-vertex graph G of maximum degree at most Δ, the mixing time of the Glauber dynamics for the 2-spin system on G with parameters (β, γ, λ) is $O(n^{2+c/\delta})$, for a universal constant $c > 0$.

Remark 58. Item 1 includes both the ferromagnetic case $1 < \sqrt{\beta\gamma} \le \frac{\Delta-\delta}{\Delta-2+\delta}$ and the antiferromagnetic case $\frac{\Delta-2+\delta}{\Delta-\delta} \le \sqrt{\beta\gamma} < 1$. 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 $c \le 1.5$ and a better contraction rate $\alpha \ge \delta$, compared with the bound $\alpha \ge \delta/2$ of the potential Ξ given by Eq. (A.6) in Proposition A.2.2. For the ferromagnetic case with $\beta = \gamma > 1$ (Ising model), a stronger result by [MS13] was known, which gives $O(n \log n)$ mixing.

The potential function from [GL18] 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 [GL18] for unbounded degree graphs.

Theorem A.3.2. *Fix an integer $\Delta \ge 3$, and nonnegative real numbers β, γ, λ satisfying $\beta \le 1 \le \gamma$, $\sqrt{\beta\gamma} \ge \frac{\Delta}{\Delta-2}$, 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 at most Δ, the mixing time of the Glauber dynamics for the ferromagnetic 2-spin system on G with parameters (β, γ, λ) is $O(n^C)$, for a constant C depending only on $\beta, \gamma, \lambda, \Delta$, but not n.*

Proofs of these theorems are provided in Appendix A.5.

A.4 Verifying (Relaxed) Boundedness for the Antiferromagnetic Case

In this section, we show the Boundedness or Relaxed Boundedness condition for our potential function Ξ defined by Eq. (A.6) in different ranges of parameters. As we have already proved Contraction, this completes the proof of Proposition A.2.2.

In Appendix A.4.1 we give background on the uniqueness region of the parameters (β, γ, λ), based on the work of [LLY13]. We then show Boundedness and Relaxed Boundedness in Appendix A.4.2. Proofs of technical lemmas are left to Appendix A.4.3.

A.4.1 Preliminaries on the Uniqueness Region

In this subsection we give a brief description of the uniqueness region of parameters (β, γ, λ). All the results here, and also their proofs, can also be found in Lemma 21 from the latest version of [LLY13].

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 $1 \leq d \leq \Delta$ recall that

$$f_d(R) \stackrel{\text{def}}{=} \lambda \left(\frac{\beta R + 1}{R + \gamma} \right)^d$$

and denote the unique fixed point of f_d by $\hat{R}_d$. Recall that the parameters (β, γ, λ) are up-to-Δ unique with gap $0 < \delta < 1$ if $|f_d'(\hat{R}_d)| < 1 - \delta$ 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

$$x_1(d) \stackrel{\text{def}}{=} \frac{\theta(d) - \sqrt{\theta(d)^2 - 4\beta\gamma}}{2\beta} \quad \text{and} \quad x_2(d) \stackrel{\text{def}}{=} \frac{\theta(d) + \sqrt{\theta(d)^2 - 4\beta\gamma}}{2\beta}$$

where

$$\theta(d) \stackrel{\text{def}}{=} d(1 - \beta\gamma) - (1 + \beta\gamma).$$

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

$$\mathcal{A} \stackrel{\text{def}}{=} \bigcap_{\overline{\Delta} \leq d < \Delta} \Big[(0, \lambda_1(d)) \cup (\lambda_2(d), \infty) \Big]. \tag{A.7}$$

In particular, when $\gamma \le 1$, 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

$$\lambda_c = \lambda_c(\beta, \gamma, \Delta) = \min_{\overline{\Delta} \le d < \Delta} \lambda_1(d)$$
$$\text{and} \quad \overline{\lambda}_c = \overline{\lambda}_c(\beta, \gamma, \Delta) = \max_{\overline{\Delta} \le d < \Delta} \lambda_2(d) = \lambda_2(\Delta - 1).$$

The following bounds on the critical fields are helpful for our proofs later.

Lemma A.4.1. *1. If $\beta = 0$, then for every integer d such that $1 < d < \Delta$, we have*

$$\lambda_c \le \frac{4\gamma^{d+1}}{d-1}.$$

2. If $\beta > 0$ and $\sqrt{\beta\gamma} \le \frac{\Delta-2}{\Delta}$, then for every integer d such that $\overline{\Delta} \le d < \Delta$ we have

$$\lambda_1(d) \le \frac{18\gamma^{d+1}}{\theta(d)} \quad \text{and} \quad \lambda_2(d) \ge \frac{\theta(d)}{18\beta^{d+1}}$$

where $\theta(d) = d(1-\beta\gamma) - (1+\beta\gamma)$.

The proof of Lemma A.4.1 is postponed to Appendix A.4.3.

A.4.2 Proofs of Boundedness

In this subsection we complete the proof of Proposition A.2.2 by establishing Boundedness and Relaxed Boundedness in the corresponding range of parameters.

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 the potential function Ξ is defined implicitly from its derivative by

$$\Xi'(y) = \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y + 1)(e^y + \gamma)}} = \sqrt{|h(y)|}, \quad \Xi(0) = 0.$$

It is surprising to find out that $\Xi' = \sqrt{|h|}$, as the potential Ξ is exactly the one from [LLY13] as indicated by Eq. (A.6). This seems not to be a coincidence, and it provides some intuition why the potential from [LLY13] works. More importantly, the fact that $\Xi' = \sqrt{|h|}$ is helpful in our proof of Boundedness and Relaxed Boundedness. Recall that for $0 \le d < \Delta$ 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 $0 \le d_i < \Delta$ and $y_i \in J_{d_i}$ where $i = 1, 2$, we have

$$\frac{\Xi'(y_2)}{\Xi'(y_1)} \cdot |h(y_1)| = \sqrt{|h(y_1)| \cdot |h(y_2)|}. \tag{A.8}$$

The following lemma gives upper bounds on $\sqrt{|h(y_1)| \cdot |h(y_2)|}$, from which we deduce Boundedness and Relaxed Boundedness immediately using Eq. (A.8). The brackets in the lemma indicate which lemma the bound is applied to.

Lemma A.4.2. *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 $\lambda > 0$. Assume that the parameters (β, γ, λ) are up-to-Δ unique with gap $0 < \delta < 1$. Then for all integers d_1, d_2 such that $0 \le d_1, d_2 < \Delta$, 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. (Proposition A.2.2 Item 2) If $\gamma \le 1$, then

$$|h(y_1)| \le \frac{4}{\Delta}.$$

H.2. (Proposition A.2.2 Item 3) If $\gamma > 1$, then

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \le \frac{8}{d_1 + d_2 + 2}.$$

S. ***Soft-constraint models:*** *$\beta > 0$ and $\lambda \in \mathcal{A}$.*

S.1. (Proposition A.2.2 Item 1) If $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$, then

$$|h(y_1)| \le \frac{1.5}{\Delta}.$$

S.2. (Proposition A.2.2 Item 2) If $\sqrt{\beta\gamma} \le \frac{\Delta-2}{\Delta}$ and $\gamma \le 1$, then

$$|h(y_1)| \le \frac{18}{\Delta}.$$

S.3. (Proposition A.2.2 Item 3) If $\sqrt{\beta\gamma} \le \frac{\Delta-2}{\Delta}$ and $\gamma > 1$, then

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} \le \frac{36}{d_1 + d_2 + 2}.$$

The following lemma, whose proof can be found in Appendix A.4.3, is helpful.

Lemma A.4.3. *The function*

$$|h(y)| = \frac{|1-\beta\gamma|e^y}{(\beta e^y + 1)(e^y + \gamma)}$$

is increasing on $[-\infty, \log\sqrt{\gamma/\beta}]$ and decreasing on $[\log\sqrt{\gamma/\beta}, +\infty]$. In particular, $|h(y)|$ is maximized at $y^* = \log\sqrt{\gamma/\beta}$ and we have the following inequality for all $y \in [-\infty, +\infty]$:

$$|h(y)| \le |h(y^*)| = \frac{|1-\sqrt{\beta\gamma}|}{1+\sqrt{\beta\gamma}}.$$

We present here the proof of Lemma A.4.2.

Proof of Lemma A.4.2. We use notations and results from Appendix A.4.1.

H. Hard-constraint models: $\beta = 0$ and $\lambda < \lambda_c$.

H.1. $\gamma \le 1$:

For every $y_1 \in J_{d_1}$ we deduce from Lemma A.4.1 that

$$e^{y_1} \le \frac{\lambda}{\gamma^{d_1}} \le \frac{\lambda_c}{\gamma^{\Delta-1}} \le \frac{4\gamma}{\Delta-2}.$$

Hence,

$$|h(y_1)| = \frac{e^{y_1}}{e^{y_1}+\gamma} \le \frac{\frac{4\gamma}{\Delta-2}}{\frac{4\gamma}{\Delta-2}+\gamma} = \frac{4}{\Delta+2} \le \frac{4}{\Delta}.$$

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

$$\sqrt{|h(y_1)| \cdot |h(y_2)|} = \sqrt{\frac{e^{y_1}}{e^{y_1}+\gamma}} \cdot \sqrt{\frac{e^{y_2}}{e^{y_2}+\gamma}} = \frac{1}{\sqrt{(1+\gamma e^{-y_1})(1+\gamma e^{-y_2})}} \le \frac{1}{1+\gamma e^{-\bar{y}}},$$

where the last inequality follows from the inequality of arithmetic and geometric means by

$$(1+\gamma e^{-y_1})(1+\gamma e^{-y_2}) = 1 + \gamma(e^{-y_1}+e^{-y_2}) + \gamma^2 e^{-2\bar{y}} \ge 1 + 2\gamma e^{-\bar{y}} + \gamma^2 e^{-2\bar{y}} = (1+\gamma e^{-\bar{y}})^2.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$, we have

$$e^{\bar{y}} = \sqrt{e^{y_1} \cdot e^{y_2}} \le \sqrt{\frac{\lambda}{\gamma^{d_1}} \cdot \frac{\lambda}{\gamma^{d_2}}} = \frac{\lambda}{\gamma^{\bar{d}}}.$$

If $\bar{d} \ge 2$, then we deduce from Lemma A.4.1 and $\gamma > 1$ that

$$e^{\bar{y}} \le \frac{\lambda_c}{\gamma^{\lfloor \bar{d} \rfloor}} \le \frac{4\gamma}{\lfloor \bar{d} \rfloor - 1}.$$

It follows that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1}{1+\gamma e^{-\bar{y}}} \le \frac{1}{1+\frac{\lfloor \bar{d}\rfloor -1}{4}} = \frac{4}{\lfloor \bar{d}\rfloor +3} \le \frac{8}{d_1+d_2+2}.$$

If $\bar{d}<2$, then it is easy to see that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le 1 \le \frac{8}{d_1+d_2+2}.$$

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 A.4.3 that

$$|h(y_1)| \le \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} \le \frac{1}{\Delta-1} \le \frac{1.5}{\Delta}.$$

S.2. $\sqrt{\beta\gamma} \le \frac{\Delta-2}{\Delta}$ and $\gamma\le 1$:

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 A.4.1 and $\beta<1$ that

$$e^{y_1} \ge \lambda\beta^{d_1} \ge \overline{\lambda}_c\beta^{\Delta-1} \ge \frac{\theta(\Delta-1)}{18\beta}$$

where $\theta(d)=d(1-\beta\gamma)-(1+\beta\gamma)$. Hence,

$$\begin{aligned}|h(y_1)| &= \frac{(1-\beta\gamma)e^{y_1}}{(\beta e^{y_1}+1)(e^{y_1}+\gamma)} = \frac{1-\beta\gamma}{\beta e^{y_1}+\gamma e^{-y_1}+(1+\beta\gamma)}\\ &\le \frac{1-\beta\gamma}{\frac{\theta(\Delta-1)}{18}+(1+\beta\gamma)} = \frac{18(1-\beta\gamma)}{(\Delta-1)(1-\beta\gamma)+17(1+\beta\gamma)} \le \frac{18}{\Delta}.\end{aligned}$$

Next we consider $\lambda<\lambda_c$. For every $y_1\in J_{d_1}$ we deduce from Lemma A.4.1 and $\gamma\le 1$ that

$$e^{y_1} \le \frac{\lambda}{\gamma^{d_1}} \le \frac{\lambda_c}{\gamma^{\Delta-1}} \le \frac{18\gamma}{\theta(\Delta-1)}.$$

Hence,

$$|h(y_1)| = \frac{1-\beta\gamma}{\beta e^{y_1}+\gamma e^{-y_1}+(1+\beta\gamma)} \le \frac{1-\beta\gamma}{\frac{\theta(\Delta-1)}{18}+(1+\beta\gamma)} \le \frac{18}{\Delta}.$$

S.3. $\sqrt{\beta\gamma} \le \frac{\Delta-2}{\Delta}$ 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 $\bar{d}\le 2$ then it is easy to see that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le 1 \le \frac{6}{d_1+d_2+2}.$$

If $\bar{d}>2$ and $d_L\le\overline{\Delta}$, then we deduce from Lemma A.4.3 that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} = \frac{1}{\overline{\Delta}} \le \frac{2}{d_1+d_2-2} \le \frac{6}{d_1+d_2+2}.$$

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. (A.7). 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

$$\begin{aligned}\sqrt{|h(y_1)|\cdot|h(y_2)|} &= \sqrt{\frac{1-\beta\gamma}{\beta e^{y_1}+\gamma e^{-y_1}+(1+\beta\gamma)}}\cdot\sqrt{\frac{1-\beta\gamma}{\beta e^{y_2}+\gamma e^{-y_2}+(1+\beta\gamma)}} \\ &\le \frac{1-\beta\gamma}{\sqrt{(\beta e^{y_1}+\gamma e^{-y_1})(\beta e^{y_2}+\gamma e^{-y_2})}+(1+\beta\gamma)}. \end{aligned} \tag{A.9}$$

Therefore, we get

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1-\beta\gamma}{\gamma e^{-\bar{y}}+(1+\beta\gamma)}.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$ and $\gamma > 1$, we deduce from Lemma A.4.1 that

$$e^{\bar{y}} \le \frac{\lambda}{\gamma^{\bar{d}}} \le \frac{\lambda_1(d_L)}{\gamma^{d_L}} \le \frac{18\gamma}{\theta(d_L)},$$

where $\theta(d_L) = d_L(1-\beta\gamma)-(1+\beta\gamma)$. It follows that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1-\beta\gamma}{\gamma e^{-\bar{y}}+(1+\beta\gamma)} \le \frac{1-\beta\gamma}{\frac{\theta(d_L)}{18}+(1+\beta\gamma)} \le \frac{36}{d_1+d_2+2}.$$

Case 2: $\lambda > \lambda_2(d_R)$. Similarly, we obtain from Eq. (A.9) that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1-\beta\gamma}{\beta e^{\bar{y}}+(1+\beta\gamma)}.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$ and $\beta < 1$, we deduce from Lemma A.4.1 that

$$e^{\bar{y}} \ge \lambda\beta^{\bar{d}} \ge \lambda_2(d_R)\beta^{d_R} \ge \frac{\theta(d_R)}{18\beta},$$

where $\theta(d_R) = d_R(1-\beta\gamma)-(1+\beta\gamma)$. It follows that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1-\beta\gamma}{\beta e^{\bar{y}}+(1+\beta\gamma)} \le \frac{1-\beta\gamma}{\frac{\theta(d_R)}{18}+(1+\beta\gamma)} \le \frac{36}{d_1+d_2+2}.$$

Case 3: $\lambda_2(d_L) < \lambda < \lambda_1(d_R)$. We may assume that $d_1 \ge d_2$. By Eq. (A.9), we obtain

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1-\beta\gamma}{\sqrt{\beta\gamma}e^{\frac{y_2-y_1}{2}}+(1+\beta\gamma)}.$$

Since $y_i \in J_{d_i}$ for $i = 1, 2$ and $\beta < 1 < \gamma$, we have

$$e^{y_2-y_1} \ge \beta^{d_2}\gamma^{d_1} \ge \beta^{d_L}\gamma^{d_R}.$$

Meanwhile, we deduce from Lemma A.4.1 that

$$\frac{\theta(d_L)}{18\beta^{d_L+1}} \le \lambda_2(d_L) < \lambda < \lambda_1(d_R) \le \frac{18\gamma^{d_R+1}}{\theta(d_R)},$$

which implies

$$\sqrt{\beta\gamma}e^{\frac{y_2-y_1}{2}} \ge \sqrt{\beta^{d_L+1}\gamma^{d_R+1}} \ge \frac{\sqrt{\theta(d_L)\theta(d_R)}}{18} \ge \frac{\theta(d_L)}{18}.$$

It follows that

$$\sqrt{|h(y_1)|\cdot|h(y_2)|} \le \frac{1-\beta\gamma}{\sqrt{\beta\gamma}e^{\frac{y_2-y_1}{2}}+(1+\beta\gamma)} \le \frac{1-\beta\gamma}{\frac{\theta(d_L)}{18}+(1+\beta\gamma)} \le \frac{36}{d_1+d_2+2}.$$

□

A.4.3 Proofs of Technical Lemmas

Proof of Lemma A.4.1. We first prove Item 1. For every $1 < d < \Delta$ we have

$$\lambda_c \le \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}} = \frac{\gamma^{d+1}}{d-1}\left(\frac{d}{d-1}\right)^d \le \frac{4\gamma^{d+1}}{d-1},$$

where the last inequality follows from that $\left(\frac{d}{d-1}\right)^d \le 4$ for all integer $d > 1$.

Now, we prove Item 2. For every $\overline{\Delta} \le d < \Delta$ we have

$$x_1(d) = \frac{2\gamma}{\theta(d) + \sqrt{\theta(d)^2 - 4\beta\gamma}} \le \frac{2\gamma}{\theta(d)}.$$

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 $\left(\frac{d+1}{d-1}\right)^d \le 9$ 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 $\left(\frac{d+1}{d-1}\right)^d \le 9$ for all integer $d > 1$. □

Proof of Lemma A.4.3. For convenience, define $f : [0, +\infty] \to [0, +\infty]$ by $f(x) = \frac{|1-\beta\gamma| \cdot x}{(\beta x+1)(x+\gamma)}$; note that $f(x) = |h(e^y)|$. Since e^y is monotone increasing, it suffices to show that f is increasing on $[0, \sqrt{\gamma/\beta}]$ and decreasing on $[\sqrt{\gamma/\beta}, +\infty]$. Towards this, we compute the derivative of f as

$$\begin{aligned} f'(x) &= |1-\beta\gamma| \cdot \left(\frac{1}{(\beta x+1)(x+\gamma)} - \frac{x(\beta(x+\gamma)+(\beta x+1))}{(\beta x+1)^2(x+\gamma)^2}\right) \\ &= \frac{|1-\beta\gamma|}{(\beta x+1)^2(x+\gamma)^2}\left((\beta x+1)(x+\gamma) - x(\beta(x+\gamma)+(\beta x+1))\right) \\ &= \frac{|1-\beta\gamma|}{(\beta x+1)^2(x+\gamma)^2} \cdot (\gamma - \beta x^2). \end{aligned}$$

Note that this is nonnegative on $[0, \sqrt{\gamma/\beta}]$ and nonpositive on $[\sqrt{\gamma/\beta}, +\infty]$, so we are done. □

A.5 Proofs for Ferromagnetic Cases

A.5.1 Proof of Theorem A.3.1

Proof of Theorem A.3.1. Throughout the proof, we use the trivial potential function $\Xi(y) = y$. Note that then, $\Xi'(y) = 1$ is a constant function. Now, we prove Contraction and Boundedness. We split into the three cases.

1. *Case 1:* $\frac{\Delta-2+\delta}{\Delta-\delta} \le \sqrt{\beta\gamma} \le \frac{\Delta-\delta}{\Delta-2+\delta}$, *and* $\lambda > 0$ *is arbitrary.* We first prove the Contraction part. By Lemma A.4.3, for all $y \in [-\infty, +\infty]$ we have

$$|h(y)| \le \frac{|1-\sqrt{\beta\gamma}|}{1+\sqrt{\beta\gamma}} \le \frac{1-\delta}{\Delta-1}.$$

Now let us prove the Boundedness condition. From the above inequality we have

$$|h(y)| \le \frac{1}{\Delta-1} \le \frac{1.5}{\Delta}$$

for $\Delta \ge 3$.

2. *Case 2:* $\sqrt{\beta\gamma} \ge \frac{\Delta}{\Delta-2}$ *and* $0 < \lambda \le (1-\delta)\frac{\gamma}{\max\{1,\beta^{\Delta-1}\}\cdot((\Delta-2)\beta\gamma-\Delta)}$. For the Contraction part, since $\log(\lambda\min\{1, 1/\gamma^{\Delta-1}\}) \le y_i \le \log(\lambda\max\{1,\beta^{\Delta-1}\})$, we have

$$\begin{aligned}\left|\frac{\partial H_d(\boldsymbol{y})}{\partial y_i}\right| = |h(y_i)| &= \frac{\beta\gamma-1}{1+\beta\gamma+\gamma e^{-y_i}+\beta e^{y_i}} \le \frac{\beta\gamma-1}{1+\beta\gamma+\gamma e^{-y_i}} \\ &\le \frac{\beta\gamma-1}{1+\beta\gamma+\frac{\gamma}{\lambda\max\{1,\beta^{\Delta-1}\}}}.\end{aligned}$$

Since we assumed $\lambda \le (1-\delta)\frac{\gamma}{\max\{1,\beta^{\Delta-1}\}\cdot((\Delta-2)\beta\gamma-\Delta)}$, 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. Note that 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 A.4.3, 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. *Case 3:* $\sqrt{\beta\gamma} \ge \frac{\Delta}{\Delta-2}$ *and* $\lambda \ge \frac{1}{1-\delta}\cdot\frac{(\Delta-2)\beta\gamma-\Delta}{\beta\cdot\min\{1,1/\gamma^{\Delta-1}\}}$. For the Contraction part, since $\log(\lambda\min\{1, 1/\gamma^{\Delta-1}\}) \le y_i \le \log(\lambda\max\{1,\beta^{\Delta-1}\})$, we have

$$\begin{aligned}\left|\frac{\partial H_d(\boldsymbol{y})}{\partial y_i}\right| = |h(y_i)| &= \frac{\beta\gamma-1}{1+\beta\gamma+\gamma e^{-y_i}+\beta e^{y_i}} \le \frac{\beta\gamma-1}{1+\beta\gamma+\beta e^{y_i}} \\ &\le \frac{\beta\gamma-1}{1+\beta\gamma+\beta\lambda\min\{1,1/\gamma^{\Delta-1}\}}.\end{aligned}$$

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 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

$$\lambda \geq \frac{(\Delta-2)\beta\gamma-\Delta}{\beta\min\{1,1/\gamma^{\Delta-1}\}},$$

it follows that $y \geq \log(\lambda\min\{1,1/\gamma^{\Delta-1}\}) \geq \log\left(\frac{(\Delta-2)\beta\gamma-\Delta}{\beta}\right)$. A simple calculation reveals that $\frac{(\Delta-2)\beta\gamma-\Delta}{\beta} \geq \sqrt{\frac{\gamma}{\beta}}$ and so by Lemma A.4.3, we have

$$\begin{aligned}|h(y)| &\leq \left|h\left(\log\left(\frac{(\Delta-2)\beta\gamma-\Delta}{\beta}\right)\right)\right| \leq (\beta\gamma-1)\frac{1}{\beta\cdot\frac{(\Delta-2)\beta\gamma-\Delta}{\beta}+1}\\ &= \frac{\beta\gamma-1}{(\Delta-2)(\beta\gamma-1)-1} \leq O(1/\Delta).\end{aligned}$$

□

A.5.2 Proof of Theorem A.3.2

In this subsection, we use results from [GL18] to prove Theorem A.3.2. Their potential function is implicitly defined by its derivative for the marginal ratios as

$$\Phi'(R) = \min\left\{\frac{\beta\gamma-1}{\alpha\gamma\log\frac{\lambda+\gamma}{\beta\lambda+1}}, \frac{1}{R\log\frac{\lambda}{R}}\right\}$$

for a constant $0 \leq \alpha \leq 1$ depending only on β,γ,λ (see [GL18] for a precise definition). In our context, the corresponding potential for the log ratios is

$$\Xi'(y) = e^y\Phi(e^y) = \min\left\{\frac{\beta\gamma-1}{\alpha\gamma\log\frac{\lambda+\gamma}{\beta\lambda+1}}e^y, \frac{1}{\log\frac{\lambda}{e^y}}\right\}$$

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 [GL18] is showing that the tree recursion F_d 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 $0 \leq R \leq \lambda$. [GL18] refers 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 50, but with a constant c that depends on γ,Δ. Indeed, worst case, we have

$$\max_{y_1,y_2}\frac{\Xi'(y_2)}{\Xi'(y_1)} \geq \frac{\Xi'(\log\lambda)}{\Xi'(\log(\lambda/\gamma^{\Delta-1}))} = \frac{\lambda\frac{\beta\gamma-1}{\alpha\gamma\log\frac{\lambda+\gamma}{\beta\lambda+1}}}{\frac{\beta\gamma-1}{\alpha\log\frac{\lambda+\gamma}{\beta\lambda+1}}\cdot\frac{\lambda}{\gamma^\Delta}} = \gamma^{\Delta-1}.$$

More precisely, we have the following result from [GL18], stated in terms of the log marginal ratios.

Theorem A.5.1. *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 A.1.1, this gives $O(n^C)$ mixing with a constant C depending only on $\beta,\gamma,\lambda,\Delta$. We note this is weaker than the correlation decay result in [GL18], since there, C does not depend on Δ, and hence is efficient for arbitrary graphs.

A.6 Easy Spectral Independence Bounds Based on Marginal Bounds

Fact A.6.1 (Antiferromagnetic Case)**.** *Fix an integer $\Delta \geq 3$ and real numbers β,γ,λ, and assume $0 \leq \beta \leq \gamma$, $\gamma > 0$, $\beta\gamma < 1$ and $\lambda > 0$. Then for every n-vertex graph G of maximum degree at most Δ, the antiferromagnetic 2-spin system on G with parameters (β,γ,λ) is Cn-spectrally independent, for a constant $0<C<1$ depending only on $\beta,\gamma,\lambda,\Delta$. Furthermore, if (β,γ,Δ) is up-to-Δ unique, then we can drop the dependence on Δ.*

Proof. If R_v denotes the marginal ratio of a vertex $v \in G$, then $R_v \geq \lambda\beta^\Delta$. In the case $\gamma \leq 1$, we have $R_v \leq \lambda/\gamma^\Delta$; however, if $\gamma > 1$, we have $R_v \leq \lambda$ where the equality holds for v isolated. It follows that we immediately have the bounds

$$|\Psi_G(u \to v)| \leq \begin{cases} \left|\frac{\lambda}{\lambda+\gamma^\Delta} - \frac{\lambda\beta^\Delta}{1+\lambda\beta^\Delta}\right| = \frac{\lambda(1-\beta^\Delta\gamma^\Delta)}{(\lambda+\gamma^\Delta)(1+\lambda\beta^\Delta)}, & \text{if } \gamma \leq 1 \\ \left|\frac{\lambda}{1+\lambda} - \frac{\lambda\beta^\Delta}{1+\lambda\beta^\Delta}\right| = \frac{\lambda(1-\beta^\Delta)}{(\lambda+1)(1+\lambda\beta^\Delta)}, & \text{otherwise} \end{cases}$$

for all $u, v \in G$. Note that these constants are less than 1, and only depend on $\beta, \gamma, \lambda, \Delta$, yielding the first claim.

Now, we proceed to remove the dependence on Δ when up-to-Δ uniqueness holds. We have the following cases:

1. If $\gamma > 1$, we immediately obtain a bound of $\frac{\lambda}{1+\lambda}$ which is independent of Δ.
2. If $\beta = 0$ and $\gamma \leq 1$, then $\frac{\lambda(1-\beta^\Delta\gamma^\Delta)}{(\lambda+\gamma^\Delta)(1+\lambda\beta^\Delta)} = \frac{\lambda}{\lambda+\gamma^\Delta} \leq \frac{\lambda}{\gamma^\Delta}$. Since (β, γ, λ) is up-to-Δ unique, we must have $\lambda \leq \lambda_c(\gamma, \Delta) = \min_{1<d<\Delta} \frac{\gamma^{d+1}d^d}{(d-1)^{d+1}} \leq \frac{\gamma^\Delta(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta} \leq \gamma^\Delta \cdot O(1/\Delta)$. It follows that $\frac{\lambda}{\gamma^\Delta} \leq O(1/\Delta)$.
3. If $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$ and $\gamma \leq 1$, then
$$\frac{\lambda(1-\beta^\Delta\gamma^\Delta)}{(\lambda+\gamma^\Delta)(1+\lambda\beta^\Delta)} \leq 1 - \beta^\Delta\gamma^\Delta \approx 1 - e^{-2}.$$
4. If $\sqrt{\beta\gamma} \leq \frac{\Delta-2}{\Delta}$, then let Δ_0 be the maximal $1 < d < \Delta$ such that $\sqrt{\beta\gamma} > \frac{d-2}{d}$. If $\lambda \leq \lambda_c(\beta, \gamma, \Delta)$, then by Lemma A.4.1, we have
$$\frac{\lambda(1-\beta^\Delta\gamma^\Delta)}{(\lambda+\gamma^\Delta)(1+\lambda\beta^\Delta)} \leq \frac{\lambda}{\gamma^\Delta} \leq O(\Delta_0/\Delta).$$
If $\lambda \geq \overline{\lambda}_c(\beta, \gamma, \Delta)$, then again by Lemma A.4.1, we have
$$\frac{\lambda(1-\beta^\Delta\gamma^\Delta)}{(\lambda+\gamma^\Delta)(1+\lambda\beta^\Delta)} \leq \frac{1}{\lambda\beta^\Delta} \leq O(\Delta_0/\Delta).$$

□

Fact A.6.2 (Ferromagnetic Case)**.** *Fix an integer $\Delta \geq 3$ and positive real numbers β, γ, λ, and assume $\beta \leq \gamma$ and $\beta\gamma > 1$. Then for every n-vertex graph G of maximum degree at most Δ, the ferromagnetic 2-spin system on G with parameters (β, γ, λ) is Cn-spectrally independent, for a constant $0 < C < 1$ depending only on $\beta, \gamma, \lambda, \Delta$.*

Proof. The proof is identical to the antiferromagnetic case (see Fact A.6.1) and we omit it here. □

A.7 Additional Technical Lemmas

The following lemma implies that the potential Ξ given by Eq. (A.6) is well-defined.

Lemma A.7.1. *For all $\beta, \gamma > 0$ such that $\beta\gamma < 1$, we have*

$$\int_{-\infty}^{+\infty} \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}}\,\mathrm{d}y < +\infty.$$

Proof. For the $+\infty$ side we have

$$\begin{aligned}\int_0^{+\infty} \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}}\,\mathrm{d}y &= \int_0^{+\infty} \sqrt{\frac{1-\beta\gamma}{\beta e^y+\gamma e^{-y}+\beta\gamma+1}}\,\mathrm{d}y \\ &< \int_0^{+\infty} \frac{1}{\sqrt{\beta e^y}}\,\mathrm{d}y < +\infty.\end{aligned}$$

Similarly, for the $-\infty$ side we have

$$\int_{-\infty}^0 \sqrt{\frac{(1-\beta\gamma)e^y}{(\beta e^y+1)(e^y+\gamma)}}\,\mathrm{d}y < \int_{-\infty}^0 \frac{1}{\sqrt{\gamma e^{-y}}}\,\mathrm{d}y < +\infty.$$

□

A.7.1 Uniqueness Gaps in Terms of Parameter Gaps

In this subsection, we relate the parameter gaps with the uniqueness gaps.

Claim A.7.2 (Hardcore Constraint Model; Generalization of Lemma 7.1.4)**.** *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 $\delta/4$.*

Proof. The proof is the same as the proof of Lemma 7.1.4. □

Claim A.7.3 (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

$$\begin{aligned} f_d'(R) &= d\lambda\left(\frac{\beta R+1}{R+\gamma}\right)^{d-1} \cdot \left(\frac{\beta}{R+\gamma} - \frac{\beta R+1}{(R+\gamma)^2}\right) \\ &= -d(1-\beta\gamma)\lambda\left(\frac{\beta R+1}{R+\gamma}\right)^{d} \cdot \frac{1}{(\beta R+1)(R+\gamma)} \\ &= -d(1-\beta\gamma) \cdot \frac{f_d(R)}{(\beta R+1)(R+\gamma)}. \end{aligned}$$

At the unique fixed point $\hat{R}_d$, we have $f_d(\hat{R}_d) = \hat{R}_d$ so

$$\left|f_d'(\hat{R}_d)\right| = d(1-\beta\gamma)\frac{\hat{R}_d}{(\beta\hat{R}_d+1)(\hat{R}_d+\gamma)}.$$

By Lemma A.4.3, we have the upper bound

$$\left|f_d'(\hat{R}_d)\right| \leq d \cdot \frac{1-\beta\gamma}{(1+\sqrt{\beta\gamma})^2} = d \cdot \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}}.$$

Since we assumed $\sqrt{\beta\gamma} \geq \frac{\Delta-2(1-\delta)}{\Delta}$, we obtain

$$d \cdot \frac{1-\sqrt{\beta\gamma}}{1+\sqrt{\beta\gamma}} \leq d \cdot \frac{\Delta - (\Delta - 2(1-\delta))}{\Delta + (\Delta - 2(1-\delta))} = d \cdot \frac{1-\delta}{\Delta - 1 + \delta} \leq (1-\delta)\frac{d}{\Delta - 1}.$$

As this is at most $1 - \delta$ for all $d < \Delta$, we have up-to-Δ uniqueness with gap δ. □

Appendix B

The Monomer-Dimer Model via Correlation Decay

In this chapter, we use the correlation decay methods developed in Chapter 7 and Appendix A to establish spectral independence for the *monomer-dimer model* over all matchings of the input graph G. We already studied the monomer-dimer model in Chapter 6 through the lens of zero-freeness in the much more general setting of Holant problems. There we also established spectral independence, although the bounds were less explicit. In this chapter, we will be able to give very natural upper bounds on the spectral independence of the monomer-dimer model, which are likely the best possible using any method which upper bounds the ℓ_∞-norm of the influence matrix; see Theorem E.2.1. The methods in this chapter were also the first way spectral independence was established for this model.

Let us first recall what is the monomer-dimer model. Given a graph $G = (V, E)$, recall that a *matching* is a subset $M \subseteq E$ of edges such that every vertex $v \in V$ is incident to at most one edge in M. For $\lambda \geq 0$, the Gibbs distribution $\mathrm{dist} = \mu_{G,\lambda}$ of the monomer-dimer model on G with fugacity λ is defined as

$$\mu(M) \propto \lambda^{|M|}, \quad \forall \text{ matchings } M \subseteq E.$$

Subsets of edges which aren't matchings have probability zero. The corresponding univariate partition function

$$\mathcal{M}_G(\lambda) \stackrel{\mathsf{def}}{=} \sum_{M \subseteq E \text{ matching}} \lambda^{|M|},$$

which is also known as the *(edge) matching polynomial*. The matching constraint is a hard constraint, and indeed, it is easy to see that the monomer-dimer model on G is simply the hardcore gas model on the line graph $L(G)$ of G. Hence, may also be viewed as a two-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.

The results of this chapter are based on [CLV21a]. We prove the following.

Theorem B.0.1. *For all fixed $\Delta \geq 3$ and all $\lambda > 0$, there exists $C = C(\Delta, \lambda)$ such that for every n-vertex graph $G = (V, E)$ of maximum degree at most Δ, the mixing time of the Glauber dynamics for the monomer-dimer model on G with fugacity λ is at most $Cn \log n$.*

Theorem B.0.2 (Spectral Independence for Monomer-Dimer)**.** *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 $m = |E|$, every $\Lambda \subseteq E_G$, and every feasible boundary condition $\sigma_\Lambda : \Lambda \to \{0, 1\}$ on $\Lambda \subseteq V$ with $|\Lambda| = k$, the Gibbs distribution $\mu = \mu_{G,\lambda}$ of the monomer-dimer model on G with fugacity λ is $(\eta_0, \ldots, \eta_{m-2})$-spectrally independent with*

$$\eta_k \leq \min\left\{2\lambda\Delta, \frac{\lambda\Delta}{1+\lambda\Delta} \cdot \left(\sqrt{1+\lambda\Delta}+1\right) + \left(\sqrt{1+\lambda\Delta}-1\right), \frac{\lambda}{1+\lambda}(m-k-1)\right\}.$$

Remark 59. We note that [Ali+21] independently gave an $O(1)$-spectral independence bound for *vertex-to-vertex influences* using Hurwitz stability of the *vertex matching polynomial*, which is incomparable to our result here.

Remark 60. For all λ, Δ, the bound $\frac{\lambda\Delta}{1+\lambda\Delta} \cdot \left(\sqrt{1+\lambda\Delta}+1\right) + \left(\sqrt{1+\lambda\Delta}-1\right)$ is strictly better than the bound $2\sqrt{1+\lambda\Delta}$ originally proved in [CLV21a]. We further fix a bug which was discovered in the original argument of [CLV21a].

Our proof follows the strategy developed in Chapter 7 and Appendix A, which uses correlation decay. This was first done in [ALO21] and subsequently in [CLV20; Che+21d; Fen+21] for other model. We prove Theorem B.0.2 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, much like Theorem 7.2.2 and Corollary 7.2.3. To do this, we extend known divisibility results on the univariate matching polynomial [God93], following the proof of Theorem 7.2.2, although the analysis in this case is simpler. 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 and the potential method.

We formalize these in the following two intermediate theorems. Throughout this chapter, rather than subscripting our influence matrix Ψ with the distribution μ, we subscript them with the underlying graph G to emphasize which graph is being considered.

Theorem B.0.3 (Reduction from Graphs to Trees)**.** *Fix a graph $G = (V_G, E_G)$, an arbitrary "root" vertex $r \in V_G$, an edge $e \sim r$ incident to r, and $\lambda > 0$. Then there exists a tree $T = \mathrm{T}_{\mathsf{SAW}}(G, r) = (V_T, E_T)$ 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 B.0.4 (Total Edge Influence in Trees)**.** *Let $T = (V_T, E_T)$ be any tree of maximum degree $\leq \Delta$, and fix $\lambda > 0$. Then we have the bound*

$$\sum_{f \in E_T : f \neq e} |\Psi_T(e \to f)| \leq \min\left\{2\lambda\Delta, \frac{\lambda\Delta}{1+\lambda\Delta} \cdot \left(\sqrt{1+\lambda\Delta}+1\right) + \left(\sqrt{1+\lambda\Delta}-1\right)\right\}, \quad \forall e \in E_T.$$

Remark 61. It is interesting to note that Theorem E.2.1 predicts this upper bound is tight for the ℓ_∞-norm $\|\Psi_\mu\|_\infty$ by considering the influence matrix for the distribution on the infinite Δ-regular tree. However, it also predicts that $\boldsymbol{\lambda}_{\max}(\Psi_\mu)$ can be significantly smaller.

Assuming the truth of these two theorems, we now give a straightforward proof of Theorem B.0.2.

Proof of Theorem B.0.2. Fix G, $\Lambda \subseteq E_G$ and $\sigma_\Lambda : \Lambda \to \{0, 1\}$. Let $H = (V_H, E_H)$ be the graph obtained from G by deleting all edges $e \in \Lambda$ such that $\sigma_\Lambda(e) = 1$, and deleting all edges $f \in \Lambda$ along with edges incident to them such that $\sigma_\Lambda(f) = 0$. Observe that H is a subgraph of G with maximum degree at most Δ, and crucially, the conditional distribution $\mu_{G,\lambda}^{\sigma_\Lambda}$ is precisely $\mu_{H,\lambda}$. By Theorems B.0.3 and B.0.4, we have the bound

$$\boldsymbol{\lambda}_{\max}(\Psi_G^{\sigma_\Lambda}) = \boldsymbol{\lambda}_{\max}(\Psi_H) \leq \|\Psi_H\|_\infty \leq \min\left\{2\lambda\Delta, \frac{\lambda\Delta}{1+\lambda\Delta} \cdot \left(\sqrt{1+\lambda\Delta}+1\right) + \left(\sqrt{1+\lambda\Delta}-1\right)\right\}.$$

The upper bound of $\frac{\lambda}{1+\lambda}(m - |\Lambda| - 1)$ follows simply from Fact 7.1.5. As $G, \Lambda, \sigma_\Lambda$ were arbitrary, the claim follows. □

All that remains is to prove Theorems B.0.3 and B.0.4. We do this in Appendices B.0.1 and B.0.2, respectively, noting that the arguments are completely independent of one another.

B.0.1 Reducing Influences in Graphs to Influences in Trees: Proof of Theorem B.0.3

Fix a graph $G = (V_G, E_G)$ with maximum degree $\leq \Delta$, and a vertex $r \in V_G$. Let $T = \mathrm{T}_{\mathsf{SAW}}(G, r) = (V_T, E_T)$ denote the self-avoiding walk tree in G rooted at r; in the context of matchings, this is also known as the *path tree*, and we refer to [God93] for formal definitions. Note that we do not impose any boundary conditions on $\mathrm{T}_{\mathsf{SAW}}(G, r)$ like in Weitz's self-avoiding walk tree [Wei06] (see Definition 35). For every vertex $u \in V_G$, we write $\mathcal{C}(u)$ to be the set of copies of u in T; we often write such copies as $\hat{u} \in \mathcal{C}(u)$. Similarly, for every edge $e \in E_G$, we write $\mathcal{C}(e)$ to be the set of

copies of e in T; we often write such copies as $\hat{e} \in \mathcal{C}(e)$. If $\boldsymbol{\lambda} = \boldsymbol{\lambda}_G : E_G \to \mathbb{R}_{\geq 0}$ is a collection of edge activities associated to each edge (so that $\mu_{G,\boldsymbol{\lambda}}(M) \propto \prod_{e \in M} \lambda_e$), then we associate activities $\boldsymbol{\lambda}_T : E_T \to \mathbb{R}_{\geq 0}$ where $\boldsymbol{\lambda}_T(\hat{e}) = \boldsymbol{\lambda}_G(e)$ for every $e \in G$ and $\hat{e} \in \mathcal{C}(e) \subseteq E_T$.

We prove the following more fine-grained relationship between pairwise influences in G and pairwise influences in $T = \mathrm{T}_{\mathsf{SAW}}(G, r)$.

Proposition B.0.5 (Influence in G to Influence in $\mathrm{T}_{\mathsf{SAW}}(G,r)$)**.** *For every graph $G = (V_G, E_G)$, $r \in V_G$, $e \sim r$, $f \in E_G$, and edge activities $\boldsymbol{\lambda} : E_G \to \mathbb{R}_{\geq 0}$, if we let $T = \mathrm{T}_{\mathsf{SAW}}(G, r) = (V_T, E_T)$, then we have the identity*

$$\Psi_G(e \to f) = \sum_{\hat{f} \in \mathcal{C}(f)} \Psi_T(\hat{e} \to \hat{f}).$$

Note that there is only one copy $\hat{e}$ of e in T.

This is a direct analog of Corollary 7.2.3. We note that by the Triangle Inequality, Proposition B.0.5 immediately implies Theorem B.0.3. Hence, it suffices to prove Proposition B.0.5, which we do by generalizing properties of the univariate matching polynomial.

Define the following multivariate (edge) matching polynomial

$$\mathcal{M}_G(\lambda_e : e \in E_T) = \sum_{M \subseteq E \text{ matching}} \prod_{e \in M} \lambda_e.$$

$\mathcal{M}_G$ is also the partition function of the monomer-dimer model on G with edge activities $\lambda_e \geq 0$. Furthermore, if $r \in V_G$ is arbitrary, and we denote $T = \mathrm{T}_{\mathsf{SAW}}(G, r)$, then define

$$\widehat{\mathcal{M}}_T(\lambda_e : e \in E_G) \stackrel{\mathsf{def}}{=} \mathcal{M}_T(\widehat{\lambda}_{\hat{e}} : \hat{e} \in E_T)$$

where $\widehat{\lambda}_{\hat{e}} = \lambda_e$ for all $\hat{e} \in \mathcal{C}(e)$ and all $e \in E_G$. We note that while $\mathcal{M}_G$ is always multiaffine, $\widehat{\mathcal{M}}_T$ 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 $\widehat{\mathcal{M}}_T$ since no self-avoiding walk can reuse e after using e to leave r. In particular, $|\mathcal{C}(e)| = 1$. We will crucially need the following decomposition of $\mathcal{M}_G$.

Lemma B.0.6. *For every graph $G = (V_G, E_G)$ and any vertex $v \in V_G$, we have the identity*

$$\mathcal{M}_G(\boldsymbol{\lambda}) = \mathcal{M}_{G-r}(\boldsymbol{\lambda}) + \sum_{v \sim r} \lambda_{rv} \mathcal{M}_{G-r-v}(\boldsymbol{\lambda}).$$

Proof. Group the matchings for which r is not saturated in the term $\mathcal{M}_{G-r}(\boldsymbol{\lambda})$. Similarly, group the matchings for which a fixed edge $e = \{r, v\}$ incident to r is selected in $\mathcal{M}_{G-r-v}(\boldsymbol{\lambda})$. □

We prove the following, a univariate analog of which was already proved in [God93].

Lemma B.0.7. *For every graph $G = (V_G, E_G)$ and $r \in V_G$, taking $T = \mathrm{T}_{\mathsf{SAW}}(G, r)$, we have the identity*

$$\frac{\mathcal{M}_G(\boldsymbol{\lambda})}{\mathcal{M}_{G-r}(\boldsymbol{\lambda})} = \frac{\widehat{\mathcal{M}}_T(\boldsymbol{\lambda})}{\widehat{\mathcal{M}}_{T-r}(\boldsymbol{\lambda})}.$$

Furthermore, we may write $\widehat{\mathcal{M}}_T(\boldsymbol{\lambda}) = \mathcal{M}_G(\boldsymbol{\lambda}) \cdot q(\boldsymbol{\lambda})$ for some polynomial q which does not depend on λ_e for any $e \sim r$.

This is the analog of Theorem 7.2.2. First, let us see how to use Lemma B.0.7 to prove Proposition B.0.5.

Proof of Proposition B.0.5. The proof is almost identical to the proof of Corollary 7.2.3. Fix $r \in V_G$, and write $T = \mathrm{T}_{\mathsf{SAW}}(G, r)$. By Lemma B.0.7, we have that $\widehat{\mathcal{M}}_T(\boldsymbol{\lambda}) = \mathcal{M}_G(\boldsymbol{\lambda}) \cdot q(\boldsymbol{\lambda})$ for a polynomial q which does not depend on λ_e for all $e \sim r$. It follows that if $e \sim r$, then

$$\Pr_{\hat{M} \sim \mu_T}[\hat{e} \in \hat{M}] = (\lambda_e \partial_e \log \widehat{\mathcal{M}}_T)(\boldsymbol{\lambda}) = (\lambda_e \partial_e \log \mathcal{M}_G)(\boldsymbol{\lambda}) = \Pr_{M \sim \mu_G}[e \in M]. \tag{B.1}$$

It then also follows that for any $e \sim r$ and any edge $f \in E_G$, we have the identity

$$(\lambda_f \lambda_e \cdot \partial_f \partial_e \log \widehat{\mathcal{M}}_T)(\boldsymbol{\lambda}) = (\lambda_f \lambda_e \cdot \partial_f \partial_e \log \mathcal{M}_G)(\boldsymbol{\lambda}).$$

Now, let us understand the left-hand and right-hand sides separately as influences. For the right-hand side, we have that

$$\begin{aligned}
&(\lambda_f \lambda_e \cdot \partial_f \partial_e \log \mathcal{M}_G)(\boldsymbol{\lambda}) \\
&= \lambda_e \lambda_f \cdot \partial_f \frac{(\partial_e \mathcal{M}_G)(\boldsymbol{\lambda})}{\mathcal{M}_G(\boldsymbol{\lambda})} \\
&= \lambda_f \lambda_e \cdot \left(\frac{(\partial_f \partial_e \mathcal{M}_G)(\boldsymbol{\lambda})}{\mathcal{M}_G(\boldsymbol{\lambda})} - \frac{(\partial_f \mathcal{M}_G)(\boldsymbol{\lambda}) \cdot (\partial_e \mathcal{M}_G)(f)}{\mathcal{M}_G(\boldsymbol{\lambda})^2} \right) \\
&= \Pr_{M \sim \mu_G}[e, f \in M] - \Pr_{M \sim \mu_G}[e \in M] \cdot \Pr_{M \sim \mu_G}[f \in M] \\
&= \Pr_{M \sim \mu_G}[e \in M] \cdot \left(\Pr_{M \sim \mu_G}[f \in M \mid e \in M] - \Pr_{M \sim \mu_G}[f \in M] \right) \\
&= \Pr_{M \sim \mu_G}[e \in M] \cdot \Pr_{M \sim \mu_G}[e \notin M] \cdot \Psi_G(e \to f).
\end{aligned}$$

For the left-hand side, we have by the Chain Rule that

$$\begin{aligned}
(\lambda_f \lambda_e \cdot \partial_f \partial_e \log \widehat{\mathcal{M}}_T)(\boldsymbol{\lambda}) &= \lambda_f \lambda_e \cdot \partial_f \frac{(\partial_e \widehat{\mathcal{M}}_T)(\boldsymbol{\lambda})}{\widehat{\mathcal{M}}_T(\boldsymbol{\lambda})} \\
&= \sum_{\hat{f} \in \mathcal{C}(f)} \lambda_f \lambda_e \cdot \partial_{\hat{f}} \frac{(\partial_e \mathcal{M}_T)(\hat{\boldsymbol{\lambda}})}{\mathcal{M}_T(\hat{\boldsymbol{\lambda}})} \Bigg|_{\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}} \\
&= \sum_{\hat{f} \in \mathcal{C}(f)} \Pr_{\hat{M} \sim \mu_T}[\hat{e} \in \hat{M}] \cdot \Pr_{\hat{M} \sim \mu_T}[\hat{e} \notin \hat{M}] \cdot \Psi_T(\hat{e} \to \hat{f}) \\
&= \Pr_{M \sim \mu_G}[e \in M] \cdot \Pr_{M \sim \mu_G}[e \notin M] \sum_{\hat{f} \in \mathcal{C}(f)} \Psi_T(\hat{e} \to \hat{f}) \qquad \text{(Eq. (B.1))} \\
&= \Pr_{M \sim \mu_G}[e \in M] \cdot \Pr_{M \sim \mu_G}[e \notin M] \cdot \Psi_G(e \to f).
\end{aligned}$$

The claim immediately follows. □

All that remains is to prove Lemma B.0.7.

Proof of Lemma B.0.7. 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 $\mathcal{M}_G = \widehat{\mathcal{M}}_T$ (i.e. q is identically 1). This forms our base case. Now, by Lemma B.0.6 we may write

$$\mathcal{M}_G(\boldsymbol{\lambda}) = \mathcal{M}_{G-r}(\boldsymbol{\lambda}) + \sum_{v \sim r} \lambda_{rv} \mathcal{M}_{G-r-v}(\boldsymbol{\lambda})$$

where we note the polynomials $\mathcal{M}_{G-r}(\boldsymbol{\lambda}), \mathcal{M}_{G-r-v}(\boldsymbol{\lambda})$ do not depend on any λ_e for $e \sim r$. Therefore, we deduce that

$$\begin{aligned}
\frac{\mathcal{M}_G(\boldsymbol{\lambda})}{\mathcal{M}_{G-r}(\boldsymbol{\lambda})} &= 1 + \sum_{v \sim r} \lambda_{rv} \cdot \frac{\mathcal{M}_{G-r-v}(\boldsymbol{\lambda})}{\mathcal{M}_{G-r}(\boldsymbol{\lambda})} \\
&= 1 + \sum_{v \sim r} \lambda_{rv} \cdot \frac{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G-r,v)-v}(\boldsymbol{\lambda})}{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G-r,v)}(\boldsymbol{\lambda})} \qquad \text{(Induction)} \\
&= 1 + \sum_{v \sim r} \lambda_{rv} \cdot \frac{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G,r)-r-v}(\boldsymbol{\lambda})}{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G,r)-r}(\boldsymbol{\lambda})} \\
&\qquad (\mathsf{T}_{\mathsf{SAW}}(G-r,v) \cong \text{subtree of } \mathsf{T}_{\mathsf{SAW}}(G,r) \text{ rooted at } v) \\
&= \frac{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G,r)-r}(\boldsymbol{\lambda}) + \sum_{v \sim r} \lambda_{rv} \widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G,r)-r-v}(\boldsymbol{\lambda})}{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G,r)-r}(\boldsymbol{\lambda})} \\
&= \frac{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G,r)}(\boldsymbol{\lambda})}{\widehat{\mathcal{M}}_{\mathsf{T}_{\mathsf{SAW}}(G,r)-r}(\boldsymbol{\lambda})}.
\end{aligned}$$

Note that in the final step, we employ Lemma B.0.6. This is permitted because one can convert $\widehat{\mathcal{M}}_T$ to $\mathcal{M}_T$, use Lemma B.0.6 and the fact that $\hat{\lambda}_{rv} = \lambda_{rv}$ for $v \sim r$, and convert back. 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

$$\begin{aligned}\widehat{\mathcal{M}}_{\mathrm{T}_{\mathsf{SAW}}(G,r)}(\boldsymbol{\lambda}) &= \mathcal{M}_G(\boldsymbol{\lambda}) \cdot \frac{\widehat{\mathcal{M}}_{\mathrm{T}_{\mathsf{SAW}}(G,r)-r}(\boldsymbol{\lambda})}{\mathcal{M}_{G-r}(\boldsymbol{\lambda})} \\ &= \mathcal{M}_G(\boldsymbol{\lambda}) \cdot \frac{\prod_{v\sim u} \widehat{\mathcal{M}}_{\mathrm{T}_{\mathsf{SAW}}(G-r,v)}(\boldsymbol{\lambda})}{\mathcal{M}_{G-r}(\boldsymbol{\lambda})} \\ &= \mathcal{M}_G(\boldsymbol{\lambda}) \cdot q(\boldsymbol{\lambda}),\end{aligned}$$

where in the second step we use that deleting r from $\mathrm{T}_{\mathsf{SAW}}(G, r)$ disconnects the subtrees $\mathrm{T}_{\mathsf{SAW}}(G - r, v)$, and in the final step we use that $\mathcal{M}_{G-r}(\boldsymbol{\lambda})$ divides each $\widehat{\mathcal{M}}_{\mathrm{T}_{\mathsf{SAW}}(G-r,v)}(\boldsymbol{\lambda})$ by the induction hypothesis. This shows the lemma. □

B.0.2 The Total Influence in a Tree: Proof of Theorem B.0.4

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 marginal probability that the vertex $r \in V_T$ is saturated in a random matching drawn from the Gibbs distribution, and let $\mu_T(\overline{r}) \stackrel{\mathsf{def}}{=} 1 - \mu_T(r)$. Similarly, for an edge $e \in E_T$, we write $\mu_T(e)$ for the marginal probability that e is in a random matching drawn from the Gibbs distribution, and let $\mu_T(\overline{e}) \stackrel{\mathsf{def}}{=} 1 - \mu_T(e)$. Note that since at most one edge incident to r may be in any given matching, we have

$$\mu_T(r) = \sum_{e \sim r} \mu_T(e). \tag{B.2}$$

We now prove several intermediate technical results which we will use to deduce Theorem B.0.4. 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 $\lambda \geq 0$:

$$\mu_G(\overline{r}) = F_\lambda(\mu_{G-r}(\overline{v}) : v \in L_r(1)) \quad \text{where} \quad F_\lambda(\boldsymbol{p}) = F_{d,\lambda}(\boldsymbol{p}) \stackrel{\mathsf{def}}{=} \frac{1}{1 + \lambda \sum_{i=1}^d p_i}. \tag{B.3}$$

We note this is immediate from Lemma B.0.6 and using that $\mu_G(\overline{r}) = \frac{\mathcal{M}_{G-r}(\boldsymbol{\lambda})}{\mathcal{M}_G(\boldsymbol{\lambda})}$ where we take $\boldsymbol{\lambda} = \lambda \mathbf{1}$.

Proposition B.0.8. *Consider the potential function $\Phi(x) \stackrel{\mathsf{def}}{=} \log x$. Then for every tree $T = (V_T, E_T)$, every edge $e = \{r_1, r_2\} \in E_T$, every positive integer $k \geq 1$, and every $r \in e = \{r_1, r_2\}$, we have the inequality*

$$\begin{aligned}&\sum_{f \in E_{T(r)} : \mathrm{dist}(e,f) = k} |\Psi_T(e \to f)| \\ &\leq \min\left\{ \left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k, \left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^{k \bmod 2} \cdot \left(\sup_{\boldsymbol{y}} \left\| \nabla(\Phi \circ F_\lambda^{\circ 2} \circ \Phi^{-1})(\boldsymbol{y}) \right\|_1 \right)^{\lfloor k/2 \rfloor} \right\}.\end{aligned}$$

Given Proposition B.0.8, we will also need a bound on the gradient norm. Conveniently, this gradient norm was already analyzed in [Bay+07] to establish the correlation decay property.

Lemma B.0.9 ([Bay+07, 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:*

$$\sup_{\boldsymbol{y}} \left\| \nabla(\Phi \circ F_\lambda^{\circ 2} \circ \Phi^{-1})(\boldsymbol{y}) \right\|_1 \leq 1 - \frac{2}{\sqrt{1+\lambda\Delta} + 1}.$$

We now show how to use intermediate technical results to prove Theorem B.0.4. Immediately following, we will prove Propositions B.0.8 and B.0.12.

Proof of Theorem B.0.4. By combining Propositions B.0.8 and B.0.12 and Lemma B.0.9, we have the following two bounds

$$\begin{aligned}
\sum_{f\in E_T: f\neq e} |\Psi_T(e\to f)| &= \left(\sum_{f\in E_{T(r_1)}} + \sum_{f\in E_{T(r_2)}}\right)|\Psi_T(e\to f)| \\
&\leq 2\sum_{k=1}^{\infty}\left(1-\frac{2}{\sqrt{1+\lambda\Delta}+1}\right)^{\lfloor k/2\rfloor}\cdot\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^{k \bmod 2} \\
&= 2\frac{\lambda\Delta}{1+\lambda\Delta}\sum_{k=0}^{\infty}\left(1-\frac{2}{\sqrt{1+\lambda\Delta}+1}\right)^{k} + 2\sum_{k=1}^{\infty}\left(1-\frac{2}{\sqrt{1+\lambda\Delta}+1}\right)^{k} \\
&= \frac{\lambda\Delta}{1+\lambda\Delta}\cdot\left(\sqrt{1+\lambda\Delta}+1\right) + \left(\sqrt{1+\lambda\Delta}-1\right) \\
\sum_{f\in E_T: f\neq e} |\Psi_T(e\to f)| &\leq 2\sum_{k=1}^{\infty}\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^{k} = 2\cdot\frac{\lambda\Delta}{1+\lambda\Delta}\cdot\frac{1}{1-\frac{\lambda\Delta}{1+\lambda\Delta}} = 2\lambda\Delta.
\end{aligned}$$

This proves the theorem. □

Proof of Proposition B.0.8

The crux of the proof rests on the following two lemmas. The first is a convenient factorization of the pairwise influence in trees, which was already observed in prior work [ALO21; CLV20] in the context of vertex-spin systems. It is the analog of Lemma 7.3.2.

Lemma B.0.10 (Factorization of Pairwise Influence in Trees)**.** *Fix two edges $e, f \in E_T$. Let $e = e_1, r = u_1, e_2, u_2, \ldots, u_k, e_{k+1} = f$ 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 B.0.10. 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. □

Lemma B.0.11. *Let T be a tree rooted at r and let $v \in L_r(1)$. Then we have the identity*

$$\lambda\cdot\mu_T(\overline{r})\cdot\mu_{T(v)}(\overline{v}) = \mu_T(\{r,v\}).$$

Proof. More generally, if G is any graph (not necessarily a tree), and $e = \{r, v\}$, then

$$\mu_G(\{r,v\}) = \lambda\cdot\mu_G(\overline{r}\wedge\overline{v}) = \lambda\cdot\mu_G(\overline{r})\cdot\mu_{G-r}(\overline{v}),$$

because any matching in G containing $\{r, v\}$ is precisely a matching in $G - r - v$ adjoined with the edge $\{r, v\}$ (which introduces the multiplicative factor of λ in the right-hand side). If G is a tree T, then $T - r$ is the union of subtrees which are disconnected from each other. Hence, $\mu_{T-r}(\overline{v}) = \mu_{T(v)}(\overline{v})$. □

We are now ready to prove Proposition B.0.8. First, observe that if e, f are neighboring edges in T, then $\Psi_T(e\to f) = -\mu_{T-e}(f)$ due to the hard constraints of the monomer-dimer model.

Hence,

$$\begin{aligned}\sum_{f\in E_{T(r)}:\mathrm{dist}(e,f)=k}|\Psi_T(e\to f)| &= \sum_{f\in E_{T(r)}:\mu(e,f)=k}\prod_{i=1}^{k}|\Psi_T(e_i\to e_{i+1})| && \text{(Lemma B.0.10)}\\ &= \sum_{f\in E_{T(r)}:\mu(e,f)=k}\prod_{i=1}^{k}\mu_{T-e_i}(e_{i+1})\\ &= \sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k]}\prod_{i=1}^{k}\mu_{T(u_i)}(\{u_i,u_{i+1}\})\\ &= \sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k]}\prod_{i=1}^{k}\lambda\cdot\mu_{T(u_i)}(\overline{u_i})\cdot\mu_{T(u_{i+1})}(\overline{u_{i+1}}),\\ & && \text{(Lemma B.0.11)}\end{aligned}$$

where again we write $e=e_1, r=u_1, e_2, u_2, \ldots, u_k, e_{k+1}=f$ for the unique path in T from e to f, for each f.

Note that for the first bound $\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k$, one could have stopped at the penultimate step. To see this, observe that by directly applying the tree recursion F_λ from Eq. (B.3),

$$\mu_T(\overline{r})=\frac{1}{1+\lambda\sum_{v\in L_r(1)}\mu_{T(v)}(\overline{v})}\geq\frac{1}{1+\lambda\Delta}$$

since $\mu_{T(v)}(\overline{v})\leq 1$ trivially for all $v\in L_r(1)$. Hence $\mu_T(r)\leq\frac{\lambda\Delta}{1+\lambda\Delta}$. This upper bound applies to general rooted trees, in particular, the root u_i of $T(u_i)$ for each $i=1,\ldots,k$. It follows that

$$\begin{aligned}&\sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k]}\prod_{i=1}^{k}\mu_{T(u_i)}(\{u_i,u_{i+1}\})\\ &=\sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k-1]}\prod_{i=1}^{k-1}\mu_{T(u_i)}(\{u_i,u_{i+1}\})\underbrace{\sum_{u_{k+1}\in L_{u_k}(1)}\mu_{T(u_k)}(\{u_k,u_{k+1}\})}_{=\mu_{T(u_k)}(u_k)}\\ &\leq\frac{\lambda\Delta}{1+\lambda\Delta}\sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k-1]}\prod_{i=1}^{k-1}\mu_{T(u_i)}(\{u_i,u_{i+1}\})\\ &\leq\cdots && \text{(Induction)}\\ &\leq\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k.\end{aligned}$$

It turns out, one may view this simple analysis as a "one-step" analysis, in the sense that we only applied the tree recursion F_λ one step at a time. We will establish the second upper bound via a "two-step analysis".

We have shown the identity above

$$\sum_{f\in E_{T(r)}:\mathrm{dist}(e,f)=k}|\Psi_T(e\to f)|=\sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k]}\prod_{i=1}^{k}\lambda\cdot\mu_{T(u_i)}(\overline{u_i})\cdot\mu_{T(u_{i+1})}(\overline{u_{i+1}}). \tag{B.4}$$

Let us now reinterpret the terms $\lambda\cdot\mu_{T(u_i)}(\overline{u_i})\cdot\mu_{T(u_{i+1})}(\overline{u_{i+1}})$ appearing in the right-hand side of Eq. (B.4) as *derivatives* of the tree recursion F_λ after composing with our potential Φ. Composing with the recursion F_λ yields the following recursion for the logarithm of the marginals $y_i=\log p_i=\Phi(p_i)$:

$$(\Phi\circ F_\lambda\circ\Phi^{-1})(\boldsymbol{y})=-\log\left(1+\lambda\sum_{i=1}^{d}\exp(y_i)\right).$$

Differentiating and applying the Inverse Function Theorem, we obtain the the identities

$$\begin{aligned}(\partial_{p_i} F_\lambda)(\boldsymbol{p}) &= -\frac{\lambda}{\left(1+\lambda\sum_{j=1}^d p_j\right)^2} = -\lambda F_\lambda(\boldsymbol{p})^2;\\ \partial_{y_i}(\Phi\circ F_\lambda\circ\Phi^{-1})(\boldsymbol{y}) &= \frac{(\Phi'\circ F_\lambda)(\boldsymbol{p})}{\Phi'(p_i)}\cdot(\partial_{p_i}F_\lambda)(\boldsymbol{p}) = -\frac{p_i}{F_\lambda(\boldsymbol{p})}\cdot\lambda F_\lambda(\boldsymbol{p})^2\\ &= -\lambda\cdot p_i\cdot F_\lambda(\boldsymbol{p}).\end{aligned}$$

Since for any $\boldsymbol{y}=\log\boldsymbol{p}$, we have

$$\begin{aligned}\left\|\nabla(\Phi\circ F_\lambda\circ\Phi^{-1})(\boldsymbol{y})\right\|_1 &= F_\lambda(\boldsymbol{p})\cdot\lambda\sum_{i=1}^d p_i\\ &= \frac{\lambda\sum_{i=1}^d p_i}{1+\lambda\sum_{i=1}^d p_i}\\ &= 1-F_\lambda(\boldsymbol{p})\\ &\le \frac{\lambda\Delta}{1+\lambda\Delta},\end{aligned}$$

we see immediately see via the same kind of inductive argument that the right-hand side of Eq. (B.4) is upper bounded by $\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^k$. However, we can do better. Following [Bay+07], we consider the two-step recursion $F_\lambda^{\circ 2}$, which admits faster contraction rates when $\lambda\Delta\ge\frac{3}{4}$ by Lemma B.0.9. First, we calculate that

$$\begin{aligned}(\Phi\circ F_\lambda^{\circ 2}\circ\Phi^{-1})(\boldsymbol{y}) &= -\log\left(1+\lambda\sum_{i=1}^d\frac{1}{1+\lambda\sum_{j=1}^d\exp(y_{ij})}\right)\\ \partial_{y_{ij}}(\Phi\circ F_\lambda^{\circ 2}\circ\Phi^{-1})(\boldsymbol{y}) &= \frac{(\Phi'\circ F_\lambda^{\circ 2})(\boldsymbol{p})}{\Phi'(p_{ij})}\cdot(\partial_{p_{ij}}F_\lambda^{\circ 2})(\boldsymbol{y})\\ &= \frac{p_{ij}}{F_\lambda^{\circ 2}(\boldsymbol{p})}\cdot((\partial_{p_i}F_\lambda)\circ F_\lambda)(\boldsymbol{p})\cdot(\partial_{p_j}F_\lambda)(\boldsymbol{p}_i)\\ &= \frac{p_{ij}}{F_\lambda^{\circ 2}(\boldsymbol{p})}\cdot\left(-\lambda(F_\lambda\circ F_\lambda)(\boldsymbol{p})^2\right)\cdot\left(-\lambda F_\lambda(\boldsymbol{p}_i)^2\right)\\ &= \left(\lambda\cdot p_{ij}\cdot F_\lambda(\boldsymbol{p}_i)\right)\cdot\left(\lambda\cdot F_\lambda(\boldsymbol{p}_i)\cdot F_\lambda^{\circ 2}(\boldsymbol{p})\right),\end{aligned}$$

where we imagine that j is a child of i which is a child of the root. With this, we can then peel off two levels at a time. In particular, the right-hand side of Eq. (B.4) satisfies

$$\begin{aligned}&\sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k]}\prod_{i=1}^k\lambda\cdot\mu_{T(u_i)}(\overline{u_i})\cdot\mu_{T(u_{i+1})}(\overline{u_{i+1}})\\ &= \sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k-2]}\prod_{i=1}^{k-2}\lambda\cdot\mu_{T(u_i)}(\overline{u_i})\cdot\mu_{T(u_{i+1})}(\overline{u_{i+1}})\\ &\quad\times\underbrace{\left(\sum_{u_k\in L_{u_{k-1}}(1)}\sum_{u_{k+1}\in L_{u_k}(1)}\left(\lambda\mu_{T(u_{k-1})}(\overline{u_{k-1}})\mu_{T(u_k)}(\overline{u_k})\right)\cdot\left(\lambda\mu_{T(u_k)}(\overline{u_k})\mu_{T(u_{k+1})}(\overline{u_{k+1}})\right)\right)}_{=\left\|\nabla(\Phi\circ F_\lambda^{\circ 2}\circ\Phi^{-1}(\boldsymbol{y})\right\|_1\text{ where }\boldsymbol{y}\text{ is the logarithm of the }\mu_{T(u_{k+1})}(\overline{u_{k+1}})}\\ &\le\left(\sup_{\boldsymbol{y}}\left\|\nabla(\Phi\circ F_\lambda^{\circ 2}\circ\Phi^{-1})(\boldsymbol{y})\right\|_1\right)\cdot\sum_{u_{i+1}\in L_{u_i}(1),\forall i\in[k-2]}\prod_{i=1}^{k-2}\lambda\cdot\mu_{T(u_i)}(\overline{u_i})\cdot\mu_{T(u_{i+1})}(\overline{u_{i+1}})\\ &\le\cdots\qquad\text{(Induction)}\\ &\le\left(\sup_{\boldsymbol{y}}\left\|\nabla(\Phi\circ F_\lambda^{\circ 2}\circ\Phi^{-1})(\boldsymbol{y})\right\|_1\right)^{\lfloor k/2\rfloor}\cdot\left(\frac{\lambda\Delta}{1+\lambda\Delta}\right)^{k\bmod 2}\end{aligned}$$

as desired.

An Error in the Original Proof Provided in [CLV21a]

In the process of writing this thesis, we discovered a bug in the original proof of Theorem B.0.4 in [CLV21a]. The original proof proceeds by first establishing the following bound.

Proposition B.0.12. *Fix a tree T and an edge $e \in E_T$ with endpoints r_1, r_2. Then we have the bound*

$$\sum_{f \in E_{T(r_i)}} |\Psi_T(e \to f)| \leq 2 \sum_{k=1}^{\infty} \max \left\{ \prod_{i=1}^{k} \mu_{T(u_i)}(u_i) \;\middle|\; \substack{u_1 \in e = \{r_1, r_2\}, \\ u_{i+1} \in L_{u_i}(1), \forall i \in [k-1]} \right\},$$

where we write $e = e_1, r = u_1, e_2, u_2, \ldots, u_k, e_{k+1} = f$ for the unique path from e to f such that edge e_i connects vertices u_{i-1} and u_i.

This bound is correct, and one can use it to deduce the $2\lambda\Delta$ upper bound in Theorem B.0.4, since each $\mu_{T(u_i)}(u_i)$ is upper bounded by $\frac{\lambda\Delta}{1+\lambda\Delta}$.

However, the proof of the second bound based on this proposition is erroneous. Indeed, in the ensuing analysis which bounds the right-hand side of the inequality in Proposition B.0.12, the following inequality is used without justification: For every tree T rooted at r, and every $u \in L_r(1)$,

$$\begin{aligned} &\left(\sum_{v \in L_r(1)} \lambda \mu_T(\overline{r}) \mu_{T(v)}(\overline{v}) \right) \left(\sum_{w \in L_u(1)} \lambda \mu_{T(u)}(\overline{u}) \mu_{T(w)}(\overline{w}) \right) \\ &\leq \sum_{v \in L_r(1), w \in L_v(1)} \left(\lambda \mu_T(\overline{r}) \mu_{T(v)}(\overline{v}) \right) \cdot \left(\lambda \mu_{T(v)}(\overline{v}) \mu_{T(w)}(\overline{w}) \right), \end{aligned} \tag{B.5}$$

where we have essentially swapped out u with v in one of the terms. However, since the right-hand side of Eq. (B.5) satisfies

$$\begin{aligned} &\sum_{v \in L_r(1), w \in L_v(1)} \left(\lambda \mu_T(\overline{r}) \mu_{T(v)}(\overline{v}) \right) \cdot \left(\lambda \mu_{T(v)}(\overline{v}) \mu_{T(w)}(\overline{w}) \right) \\ &= \sum_{v \in L_r(1)} \left(\left(\lambda \mu_T(\overline{r}) \mu_{T(v)}(\overline{v}) \right) \cdot \sum_{w \in L_v(1)} \left(\lambda \mu_{T(v)}(\overline{v}) \mu_{T(w)}(\overline{w}) \right) \right) \\ &\leq \left(\sum_{v \in L_r(1)} \left(\lambda \mu_T(\overline{r}) \mu_{T(v)}(\overline{v}) \right) \right) \cdot \max_{u \in L_r(1)} \left\{ \sum_{w \in L_u(1)} \left(\lambda \mu_{T(u)}(\overline{u}) \mu_{T(w)}(\overline{w}) \right) \right\}, \end{aligned}$$

Eq. (B.5) cannot hold unless $\sum_{w \in L_u(1)} \lambda \mu_{T(u)}(\overline{u}) \mu_{T(w)}(\overline{w})$ is the same over all $u \in L_r(1)$. This is false even for small examples of rooted trees.

Appendix C

ℓ_1-Independence for the Hardcore Model

In this chapter, use correlation decay to bound the induced ℓ_1-norm of the influence matrix for the hardcore model. The ℓ_1-norm captures the maximum total influence *on* a vertex, whereas the ℓ_∞-captures the maximum total influence *of* a vertex. While we saw how to obtain ℓ_∞-norm bounds in Chapter 7 and Appendices A and B, ℓ_1-norm bounds appear to be more difficult to obtain, since we lose the beautiful and direct connection between influences in the input graph $G = (V, E)$ and influences on Weitz's self-avoiding walk tree; see Theorem 7.2.2 and Corollary 7.2.3. The most we get is preservation of the marginal probability and marginal ratio of the root, as originally proved by Weitz [Wei06].

One additional drawback of the analysis presented in this chapter is that it leads to an exponentially suboptimal spectral independence upper bound of $\exp(O(1/\delta))$ for the hardcore model when $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Nevertheless, we record the techniques here, with the hope that it can be useful in the event that bounding the ℓ_1-norm of the influence matrix is necessary. This was also the first way in which spectral independence was established for the hardcore model in the tree uniqueness regime. The results of this chapter are based on [ALO21].

We prove the following.

Theorem C.0.1 (ℓ_1-Independence for Tree Unique Hardcore Model)**.** *There exists a function $C : [0,1] \to \mathbb{R}_{>0}$ such that for every graph $G = (V, E)$ of maximum degree $\leq \Delta$, every boundary condition $\sigma_\Lambda : \Lambda \to \{0,1\}$ on a subset of vertices $\Lambda \subseteq V$, every $0 < \delta < 1$, and every $\lambda \leq (1-\delta)\lambda_c(\Delta)$,*

$$\left\|\Psi_\mu^{\sigma_\Lambda}\right\|_1 = \max_{v \in V \setminus \Lambda} \sum_{u \in V \setminus \Lambda : u \neq v} \left|\Psi_\mu^{\sigma_\Lambda}(u \to v)\right| \leq C(\delta).$$

Furthermore, $C(\delta)$ satisfies $C(\delta) \leq \exp(O(1/\delta))$.

The key challenge in establishing this is that we no longer have tools like Theorem 7.2.2 and Corollary 7.2.3 at our disposal. However, at a high level, we follow the same strategy as we did for bounding the ℓ_∞-norm $\|\Psi_\mu\|_\infty$. Namely, we still reduce to bounding influences in self-avoiding walk trees, and then bound the total influence *on* the root of the tree. For the first step, follows from a kind of decoupling lemma.

To control influences on the root, we again leverage the tree recursions for the hardcore model. We also amortize the total influence using the potential method. This allows us to show a strong kind of correlation decay, where the total pairwise influence of all vertices at a fixed distance decays as the distance grows. Following the release of a preliminary draft of [ALO21], it was pointed out by Eric Vigoda and Zongchen Chen that the notion of correlation decay proved here is extremely similar to the notion of *aggregate strong spatial mixing* (for trees) studied in [MS13; BCV20]. Two perhaps interesting open problems here are to extend Theorem C.0.1 to other two-state spin systems in the tree uniqueness regime, as well as improve the bound to $C(\delta) \leq O(1/\delta)$, which would match the ℓ_∞-independence upper bounds.

Some Additional Preliminaries on Correlation Decay For our analysis on *trees*, we will need a more flexible notion of "boundary condition", which is given by a function $p : A \to [0,1]$,

where A is a subset of vertices excluding the root r of the tree. For each $v \in A$, the intuitive meaning of $p(v)$ is that the marginal probability of v w.r.t. the subtree of T rooted at v is pinned to the value $p(v)$. In the case p maps all vertices of A to 0 or 1, then p is a bonafide boundary condition in the traditional sense, where vertices of A are pinned to be 0 or 1.

Now, given such a generalized boundary condition $p : A \to [0, 1]$, we may define the conditional marginal $\Pr_{T,r}[r \mid p]$ by simply iteratively applying the tree recursions from the bottom-up, starting with the marginals of leaves of T fixed to the usual $\frac{\lambda}{1+\lambda}$, and the marginals of vertices of A fixed under p. We can then define the ratio of conditional marginals by $R^p_{T,r} = \frac{\Pr_{T,r}[r|p]}{1-\Pr_{T,r}[r|p]}$. By construction, these ratios of conditional marginals satisfy the same tree recurrence, i.e. we have the equality $R^p_{T,r} = F(R^p_{T_u,u} : u \in L_r(1))$. Here, we make a slight abuse of notation by writing p even when considering a subtree T_u; this should be understood as the restriction of p to this subtree.

For the second step of our analysis, in the case of the hardcore model, [Wei06] showed that weak spatial mixing on the infinite Δ-regular tree implies strong spatial mixing on all trees of maximum degree $\leq \Delta$, and hence, on all graphs of maximum degree $\leq \Delta$. To conveniently state the strong spatial mixing result proved in [Wei06], we make the following definition.

Definition 51. *If T is a tree rooted at $r \in T$, we define $R^{\min}_{T,r}(\ell)$ to be the minimum conditional probability ratio that r is assigned 1 in a random configuration, over all possible marginals of vertices at depth ℓ in the subtree rooted at r. That is, $R^{\min}_{T,r}(\ell) = \min_p R^p_{T,r}$, where p is an assignment of marginals of vertices at depth ℓ in T_u. Similarly, define $R^{\max}_{T,r}(\ell)$ to be the maximum such conditional probability ratio. Finally, define $R^{\min}(\ell) = \min_{T,r} R^{\min}_{T,r}(\ell)$ and $R^{\max}(\ell) = \max_{T,r} R^{\max}_{T,r}(\ell)$, where the minimum and maximum are over all trees T rooted at r of maximum degree $\leq \Delta$.*

Remark 62. Note that the map $x \mapsto \frac{x}{1-x}$ is monotone increasing. Hence, the boundary condition p which achieves $R^{\min}_{T,r}(\ell)$ is the one which minimizes $\Pr[r \mid p]$. Essentially, due to the antiferromagnetic nature of the hardcore model, the level-ℓ boundary condition minimizing p^τ_r is the all-1 configuration if ℓ is odd, and the all-0 configuration if ℓ is even. Determining the configuration achieving $R^{\max}_r(\ell)$ can be done a similar way.

Note, in particular, that the boundary condition p achieving the minimum or maximum maps the marginals of vertices to $\{0, 1\}$; there is no advantage to allowing fractional marginals. However, this formulation of $R^{\min}_{T,r}(\ell)$ and $R^{\max}_{T,r}(\ell)$ will be convenient later.

Fact C.0.2. *We have the inequalities*

1. $0 = R^{\min}(1) \leq R^{\max}(1) = \lambda$,
2. $\frac{\lambda}{(1+\lambda)^\Delta} = R^{\min}(2) \leq R^{\max}(2) = \lambda$,
3. $R^{\min}(\ell) \leq R^{\min}(\ell + 1)$ *and* $R^{\max}(\ell) \geq R^{\max}(\ell + 1)$ *for any* $\ell \geq 1$.

Proof. By monotonicity of F Fact 7.1.3, $R^p_{T,r} \leq F(0, \ldots, 0) = \lambda$ for any tree T rooted at r and any boundary condition p. Thus, we have $0 \leq R^{\min}(\ell) \leq R^{\max}(\ell) \leq \lambda$ for all $\ell \geq 1$.

1. $R^{\min}(1) = 0$ since any boundary condition which pins a neighbor of the root to "in" forces the root to be "out". $R^{\max}(1) = \lambda$ holds since one can consider the tree consisting of a single vertex.

2. $R^{\min}(2) \geq \frac{\lambda}{(1+\lambda)^\Delta}$ since by monotonicity of the tree recursion F, to minimize the marginal ratio of the root, it suffices to maximize the marginal ratios of the neighbors, which are most at λ. $R^{\min}(2) = \frac{\lambda}{(1+\lambda)^\Delta}$ holds by considering the star graph with Δ leaves (and no boundary conditions). We also have $R^{\max}(2) = \lambda$ by pinning all vertices at distance-2 from the root to "in", so that all neighbors of the root are forced to be "out".

3. $R^{\min}(\ell) \leq R^{\min}(\ell + 1)$ holds because any boundary condition $p_{\ell+1}$ on vertices at level $\ell + 1$ gives rise to a boundary condition p_ℓ on vertices at level ℓ using the tree recurrence. The same observations implies $R^{\max}(\ell) \geq R^{\max}(\ell + 1)$.

□

Theorem C.0.3 (Weak Spatial Mixing Implies Strong Spatial Mixing; [Wei06])**.** *Assume $\lambda = (1-\delta)\lambda_c(\Delta)$ for some $0 < \delta < 1$. Then there exist constants $C > 0$ and $0 < \alpha < 1$ such that for every tree T of maximum degree $\leq \Delta$ rooted at some $r \in T$, and every level ℓ, we have the bound*

$$|R^{\min}_{T,r}(\ell) - R^{\max}_{T,r}(\ell)| \leq C \cdot \alpha^{\ell}$$

Later on in the paper, we will need more precise control over C, α. However, the above result is sufficient for the present discussion

C.1 Influence Decoupling in Weitz's Self-Avoiding Walk Tree

In this section, we take a step towards proving Theorem C.0.1. Specifically, we focus on bounding

$$\sum_{u \in V : u \neq v} |\Psi_\mu(u, v)|$$

where from now on, we take μ to be the distribution corresponding to the hardcore distribution on input graph $G = (V, E)$ with parameter $\lambda > 0$. Here, the relevant uniqueness threshold is given by $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$.

Before we proceed to bound this quantity for general graphs, we note that one can easily deduce an $O(1)$ upper bound for amenable graphs (i.e. graphs such that the balls around any vertex grows subexponentially fast in the radius) in a black-box fashion directly using strong spatial mixing Definition 33, thus recovering some of the previously known connections between spatial mixing properties of the hardcore distribution, and temporal mixing of the Glauber dynamics [Dye+04b; Wei04; CP21a]. This follows from the straightforward observation that the exponential decay of correlations in distance dominates the subexponential growth of the number of vertices at each distance scale. The class of amenable graphs notably includes lattices such as $\mathbb{Z}^d$, but exclude most graphs such as expanders. Hence, instead of applying strong spatial mixing as a black-box, we revisit its proof, and modify it as necessary.

The high-level strategy is to convert this problem on general graphs to bounding a similar quantity for trees. We do this by leveraging the self-avoiding walk tree construction of [Wei06]. However, since a vertex $u \in G$ may have many copies in the corresponding self-avoiding walk tree $T = T_{\mathsf{SAW}}(G, r)$, we need to "decouple" these copies so as to obtain single-vertex influences again.

Definition 52 (R-Pseudoinfluence)**.** *Recall that for a fixed tree T rooted at r with boundary condition $p : A \to [0,1]$ (where A is a subset of vertices not containing r), we write $R^{p}_{T,r} = \frac{\Pr[r|p]}{1-\Pr[r|p]}$. For a distinguished vertex $v \in T$ with $v \neq r$ and $v \notin A$, and any $s \in [0,1]$, we further write $R^{v^s,p}_{T,r}$ in place of $R^{\tilde{p}}_{T,r}$ where $\tilde{p} : A \cup \{v\} \to [0,1]$ is the boundary condition with $\tilde{p}(u) = p(u)$ for all $u \in A$ and $\tilde{p}(v) = s$. With this, we define the R-pseudoinfluence of v on the root r by the quantity*

$$\mathcal{R}^{v}_{T,r} = \max_{p} \mathcal{R}^{v,p}_{T,r} \qquad \text{where} \qquad \mathcal{R}^{v,p}_{T,r} = \left| R^{v^0,p}_{T,r} - R^{v^1,p}_{T,r} \right|$$

and the maximum is taken over all partial assignments $p : L_r(\ell(v)) \setminus \{v\} \to [0,1]$ of marginal values. Again, we drop the subscript T when the tree is clear from context.

Remark 63. It was pointed out to us by Zongchen Chen and Eric Vigoda that our notion of R-pseudoinfluence is very related to the notion of "aggregate strong spatial mixing" used in [MS13] to analyze the Glauber dynamics, and in [BCV20] to analyze the Swendsen-Wang dynamics, both for the ferromagnetic Ising model. In fact, it turns out our result also directly implies aggregate strong spatial mixing for arbitrary trees in the uniqueness regime $\lambda < \lambda_c(\Delta)$.

Our first step is to do the decoupling using the R-pseudoinfluence. The second step is to bound the total R-pseudoinfluence of vertices in a tree on the root. These steps are captured in the following two results. We emphasize Lemma C.1.1 is generic, and holds for any two-state spin system.

Lemma C.1.1 (Decoupling)**.** *Consider the hardcore distribution μ on a graph $G = (V, E)$ with parameter $\lambda > 0$. Fix a vertex $r \in G$ and let $T = T_{\mathsf{SAW}}(G, r)$. Then the following inequality holds:*

$$\sum_{v \in G : v \neq r} |\Psi_\mu(v, r)| \leq 2 \sum_{v \in T : v \neq r} \mathcal{R}^{v}_{r}$$

In particular, to bound $\sum_{v\in G:v\neq r}|\Psi_\mu(v,r)|$, it suffices to bound $\sum_{v\in T:v\neq r}\mathcal{R}_r^v$ for every tree T of maximum degree $\leq\Delta$ rooted at r. This motivates the next result.

Proposition C.1.2 (R-Pseudoinfluence Bound)**.** *Assume λ is up-to-Δ unique with gap $0<\delta<1$. Then for every tree T of maximum degree $\leq\Delta$ rooted at r, we have the bound*

$$\sum_{v\in T:v\neq r}\mathcal{R}_r^v\leq\exp(O(1/\delta))$$

With these two lemmas in hand, we may prove Theorem C.0.1.

Proof of Theorem C.0.1. Let $0<\delta<1$ and take $\lambda=(1-\delta)\lambda_c(\Delta)$. By Lemma 7.1.4, λ is up-to-Δ unique. Let $G=(V,E)$ be a graph with maximum degree $\leq\Delta$ and let $T=T_{\mathsf{SAW}}(G,r)$ be the self-avoiding walk tree rooted at an arbitrary vertex $r\in V$. By Lemma C.1.1

$$\sum_{u\in V:u\neq v}|\Psi_\mu(u,v)|\leq 2\sum_{v\in T:v\neq r}\mathcal{R}_r^v$$

By Proposition C.1.2, the right-hand side is bounded above by $\exp(O(1/\delta))$ as desired. □

It remains to prove Lemma C.1.1 and Proposition C.1.2. The rest of the section is devoted to proving the former. The proof of the latter is contained in the following section.

Proof of Lemma C.1.1. Recall that for a vertex $v\in G$ with $v\neq r$, $C(v)$ denotes the set of all copies of v in $T=T_{\mathsf{SAW}}(G,r)$. In particular, $\{C(v):v\in G,v\neq r\}$ is a partition of the vertices of T (excluding the root r of T). Hence, to prove the claim, it suffices to show that for each $v\in G$ with $v\neq r$, we have

$$|\Psi_\mu(v,r)|\leq 2\sum_{u\in C(v)}\mathcal{R}_r^u$$

Towards this, we define an intermediate quantity which we use only for the purposes of this proof. Specifically, we define the pseudoinfluence as

$$\mathcal{I}_r^u=\max_p\left|\Pr\left[r\mid u^0,p\right]-\Pr\left[r\mid u^1,p\right]\right|$$

where again the maximum is taken over all partial assignments $p:L_r(\ell(u))\setminus\{u\}\to[0,1]$. Observe that $\mathcal{I}_r^u\leq\mathcal{R}_r^u$ trivially since

$$\begin{aligned}\left|R_r^{u^0,p}-R_r^{u^1,p}\right|&=\left|\frac{\Pr\left[r\mid u^0,p\right]}{1-\Pr\left[r\mid u^0,p\right]}-\frac{\Pr\left[r\mid u^1,p\right]}{1-\Pr\left[r\mid u^1,p\right]}\right|\\&=\frac{\left|\Pr\left[r\mid u^0,p\right]-\Pr\left[r\mid u^1,p\right]\right|}{(1-\Pr\left[r\mid u^0,p\right])(1-\Pr\left[r\mid u^1,p\right])}\\&\geq\left|\Pr\left[r\mid u^0,p\right]-\Pr\left[r\mid u^1,p\right]\right|\end{aligned}$$

holds for any p. Hence, it suffices to prove

$$|\Psi_\mu(v,r)|\leq 2\sum_{u\in C(v)}\mathcal{I}_r^u$$

Now, recall that an assignment of v in G lifts to an assignment of all vertices in $C(v)$ in T; if v is assigned 0 (resp. 1), then all vertices in $C(v)$ are assigned 0 (resp. 1). By Corollary 7.2.3, we have that

$$|\Psi_\mu(v,r)|=\left|\Pr[r\mid\sigma_{C(v),0}]-\Pr[r\mid\sigma_{C(v),1}]\right|$$

We may deduce the desired inequality via the following more general claim.

Claim C.1.3. *Fix a set of vertices A in T such that no vertex of A has an ancestor also in A. Then*

$$|\Pr[r\mid\sigma_{A,0}]-\Pr[r\mid\sigma_{A,1}]|\leq 2\sum_{v\in A}\mathcal{I}_r^v$$

where $\sigma_{A,0}$ assigns all vertices in A to 0, and $\sigma_{A,1}$ assigns all vertices in A to 1.

In particular, taking A to be the set of vertices in $C(v)$ such that no ancestor is also in $C(v)$, we obtain our main decoupling result. □

All that remains is to prove the claim.

Proof of Claim C.1.3. We go by induction on the size of A. The base case is obvious. Let v denote the closest vertex of A to r. More specifically, let $v \in A$ be such that $d(r, v) = d(r, A)$. Let $B = A \setminus \{v\}$. For the inductive step, observe by Triangle Inequality that

$$\begin{aligned}|\Pr[r \mid \sigma_{A,0}] - \Pr[r \mid \sigma_{A,1}]| \le &\left|\Pr[r \mid \sigma_{B,0}, v^0] - \Pr[r \mid \sigma_{B,0}]\right| \\ &+ |\Pr[r \mid \sigma_{B,0}] - \Pr[r \mid \sigma_{B,1}]| \\ &+ \left|\Pr[r \mid \sigma_{B,1}] - \Pr[r \mid \sigma_{B,1}, v^1]\right|\end{aligned}$$

By the induction, $|\Pr[r \mid \sigma_{B,0}] - \Pr[r \mid \sigma_{B,1}]| \le 2\sum_{u \in B} \mathcal{I}_r^u = 2\sum_{u \in A : u \neq v} \mathcal{I}_r^u$. Hence, it suffices to bound the first and last terms by $\mathcal{I}_r^v$. We focus on the first term in the inequality $\left|\Pr[r \mid \sigma_{B,0}, v^0] - \Pr[r \mid \sigma_{B,0}]\right|$; the same argument works for the last term $\left|\Pr[r \mid \sigma_{B,1}] - \Pr[r \mid \sigma_{B,1}, v^1]\right|$. Since the tree recursion is a monotone function in each individual variable Fact 7.1.3, $\Pr[r \mid \sigma_{B,0}]$ is a monotone function in the marginal of v. It follows that

$$\left|\Pr[r \mid \sigma_{B,0}, v^0] - \Pr[r \mid \sigma_{B,0}]\right| \le \left|\Pr[r \mid \sigma_{B,0}, v^0] - \Pr[r \mid \sigma_{B,0}, v^1]\right|$$

But

$$\left|\Pr[r \mid \sigma_{B,0}, v^0] - \Pr[r \mid \sigma_{B,0}, v^1]\right| \le \mathcal{I}_r^v$$

clearly holds simply because all vertices in B are at least as far from r as v is, and hence, the effect of B may be simulated by enforcing appropriate marginals on all vertices distance $d(r, v)$. The same argument shows that $\left|\Pr[r \mid \overline{B}] - \Pr[r \mid \overline{B}, \overline{v}]\right| \le \mathcal{I}_v^r$, and so the claim follows. □

C.1.1 R-Pseudoinfluence Decay

Our goal is now to prove Proposition C.1.2. More specifically, we aim to upper bound $\sum_{v \in T : v \neq r} \mathcal{R}_r^v$ by $\exp(O(1/\delta))$. To do this, we write

$$\sum_{v \in T : v \neq r} \mathcal{R}_r^v = \sum_{\ell=1}^{\infty} \sum_{v \in L_r(\ell)} \mathcal{R}_r^v$$

Thus, it suffices to bound $\sum_{v \in L_r(\ell)} \mathcal{R}_r^v$ for each level ℓ. We show that this quantity in fact decays exponentially fast as ℓ increases when $\lambda < \lambda_c(\Delta)$. Specifically, to prove Proposition C.1.2, we use the following two lemmas, which precisely quantify the decay rate.

Proposition C.1.4 (Decay Rate Bound)**.** *Assume λ is up-to-Δ unique with gap $0 < \delta < 1$. Then there exists $\ell_0 = \Theta(1/\delta)$ such that for every tree T of maximum degree $\le \Delta$ rooted at r and any $\ell > \ell_0$, we have the bound*

$$\frac{\sum_{v \in L_r(\ell)} \mathcal{R}_r^v}{\max_{u \in L_r(\ell - \ell_0)} \left\{\sum_{v \in L_u(\ell_0)} \mathcal{R}_u^v\right\}} \le O(1) \cdot \sqrt{1-\delta}^{\ell - \ell_0}$$

We prove Proposition C.1.4 in the next section. Roughly speaking, the reason for the assumption $\ell > \ell_0$ above is that we can exploit spatial mixing to argue that the marginals of the root is independent of the boundary condition at level ℓ, for a large enough ℓ_0; see Appendix C.2 for more details. For $\ell < \ell_0$ we use the following lemma.

Lemma C.1.5 (Trivial "Decay" Rate)**.** *Assume λ up-to-Δ unique with gap $0 < \delta < 1$. Then for any tree T of maximum degree $\le \Delta$ rooted at r and any $\ell > 0$, we have*

$$\frac{\sum_{v \in L_r(\ell)} \mathcal{R}_r^v}{\max_{u \in L_r(1)} \left\{\sum_{v \in L_u(\ell-1)} \mathcal{R}_u^v\right\}} \le O(1)$$

Furthermore, for the first level, we have the inequality

$$\sum_{v \in L_r(1)} \mathcal{R}_r^v \le O(1)$$

Proof. Since $\lambda \le \lambda_c(\Delta) \le O(1/\Delta)$, it suffices to show an upper bound of $(\Delta-1)\lambda$ on the ratio.

We proceed via the Mean Value Theorem. For any fixed $v \in L_u(\ell-1)$ where $u \in L_r(1)$, and for any $p : L_r(\ell(v)) \setminus \{v\} \to [0,1]$, we have there exists $\tilde{\mathbf{R}}$ such that

$$\left|R_r^{v^0,p} - R_r^{v^1,p}\right| = \left|\partial_{R_u} F(\tilde{\mathbf{R}})\right| \cdot \left|R_u^{v^0,p} - R_u^{v^1,p}\right|$$

Since

$$|\partial_{R_u} F(\mathbf{R})| = \frac{F(\mathbf{R})}{R_u + 1}$$

is monotone decreasing in each coordinate, we obtain an upper bound of

$$\left|R_r^{v^0,p} - R_r^{v^1,p}\right| \le |\partial_{R_u} F(\mathbf{0})| \cdot \mathcal{R}_u^v = \lambda \cdot \mathcal{R}_u^v$$

As this holds for all $p : L_r(\ell(v)) \setminus \{v\} \to [0,1]$, it follows that $\mathcal{R}_r^v \le \lambda \cdot \mathcal{R}_u^v$. Summing over all $v \in L_r(\ell)$, we have

$$\sum_{v \in L_r(\ell)} \mathcal{R}_r^v \le \lambda \sum_{u \in L_r(1)} \sum_{v \in L_u(\ell-1)} \mathcal{R}_u^v \le (\Delta-1)\lambda \max_{u \in L_r(1)} \left\{ \sum_{v \in L_u(\ell-1)} \mathcal{R}_u^v \right\}$$

as desired.

For the bound on $\sum_{v \in L_r(1)} \mathcal{R}_r^v$, note that it suffices to bound $\mathcal{R}_r^v$ by λ for any $v \in L_r(1)$. To do this, fix any $p : L_r(1) \setminus \{v\} \to [0,1]$. We have $R_r^{v^1,p} = 0$ simply because conditioning v to be "in" forces r to be "out". Hence,

$$\left|R_r^{v^0,p} - R_r^{v^1,p}\right| = R_r^{v^0,p}$$

Since the tree recursion is a monotone function in each coordinate, $R_r^{v^0,p}$ is maximized when p is identically zero on $L_r(1) \setminus \{v\}$, in which case, $R_r^{v^0,p} = \lambda$. It follows that $\mathcal{R}_r^v \le \lambda$ as desired. □

These two results together immediately imply Proposition C.1.2.

Proof of Proposition C.1.2. Using Lemma C.1.5, we have for any $\ell \le \ell_0$ that

$$\sum_{v \in L_r(\ell)} \mathcal{R}_r^v \le O(C^\ell)$$

for a universal constant $C > 0$. When $\ell > \ell_0$, we have

$$\sum_{v \in L_r(\ell)} \mathcal{R}_r^v \le O(1) \cdot C^{\ell_0} \cdot \sqrt{1-\delta}^{\ell-\ell_0}$$

Hence, summing over all ℓ and using $\ell_0 \le O(1/\delta)$, we obtain that

$$\sum_{v \in T : v \ne r} \mathcal{R}_r^v \le O(1) \cdot C^{\ell_0} \cdot \exp(O(1/\delta)) \cdot \underbrace{\sum_{\ell=1}^{\infty} \sqrt{1-\delta}^{\ell}}_{\le O(1/\delta)} \le O(1) \cdot \exp(O(1/\delta))$$

The claim follows. □

C.2 Bounding the R-Pseudoinfluence Decay: The Potential Method

Our goal in this section is to prove Proposition C.1.4, which states that the total R-pseudoinfluence $\sum_{v \in L_r(\ell)} \mathcal{R}_r^v$ of a given level ℓ of a tree decays exponentially fast as ℓ increases. We use the potential method (otherwise known as the message decay argument), which has been successfully used in [LLY12; LLY13; Res+13; SST14; Sin+15] to establish strong spatial mixing all the way up to the uniqueness threshold.

Definition 53 (φ-Pseudoinfluence)**.** *We say a function $\varphi : [0,\infty) \mapsto [a,b]$, for $a < b$, with derivative $\Phi = \varphi'$, is a potential function if it is*

1. *continuously differentiable,*
2. *strictly monotone increasing, i.e., Φ is strictly positive,*
3. *concave, i.e., Φ is decreasing.*

For a boundary condition $p : A \to [0,1]$, where A is a subset of vertices not containing r, let $K_r^p = \varphi(R_r^p)$. Again, we define

$$\mathcal{K}_r^{v,p} \stackrel{\text{def}}{=} \left|K_r^{v^0,p} - K_r^{v^1,p}\right| = \left|\varphi(R_r^{v^0,p}) - \varphi(R_r^{v^1,p})\right|$$

Define the φ-pseudoinfluence of a vertex v on r as

$$\mathcal{K}_r^v \stackrel{\text{def}}{=} \max_{p:L_r(\ell(v))\setminus\{v\}\to[0,1]} K_r^{v,p}$$

Finally, we define

$$K_r^{\min}(\ell) \stackrel{\text{def}}{=} \min_{p:L_r(\ell)\to[0,1]} K_r^p = \varphi(R_r^{\min}(\ell))$$
$$K_r^{\max}(\ell) \stackrel{\text{def}}{=} \max_{p:L_r(\ell)\to[0,1]} K_r^p = \varphi(R_r^{\max}(\ell))$$

Let us now fix the potential function that we will use. In this work, we use the potential function introduced in [LLY13]. We define φ as

$$\varphi(R) \stackrel{\text{def}}{=} 2\log(\sqrt{R} + \sqrt{R+1})$$
$$\Phi(R) \stackrel{\text{def}}{=} \varphi'(R) = \frac{1}{\sqrt{R(R+1)}}$$

We note that since Φ is continuous, positive, and decreasing, we have φ is continuously differentiable, strictly monotone increasing and concave as desired. One additional feature of this potential function is that it has no dependence on λ or Δ. While it may be comforting to have an explicit expression for φ, all of our proofs rely at most on the explicit expression for Φ, rather than φ. For the derivation and further discussion of this potential function, we refer the reader to [LLY13].

To control $\sum_{v\in L_r(\ell)} \mathcal{R}_r^v$, it turns out it suffices to control the decay of $\sum_{v\in L_r(\ell)} \mathcal{K}_r^v$ as ℓ increases, as we will see later.

Proposition C.2.1 (φ-Pseudoinfluence Decay Rate Bound)**.** *Assume λ is up-to-Δ unique with gap $0 < \delta < 1$ (see Definition 34). For $\ell \geq 2$, assume that there exists $\eta \leq 1/2$ such that for all $u \in L_r(1)$, we have the inequality $\left|R_u^{\min}(\ell-1) - R_u^{\max}(\ell-1)\right| \leq \eta$. Then,*

$$\frac{\sum_{v\in L_r(\ell)} \mathcal{K}_r^v}{\max_{u\in L_r(1)}\left\{\sum_{v\in L_u(\ell-1)} \mathcal{K}_u^v\right\}} \leq (1+2\eta)^{\Delta+1}\sqrt{1-\delta}.$$

Unfortunately, due to the additional error factor of $(1+2\eta)^{\Delta+1}$, we must control $\eta = \eta(\ell)$. To do this, we leverage the strong spatial mixing result proved in [Wei06]. We state a "precise" version here, where the constant in front of the decay is stated explicitly.

Definition 54. *Define $\eta^* = \frac{R^{\max}(2)}{R^{\min}(2)} \cdot \left|R^{\min}(2) - R^{\max}(2)\right|$. Note by Fact C.0.2 and the fact that $\lambda \leq O(1/\Delta)$, we have*

$$\eta^* \leq \frac{\lambda}{\frac{\lambda}{(1+\lambda)^\Delta}} \cdot \left|\lambda - \frac{\lambda}{(1+\lambda)^\Delta}\right| \leq O(1/\Delta). \tag{C.1}$$

Proposition C.2.2 (Strong Spatial Mixing [Wei06])**.** *Assume that λ is up-to-Δ unique with gap $0 < \delta < 1$. Then for all trees T rooted at r of maximum degree $\leq \Delta$, we have*

$$\left|R_{T,r}^{\min}(\ell) - R_{T,r}^{\max}(\ell)\right| \leq \sqrt{1-\delta}^{\ell-2} \cdot \eta^*$$

For the sake of completeness, we provide a proof of this specific bound in Appendix C.3.

With these results in hand, we may deduce Proposition C.1.4.

C.2.1 Proof of Proposition C.1.4

Our goal is to use exponential decay of $\sum_{v\in L_r(\ell)} \mathcal{K}_r^v$ in ℓ to prove exponential decay of $\sum_{v\in L_r(\ell)} \mathcal{R}_r^v$. In order to apply Proposition C.2.1 and Proposition C.2.2, we must relate $\mathcal{R}_r^v$ to $\mathcal{K}_r^v$. This is done in the following lemma.

Lemma C.2.3 (Relating R-Pseudoinfluence to φ-Pseudoinfluence)**.** *Let T be a tree rooted at r. For any $\ell \geq 1$ and any vertex $v \in L_r(\ell)$, we have the bound*

$$\Phi(R^{\max}(\ell)) \cdot \mathcal{R}_r^v \leq \mathcal{K}_r^v \leq \Phi(R^{\min}(\ell)) \cdot \mathcal{R}_r^v.$$

Proof. First, observe by the Mean Value Theorem and monotonicity of Φ that for any $R_0 \leq R_1$, we have the inequalities

$$\Phi(R_1) \cdot |R_1 - R_0| \leq |\varphi(R_1) - \varphi(R_0)| \leq \Phi(R_0) \cdot |R_1 - R_0| .$$

Now, fix $v \in L_r(\ell)$; we prove the RHS inequality in lemma's statement. For any boundary condition $p : L_r(\ell(v)) \setminus \{v\} \to [0,1]$, we have

$$\begin{aligned}\left|\varphi(R_r^{v^0,p}) - \varphi(R_r^{v^1,p})\right| &\leq \max\{\Phi(R_r^{v^0,p}), \Phi(R_r^{v^1,p})\} \cdot \left|R_r^{v^0,p} - R_r^{v^1,p}\right| \\ &\leq \Phi(R_r^{\min}(\ell)) \cdot \mathcal{R}_r^v \leq \Phi(R^{\min}(\ell)) \cdot \mathcal{R}_r^v.\end{aligned}$$

As this holds for any p, we have $\mathcal{K}_r^v \leq \Phi(R_r^{\min}(\ell)) \cdot \mathcal{R}_r^v$. The reverse inequality can be proved analogously. □

Furthermore, we must show that for $\ell_0 = \Theta(1/\delta)$, Proposition C.2.1 is applicable. For this, we appeal to Proposition C.2.2 and the fact that $\eta^* \leq O(1/\Delta)$. Observe that $\sqrt{1-\delta}^{\ell_0 - 2} \cdot \eta^* \leq 1/2$ holds for $\ell_0 = \Theta(1/\delta)$.

With these in hand, we may apply Proposition C.2.1, Proposition C.2.2 and Lemma C.2.3 for

$\ell > \ell_0$ to deduce that

$$\begin{aligned}
&\sum_{v \in L_r(\ell)} \mathcal{R}_r^v \\
&\leq \frac{1}{\Phi(R^{\max}(\ell))} \cdot \sum_{v \in L_r(\ell)} \mathcal{K}_r^v && \text{(Lemma C.2.3)} \\
&\leq \frac{1}{\Phi(R^{\max}(\ell))} \cdot \max_{u \in L_r(1)} \left\{ \sum_{v \in L_u(\ell-1)} \mathcal{K}_u^v \right\} \cdot \sqrt{1-\delta} \cdot \left(1 + 2\eta^* \sqrt{1-\delta}^{\ell-3}\right)^{\Delta+1} \\
& && \text{(Propositions C.2.1 and C.2.2)} \\
&\leq \dots && \text{(Induction)} \\
&\leq \frac{1}{\Phi(R^{\max}(\ell))} \cdot \max_{u \in L_r(\ell-\ell_0)} \left\{ \sum_{v \in L_u(\ell_0)} \mathcal{K}_u^v \right\} \cdot \sqrt{1-\delta}^{\ell-\ell_0} \prod_{k=\ell_0}^{\ell-1} \left(1 + 2\eta^* \sqrt{1-\delta}^{k-2}\right)^{\Delta+1} \\
& && \text{(Propositions C.2.1 and C.2.2)} \\
&\leq \frac{\Phi(R^{\min}(\ell_0))}{\Phi(R^{\max}(\ell))} \cdot \max_{u \in L_r(\ell-\ell_0)} \left\{ \sum_{v \in L_u(\ell_0)} \mathcal{R}_u^v \right\} \cdot \sqrt{1-\delta}^{\ell-\ell_0} \prod_{k=\ell_0}^{\ell-1} \left(1 + 2\eta^* \sqrt{1-\delta}^{k-2}\right)^{\Delta+1} \\
& && \text{(Lemma C.2.3)} \\
&\leq \frac{\Phi(R^{\min}(2))}{\Phi(R^{\max}(2))} \cdot \max_{u \in L_r(\ell-\ell_0)} \left\{ \sum_{v \in L_u(\ell_0)} \mathcal{R}_u^v \right\} \cdot \sqrt{1-\delta}^{\ell-\ell_0} \exp\left(O(1) \cdot \sum_{k=\ell_0}^{\ell-1} \sqrt{1-\delta}^{k-2} \right) \\
& && \text{(Using } 1+x \leq e^x \text{, Eq. (C.1) and Fact C.0.2, and Monotonicity of } \Phi) \\
&\leq \frac{R^{\max}(2)}{R^{\min}(2)} \cdot \max_{u \in L_r(\ell-\ell_0)} \left\{ \sum_{v \in L_u(\ell_0)} \mathcal{R}_u^v \right\} \cdot \sqrt{1-\delta}^{\ell-\ell_0} \cdot O(1) \\
& && \text{(Using } \tfrac{\Phi(R_0)}{\Phi(R_1)} = \sqrt{\tfrac{R_1(R_1+1)}{R_0(R_0+1)}} \leq \tfrac{R_1}{R_0} \text{ for } R_0 \leq R_1 \text{, and } \ell_0 = \Theta(1/\delta)) \\
&\leq \frac{\lambda}{\frac{\lambda}{(1+\lambda)^\Delta}} \cdot \max_{u \in L_r(\ell-\ell_0)} \left\{ \sum_{v \in L_u(\ell_0)} \mathcal{R}_u^v \right\} \cdot \sqrt{1-\delta}^{\ell-\ell_0} \cdot O(1) && \text{(Fact C.0.2)} \\
&\leq O(1) \cdot \sqrt{1-\delta}^{\ell-\ell_0} \cdot \max_{u \in L_r(\ell-\ell_0)} \left\{ \sum_{v \in L_u(\ell_0)} \mathcal{R}_u^v \right\}. && \text{(Using } \lambda \leq O(1/\Delta))
\end{aligned}$$

At this point, all that is left is to prove Proposition C.2.1 and Proposition C.2.2. We prove Proposition C.2.1 in the following subsection, and Proposition C.2.2 in Appendix C.3.

C.2.2 The φ-Pseudoinfluence Decay: Proof of Proposition C.2.1

In the previous subsection, we reduced exponential decay of the total R-pseudoinfluence $\sum_{v \in L_r(\ell)} \mathcal{R}_r^v$ to exponential decay of the total φ-pseudoinfluence $\sum_{v \in L_r(\ell)} \mathcal{K}_r^v$. Our goal in this subsection is now to prove Proposition C.2.1. While initially this appears to be a more daunting task, it is made feasible by the fact that the tree recurrence F for R induces a corresponding tree recurrence for K given by

$$K_r^\sigma = (\varphi \circ F \circ \varphi^{-1})(K_u^\sigma : u \in L_r(1)).$$

Using this tree recurrence for K_r^σ, we prove Lemma C.2.4 and Lemma C.2.6. Chained together with Lemma C.2.5, which lies at the heart of the results in [LLY13], we immediately obtain Proposition C.2.1.

Throughout, we will let $\mathbf{R} = (R_u : u \in L_r(1))$, $\mathbf{R}^{\max}(\ell) = (R_u^{\max}(\ell-1) : u \in L_r(1))$ and $\mathbf{R}^{\min}(\ell) = (R_u^{\min}(\ell-1) : u \in L_r(1))$ denote vectors with $|L_r(1)|$ many entries. We define $\mathbf{K}, \mathbf{K}^{\max}(\ell), \mathbf{K}^{\min}(\ell)$ analogously. Finally, if $\mathbf{x}, \mathbf{y}$ are two vectors of the same dimension, then we write $\mathbf{x} \leq \mathbf{y}$ for entrywise inequality; if $y \in [-\infty, \infty]$, we write $\mathbf{x} \leq y$ if all entries of x are upper bounded by y.

Lemma C.2.4 (True Decay). *For every λ, and every tree T rooted at r, we have the inequality*

$$\frac{\sum_{v\in L_r(\ell)}\mathcal{K}_r^v}{\max_{u\in L_r(1)}\left\{\sum_{v\in L_u(\ell-1)}\mathcal{K}_u^v\right\}} \le \sum_{u\in L_r(1)}\max_{\mathbf{K}^{\min}(\ell)\le\mathbf{K}\le\mathbf{K}^{\max}(\ell)}\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right|.$$

Lemma C.2.5 (Ideal Decay; [LLY13] Lemmas 12, 13, 14). *Assume λ is up-to-Δ unique with gap $0<\delta<1$. Let T be any tree of maximum degree $\le\Delta$ rooted at r. Then we have the bound*

$$\max_{0\le\mathbf{K}\le\infty}\left\|\nabla(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right\|_1 \le \sqrt{1-\delta}$$

Lemma C.2.6 (Relating True Decay to Ideal Decay). *Assume we have the inequality $\left|R_u^{\max}(\ell-1)-R_u^{\min}(\ell-1)\right|\le\eta$ for all $u\in L_r(1)$, where $\eta\le\frac{1}{2}$. Then for every λ, and every tree T with maximum degree $\le\Delta$ rooted at r, we have the inequality*

$$\begin{aligned}&\sum_{u\in L_r(1)}\max_{\mathbf{K}^{\min}(\ell)\le\mathbf{K}\le\mathbf{K}^{\max}(\ell)}\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right|\\&\le(1+2\eta)^{\Delta+1}\left\|\nabla(\varphi\circ F\circ\varphi^{-1})(\mathbf{K}^{\max}(\ell))\right\|_1\end{aligned}$$

Proof of Lemma C.2.4. To prove the claim, it suffices to show that if $v\in L_u(\ell-1)$ for $u\in L_r(1)$, then

$$\mathcal{K}_r^v \le \max_{\mathbf{K}^{\min}(\ell)\le\mathbf{K}\le\mathbf{K}^{\max}(\ell)}\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right|\cdot\mathcal{K}_u^v \tag{C.2}$$

since it then follows that

$$\begin{aligned}&\sum_{v\in L_r(\ell)}\mathcal{K}_r^v=\sum_{u\in L_r(1)}\sum_{v\in L_u(\ell-1)}\mathcal{K}_r^v\\&\le\sum_{u\in L_r(1)}\sum_{v\in L_u(\ell-1)}\max_{\mathbf{K}^{\min}(\ell)\le\mathbf{K}\le\mathbf{K}^{\max}(\ell)}\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right|\cdot\mathcal{K}_u^v\\&\le\sum_{u\in L_r(1)}\left[\max_{\mathbf{K}^{\min}(\ell)\le\mathbf{K}\le\mathbf{K}^{\max}(\ell)}\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right|\right]\cdot\left[\sum_{v\in L_u(\ell-1)}\mathcal{K}_u^v\right]\\&\le\left[\sum_{u\in L_r(1)}\max_{\mathbf{K}^{\min}(\ell)\le\mathbf{K}\le\mathbf{K}^{\max}(\ell)}\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right|\right]\cdot\max_{u\in L_r(1)}\left\{\sum_{v\in L_u(\ell-1)}\mathcal{K}_u^v\right\}\end{aligned}$$

as desired. Now, it remains to prove Eq. (C.2).

Fix an arbitrary partial assignment $p: L_r(\ell)\setminus\{v\}\to[0,1]$. By the Mean Value Theorem, there exists $\mathbf{K}^{\min}(\ell)\le\tilde{\mathbf{K}}\le\mathbf{K}^{\max}(\ell)$ such that

$$\begin{aligned}\mathcal{K}_r^{v,p}&=\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\tilde{\mathbf{K}})\right|\cdot\mathcal{K}_u^{v,p}\\&\le\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\tilde{\mathbf{K}})\right|\cdot\mathcal{K}_u^{v}\\&\le\max_{\mathbf{K}^{\min}(\ell)\le\mathbf{K}\le\mathbf{K}^{\max}(\ell)}\left|\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})\right|\cdot\mathcal{K}_u^{v}\end{aligned}$$

Since this holds for all p, we obtain the desired bound. □

Proof of Lemma C.2.6. Fix $u\in L_r(1)$. We have by the Chain Rule that

$$\begin{aligned}\partial_{K_u}(\varphi\circ F\circ\varphi^{-1})(\mathbf{K})&=(\varphi'\circ F\circ\varphi^{-1})(\mathbf{K})\cdot(\partial_{R_u}F\circ\varphi^{-1})(\mathbf{K})\cdot(\varphi^{-1})'(K_u)\\&=(\Phi\circ F)(\mathbf{R})\cdot(\partial_{R_u}F)(\mathbf{R})\cdot\frac{1}{\Phi(R_u)}\end{aligned}$$

where we recall $\varphi'=\Phi$. Note that $(\varphi^{-1})'=\frac{1}{\Phi\circ\varphi^{-1}}$ follows by the Inverse Function Theorem.

We now bound each term separately under the restriction $\mathbf{R}^{\min}(\ell)\le\mathbf{R}\le\mathbf{R}^{\max}(\ell)$. We claim the following.

1. $|(\Phi \circ F)(\mathbf{R})| \le |(\Phi \circ F)(\mathbf{R}^{\max}(\ell))|$: To see this, observe that F is monotone decreasing in each coordinate. Furthermore, Φ is monotone decreasing. Hence, $\Phi \circ F$ is monotone increasing in each coordinate.

2. $|\partial_{R_u} F(\mathbf{R})| \le (1+2\eta)^{\Delta+1} |\partial_{R_u} F(\mathbf{R}^{\max}(\ell))|$: To see this, observe that

$$\partial_{R_u} F(\mathbf{R}) = -\lambda \prod_{w \in L_r(1): w \neq u} \frac{1}{R_w + 1} \cdot \frac{1}{(R_u+1)^2}$$

is negative and monotone increasing. Hence, $|\partial_{R_u} F(\mathbf{R})|$ is positive and monotone decreasing. With this observation, define $\boldsymbol{\eta} = \mathbf{R}^{\max}(\ell) - \mathbf{R}^{\min}(\ell) = (\eta_u : u \in L_r(1))$ for convenience. Note that $\boldsymbol{\eta} \le \eta$. Then we have

$$\begin{aligned}
&|\partial_{R_u} F(\mathbf{R})| \le \left|\partial_{R_u} F(\mathbf{R}^{\min}(\ell))\right| = |\partial_{R_u} F(\mathbf{R}^{\max}(\ell) - \boldsymbol{\eta})| \\
&= \lambda \prod_{w \in L_r(1): w \neq u} \frac{1}{(R_w^{\max}+1) - \eta_w} \cdot \frac{1}{((R_u^{\max}+1) - \eta_u)((R_u^{\max}+1) - \eta_u)}
\end{aligned}$$

Our goal is to control this latter inequality by $(1+2\eta)^{\Delta+1} |\partial_{R_u} F(\mathbf{R}^{\max}(\ell))|$. To do this, we use the following claim.

Claim C.2.7. *Assume $\eta \le \frac{1}{2}$. Then for every $x \ge 0$, we have*

$$\frac{1}{(x+1)-\eta} \le (1+2\eta) \cdot \frac{1}{x+1}$$

Proof. Rearranging, the claim is equivalent to

$$\begin{aligned}
&x+1 \le (1+2\eta)((x+1)-\eta) = ((x+1)-\eta) + 2\eta((x+1)-\eta) \\
&\iff \eta \le 2\eta((x+1)-\eta) \\
&\iff \eta \le \frac{1}{2} + x
\end{aligned}$$

□

With this claim in hand, we see that

$$\begin{aligned}
&\lambda \prod_{w \in L_r(1): w \neq u} \frac{1}{(R_w^{\max}+1) - \eta_w} \cdot \frac{1}{((R_u^{\max}+1) - \eta_u)((R_u^{\max}+1) - \eta_u)} \\
&\le \underbrace{(1+2\eta_u)^2 \prod_{w \in L_r(1): w \neq u} (1+2\eta_w)}_{\text{now recall } \boldsymbol{\eta} \le \eta} \cdot \underbrace{\lambda \prod_{w \in L_r(1): w \neq u} \frac{1}{R_w^{\max}+1} \cdot \frac{1}{(R_u^{\max}+1)^2}}_{= |\partial_{R_u} F(\mathbf{R}^{\max}(\ell))|} \\
&\le (1+2\eta)^{\Delta+1} |\partial_{R_u} F(\mathbf{R}^{\max}(\ell))|
\end{aligned}$$

3. $\left|\frac{1}{\Phi(R_u)}\right| \le \left|\frac{1}{\Phi(R_u^{\max})}\right|$: This just follows by the fact that Φ is positive and monotone decreasing, so that $\frac{1}{\Phi}$ is positive and monotone increasing.

From this, we obtain

$$\max_{\mathbf{K}^{\min}(\ell) \le \mathbf{K} \le \mathbf{K}^{\max}(\ell)} \left|\partial_{K_u}(\varphi \circ F \circ \varphi^{-1})(\mathbf{K})\right| \le (1+2\eta)^{\Delta+1} \cdot \left|\partial_{K_u}(\varphi \circ F \circ \varphi^{-1})(\mathbf{K}^{\max})\right|$$

Hence

$$\begin{aligned}
&\sum_{u \in L_r(1)} \max_{\mathbf{K}^{\min}(\ell) \le \mathbf{K} \le \mathbf{K}^{\max}(\ell)} \left|\partial_{K_u}(\varphi \circ F \circ \varphi^{-1})(\mathbf{K})\right| \\
&\le (1+2\eta)^{\Delta+1} \sum_{u \in L_r(1)} \left|\partial_{K_u}(\varphi \circ F \circ \varphi^{-1})(\mathbf{K}^{\max}(\ell))\right| \\
&= (1+2\eta)^{\Delta+1} \left\|\nabla(\varphi \circ F \circ \varphi^{-1})(\mathbf{K}^{\max}(\ell))\right\|_1
\end{aligned}$$

as desired. □

Remark 64. We note that the proofs of Lemma C.2.4 and Lemma C.2.6 did not truly rely on the fact that $\Phi(R)$ had the form $\frac{1}{\sqrt{R(R+1)}}$. The arguments go through for any continuously differentiable, monotone increasing, concave potential function. Where we needed the definition of Φ itself is in the bound on $\left\|\nabla(\varphi \circ F \circ \varphi^{-1})(\mathbf{K})\right\|_1$ given in Lemma C.2.5, which was proved in [LLY13].

C.3 Precise Strong Spatial Mixing: Proof of Proposition C.2.2

In this subsection, our goal is to prove Proposition C.2.2. We use the following strong spatial mixing result proved in [LLY13].

Theorem C.3.1 (Theorem 9 from [LLY13]). *Assume that λ is up-to-Δ unique with rate δ, that is, $\delta = 1 - \max_{1\le d<\Delta}\left|f_d'(\hat{R}_d)\right|$ satisfies $0 < \delta < 1$. For every T rooted at r and every level ℓ, we have*

$$\left|K_r^{\min}(\ell) - K_r^{\max}(\ell)\right| \le \sqrt{1-\delta}\cdot \max_{u\in L_r(1)}\left|K_u^{\min}(\ell-1) - K_u^{\max}(\ell-1)\right|$$

Proof of Proposition C.2.2. First, observe that

$$\Phi(R^{\max}(\ell))\cdot\left|R_r^{\min}(\ell) - R_r^{\max}(\ell)\right| \le \left|K_r^{\min}(\ell) - K_r^{\max}(\ell)\right| \tag{C.3}$$
$$\le \Phi(R^{\min}(\ell))\cdot\left|R_r^{\min}(\ell) - R_r^{\max}(\ell)\right| \tag{C.4}$$

This holds via a nearly identical argument to the proof of Lemma C.2.3. With these inequalities in hand, we have

$$\begin{aligned}
\left|R_r^{\min}(\ell) - R_r^{\max}(\ell)\right| &\le \frac{1}{\Phi(R^{\max}(\ell))}\cdot\left|K_r^{\min}(\ell) - K_r^{\max}(\ell)\right| && \text{(Eq. (C.3))}\\
&\le \frac{1}{\Phi(R^{\max}(\ell))}\cdot\sqrt{1-\delta}^{\ell-2}\cdot\max_{u\in L_r(\ell-2)}\left\{\left|K_u^{\min}(2) - K_u^{\max}(2)\right|\right\} && \text{(Theorem C.3.1)}\\
&\le \frac{\Phi(R^{\min}(2))}{\Phi(R^{\max}(\ell))}\cdot\sqrt{1-\delta}^{\ell-2}\cdot\max_{u\in L_r(\ell-2)}\left\{\left|R_u^{\min}(2) - R_u^{\max}(2)\right|\right\} && \text{(Eq. (C.3))}\\
&\le \frac{\Phi(R^{\min}(2))}{\Phi(R^{\max}(2))}\cdot\sqrt{1-\delta}^{\ell-2}\cdot\max_{u\in L_r(\ell-2)}\left\{\left|R_u^{\min}(2) - R_u^{\max}(2)\right|\right\} && \text{(Fact C.0.2)}\\
&\le \frac{R^{\max}(2)}{R^{\min}(2)}\cdot\sqrt{1-\delta}^{\ell-2}\cdot\max_{u\in L_r(\ell-2)}\left\{\left|R_u^{\min}(2) - R_u^{\max}(2)\right|\right\} &&\\
& && \left(\tfrac{\Phi(R_0)}{\Phi(R_1)} = \sqrt{\tfrac{R_1(R_1+1)}{R_0(R_0+1)}} \le \tfrac{R_1}{R_0} \text{ for } R_0 \le R_1\right)\\
&\le \sqrt{1-\delta}^{\ell-2}\cdot\eta^* && \text{(Definition 54)}
\end{aligned}$$

□

Appendix D

Graph Homomorphisms, Tensor Networks, and More

In this chapter, 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 [JS89], Ising model on line graphs [Dye+21], edge covers [HLZ16], and weighted even subgraphs [JS93], in the setting we consider here, rapid mixing for any local Markov chain was not known before. Prior works [BS16; BS17a; Reg18; PR17] had studied these problems but only from the perspective of deterministic approximation algorithms using Barvinok's polynomial interpolation method [Bar16a]. 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 10.0.1, and then reduce spectral independence to the existence of a sufficiently large zero-free region for the multivariate partition function via Theorems 6.1.4 and 6.4.1. 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.

Beyond graph homomorphism and tensor networks, which are also examples of graphical models, we further study arbitrary measures on the discrete hypercube in terms of discrete Fourier analysis. This is provided in Appendix D.2. This chapter is based on [CLV21b].

D.1 Weakly Interacting Graphical Models

In this section, we state and prove our main results on weighted graph homomorphisms and tensor network contractions. All of the models we study in this section should be thought of as in the *weak interaction regime* (or "high-temperature regime"), as this is the natural setting in which one would expect correlations to be bounded in the sense of spectral independence.

D.1.1 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 D.1.1 (Spectral Independence and Mixing for Weighted Graph Homomorphisms)**.** *Fix a positive integer $q \geq 2$, let $G = (V, E)$ be a n-vertex 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 $\zeta \approx 0.56$ independent of $q, G, \{A^{uv}\}_{uv \in E}$ such that if $|A^{uv}(j,k) - 1| \leq \frac{\zeta}{\Delta + \zeta} - \epsilon$ for some fixed $\epsilon > 0$, all $uv \in E$ and all $j, k \in [q]$, then the associated graph homomorphism*

distribution μ on vertex configurations $\tau : V \to [q]$ given by

$$\mu(\tau) \propto \prod_{uv \in E} A^{uv}(\tau(u), \tau(v))$$

is $(\eta, \dots, \eta)$-spectrally independent for some constant $\eta = \eta(\Delta, \epsilon)$. In particular, if $\Delta, \epsilon = \Theta(1)$, then the Glauber dynamics for sampling from μ mixes in $O(n \log n)$ steps.

Remark 65. 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 [BS17a]. We state here a slightly more general theorem, the proof of which is included in Appendix D.3.1 for completeness.

Theorem D.1.2 (Zero-Freeness for Weighted Graph Homomorphisms; [BS17a])**.** *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 $\zeta \approx 0.56$ independent of $q, G, \{A^{uv}\}_{uv \in E}$ such that if $|A^{uv}(j,k) - 1| < \frac{\zeta}{\Delta + \zeta}$ for all $uv \in E$ and all $j, k \in [q]$, then for every $\Lambda \subseteq V$ and every boundary condition $\sigma_\Lambda : \Lambda \to [q]$, the graph homomorphism partition function*

$$\sum_{\substack{\tau : V \to [q] \\ \tau|_\Lambda = \sigma_\Lambda}} \prod_{uv \in E} A^{uv}(\tau(u), \tau(v))$$

with pinning σ_Λ is nonzero.

We give below the proof of Theorem D.1.1.

Proof of Theorem D.1.1. By Theorem 6.1.4, it suffices to prove that the multivariate partition function

$$\sum_{\substack{\tau : V \to [q] \\ \tau|_\Lambda = \sigma_\Lambda}} \prod_{uv \in E} A^{uv}(\tau(u), \tau(v)) \prod_{v \in V} \lambda_{v, \tau(v)} \tag{D.1}$$

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 σ_Λ, where $c = c(\Delta, \epsilon) > 0$ is some constant depending only on Δ, ϵ but not G. Define a new set of matrices $\{\tilde{A}^{uv}\}_{uv \in E}$ by

$$\tilde{A}^{uv}(j,k) \stackrel{\text{def}}{=} A^{uv}(j,k) \cdot \lambda_{u,j}^{1/\deg(u)} \cdot \lambda_{v,k}^{1/\deg(v)}, \quad \forall uv \in E, \forall j, k \in [q].$$

Note that the partition function for $G, \{\tilde{A}^{uv}\}_{uv \in E}$ is precisely given in Eq. (D.1).

Since $|A^{uv}(j,k) - 1| \leq \frac{\zeta}{\Delta + \zeta} - \epsilon$, there exists our desired $c(\Delta, \epsilon) > 0$ such that $|\lambda_{u,j} - 1|, |\lambda_{v,k} - 1| < c(\Delta, \epsilon)$ implies $|\tilde{A}^{uv}(j,k) - 1| < \frac{\zeta}{\Delta + \zeta}$, for all $uv \in E$ and all $j, k \in [q]$. It follows from Theorem D.1.2 that the multivariate partition function Eq. (D.1) is nonzero. As this holds for all $\lambda \in \mathcal{D}$, we are done. □

D.1.2 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 [MS08a; AL10; Orú14] and are also used to model Holant problems [CHL12; CLX09; CLX11]. 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 $\tau : E \to [q]$ to be given by

$$\mu(\tau) \propto \prod_{v \in V} f_v(\tau \mid_{E(v)}). \tag{D.2}$$

The associated partition function, also known as a *tensor network contraction*, is given by

$$\sum_{\tau:E\to[q]}\prod_{v\in V} f_v(\tau\mid_{E(v)}).$$

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 6.6 (see, for instance, Eq. (6.9)), where $q=2$ and each local function f_v is symmetric. Zeros for tensor network contractions were analyzed in [Reg18] in the symmetric case.

Theorem D.1.3 (Spectral Independence and Mixing for Tensor Network Distribution)**.** *Fix a positive integer $q\geq 2$, let $G=(V,E)$ be a n-vertex 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 $\zeta\approx 0.56$ independent of $q,G,\{f_v\}_{v\in V}$ such that if $|f_v(\alpha)-1|\leq\frac{\zeta}{\Delta+1+\zeta}-\epsilon$ for some fixed $\epsilon>0$, all $v\in V$ and all $\alpha:E(v)\to[q]$, then the tensor network distribution μ on edge configurations $\tau:E\to[q]$ given by Eq. (D.2) is $(\eta,\dots,\eta)$-spectrally independent for some constant $\eta=\eta(\Delta,\epsilon)$. In particular, if $\Delta,\epsilon=\Theta(1)$, 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 [Reg18] 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 Appendix D.3.2 for completeness.

Theorem D.1.4 (Zero-Freeness for Tensor Network Contractions; [Reg18])**.** *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 $\zeta\approx 0.56$ independent of $q,G,\{f_v\}_{v\in V}$ such that if $|f_v(\alpha)-1|<\frac{\zeta}{\Delta+1+\zeta}$ for all $v\in V$ and all $\alpha:E(v)\to[q]$, then for every $F\subseteq E$ and every boundary condition $\sigma_F:F\to[q]$, the tensor network contraction*

$$\sum_{\substack{\tau:E\to[q]\\ \tau|_F=\sigma_F}}\prod_{v\in V} f_v(\tau\mid_{E(v)})$$

with pinning σ_F is nonzero.

We give below the proof of Theorem D.1.3.

Proof of Theorem D.1.3. By Theorem 6.1.4, it suffices to prove that the multivariate partition function

$$\sum_{\substack{\tau:E\to[q]\\ \tau|_F=\sigma_F}}\prod_{v\in V} f_v(\tau\mid_{E(v)})\prod_{e\in E}\lambda_{e,\tau(e)} \tag{D.3}$$

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 σ_F, where $c=c(\Delta,\epsilon)>0$ 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. (D.3).

Since $|f_v(\alpha)-1|\leq\frac{\zeta}{\Delta+1+\zeta}-\epsilon$, there exists our desired $c(\Delta,\epsilon)>0$ such that $|\lambda_{e,k}-1|<c(\Delta,\epsilon)$ for all $e\in E(v)$ implies $|\tilde{f}_v(\alpha)-1|<\frac{\zeta}{\Delta+1+\zeta}$, for all $v\in V$ and all $\alpha:E(v)\to[q]$. It follows from Theorem D.1.4 that the multivariate partition function Eq. (D.3) is nonzero. As this holds for all $\lambda\in\mathcal{D}$, we are done. □

D.2 Arbitrary Measures on the Discrete Cube

In this section, we state a general result for mixing of an arbitrary measure on the discrete cube $\{-1,1\}^n$. For this, we fix an arbitrary potential $f : \{-1,1\}^n \to \mathbb{R}$. A standard result from analysis of Boolean functions says that f admits a unique representation as a multilinear polynomial $f(x) = \sum_{S\subseteq[n]} \hat{f}(S) \prod_{i\in S} x_i$. This representation is known as the *Fourier-Walsh transform* of f (see [ODo14] and references therein), and the coefficients $\hat{f}(S)$ are known as the *($\mathbb{F}_2$-)Fourier coefficients*. [Bar17b] showed that when the Fourier coefficients, as well as the degree $\deg f$ of f as a multilinear polynomial, are sufficiently small, then one has a zero-free disk for the corresponding partition function $\sum_{x\in\{-1,1\}^n} \exp(f(x))$. We convert this via Theorem 6.1.4 into a corresponding statement for the spectral independence of the distribution. Since we do not assume that f arises from a spin system (or, more generally, tensor network) on a bounded-degree graph, we only obtain a spectral gap bound with a relatively large exponent using Theorem 2.3.1.

Theorem D.2.1. *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\epsilon > 0$ be given, and assume that*

$$\sqrt{\deg f} \cdot L(f) \le C - \epsilon,$$

where $C \approx 0.55$ is an absolute constant, and $L(f) \stackrel{\text{def}}{=} \max_{i\in[n]} \sum_{S\subseteq[n]:S\ni i} \left|\hat{f}(S)\right|$. Further assume that the associated Gibbs distribution μ on $\{-1,1\}^n$ given by

$$\mu(x) \propto \exp(f(x))$$

is $\mathfrak{b}$-marginally bounded for some $\mathfrak{b} > 0$. Let $\mathfrak{h} \stackrel{\text{def}}{=} \frac{2C}{\sqrt{\deg f}} - 2L(f)$. Then μ is $(\eta,\dots,\eta)$-spectrally independent where η is a constant depending only on $\mathfrak{b}$ and $\mathfrak{h}$. In particular, if $\mathfrak{b},\mathfrak{h} = \Theta(1)$, then the Glauber dynamics for sampling from μ has spectral gap $n^{-O(1)}$.

Remark 66. One may also view $L(f)$ as bounding the *Lipschitz constant* of f.

Remark 67. A standard calculation using Dobrushin uniqueness condition yields that the Glauber dynamics is rapidly mixing when

$$\max_{i\in[n]} \sum_{S\subseteq[n]:S\ni i} (|S|-1) \cdot \left|\hat{f}(S)\right| < 1,$$

which can be weakened to $(\deg(f)-1)\cdot L(f) < 1$. These bounds are in general not comparable with the above due to the square root. While this bound is stronger when $\deg(f)$ is small, the above is stronger when most of the Fourier mass of f is on high-degree monomials.

Remark 68. A standard notion in analysis of Boolean functions is also that of "influence", which to avoid confusion with the notion of influence used to define spectral independence, we refer to as "voter influence". This terminology is consistent with the traditional applications of analysis of Boolean functions to social choice theory and voting systems; see [ODo14] and references therein. A standard result in analysis of Boolean functions says that the "voter influence" of coordinate i is precisely

$$\sum_{S\subseteq[n]:S\ni i} \left|\hat{f}(S)\right|^2.$$

Hence, while we do not establish a formal connection between small "voter influence" and strong spectral independence guarantees, our result Theorem D.2.1 says this is true at least morally.

We need the following zero-free result from [Bar17b].

Theorem D.2.2 ([Bar17b])**.** *Let $f : \{-1,1\}^n \to \mathbb{C}$ be given, and assume that*

$$\sqrt{\deg(f)} \cdot L(f) < C,$$

where $C \approx 0.55$ is an absolute constant, and $L(f) \stackrel{\text{def}}{=} \max_{i\in[n]} \sum_{S\subseteq[n]:S\ni i} \left|\hat{f}(S)\right|$. Then for every $\Lambda \subseteq [n]$ and every pinning $\sigma_\Lambda : \Lambda \to \{-1,1\}$, we have that the partition function of the associated Gibbs measure on $\{-1,1\}^n$ with pinning σ_Λ is nonzero:

$$\sum_{x\in\{-1,1\}^n : x|_\Lambda = \sigma_\Lambda} \exp(f(x)) \neq 0.$$

We now prove Theorem D.2.1.

Proof of Theorem D.2.1. By Theorem 6.4.1, it suffices to prove that the multivariate partition function

$$\sum_{x\in\{-1,1\}^n : x|_\Lambda = \sigma_\Lambda} \exp(f(x)) \prod_{i\in[n]:x_i=1} \lambda_i \tag{D.4}$$

is nonzero whenever λ lies in the set $\mathcal{D} = \{\lambda \in \mathbb{C}^n : |\lambda_i - 1| < c, \forall i \in [n]\}$ for all pinnings σ_Λ, where $c = c(\mathfrak{h}) > 0$ is a constant depending only on $\mathfrak{h}$ but not n. Define a new function $g : \{-1,1\}^n \to \mathbb{C}$ by

$$g(x) = f(x) + \sum_{i=1}^n \frac{1+x_i}{2} \log \lambda_i = f(x) + \sum_{i\in[n]:x_i=1} \log \lambda_i.$$

Then $\exp(g(x)) = \exp(f(x)) \prod_{i\in[n]:x_i=1} \lambda_i$ and the partition function $\sum_{x\in\{-1,1\}^n : x|_\Lambda=\sigma_\Lambda} \exp(g(x))$ associated with g is precisely our desired multivariate partition function Eq. (D.4). Our goal is to apply Theorem D.2.2 to g and deduce our desired stability statement.

First, it is clear from the definition of g that the Fourier coefficients of g are given by

$$\hat{g}(S) = \begin{cases} \hat{f}(S), & \text{if } |S| > 1; \\ \hat{f}(i) + \frac{1}{2}\log\lambda_i, & \text{if } S = \{i\} \text{ for some } i \in [n]; \\ \hat{f}(\emptyset) + \frac{1}{2}\sum_{i=1}^n \log\lambda_i, & \text{if } S = \emptyset. \end{cases}$$

It follows that

$$L(g) \le L(f) + \frac{1}{2} \max_{i\in[n]} |\log \lambda_i| .$$

Note that $\deg(g) = \deg(f)$ (unless $\deg(f) \le 1$, in which case spectral independence and rapid mixing is trivial). Hence, if $\lambda \in \mathbb{C}^n$ satisfies $|\log\lambda_i| < \mathfrak{h}$ for all $i \in [n]$, then rearranging yields precisely that $\sqrt{\deg(g)} \cdot L(g) < C$ and the zero-freeness follows from Theorem D.2.2. Furthermore, it is clear that the set $\{\lambda \in \mathbb{C}^n : |\log\lambda_i| < \mathfrak{h}, \forall i \in [n]\}$ contains $\mathcal{D}$ for a value of $c(\mathfrak{h}) > 0$ which depends only on $\mathfrak{h}$, just by continuity of the logarithm. □

D.3 Proofs of Zero-Free Results

In this section, we supply proofs of the two main zero-free statements Theorems D.1.2 and D.1.4 used for graph homomorphisms and tensor network contractions. 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 [Bar15; Bar17a; BD20; BB21] for applications besides those discussed in this paper.

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 [Bar16b]. See [Bar16a] for a proof.

Lemma D.3.1 (Angle Lemma)**.** *Let $x_1, \dots, x_n \in \mathbb{C}$ be nonzero complex numbers viewed as vectors in $\mathbb{R}^2$. Suppose there is an angle $0 \le \theta < 2\pi/3$ such that for all i, j, the angle between x_i, x_j is at most θ. Then we have the lower bound $|\sum_{i=1}^n x_i| \ge \cos(\theta/2) \sum_{i=1}^n |x_i|$.*

D.3.1 Proofs for Weighted Graph Homomorphisms

Our goal in this subsection is to prove Theorem D.1.2, i.e. that the weighted graph homomorphism partition function

$$\mathcal{Z}_\phi^S(A) \stackrel{\text{def}}{=} \sum_{\substack{\sigma:V\to[q]\\ \sigma|_S=\phi}} \prod_{uv\in E} A^{uv}(\sigma(u),\sigma(v))$$

is nonzero in a large polydisk around 1, where $S \subseteq V, \phi : S \to [q]$, and we view $\mathcal{Z}_\phi^S(A)$ 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 D.3.2 (Lemma 3.3 from [BS17a])**.** *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) $\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A) \neq 0$ 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

$$|\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A)| \geq \frac{\tau}{\Delta} \sum_{v\sim u} \sum_{k\in[q]} |A^{uv}(j,k)| \cdot \left| \frac{\partial}{\partial A^{uv}(j,k)} \mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A) \right|.$$

Then for every $u \in V \setminus S$ and every $j,k \in [q]$, the angle between $\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A)$ and $\mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A)$ 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 $\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A)$ and $\mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A)$ by

$$\left| \log \mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A) - \log \mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A) \right|. \tag{D.5}$$

The strategy is to write $\mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A)$ as $\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(\tilde{A})$ 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 $\mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A) = \mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(\tilde{A})$. By the Fundamental Theorem of Calculus, we may upper bound Eq. (D.5) by

$$\begin{aligned}
&\max_{B\in\mathcal{U}(\delta)} \sum_{v\sim u} \sum_{c\in[q]} \left| \frac{\partial}{\partial A^{uv}(j,c)} \log \mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(B) \right| \cdot \underbrace{|A^{uv}(j,c) - \tilde{A}^{uv}(j,c)|}_{\leq 2\delta \text{ since } A,\tilde{A}\in\mathcal{U}(\delta)} \\
&\leq \frac{2\delta}{1-\delta} \max_{B\in\mathcal{U}(\delta)} \underbrace{\sum_{v\sim u} \sum_{c\in[q]} |A^{uv}(j,c)| \cdot \frac{1}{|\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(B)|} \cdot \left| \frac{\partial}{\partial A^{uv}(j,c)} \mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(B) \right|}_{\leq \Delta/\tau \text{ by assumption (2)}} \\
&\leq \frac{2\delta\Delta}{\tau(1-\delta)}.
\end{aligned}$$

□

Lemma D.3.3 (Lemma 3.4 from [BS17a])**.** *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) $\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A) \neq 0$ for every $u \in V \setminus S$ and every $j \in [q]$;

(2) The angle between $\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A)$ *and* $\mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A)$ *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*

$$|\mathcal{Z}_\phi^S(A)| \geq \frac{\cos(\theta/2)}{\Delta} \sum_{v\sim u} \sum_{k\in[q]} |A^{uv}(\phi(u),k)| \cdot \left| \frac{\partial}{\partial A^{uv}(\phi(u),k)} \mathcal{Z}_\phi^S(A) \right|.$$

Proof. If $v \in S$ as well, then there is a unique $k \in [q]$ for which $\frac{\partial}{\partial A^{uv}(\phi(u),k)} \mathcal{Z}_\phi^S(A) \neq 0$, namely $k = \phi(v)$. In this case, $A^{uv}(\phi(u),k) \cdot \frac{\partial}{\partial A^{uv}(\phi(u),k)} \mathcal{Z}_\phi^S(A) = \mathcal{Z}_\phi^S(A)$. Otherwise, $v \notin S$ and $\frac{\partial}{\partial A^{uv}(\phi(u),k)} \mathcal{Z}_\phi^S(A) = \frac{1}{A^{uv}(\phi(u),k)} \cdot \mathcal{Z}_{\phi_{v,k}}^{S\cup\{v\}}(A)$, where $\phi_{v,k}$ is the unique extension of ϕ mapping v to k.

Combining these two observations, we obtain

$$\begin{aligned}
&\sum_{v\sim u} \sum_{k\in[q]} |A^{uv}(\phi(u),k)| \cdot \left| \frac{\partial}{\partial A^{uv}(\phi(u),k)} \mathcal{Z}_\phi^S(A) \right| \\
&= |N(u)\cap S| \cdot |\mathcal{Z}_\phi^S(A)| + \sum_{v\sim u: v\notin S} \sum_{k\in[q]} |\mathcal{Z}_{\phi_{v,k}}^{S\cup\{v\}}(A)| \\
&\leq |N(u)\cap S| \cdot |\mathcal{Z}_\phi^S(A)| + \frac{1}{\cos(\theta/2)} \underbrace{\left| \sum_{v\sim u: v\notin S} \sum_{k\in[q]} \mathcal{Z}_{\phi_{v,k}}^{S\cup\{v\}}(A) \right|}_{=|N(u)\setminus S|\cdot|\mathcal{Z}_\phi^S(A)|} \qquad \text{(Lemma D.3.1)} \\
&\leq \frac{\Delta}{\cos(\theta/2)} \cdot |\mathcal{Z}_\phi^S(A)|.
\end{aligned}$$

Rearranging yields the desired result. □

With these lemmas in hand, we can now prove the main zero-free result.

Proof of Theorem D.1.2. 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 $\mathcal{Z}_\phi^S(A) \neq 0$.

(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 $\mathcal{Z}_{\phi_{u,j}}^{S\cup\{u\}}(A)$ and $\mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A)$ 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

$$|\mathcal{Z}_\phi^S(A)| \geq \frac{\cos(\theta/2)}{\Delta} \sum_{v\sim u} \sum_{k\in[q]} |A^{uv}(\phi(u),k)| \cdot \left| \frac{\partial}{\partial A^{uv}(\phi(u),k)} \mathcal{Z}_\phi^S(A) \right|.$$

The base case $S = V$ is easily verified since $\mathcal{Z}_\phi^S(A) = \prod_{uv\in E} A^{uv}(\phi(u),\phi(v))$, a product of nonzero complex numbers.

Now, let $S \subseteq V$ with $|S| < |V|$.

Proof of (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 $\mathcal{Z}_\phi^S(A) = \sum_{k\in[q]} \mathcal{Z}_{\phi_{u,k}}^{S\cup\{u\}}(A)$, Lemma D.3.1 applied to $\mathcal{Z}_\phi^S(A)$ yields (i) assuming that (ii) holds. We prove (ii) below.

Proof of (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 D.3.2.

Proof of (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 D.3.3, 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 [Reg18], 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. □

D.3.2 Proofs for Tensor Network Contractions

Our goal in this subsection is to prove Theorem D.1.4, i.e. that the tensor network partition function

$$\mathcal{Z}_\phi^F(h) = \sum_{\substack{\sigma: E\to[q] \\ \sigma|_F=\phi}} \prod_{v\in V} h_v(\sigma\,|_{E(v)})$$

is nonzero in a large polydisk around 1, where $F \subseteq E, \phi : F \to [q]$, and we view $\mathcal{Z}_\phi^F(\cdot)$ as a polynomial with variables $\{h_v(\alpha)\}_{v,\alpha}$. We prove the following stronger result.

Theorem D.3.4 (Generalization of Theorem 6 from [Reg18]). *Let $G = (V, E)$ be a graph of maximum degree $\le \Delta$. Then for every $F \subseteq E$, $\phi : F \to [q]$, $\eta > 0$, and $0 \le \theta < 2\pi/3$, the function $\mathcal{Z}_\phi^F(h)$ is nonzero whenever $h \in \prod_{v\in V} S_v(\delta, \eta)$, where*

$$S_v(\delta,\eta) = \left\{h_v : [q]^{E(v)} \to \mathbb{C} : \substack{|h_v(\alpha)-h_v(\beta)|<\delta,\forall\alpha,\beta:E(v)\to[q], \\ |h_v(\alpha)|\ge\eta,\forall\alpha:E(v)\to[q]}\right\}$$

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 D.1.4.

Proof of Theorem D.1.4. Observe that $S_v(\delta,\eta)$ contains a disk around 1 of radius $\min\{\delta/2, 1-\eta\}$. Using Theorem D.3.4 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 [Reg18], 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 D.3.4. We will need the following lemmas to implement an inductive approach.

Lemma D.3.5 (Lemma 8 from [Reg18]). *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) $\mathcal{Z}_\psi^{F\cup E(u)}(h) \neq 0$;

(2) For all $v \in N(u) \cup \{u\}$, we have

$$|\mathcal{Z}_\psi^{F\cup E(u)}(h)| \ge \tau \sum_{\substack{\alpha: E(v)\to[q] \\ \text{compatible with } \psi}} |h_v(\alpha)| \cdot \left|\frac{\partial}{\partial h_v(\alpha)}\mathcal{Z}_\psi^{F\cup E(u)}(h)\right|.$$

Then for all extensions $\psi, \tilde\psi : F \cup E(u) \to \mathbb{C}$ of ϕ, the angle between $\mathcal{Z}_\psi^{F\cup E(u)}$ and $\mathcal{Z}_{\tilde\psi}^{F\cup E(u)}(h)$ 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 $\mathcal{Z}_\psi^{F\cup E(u)}(h)$ and $\mathcal{Z}_{\tilde\psi}^{F\cup E(u)}(h)$ by

$$\left|\log \mathcal{Z}_\psi^{F\cup E(u)}(h) - \log \mathcal{Z}_{\tilde\psi}^{F\cup E(u)}(h)\right|. \tag{D.6}$$

The strategy is to write $\mathcal{Z}_{\tilde\psi}^{F\cup E(u)}(h)$ as $\mathcal{Z}_\psi^{F\cup E(u)}(\tilde h)$ 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 $\mathcal{Z}_{\psi}^{F \cup E(u)}(h)$ 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 $\mathcal{Z}_{\tilde{\psi}}^{F \cup E(u)}(h) = \mathcal{Z}_{\psi}^{F \cup E(u)}(\tilde{h})$. By the Fundamental Theorem of Calculus , we may upper bound Eq. (D.6) by

$$\begin{aligned}
&\max_{x \in \prod_{v \in V} S_v(\delta,\eta)} \sum_{v \in N(u) \cup \{u\}} \sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \psi}} \left| \frac{\partial}{\partial h_v(\alpha)} \log \mathcal{Z}_{\psi}^{F \cup E(u)}(x) \right| \cdot \left| h_v(\alpha) - \tilde{h}_v(\alpha) \right| \\
&\leq \frac{\delta}{\eta} \max_{x \in \prod_{v \in V} S_v(\delta,\eta)} \underbrace{\sum_{v \in N(u) \cup \{u\}}}_{\leq \Delta + 1} \underbrace{\sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \psi}} |h_v(\alpha)| \cdot \frac{1}{|\mathcal{Z}_{\psi}^{F \cup E(u)}(x)|} \cdot \left| \frac{\partial}{\partial h_v(\alpha)} \mathcal{Z}_{\psi}^{F \cup E(u)}(x) \right|}_{\leq 1/\tau \text{ by assumption (2)}} \\
&\qquad \text{(Definition of } S_v(\delta,\eta)) \\
&\leq \frac{\delta(\Delta+1)}{\tau \eta}.
\end{aligned}$$

□

Lemma D.3.6 (Lemma 9 from [Reg18]). *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) $\mathcal{Z}_{\psi}^{F \cup E(v)}(h) \neq 0$;

(2) The angle between $\mathcal{Z}_{\psi}^{F \cup E(v)}(h)$ and $\mathcal{Z}_{\tilde{\psi}}^{F \cup E(v)}(h)$ 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

$$|\mathcal{Z}_{\phi}^{F}(h)| \geq \cos(\theta/2) \sum_{\substack{\alpha : E(v) \to [q] \\ \textit{compatible with } \phi}} |h_v(\alpha)| \cdot \left| \frac{\partial}{\partial h_v(\alpha)} \mathcal{Z}_{\phi}^{F}(h) \right|.$$

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 $\mathcal{Z}_{\phi}^{F}(h)$ and we can replace $\cos(\theta/2)$ by 1.

Suppose $v \in N(u)$. Since $\mathcal{Z}_{\phi}^{F}(h) = \sum_{\substack{\psi : F \cup E(v) \to [q] \\ \psi|_F = \phi}} \mathcal{Z}_{\psi}^{F \cup E(v)}(h)$, assumptions (1) and (2) make Lemma D.3.1 applicable, yielding

$$\begin{aligned}
|\mathcal{Z}_{\phi}^{F}(h)| &\geq \cos(\theta/2) \sum_{\substack{\psi : F \cup E(v) \to [q] \\ \psi|_F = \phi}} |\mathcal{Z}_{\psi}^{F \cup E(v)}(h)| \\
&= \cos(\theta/2) \sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \psi}} |h_v(\alpha)| \cdot \left| \frac{\partial}{\partial h_v(\alpha)} \mathcal{Z}_{\phi}^{F}(h) \right|
\end{aligned}$$

as desired. □

With these lemmas in hand, we may now proceed with the proof of Theorem D.3.4.

Proof of Theorem D.3.4. 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 $\mathcal{Z}_{\phi}^{F}(h) \neq 0$.

(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 $\mathcal{Z}_{\psi}^{F \cup E(u)}(h)$ and $\mathcal{Z}_{\tilde{\psi}}^{F \cup E(u)}(h)$ 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

$$|\mathcal{Z}_{\phi}^{F}(h)| \geq \cos(\theta/2) \sum_{\substack{\alpha : E(v) \to [q] \\ \text{compatible with } \phi}} |h_v(\alpha)| \cdot \left| \frac{\partial}{\partial h_v(\alpha)} \mathcal{Z}_{\phi}^{F}(h) \right|.$$

The base case $F = E$ is easily verified since $\mathcal{Z}_{\phi}^{F}(h) = \prod_{v \in V} h_v(\phi \mid_{E(v)})$, a product of nonzero complex numbers.

Now, let $F \subseteq E$ with $|F| < |E|$.

Proof of (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 $\mathcal{Z}_{\phi}^{F}(h) = \sum_{\substack{\psi : F \cup E(v) \to [q] \\ \psi|_F = \phi}} \mathcal{Z}_{\psi}^{F \cup E(v)}(h)$, Lemma D.3.1 applied to $\mathcal{Z}_{\psi}^{F \cup E(v)}(h)$ yields (i) assuming that (ii) holds. We prove (ii) below.

Proof of (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 D.3.5 to $F \cup E(u)$ then yields (ii).

Proof of (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 D.3.6, yielding (iii) for F. □

Appendix E

Influence Calculations for Spin Systems on Bethe Lattices

In this chapter, we calculate the influence and spectral independence of Gibbs distributions of various notable spin systems on *Bethe lattices*, i.e. the infinite d-regular tree $\mathbb{T}_d$ for $d \geq 2$. For many models, the infinite Δ-regular tree is often considered the "worst case" out of all graphs of maximum degree Δ. This has been made rigorous in several instances, most notably for the hardcore model due to Weitz [Wei06], and subsequently for antiferromagnetic two-state spin systems and more [LLY13]. The formulas we establish in this chapter can be thought of as spectral independence *lower bounds* for various models, although technically, our methods are not completely rigorous. At the very least, they yield highly plausible, quantitative predictions.

The infinite regular tree displays many beautiful properties. For instance, while it is a tree, it has ℓ_2-spectrum contained in the interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ [Car72; Fri91] (see also [MW89; HLW06]). They are, in a certain sense, optimal *expanders*, even though finite-sized trees are terrible expanders. The beautiful theory of *Ramanujan graphs* is derived from them. For us, a particularly convenient aspect of performing calculations on $\mathbb{T}_d$ is that, due to symmetry, it suffices to look only at the *univariate* tree recursion rather than the full multivariate tree recursion. Furthermore, we already previously saw in Chapter 7 that influences in trees factorize along paths; see Lemma 7.3.2. This will be crucial for our analysis.

Since we are working with infinite graphs and manipulating infinite matrices in this chapter, our calculations will not be fully rigorous. We have made no effort to rigorously define what spectrum means in this context. For instance, even though the ℓ_2-spectrum is exactly $[-2\sqrt{d-1}, 2\sqrt{d-1}]$, we will also consider $\pm d$ as "eigenvalues" of the adjacency operator of $\mathbb{T}_d$, since this is the case for finite (bipartite) d-regular graphs. The reason these "eigenvalues" are not in the ℓ_2-spectrum is that they naturally corresponding to (eigen)functions with ± 1 entries, which are not "square integrable" (or "ℓ_2-integrable"). We will treat all infinite matrices arising in this chapter as if they are finite matrices; we will only loosely justify the operations we perform.

Throughout, we write $\mathbb{T}_d = (V_{\mathbb{T}_d}, E_{\mathbb{T}_d})$ for the infinite d-regular tree with adjacency operator $A_{\mathbb{T}_d}$. By deleting a (root) vertex $r \in V_{\mathbb{T}_d}$, we are left with d copies of the infinite $(d-1)$-ary tree $\hat{\mathbb{T}}_{d-1} = (V_{\hat{\mathbb{T}}_{d-1}}, E_{\hat{\mathbb{T}}_{d-1}})$ rooted at the respective neighbors of r. For all models we consider in this chapter, we will work in the *tree uniqueness regime*, i.e. the regime of parameters in which there is a unique Gibbs measure for the system on $\mathbb{T}_d$. We will often denote this unique Gibbs measure simply as μ. This is (hopefully) with little risk of confusion, since we have divided the chapter into sections corresponding to each model we analyze, which are all completely independent of each other.

E.1 Vertex Two-Spin Systems

Recall that a two-spin system on vertices of a graph $G = (V, E)$ is specified by the parameters β, γ, λ, with univariate partition function

$$\mathcal{Z}_G(\lambda) = \mathcal{Z}_{G,\beta,\gamma}(\lambda) = \sum_{\sigma: V \to \{0,1\}} \beta^{m_0(\sigma)} \gamma^{m_1(\sigma)} \lambda^{n_0(\sigma)}$$

where $m_0(\sigma)$ (resp. $m_1(\sigma)$) denotes the number of edges with endpoints both mapped to 0 (resp. 1) under σ, and $n_0(\sigma)$ denotes the number of vertices mapped to 0. Furthermore, recall the univariate tree recursion

$$f_d(R) = \lambda \left(\frac{\beta R + 1}{R + \gamma} \right)^d$$

where recall the variables R denote the *marginal ratios*. It is easy to see by monotonicity of f_d that it has a unique fixed point $\hat{R}_d$, and it is well-known that $\left| f'_{d-1}(\hat{R}_{d-1}) \right|$ is intimately connected with the decay of correlation phenomenon. In particular, it is known that if $\left| f'_{d-1}(\hat{R}_{d-1}) \right| < 1$, then there is a *unique* Gibbs measure $\mu = \mu_{\mathbb{T}_d,\beta,\gamma,\lambda}$ for the two-spin system on the infinite d-regular tree $\mathbb{T}_d$. With the probability measure specified, it makes sense to speak of marginal probabilities of vertices, marginal ratios, influences, etc.

We prove the following.

Theorem E.1.1 (Two-Spin Systems on $\mathbb{T}_d$)**.** *Fix integers $d \geq 2$ and any (β, γ, λ) such that $\left| f'_{d-1}\left(\hat{R}_{d-1}\right) \right| = 1 - \delta < 1$ for some $0 < \delta < 1$. Then the Gibbs distribution μ of the two-spin system on $\mathbb{T}_d$ with parameters (β, γ, λ) satisfies*

$$\lambda_{\max}(\Psi_\mu) - 1 = \|\Psi_\mu\|_\infty - 1 = \frac{d}{d-1} \cdot \left(\frac{1}{\delta} - 1 \right).$$

Note that the corresponding correlation decay prediction from Chapter 7 and Appendix A is optimal up to (small) constant factors. The corresponding "(eigen)function" (which is not square integrable) is the all-ones function when the system is ferromagnetic. When the system is antiferromagnetic, this function is instead the ± 1-function which alternates in sign based on the parity of the distance to some arbitrarily fixed reference vertex.

It turns out that we can, in fact, completely characterize the spectrum of Ψ_μ in terms of the spectrum of the adjacency matrix of $\mathbb{T}_d$. We will additionally need the following two intermediate claims.

Lemma E.1.2 (Influence Along Neighboring Vertices)**.** *If $u \sim v$ in $\mathbb{T}_d$ and μ is the unique Gibbs distribution on $\mathbb{T}_d$, then*

$$\Psi_\mu(u \to v) = \frac{1}{d-1} \cdot f'_{d-1}\left(\hat{R}_{d-1}\right).$$

Proposition E.1.3 (Inverse of Exponential Distance Matrix in Trees; Proposition 3.3 in [BLP06])**.** *Let T be a tree with adjacency matrix A_T and degree matrix D_T, and let x be a real number not equal to 1 or -1. Finally, define the matrix $C_T(x) \in \mathbb{R}^{V_T \times V_T}$ via*

$$C_T(x)_{uv} = \begin{cases} x^{\mathrm{dist}_T(u,v)}, & \text{if } u \neq v \\ 0, & \text{otherwise.} \end{cases}$$

Then we have the matrix identity

$$(\mathsf{Id} + C_T(x))^{-1} = \mathsf{Id} - \frac{x}{1-x^2} \cdot A_T + \frac{x^2}{1-x^2} \cdot D_T.$$

We will prove Lemma E.1.2 in a moment. First, let us see how these tools imply Theorem E.1.1.

Proof of Theorem E.1.1. For convenience, let $x = \frac{1}{d-1} \cdot f'_{d-1}\left(\hat{R}_{d-1}\right)$. Observe that $\Psi_\mu - \mathsf{Id} = C_{\mathbb{T}_d}(x)$ by Lemmas 7.3.2 and E.1.2. Hence, by Proposition E.1.3, we have that

$$(I + C_{\mathbb{T}_d}(x))^{-1} = \left(1 + \frac{dx^2}{1-x^2}\right) \cdot \mathsf{Id} - \frac{x}{1-x^2} \cdot A_{\mathbb{T}_d},$$

the point being that the right-hand side is significantly easier to understand spectrally. It follows that

$$C_{\mathbb{T}_d}(x) = \left(\left(1 + \frac{dx^2}{1-x^2}\right) \cdot \mathsf{Id} - \frac{x}{1-x^2} \cdot A_{\mathbb{T}_d} \right)^{-1} - \mathsf{Id}.$$

Now, let α be an eigenvalue of $A_{\mathbb{T}_d}$. Then,

$$\left(\left(1+\frac{dx^2}{1-x^2}\right)-\frac{\alpha x}{1-x^2}\right)^{-1}-1=\frac{1-x^2}{1-\alpha x+(d-1)x^2}-1=\frac{\alpha x-dx^2}{1-\alpha x+(d-1)x^2} \tag{E.1}$$

is an eigenvalue of $C_{\mathbb{T}_d}(x)$. This function is clearly monotone increasing in α if $x \geq 0$, and monotone decreasing in α if $x \leq 0$. Hence, in either case, the maximum is attained at

$$\frac{d\cdot|x|-d\cdot x^2}{1-d\cdot|x|+(d-1)\cdot x^2}=\frac{d\cdot|x|\cdot(1-|x|)}{(1-(d-1)\cdot|x|)(1-|x|)}=\frac{d\cdot|x|}{1-(d-1)\cdot|x|}$$

As $|x|=\frac{1}{d-1}\cdot\left|f'_{d-1}\left(\hat{R}_{d-1}\right)\right|=\frac{1}{d-1}\cdot(1-\delta)$, the eigenvalue formula follows.

For $\|\Psi_\mu\|_\infty$, note by symmetry that it suffices to compute the total absolute influence of an arbitrarily chosen vertex u. Again using Lemmas 7.3.2 and E.1.2, we have

$$\begin{aligned}\sum_{v\in V_{\mathbb{T}_d}:v\neq u}|\Psi_\mu(u\to v)| &= \sum_{v\in\mathbb{T}_d:v\neq u}\left|\frac{f'(\hat{R})}{d-1}\right|^{\operatorname{dist}(u,v)}\\ &=\sum_{k=1}^{\infty}\left(\frac{1-\delta}{d-1}\right)^k\cdot\underbrace{\#\{v\in V_{\mathbb{T}_d}:\operatorname{dist}(u,v)=k\}}_{=d(d-1)^{k-1}}\\ &=\frac{d}{d-1}\sum_{k=1}^{\infty}(1-\delta)^k\\ &=\frac{d}{d-1}\cdot\left(\frac{1}{\delta}-1\right).\end{aligned}$$

As previously mentioned, one can explicitly construct an "eigenfunction" $\varphi: V_{\mathbb{T}_d}\to\mathbb{R}$ such that $\Psi_\mu\cdot\varphi=\frac{d}{d-1}\cdot\left(\frac{1}{\delta}-1\right)\cdot\varphi$. This eigenfunction is not square integrable as its entries are all ±1, but is the natural analog of the top (or bottom, depending on the sign of x) eigenvector for finite bipartite d-regular graphs. If the system is ferromagnetic, i.e. $x\geq 0$, we can simply take the all-ones function. If the system is antiferromagnetic, i.e. $x\leq 0$, then fix an arbitrary root vertex r, and define

$$\varphi(v)=\begin{cases}+1, & \text{if } \operatorname{dist}(r,v) \text{ is even}\\ -1, & \text{otherwise.}\end{cases} \tag{E.2}$$

Up to a global change in sign, this function φ is independent of the choice of r. Crucially, if $\operatorname{dist}(r,u)$ is even and $\operatorname{dist}(u,v)$ is even, then $\operatorname{dist}(r,v)$ is also even, which implies $(-1)^{\operatorname{dist}(r,u)+\operatorname{dist}(u,v)}=(-1)^{\operatorname{dist}(r,v)}$. A straightforward calculation similar to the above calculation of $\|\Psi_\mu\|_\infty$ then yields $\Psi_\mu\cdot\varphi=\frac{d}{d-1}\cdot\left(\frac{1}{\delta}-1\right)\cdot\varphi$. □

Proof of Lemma E.1.2. For convenience, write $\hat{R}$ for $\hat{R}_{d-1}$ and f for f_{d-1}. We calculate that

$$\begin{aligned}R_v^{u\leftarrow 0}&=\beta\lambda\prod_{w\sim v:w\neq u}\frac{\beta\hat{R}+1}{\hat{R}+\gamma}=\beta f(\hat{R})=\beta\hat{R}\\ R_v^{u\leftarrow 1}&=\frac{1}{\gamma}\lambda\prod_{w\sim v:w\neq u}\frac{\beta\hat{R}+1}{\hat{R}+\gamma}=\frac{1}{\gamma}f(\hat{R})=\frac{1}{\gamma}\hat{R}.\end{aligned}$$

Here, we crucially used that the subtree rooted at each neighbor of v is the infinite $(d-1)$-ary tree $\hat{\mathbb{T}}_{d-1}$, from which the fixed point of tree recursion f_{d-1} gives the correct marginal ratio.

Converting the marginal ratios back into the marginal probabilities, it follows that

$$\Psi_\mu(u\to v)=\frac{R_v^{u\leftarrow 0}}{R_v^{u\leftarrow 0}+1}-\frac{R_v^{u\leftarrow 1}}{R_v^{u\leftarrow 1}+1}=\frac{\beta\hat{R}}{\beta\hat{R}+1}-\frac{\hat{R}}{\hat{R}+\gamma}=\frac{(\beta\gamma-1)\cdot\hat{R}}{(\beta\hat{R}+1)(\hat{R}+\gamma)}.$$

On the other hand, we compute that

$$f'(R)=\frac{(d-1)(\beta\gamma-1)}{(\beta R+1)(R+\gamma)}\cdot\lambda\left(\frac{\beta R+1}{R+\gamma}\right)^{d-1}=(d-1)\frac{\beta\gamma-1}{(\beta R+1)(R+\gamma)}f(R).$$

By definition of $\hat{R}$, we have $f\left(\hat{R}\right) = \hat{R}$, so

$$f'\left(\hat{R}\right) = (d-1)\frac{(\beta\gamma - 1)\cdot \hat{R}}{(\beta\hat{R}+1)(\hat{R}+\gamma)} = (d-1)\cdot \Psi_\mu(u,v).$$

The claim follows. □

E.2 Monomer-Dimer Model

Recall that the monomer-dimer model is the two-spin system with parameters $(0, 1, \lambda)$ on the line graph of $G = (V, E)$, i.e. it is a weighted distribution over all matchings (edge configurations) on G. It is well-known that no phase transition occurs in this model (e.g. via Heilmann–Lieb [HL72]), so for every $\lambda \geq 0$, there is a unique Gibbs distribution μ for the monomer-dimer over $\mathbb{T}_d$ for any $\lambda \geq 0$. Our goal is to establish a similar formula for the maximum eigenvalue of the influence matrix for this model.

First, note that since the monomer-dimer model is the hardcore model but on a special class of graphs, one could try to apply the tree recursion for the hardcore model. However, this is more challenging, since the line graph of a tree is no longer a tree. Instead, there is another tree recursion which only exists for the monomer-dimer model rather than the hardcore model more generally. Rather than being a recursion for the marginals or marginal ratios of edges, it is a recursion on the marginal probability that a given vertex has an incident edge in the random matching.

Let $\mathcal{Z}_G(\lambda) = \sum_{\text{matching } F\subseteq E} \lambda^{|F|}$ denote the univariate partition function of the Gibbs distribution $\mu = \mu_{G,\lambda}$ of the monomer-dimer model on $G = (V, E)$. For an edge $e \in E$, we abuse notation and abbreviate e for the event that $e \in F \sim \mu$. Similarly, we abbreviate $\overline{e}$ for the event that $e \notin F \sim \mu$. For an arbitrary vertex $r \in V$ with incident edges $e_1, \ldots, e_d$, define

$$p_G(\overline{r}) \stackrel{\text{def}}{=} \Pr_{\mu}[\overline{e}_1, \ldots, \overline{e}_d] = \frac{\mathcal{Z}_{G-r}(\lambda)}{\mathcal{Z}_G(\lambda)}$$

to be the probability that no edge incident to r is in a randomly drawn matching $F \sim E$. If T is a tree rooted at r, and $v_1, \ldots, v_d$ are the children of r which are the roots of corresponding subtrees $T_1, \ldots, T_d$, then Eq. (B.3) says that

$$p_T(\overline{r}) = \frac{1}{1 + \lambda\sum_{i=1}^{d} p_{T_i}(\overline{v}_i)}. \tag{E.3}$$

In particular, if $\hat{T}_{d,k}$ denotes the d-ary tree of height k, and $p_k = p_{\hat{T}_{d,k}}(\overline{r})$, then by symmetry,

$$p_k = f(p_{k-1}) \quad \text{where} \quad f(x) = \frac{1}{1 + (d-1)\lambda x}. \tag{E.4}$$

This is the tree recursion we will use in our analysis.

We prove the following for the maximum eigenvalue of the pairwise edge influence matrix, employing a very similar strategy as in Appendix E.1.

Theorem E.2.1. *For all integers $d \geq 2$ and all real numbers $\lambda \geq 0$, the Gibbs distribution μ of the monomer-dimer model on $\mathbb{T}_d$ satisfies*

$$\boldsymbol{\lambda}_{\max}(\Psi_\mu) - 1 = -\frac{2x}{1+x}, \qquad \textit{and}$$
$$\|\Psi_\mu\|_\infty - 1 = 2\cdot\frac{(d-1)\cdot|x|}{1-(d-1)\cdot|x|} = \sqrt{1+4\lambda(d-1)} - 1 = \Theta\left(\sqrt{\lambda d}\right),$$

where $x = -\frac{1}{d-1}\left(1 - \frac{2}{\sqrt{1+4\lambda(d-1)}+1}\right)$.

Remark 69. We note that in the case that $d = 2$, we obtain that $\boldsymbol{\lambda}_{\max}(\Psi_\mu) = \sqrt{1+4\lambda}-1 = \Theta(\sqrt{\lambda})$, which blows up to ∞ as $\lambda \to \infty$. This reflects the fact that the Glauber dynamics does not mix rapidly for sampling perfect matchings on a long path nor a cycle. However, curiously, we have $\boldsymbol{\lambda}_{\max}(\Psi_\mu) \leq O(1/d)$ for $d \geq 3$, independent of λ.

Remark 70. Looking at the total absolute influence of an edge yields $\|\Psi_\mu\|_\infty = \Theta\left(\sqrt{\lambda d}\right)$ for all λ, d. This suggests that the correlation decay prediction from Appendix B is very *suboptimal* except in the case $d = 2$.

To prove this, we use the following tools.

Lemma E.2.2 (Influence Along Neighboring Edges)**.** *For edges $e \sim f$ in $\mathbb{T}_d$ (i.e. e, f share a single endpoint), we have the identity*

$$\Psi_\mu(e \to f) = -\frac{1}{d-1}\left(1 - \frac{2}{\sqrt{1+4\lambda(d-1)}+1}\right)$$

Proposition E.2.3 (Inverse of Exponential Distance Matrix in Line Graphs of Trees; see Theorem 6 in [BS13])**.** *Let T be a tree with line graph $L(T)$, and for each vertex $v \in T$, let $E(v)$ denote the set of edges incident to v in T; note that $E(v)$ corresponds to a clique in $L(T)$. Now, let $x \in \mathbb{R}$, and define the matrix $C_{L(T)}(x) \in \mathbb{R}^{E_T \times E_T}$ by*

$$C_{L(T)}(x)_{ef} = \begin{cases} x^{\mathrm{dist}_{L(T)}(e,f)}, & \text{if } e \neq f \\ 0, & o.w. \end{cases}$$

Furthermore, for each $v \in T$, define $C_{L(T)}(x)_v \in \mathbb{R}^{E_T \times E_T}$ to be the submatrix of $C_{L(T)}(x)$ with rows and columns in $E(v)$, padded with zeros. We also take Id_v to be the identity matrix on $E(v)$, padded with zeros. Then we have the following identity.

$$\left(\mathsf{Id} + C_{L(T)}(x)\right)^{-1} = -\mathsf{Id} + \sum_{v \in V_T}\left(\mathsf{Id}_v + C_{L(T)}(x)_v\right)^{-1}$$

Here, we note that the matrix $\mathsf{Id}_v + C_{L(T)}(x)_v$ is meant to be inverted on the block $E(v)$, which can also be interpreted as the Moore-Penrose pseudoinverse on all of E_T.

Remark 71. This is a very special case of a more general result on product distance matrices defined over *block graphs*. We refer interested readers to [BS13].

We will prove Lemma E.2.2 in a moment. Let us first see how to prove Theorem E.2.1.

Proof of Theorem E.2.1. Again, for convenience, let $x = -\frac{1}{d-1}\left(1 - \frac{2}{\sqrt{1+4\lambda(d-1)}+1}\right)$. Then $\Psi_\mu - \mathsf{Id} = C_{L(\mathbb{T}_d)}(x)$ by Lemmas B.0.10 and E.2.2; note the latter is the analog of Lemma 7.3.2 for the monomer-dimer model. From now on, we will write $C(x)$ for $C_{L(\mathbb{T}_d)}(x)$. It follows by Proposition E.2.3 that

$$C(x) = \left(-\mathsf{Id} + \sum_{v \in V_T}\left(\mathsf{Id}_v + C(x)_v\right)^{-1}\right)^{-1} - \mathsf{Id},$$

the point being that the right-hand side is significantly easier to understand spectrally. Now, observe that $C(x)_v = x \cdot (\mathbf{1}_v\mathbf{1}_v^\top - \mathsf{Id}_v)$, where $\mathbf{1}_v$ is the indicator vector of the edges incident to v, so that $\mathsf{Id}_v + C(x)_v = (1-x)\mathsf{Id}_v + x\mathbf{1}_v\mathbf{1}_v^\top$. By the Sherman-Morrison formula, we have

$$\begin{aligned}
\left(\mathsf{Id}_v + C(x)_v\right)^{-1} &= \left(x\left(\frac{1-x}{x}\mathsf{Id}_v + \mathbf{1}_v\mathbf{1}_v^\top\right)\right)^{-1} \\
&= \frac{1}{x}\cdot\left(\frac{x}{1-x}\mathsf{Id}_v - \frac{\left(\frac{x}{1-x}\right)^2\mathbf{1}_v\mathbf{1}_v^\top}{1+\frac{x}{1-x}d}\right) \\
&= \frac{1}{1-x}\left(\mathsf{Id}_v - \frac{x}{1+(d-1)x}\mathbf{1}_v\mathbf{1}_v^\top\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
\sum_{v \in V_T}\left(\mathsf{Id}_v + C(x)_v\right)^{-1} &= \frac{2}{1-x}\mathsf{Id} - \frac{x}{(1-x)(1+(d-1)x)}\sum_{v \in V_{\mathbb{T}_d}}\mathbf{1}_v\mathbf{1}_v^\top \\
&= \frac{2}{1-x}\mathsf{Id} - \frac{x}{(1-x)(1+(d-1)x)}BB^\top
\end{aligned}$$

where B is the matrix with columns $\mathbf{1}_v$ for each $v \in V_{\mathbb{T}_d}$. It follows that

$$C(x) = \left(\frac{1+x}{1-x}\mathsf{Id} - \frac{x}{(1-x)(1+(d-1)x)}BB^\top\right)^{-1} - \mathsf{Id}.$$

Furthermore, if α is an eigenvalue of $BB^\top$, then

$$\mathfrak{h}(\alpha) = \left(\frac{1+x}{1-x} - \frac{\alpha x}{(1-x)(1+(d-1)x)}\right)^{-1} - 1$$

is an eigenvalue of $C(x)$. Furthermore, since $-\frac{1}{d-1} < x \le 0$, we have $\mathfrak{h}(\alpha)$ is a monotone decreasing function in α. Now, observe that $BB^\top$ has the same spectrum as $B^\top B = dI + A_{\mathbb{T}_d}$. Hence, $BB^\top$ has eigenvalue 0 (since $A_{\mathbb{T}_d}$ has eigenvalue $-d$ by bipartiteness and regularity of $\mathbb{T}_d$). Furthermore, 0 is the smallest eigenvalue of $BB^\top$ since $BB^\top$ is positive semidefinite. It follows that

$$\lambda_{\max}(\Psi_\mu) = \lambda_{\max}(C(x)) = \frac{1-x}{1+x} - 1 = -\frac{2x}{1+x},$$

yielding desired the eigenvalue formula.

Now, we compute $\|\Psi_\mu\|_\infty$. By symmetry, it suffices to compute the total absolute influence of an arbitrarily chosen edge e. Again using Lemmas B.0.10 and E.2.2, we have

$$\begin{aligned}\sum_{f \in E_{\mathbb{T}_d} : f \neq e} |\Psi_\mu(e \to f)| &= \sum_{k=1}^{\infty} |x|^k \cdot \underbrace{\#\{f \in E_{\mathbb{T}_d} : \operatorname{dist}(e,f) = k\}}_{=2(d-1)^k} \\ &= 2 \cdot \frac{(d-1)\cdot|x|}{1-(d-1)\cdot|x|} \\ &= \sqrt{1+4\lambda(d-1)} - 1\end{aligned}$$

as desired, where we plug in $|x| = \frac{1}{d-1}\left(1 - \frac{2}{\sqrt{1+4\lambda(d-1)}+1}\right)$ in the last step. □

Proof of Lemma E.2.2. Observe that

$$\Psi_\mu(e \to f) = -\Pr_{F \sim \mu}[f \in F \mid e \notin F]$$

since $e \sim f$. Again, we abbreviate the event $f \in F$ as simply f, and the event $e \notin F$ as simply $\overline{e}$, where $F \sim \mu$ is a random matching. The right-hand side and similar probabilities will then be written as $\Pr_G[f \mid \overline{e}]$ to highlight the underlying graph G we are working with.

As we've conditioned e to not be in the matching, $\Pr_{\mathbb{T}_d}[f \mid \overline{e}]$ is the same as $\Pr_{\hat{\mathbb{T}}_{d-1}}[f]$, where f is incident to the root vertex r of $\hat{\mathbb{T}}_{d-1}$. To compute this probability, let $e_1, \dots, e_{d-1}$ denote the edges incident to r. Since the events $e_i \in F \sim \mu$ are all pairwise disjoint, and at most one of the edges e_i can be in a given matching, by symmetry of $\hat{\mathbb{T}}_{d-1}$, we have

$$\Pr_{\hat{\mathbb{T}}_{d-1}}[f] = \frac{1}{d-1}\Pr_{\hat{\mathbb{T}}_{d-1}}[e_1 \vee \cdots \vee e_{d-1}] = \frac{1}{d-1}\left(1 - \Pr_{\hat{\mathbb{T}}_{d-1}}[\overline{e}_1 \wedge \cdots \wedge \overline{e}_{d-1}]\right).$$

Hence, it suffices to show that

$$\Pr_{\hat{\mathbb{T}}_{d-1}}[\overline{e}_1 \wedge \cdots \wedge \overline{e}_{d-1}] = \frac{2}{\sqrt{1+4\lambda(d-1)}+1}.$$

For this, recall $p_k = p_{\hat{T}_{d-1,k}}(\overline{r})$ where $\hat{T}_{d-1,k}$ is the $(d-1)$-ary tree with depth k rooted at r. We aim to show that p_k converges to the above probability as $k \to \infty$ and compute the limit.

Eq. (E.4) shows that $p_k = f(p_{k-1})$ where $f(x) = \frac{1}{1+(d-1)\lambda x}$ gives the univariate recursion. f has a unique positive fixed point $\hat{p} = \hat{p}_{d,\lambda}$ which satisfies the equation $(d-1)\lambda\hat{p}^2 + \hat{p} - 1 = 0$. Solving this quadratic, we obtain

$$\hat{p} = \frac{\sqrt{1+4\lambda(d-1)}-1}{2\lambda(d-1)} = \frac{2}{\sqrt{1+4\lambda(d-1)}+1}$$

as desired. Note that we always have convergence to the unique fixed point, since

$$|f'(\hat{p})| = 1 - \frac{\sqrt{1+4\lambda(d-1)}-1}{2\lambda(d-1)} < 1.$$

Hence, $p_k \to \hat{p} = \Pr_{\hat{\mathbb{T}}_{d-1}}[\overline{e}_1 \wedge \cdots \wedge \overline{e}_{d-1}]$ as $k \to \infty$, yielding the desired formula. □

E.3 Proper Vertex q-Colorings

In this section, consider multi-spin systems, in particular, the uniform distribution $\mu = \mu_{G,q}$ over proper vertex-colorings with q colors, where $q \geq 2$ is an integer. This is also known as the antiferromagnetic q-state Potts model at zero temperature. If $q \geq d+1$, [Jon02] proved that there is the Gibbs measure μ on $\mathbb{T}_d$ is unique.

Since we are considering multi-state spin systems, we will consider the notion of influence matrix Ψ_μ in Claim 2.5.4. For concreteness, Ψ_μ is comprised of $n \times n$ block submatrices indexed by pairs of vertices $u, v \in V$. For each pair of distinct vertices $u \neq v$, the corresponding $q \times q$ block submatrix is has entries

$$\Psi_\mu^{u\to v}(c \to c') = \Pr_{\chi\sim\mu}[\chi(v) = c' \mid \chi(u) = c] - \Pr_{\chi\sim\mu}[\chi(v) = c'], \quad \forall c, c' \in [q].$$

Recall the diagonal $q \times q$ blocks of Ψ_μ are identity matrices.

We prove the following.

Theorem E.3.1. *For integers $d \geq 2$ and $q \geq d+1$, the uniform distribution μ over proper q-colorings of $\mathbb{T}_d$ satisfies*

$$\lambda_{\max}(\Psi_\mu) - 1 = \frac{d}{q-d} \quad \text{and} \quad \|\Psi_\mu\|_\infty - 1 = \frac{2(q-1)}{q} \cdot \frac{d}{q-d}.$$

Remark 72. It is interesting to note that $\|\cdot\|_\infty$ is off by just a factor of 2, independent of d, q. Furthermore, this suggests that as q approaches the maximum degree Δ of the graph (e.g. $q = \Delta + O(1)$; see Conjecture 4), the spectral independence of the distribution should be at least on the order of Δ.

To prove this, we will need the following tools.

Lemma E.3.2 (Influence Along Neighboring Vertices)**.** *For every $u \sim v$ in $\mathbb{T}_d$, we have the identity*

$$\Psi_\mu^{u\to v} = -\frac{1}{q-1}\left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right).$$

Proof. The marginal probability of a given vertex being assigned some color is $\frac{1}{q}$ by symmetry. The same argument shows that the conditional marginal probabilities are $\frac{1}{q-1}$, except when the colors assigned to the u, v are equal, in which case the probability is 0. □

Lemma E.3.3 (Influence Factorization in Trees)**.** *Fix a tree T and two distinct vertices $u, v \in T$, and suppose $w \neq u, v$ is a vertex that is on the unique path between u, v in T. If $\mu = \mu_{T,q}$ is the uniform distribution over proper q-colorings of T, then*

$$\Psi_\mu^{u\to v} = \Psi_\mu^{u\to w} \cdot \Psi_\mu^{w\to v}.$$

Proof. This is a straightforward analog of Lemmas 7.3.2 and B.0.10 above, and can be proved in the same manner by applying conditional independence. □

With these tools in hand, we now give our first proof of Theorem E.3.1, which is much simpler than our second proof below.

First Proof of Theorem E.3.1. By Lemmas E.3.2 and E.3.3, Ψ_μ is a block matrix with $\Psi_\mu^{u\to v} = W^{\mathrm{dist}(u,v)}$ when $\mathrm{dist}(u,v) \geq 1$, where $W = -\frac{1}{q-1}\left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)$. Now, observe that $W^k = (-1)^k\frac{1}{(q-1)^k}\left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)$ for $k \geq 1$ simply because $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ is a projection matrix, so that

powering it doesn't change the matrix. It follows that we may express Ψ_μ as the tensor product $B_{\mathbb{T}_d} \otimes \left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)$ where $B_{\mathbb{T}_d}$ is the infinite matrix with the same dimensions as $A_{\mathbb{T}_d}$ and entries

$$B_{\mathbb{T}_d}(u,v) = \begin{cases} \left(\frac{-1}{q-1}\right)^{\operatorname{dist}(u,v)}, & \text{if } u \neq v \\ 0, & \text{otherwise.} \end{cases}$$

Since $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ has eigenvalues 0 and 1 (with multiplicities 1 and $q-1$ respectively), and the eigenvalues of tensor products are given by products of eigenvalues of the constituent matrices, we have

$$\lambda_{\max}(\Psi_\mu) = \lambda_{\max}\left(B_{\mathbb{T}_d} \otimes \left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)\right) = \lambda_{\max}(B_{\mathbb{T}_d}).$$

Following the proof of Theorem E.1.1, using Proposition E.2.3, we have the eigenvalues of $B_{\mathbb{T}_d}$ are precisely given by

$$\frac{-\frac{1}{q-1}\alpha - \frac{d}{(q-1)^2}}{1 + \frac{1}{q-1}\alpha + \frac{d-1}{(q-1)^2}},$$

where α is an eigenvalue of $A_{\mathbb{T}_d}$; see Eq. (E.1). Since this function is monotone decreasing in α, we plug in $\alpha = -d$ to obtain

$$\frac{\frac{d}{q-1} - \frac{d}{(q-1)^2}}{1 - \frac{d}{q-1} + \frac{d-1}{(q-1)^2}} = \frac{d(q-2)}{(q-1)^2 - d(q-1) + (d-1)} = \frac{d}{q-d}$$

as desired. Note that similar Theorem E.1.1, we can produce an "eigenfunction" with this eigenvalue by taking $\varphi \otimes \psi$, where $\varphi : V_{\mathbb{T}_d} \to \{\pm 1\}$ is defined in Eq. (E.2), and $\psi \in \mathbb{R}^q$ is any eigenvector of $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ with eigenvalue 1, i.e. any vector $\psi \in \mathbb{R}^q$ satisfying $\langle \mathbf{1}_q, \psi\rangle = 0$.

Now, we compute $\|\Psi_\mu\|_\infty$. We again use Lemmas E.3.2 and E.3.3 with the fact that $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ is a projection matrix. Now, by symmetry, it suffices to compute the total absolute influence of an arbitrarily chosen vertex-color pair uc. Again using Lemma E.3.3, we have

$$\begin{aligned}
\sum_{vc' : v \neq u} |\Psi_\mu(uc \to vc')| &= \sum_{v \in V_{\mathbb{T}_d} : v \neq u} \sum_{c' \in [q]} \left|\Psi_\mu^{u \to v}(c \to c')\right| \\
&= \sum_{v \in V_{\mathbb{T}_d} : v \neq u} \sum_{c' \in [q]} W^{\operatorname{dist}(u,v)}(c, c') \\
&= \sum_{k=1}^{\infty} \sum_{v \in V_{\mathbb{T}_d} : \operatorname{dist}(u,v) = k} \sum_{c' \in [q]} \frac{1}{(q-1)^k} \cdot \underbrace{\left|\left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)(c, c')\right|}_{= \frac{2(q-1)}{q}} \\
&= \frac{2(q-1)}{q} \sum_{k=1}^{\infty} \frac{1}{(q-1)^k} \cdot \underbrace{\#\{v \in V_{\mathbb{T}_d} : \operatorname{dist}(u,v) = k\}}_{= d(d-1)^{k-1}} \\
&= \frac{2(q-1)}{q} \cdot \frac{d}{d-1} \cdot \sum_{k=1}^{\infty} \left(1 - \frac{q-d}{q-1}\right)^k \\
&= \frac{2(q-1)}{q} \cdot \frac{d}{d-1} \cdot \left(\frac{q-1}{q-d} - 1\right) \\
&= \frac{2(q-1)}{q} \cdot \frac{d}{q-d}
\end{aligned}$$

as desired. Note the fact that $\|\Psi_\mu\|_\infty = \frac{2(q-1)}{q} \cdot \lambda_{\max}(\Psi_\mu)$ is explained by $\left\|\mathsf{Id}_q - \mathbf{1}_q\mathbf{1}_q^\top\right\|_\infty = \frac{2(q-1)}{q}$. □

E.3.1 A More Robust Proof

The calculations in our first proof of Theorem E.3.1 were greatly simplified due to the fact that the matrix $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ in $\Psi_\mu^{u \to v}$ (see Lemma E.3.2) is a projection matrix. In particular, it gave

Ψ_μ a particularly nice tensor product structure. In this subsection, we give a more robust proof which eschews this happy coincidence, with the hope that it could be useful in future influence calculations on Bethe lattices. Instead, we make use of the following useful and generic tool.

Proposition E.3.4 (Inverse of Product Distance Matrix in Trees with Matrix Weights; see Theorem 4 from [BS15a])**.** *Let T be a tree with adjacency matrix A_T and degree matrix D_T. Fix a positive integer $q \geq 1$, and suppose for each edge e in T, we have a $q \times q$ "weight matrix" W_e. Define $C \in \mathbb{R}^{nq \times nq}$ as a block matrix where*

$$C(u,v) = \prod_{i=1}^{k} W_{e_i} \in \mathbb{R}^{q \times q}, \quad \forall \text{ vertices } u, v,$$

and $e_1, \ldots, e_k$ is the unique path of edges from u to v; by convention, we set $C(u,u) = 0$ for all vertices u. Then

$$(\mathsf{Id} + C)^{-1} = \mathsf{Id}_{nq} - \mathsf{Id}_q \otimes D_T + \sum_{e \in E_{\mathbb{T}_d}} M_e$$

where for each $e = uv \in E_T$, $M_e \in \mathbb{R}^{nq \times nq}$ is the block matrix with

$$M_e(u,u) = M_e(v,v) = \left(\mathsf{Id}_q - W_e^2\right)^{-1}$$
$$M_e(u,v) = M_e(v,u) = -W_e \left(\mathsf{Id}_q - W_e^2\right)^{-1}$$

and zeros everywhere else.

With this additional tool in hand, we now give our second proof of Theorem E.3.1.

Second Proof of Theorem E.3.1. Observe that by Lemmas E.3.2 and E.3.3, we have $\Psi_\mu - \mathsf{Id} = C$ with $W_e = W = -\frac{1}{q-1}\left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)$. It follows by Proposition E.3.4 that

$$C = \left(-(d-1)\mathsf{Id} + \sum_{e \in E} M_e\right)^{-1} - \mathsf{Id},$$

where M_e is defined as in Proposition E.3.4, the point being that the right-hand side is significantly easier to understand spectrally. Let us now compute the M_e. Observe that $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ is the orthogonal projection onto the subspace orthogonal to $\mathbf{1}_q$. Combining this fact with the Sherman-Morrison formula, we have

$$\begin{aligned}
\left(\mathsf{Id}_q - W^2\right)^{-1} &= \left(\left(1 - \frac{1}{(q-1)^2}\right)\mathsf{Id}_q + \frac{1}{q(q-1)^2}\mathbf{1}_q\mathbf{1}_q^\top\right)^{-1} \\
&= \left(1 - \frac{1}{(q-1)^2}\right)^{-1}\mathsf{Id}_q - \frac{\left(1 - \frac{1}{(q-1)^2}\right)^{-2} \cdot \frac{1}{q(q-1)^2}}{1 + \frac{1}{(q-1)^2}\left(1 - \frac{1}{(q-1)^2}\right)}\mathbf{1}_q\mathbf{1}_q^\top \\
&= \frac{(q-1)^2}{q(q-2)}\left(\mathsf{Id}_q - \frac{1}{q(q-1)^2}\mathbf{1}_q\mathbf{1}_q^\top\right) \\
&= \mathsf{Id}_q + \frac{1}{q(q-2)}\left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right),
\end{aligned}$$

which implies that

$$-W\left(\mathsf{Id}_q - W^2\right)^{-1} = -\frac{q-1}{q(q-2)}\left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right).$$

With this in hand, we calculate that

$$\sum_{e \in E_{\mathbb{T}_d}} M_e = d \cdot \mathsf{Id} + \frac{q-1}{q(q-2)} \cdot \left(A_{\mathbb{T}_d} + \frac{d}{q-1}\mathsf{Id}\right) \otimes \left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)$$

Altogether, we obtain that

$$C = \left(\mathsf{Id} + \frac{q-1}{q(q-2)}\left(A_{\mathbb{T}_d} + \frac{d}{q-1}\mathsf{Id}\right) \otimes \left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)\right)^{-1} - \mathsf{Id}.$$

In particular, for each eigenvalue α of $\left(A_{\mathbb{T}_d} + \frac{d}{q-1}\mathsf{Id}\right) \otimes \left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)$, we obtain the eigenvalue $\left(1 + \frac{q-1}{q(q-2)}\alpha\right)^{-1} - 1$ of C. By monotonicity, we want to use the smallest such α. The eigenvalues

$$\left(A_{\mathbb{T}_d} + \frac{d}{q-1}\mathsf{Id}\right) \otimes \left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)$$

are the products of pairs of eigenvalues of $A_{\mathbb{T}_d} + \frac{d}{q-1}\mathsf{Id}$ and of $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$. Since $\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ is a projection matrix, it has eigenvalues 1 and 0, and so

$$\mathrm{Spec}\left(\left(A_{\mathbb{T}_d} + \frac{d}{q-1}\mathsf{Id}\right) \otimes \left(\mathsf{Id}_q - \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top\right)\right) = \{0\} \cup \left(\mathrm{Spec}\left(A_{\mathbb{T}_d}\right) + \frac{d}{q-1}\right),$$

where $\mathrm{Spec}(A)$ denotes the (multi)set of all eigenvalues of a matrix A. The minimum eigenvalue of $A_{\mathbb{T}_d}$ is $-d$ by regularity and bipartiteness of $\mathbb{T}_d$. Hence, we take $\alpha = -d + \frac{d}{q-1} = -\left(1 - \frac{1}{q-1}\right) \cdot d = -\frac{q-2}{q-1} \cdot d$. Plugging this in, we obtain

$$\boldsymbol{\lambda}_{\max}\left(\Psi_\mu\right) = \boldsymbol{\lambda}_{\max}(C) = \left(1 - \frac{q-1}{q(q-2)} \cdot \frac{q-2}{q-1} \cdot d\right)^{-1} - 1 = \frac{d}{q-d}.$$

The calculation of $\|\Psi_\mu\|_\infty$ is same as in the first proof we gave above. □

Appendix F

Miscellaneous Results and Observations

In this chapter, we collect some additional miscellaneous results and observations which we find interesting.

F.1 Characterization of Partite Strongly Log-Concave Polynomials

In this section, we show that strongly log-concave polynomials which are "partite" must factorize as a product of linear forms.[1] Formulated probabilistically, this says that for a probability distribution μ over a discrete product space, say, the discrete hypercube $\{0,1\}^n$, μ is $(0,\dots,0)$-spectrally independent if and only if μ is a product measure.

Theorem F.1.1. *Let μ be a probability distribution over a discrete product space $\prod_{i=1}^n \Sigma_i$ for some nonempty finite sets $\Sigma_1,\dots,\Sigma_n$ with $|\Sigma_i| \geq 2$ for all $i = 1,\dots,n$. Then the following are equivalent.*

1. *μ is $(0,\dots,0)$-spectrally independent. Said in the language of polynomials, the homogeneous multivariate generating polynomial*

$$g_\mu(x) = \sum_{\sigma \in \prod_{i=1}^n \Sigma_i} \mu(\sigma) \prod_{i=1}^n x_{i,\sigma(i)}$$

   for μ is strongly log-concave (at $\mathbf{1}$).

2. *μ may be expressed as a product measure $\bigotimes_{i=1}^n \nu_i$, where each ν_i is some probability distribution over Σ_i. Said in the language of polynomials, g_μ factorizes as a product of linear forms $\prod_{i=1}^n \left(\sum_{s\in\Sigma_i} \nu_i(s)\cdot x_{i,s}\right)$.*

Remark 73. One should be very careful about what "spectral independence" means in this case. Indeed, if μ is a distribution supported on $\binom{[n]}{k}$ for some $0 \leq k \leq n$, then $(0,\dots,0)$-spectral independence in the sense of $\lambda_{\max}(\mathcal{I}_\mu - \mathsf{Id}) \leq 0$ for all conditional distributions is certainly possible. A simple example would be the uniform distribution over all of $\binom{[n]}{k}$. However, μ is very far from being a product measure just by virtue of its support being homogeneous, even though one can view μ as a distribution on $\{0,1\}^n \supseteq \binom{[n]}{k}$. This seems like a contradiction.

The resolution is to observe that $(0,\dots,0)$-spectral independence for μ over $\binom{[n]}{k}$ is equivalent to log-concavity of the k-homogeneous polynomial

$$\sum_{S\in\binom{[n]}{k}} \mu(S) \prod_{i\in S} x_i,$$

[1] Here, "partiteness" just refers to the fact that there is a partition of the variables such that every monomial in the polynomial is formed by picking exactly one variable from each part of the partition.

while in order to apply Theorem F.1.1 above, one would need log-concavity of the n-homogeneous polynomial

$$\sum_{S \in \binom{[n]}{k}} \mu(S) \prod_{i \in S} x_i \prod_{i \notin S} y_i.$$

In terms of influence matrices, to apply Theorem F.1.1, one would need $\lambda_{\max}(\Psi_\mu - \mathsf{Id}) \leq 0$ by viewing μ as a nonhomogeneous distribution on all of $\{0,1\}^n$; a bound of $\lambda_{\max}(\mathcal{I}_\mu - \mathsf{Id}) \leq 0$ wouldn't be enough. This illustrates one rather important difference between Ψ_μ and $\mathcal{I}_\mu$.

To prove this, we reduce to the simplest nontrivial case. The rest of the argument will follow by induction.

Proof of the Case $n = 2$ and $|\Sigma_1| = |\Sigma_2| = 2$. For convenience, we may assume without loss of generality that $\Sigma_1 = \Sigma_2 = \{0,1\}^n$. So, μ is a distribution on $\{0,1\}^2$, and we may write

$$g_\mu(x_1, y_1, x_2, y_2) = \zeta x_1(px_2 + (1-p)y_2) + (1-\zeta)y_1(qx_2 + (1-q)y_2) \tag{F.1}$$

where $\zeta = \Pr_{\sigma \sim \mu}[\sigma_1 = 1]$ gives the marginal of the first coordinate, and $p = \Pr_{\sigma \sim \mu}[\sigma_2 = 1 \mid \sigma_1 = 1], q = \Pr_{\sigma \sim \mu}[\sigma_2 = 1 \mid \sigma_1 = 0]$ give the marginals of the second coordinate conditioned on the first coordinate. We show that if g_μ is log-concave (at $\mathbf{1}$), then $p = q$ so that we have the desired factorization $g_\mu(x_1, y_1, x_2, y_2) = (\zeta x_1 + (1-\zeta)y_1)(px_2 + (1-p)y_2)$. The reverse implication is obvious, since any product of linear forms is log-concave.

The claim is immediate if $\zeta = 0$ or $\zeta = 1$ so we may assume $0 < \zeta < 1$. Computing the Hessian of g_μ, we have

$$\nabla^2 g_\mu = \begin{bmatrix} 0 & A \\ A^\top & 0 \end{bmatrix} \qquad \text{where} \qquad A = \begin{bmatrix} \zeta p & \zeta(1-p) \\ (1-\zeta)q & (1-\zeta)(1-q) \end{bmatrix}.$$

Hence, the four eigenvalues of $\nabla^2 g_\mu$ are of the form $\pm\sqrt{\lambda}$ where λ is an eigenvalue of

$$\begin{aligned} AA^\top &= \begin{bmatrix} \zeta^2(p^2 + (1-p)^2) & \zeta(1-\zeta)(pq + (1-p)(1-q)) \\ \zeta(1-\zeta)(pq + (1-p)(1-q)) & (1-\zeta)^2(q^2 + (1-q)^2) \end{bmatrix} \\ &= \begin{bmatrix} \zeta & 0 \\ 0 & 1-\zeta \end{bmatrix} \underbrace{\begin{bmatrix} p^2 + (1-p)^2 & pq + (1-p)(1-q) \\ pq + (1-p)(1-q) & q^2 + (1-q)^2 \end{bmatrix}}_{=B} \begin{bmatrix} \zeta & 0 \\ 0 & 1-\zeta \end{bmatrix} \end{aligned}$$

Since g_μ is log-concave, $\nabla^2 g_\mu$ has at most one positive eigenvalue by Proposition 5.0.3. This forces $AA^\top$ to have rank at most 1, which in turn forces B to have rank at most 1. Since B is the Gram matrix

$$B = \begin{bmatrix} \langle u, u \rangle & \langle u, v \rangle \\ \langle u, v \rangle & \langle v, v \rangle \end{bmatrix},$$

where $u = [p, 1-p] \in \mathbb{R}^2_{\geq 0}, v = [q, 1-q] \in \mathbb{R}^2_{\geq 0}$ both are necessarily nonzero, this forces $u = v$, i.e. $p = q$ as desired. □

Proof of the Case $n = 2$. Now, we analyze the case $n = 2$ but allow Σ_1, Σ_2 to be arbitrary. Again, we may write g_μ as

$$g_\mu(x) = \sum_{s \in \Sigma_1} \mu_1(s) \cdot x_{1,s} \cdot \left(\sum_{t \in \Sigma_2} \mu_2^{1 \leftarrow s}(t) \cdot x_{2,t} \right) \tag{F.2}$$

where μ_1 is the marginal distribution of the first coordinate and each $\mu_2^{1 \leftarrow s}$ is the marginal distribution of the second coordinate conditioned on the first coordinate being assigned $s \in \Sigma_1$. We show that if g_μ is log-concave (at $\mathbf{1}$), then for every $t \in \Sigma_2$, $\mu_2^{1 \leftarrow s}(t) = \mu_2^{1 \leftarrow s'}(t)$ for all $s, s' \in \Sigma_1$. This will directly imply the desired factorization. Again, the converse is immediate.

Fix some $t \in \Sigma_2$ and some pair of distinct $s, s' \in \Sigma_1$. Set all $x_{2,t'}$ for $t' \neq t$ to the same variable x_2, and set all $x_{1,s''} = 0$ for $s'' \neq s, s'$. The resulting polynomial is log-concave and is of the form Eq. (F.1) (up to scaling by a constant). It follows from the case $|\Sigma_1| = |\Sigma_2| = 2$ analyzed above that $\mu_2^{1 \leftarrow s}(t) = \mu_2^{1 \leftarrow s'}(t)$. Since $t \in \Sigma_2$ and $s, s' \in \Sigma_1$ were arbitrary, we are done. □

Proof of the General Case. We go by induction on n. The base case $n = 2$ was already analyzed above. Suppose the claim holds for some $n \geq 2$. We prove the claim for $n + 1$. Suppose g_μ is strongly log-concave (at $\mathbf{1}$). Since for every $s \in \Sigma_1$, the polynomial $\partial_{x_{1,s}} g_\mu$ is also strongly log-concave and has degree $\leq n$, it factorizes as a product of linear forms by induction. In particular, we may write

$$g_\mu(x) = \sum_{s \in \Sigma_1} \mu_1(s) \cdot x_{1,s} \cdot \prod_{i=2}^{n+1} \left(\sum_{t \in \Sigma_i} \mu_i^{1 \leftarrow s}(t) \cdot x_{i,t} \right)$$

Again, we show that for every $i = 2, \ldots, n+1$ and every pair $s, s' \in \Sigma_1$, the conditional marginal distributions $\mu_i^{1 \leftarrow s}, \mu_i^{1 \leftarrow s'}$ on Σ_i are equal. We prove this using the same idea as above. For all $j \neq i$, set $x_{j,r} = 1$ for all $r \in \Sigma_j$. The resulting log-concave polynomial has the form Eq. (F.2), and so $\mu_i^{1 \leftarrow s} = \mu_i^{1 \leftarrow s'}$ follows from the case $n = 2$ analyzed above. Since $i = 2, \ldots, n+1$ and $s, s' \in \Sigma_1$ were arbitrary, we are done. □

Remark 74. With Pietro Caputo and Justin Salez, we observed that an alternative way to prove Theorem F.1.1 is the following. First, use the fact that $(0, \ldots, 0)$-spectral independence implies optimal $n \leftrightarrow 1$ contraction of relative entropy (i.e. entropic independence [Ana+22c]), which in the context of discrete product spaces is the same as *subadditivity of entropy*. Then, apply the main result of [CC09].

F.2 Spectral Independence via Log-Concavity of Inhomogeneous Polynomials

In this section, we again study the spectral independence of probability distributions μ on $\{0,1\}^n$. We saw from Theorem 5.0.1 in Chapter 5 that if the homogeneous multivariate generating polynomial

$$g_\mu(x_1, \ldots, x_n, y_1, \ldots, y_n) = \sum_{\sigma \subseteq [n]} \mu(\sigma) \prod_{i \in \sigma} x_i \prod_{i \notin \sigma} y_i$$

is strongly log-concave at $\mathbf{1}$, then μ is $(0, \ldots, 0)$-spectrally independent. Now, unfortunately by Theorem F.1.1, we can only have log-concavity for polynomials of the above form if the distribution μ is a product measure. However, in some settings where μ is not a product measure, we are still able to get strong log-concavity of the *inhomogeneous* polynomial

$$\tilde{g}_\mu(x_1, \ldots, x_n) = \sum_{\sigma \subseteq [n]} \mu(\sigma) \prod_{i \in \sigma} x_i.$$

Our goal in this section is to study what kind of spectral independence one can get from this assumption. An application to the random cluster measure when $0 < q < 1$ is given in Appendix F.3.

Theorem F.2.1. *Let μ be a probability distribution over $\{0,1\}^n$.*

1. *If the inhomogeneous multivariate polynomial*

$$\tilde{g}_\mu(x_1, \ldots, x_n) = \sum_{\sigma \subseteq [n]} \mu(\sigma) \prod_{i \in \sigma} x_i$$

is log-concave at $\mathbf{1}$, then we have the bound

$$\lambda_{\max}(\Psi_\mu) - 1 \leq \max_{i=1,\ldots,n} \frac{\Pr_{\sigma \sim \mu}[i \in \sigma]}{\Pr_{\sigma \sim \mu}[i \notin \sigma]}.$$

2. *Similarly, if the inhomogeneous multivariate polynomial*

$$\tilde{g}_{\overline{\mu}}(y_1, \ldots, y_n) = \sum_{\sigma \subseteq [n]} \mu(\sigma) \prod_{i \notin \sigma} y_i$$

is log-concave at $\mathbf{1}$, then we have the bound

$$\lambda_{\max}(\Psi_\mu) - 1 \leq \max_{i=1,\ldots,n} \frac{\Pr_{\sigma \sim \mu}[i \notin \sigma]}{\Pr_{\sigma \sim \mu}[i \in \sigma]}.$$

Proof. We only prove the first claim, as the proof of the second is similar. Similar to the claim and proof of Corollary 5.1.2, observe that

$$\begin{aligned}\partial_i \log \tilde{g}_\mu(\mathbf{1}) &= \Pr_{\sigma\sim\mu}[i \in \sigma], \quad \forall i = 1, \dots, n\\ \partial_i^2 \log \tilde{g}_\mu(\mathbf{1}) &= -\Pr_{\sigma\sim\mu}[i \in \sigma]^2, \quad \forall i = 1, \dots, n\\ \partial_i \partial_j \log \tilde{g}_\mu(\mathbf{1}) &= \Pr_{\sigma\sim\mu}[i, j \in \sigma] - \Pr_{\sigma\sim\mu}[i \in \sigma] \cdot \Pr_{\sigma\sim\mu}[j \in \sigma], \quad \forall i \neq j.\end{aligned}$$

Since

$$\Pr_{\sigma\sim\mu}[i, j \in \sigma] - \Pr_{\sigma\sim\mu}[i \in \sigma] \cdot \Pr_{\sigma\sim\mu}[j \in \sigma] = \Pr_{\sigma\sim\mu}[i \in \sigma] \cdot \Pr_{\sigma\sim\mu}[i \notin \sigma] \cdot \Psi_\mu(i \to j), \quad \forall i \neq j,$$

the matrix $\Psi_\mu - \mathsf{Id}$ is equal to

$$\operatorname{diag}\left(\frac{1}{\Pr_{\sigma\sim\mu}[i \in \sigma] \cdot \Pr_{\sigma\sim\mu}[i \notin \sigma]}\right)_{i=1}^{n} \cdot \nabla^2 \log \tilde{g}_\mu(\mathbf{1}) + \operatorname{diag}\left(\frac{\Pr_{\sigma\sim\mu}[i \in \sigma]}{\Pr_{\sigma\sim\mu}[i \notin \sigma]}\right)_{i=1}^{n}.$$

Since the first matrix is negative semidefinite by log-concavity of $\tilde{g}_\mu$, the upper bound follows. □

Remark 75. One unfortunate feature of Theorem F.2.1 is that even if μ is a product measure, this cannot recover 0-spectral independence.

F.3 Further Analysis of the Random Cluster Model when $0 < q < 1$

In this section, we perform additional analysis on the random cluster model when $0 < q < 1$. We first complete the proof of Theorem 5.4.1. The polarization idea we will employ was already one we saw in Chapter 12.

Proof of Theorem 5.4.1. For convenience, we write μ for $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ in this proof. By Theorem 5.4.3, the homogeneous polynomial $h_{\mathcal{M},q}$ is strongly log-concave on $\mathbb{R}^{n+1}_{\geq 0}$. Unfortunately, this polynomial is not multiaffine, so we first polarize it via Proposition 5.3.3 to obtain the homogeneous multiaffine polynomial

$$h^{\uparrow}_{\mathcal{M},q}(x_1, \dots, x_n, y_1, \dots, y_n) = \sum_{\sigma \subseteq [n]} \sum_{S \in \binom{[n]}{n-|\sigma|}} \frac{1}{\binom{n}{n-|\sigma|}} q^{-\operatorname{rank}(\sigma)} \prod_{i \in \sigma} x_i \prod_{j \in S} y_j.$$

Effectively, we have enlarged the ground set $\mathscr{U}$ by introducing n new "dummy" elements; we write $\overline{\mathscr{U}}$ for the collection of these dummy elements. Up to scaling by a constant, $h^{\uparrow}_{\mathcal{M},q}$ is the homogeneous multivariate generating polynomial of the distribution $\mu^{\uparrow} = \left(\mu^{\mathsf{RC}}_{\mathcal{M},p,q}\right)^{\uparrow}$ on $2^{\mathscr{U} \sqcup \overline{\mathscr{U}}}$ given by

$$\mu^{\uparrow}(\sigma \sqcup S) = \frac{1}{\binom{n}{n-|\sigma|}} \cdot \mu(\sigma), \quad \forall \sigma \subseteq \mathscr{U}, S \in \binom{\overline{\mathscr{U}}}{n - |\sigma|}.$$

This distribution "projects" down to μ under the natural mapping defined by $\sigma \sqcup S \mapsto \sigma$. The Markov chain $\mathcal{W}$ we will use is the action of the down-up walk for $\mu^{\uparrow}$ under this projection. More specifically, for every $\sigma, \sigma' \subseteq \mathscr{U}$, the transition probabilities are given by

$$\mathcal{W}(\sigma \to \sigma') \stackrel{\mathsf{def}}{=} \frac{1}{\mu(\sigma)} \sum_{S \in \binom{\overline{\mathscr{U}}}{n-|\sigma|}, S' \in \binom{\overline{\mathscr{U}}}{n-|\sigma'|}} \mu^{\uparrow}(\sigma \sqcup S) \cdot \mathcal{P}_{\mu^{\uparrow}}(\sigma \sqcup S \to \sigma' \sqcup S'). \tag{F.3}$$

In a moment, we will establish $O(n \log n)$-mixing for $\mathcal{W}$. Let us first roughly understand its transitions. There are three "types" of transitions under this Markov chain.

1. *Lazy Steps*: $\mathcal{W}$ doesn't change state under the projection. This stems both from the fact that the down-up walk $\mathcal{P}_{\mu^{\uparrow}}$ for $\mu^{\uparrow}$ is lazy, as well as from the moves which remove a dummy element in $\overline{\mathscr{U}}$ and add back some other dummy element in $\overline{\mathscr{U}}$.

2. *Glauber-Type Updates*: $\mathcal{W}$ selects a random element in $\mathscr{U}$ and either adds it to the current state or removes it. This is achieved in $\mathcal{P}_{\mu^\uparrow}$ by corresponding removing or adding a dummy element in $\overline{\mathscr{U}}$.

3. *Swap Updates*: $\mathcal{W}$ selects an element $i \in \sigma$ and another element $j \notin \sigma$, and transitions to $\sigma \setminus \{i\} \cup \{j\}$.

It immediately follows that $\mathcal{W}$ is a local Markov chain, in the sense that it updates the assignments of at most two elements of $\mathscr{U}$ in each step. In [Mou22], the Markov chain $\mathcal{P}_{\mu^\uparrow}$ is called the *exchange walk*, since its transitions involve not only single-site updates but also these additional "swap" (or "exchange") updates. We refer interested readers to [Mou22] for more precise calculations of the transition probabilities. We will not need these calculations for our mixing time analysis.

Observe that since $h^{\uparrow}_{\mathcal{M},q}$ is strongly log-concave on $\mathbb{R}^n_{\geq 0}$, the distribution $\mu^\uparrow$ is $(0,\dots,0)$-spectrally independent by Theorem 5.0.1. By Theorem 11.0.1, the modified log-Sobolev constant of $\mathcal{P}_{\mu^\uparrow}$ is at least $1/n$. We use this to bound the mixing time of $\mathcal{W}$ by showing that $\varrho(\mathcal{W}) \geq 1/n$ as well, and then appealing to Theorem 9.1.1.

Let $f : 2^{\mathscr{U}} \to \mathbb{R}_{\geq 0}$ be an arbitrary function, and extend this to a function $f^\uparrow : 2^{\mathscr{U} \sqcup \overline{\mathscr{U}}} \to \mathbb{R}_{\geq 0}$ via $f^\uparrow(\sigma \sqcup S) = f(\sigma)$ for all $\sigma \subseteq \mathscr{U}, S \in \binom{\overline{\mathscr{U}}}{n-|\sigma|}$. If f is normalized to satisfy $\mathbb{E}_\mu(f) = 1$ so that f is the density of some probability measure ν on $2^{\mathscr{U}}$, then $f^\uparrow$ is the density of the measure $\nu^\uparrow$ on $2^{\mathscr{U} \sqcup \overline{\mathscr{U}}}$ given by

$$\nu^\uparrow(\sigma \sqcup S) = \frac{1}{\binom{n}{n-|\sigma|}} \cdot \nu(\sigma), \quad \forall \sigma \subseteq \mathscr{U}, S \in \binom{\overline{\mathscr{U}}}{n-|\sigma|}.$$

A straightforward calculation reveals that

$$\begin{aligned} \mathcal{E}_{\mathcal{P}_{\mu^\uparrow}}\left(f^\uparrow, \log f^\uparrow\right) &= \mathcal{E}_{\mathcal{W}}(f, \log f) \\ \operatorname{Ent}_{\mu^\uparrow}\left(f^\uparrow\right) &= \operatorname{Ent}_\mu(f). \end{aligned}$$

It follows that

$$\frac{\mathcal{E}_{\mathcal{W}}(f, \log f)}{\operatorname{Ent}_\mu(f)} = \frac{\mathcal{E}_{\mathcal{P}_{\mu^\uparrow}}\left(f^\uparrow, \log f^\uparrow\right)}{\operatorname{Ent}_{\mu^\uparrow}\left(f^\uparrow\right)} \geq \varrho\left(\mathcal{P}_{\mu^\uparrow}\right) \geq \frac{1}{n}.$$

As this holds for all $f : 2^{\mathscr{U}} \to \mathbb{R}_{\geq 0}$, we conclude that $\varrho(\mathcal{W}) \geq 1/n$. □

F.3.1 Glauber Dynamics for the Random Cluster model when $0 < q < 1$

Again, one may wonder about the usual Glauber dynamics for sampling from the full random cluster measure $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$. We give a preliminary analysis here.

Theorem F.3.1 (Spectral Independence for Full Random Cluster when $0 < q \leq 1$)**.** *Let $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ be an n-element matroid with rank function $\operatorname{rank} : 2^{\mathscr{U}} \to \mathbb{Z}_{\geq 0}$, and fix parameters $0 < q \leq 1$ and $0 \leq p \leq 1$. Then the random cluster measure $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ is $(\eta_0, \dots, \eta_{n-2})$-spectrally independent where*

$$\eta_k \leq \min\left\{\frac{1}{\sqrt{q}}, \frac{p(1-p)(1-q)}{p+q(1-p)}(n-k-1)\right\}, \quad \forall 0 \leq k \leq n-2.$$

In particular, the Glauber dynamics for $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ mixes rapidly.

As far as we are aware, this result hasn't been written down anywhere. Combining this spectral independence result with Theorem 2.3.1, we see that the spectral gap of the Glauber dynamics is at least $\Omega\left(1/n^{1+1/\sqrt{q}}\right)$. Hence, the mixing time is at most $O\left(n^{2+1/\sqrt{q}}\right)$. Notice the spectral independence we establish is *independent* of the choice of p. This actually implies a notion known as *fractional log-concavity* [Ali+21], which further implies a strengthening of spectral independence called *entropic independence* [Ana+22c]. One can use entropic independence to lower bound the rate of entropy decay, as well as the modified logarithmic Sobolev constant. The techniques of [Ana+22c] combined with Theorem F.3.1 and the techniques in Chapter 9 imply the following mixing time. We give a proof sketch at the end of the section.

Corollary F.3.2. *Let $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ be an n-element matroid with rank function $\operatorname{rank} : 2^{\mathscr{U}} \to \mathbb{Z}_{\geq 0}$, and fix parameters $0 < q \leq 1$ and $0 \leq p \leq 1$. Then the Glauber dynamics for sampling from the random cluster measure $\mu = \mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ has modified log-Sobolev constant $\varrho(\mathcal{P}_\mu) \geq \Omega_{p,q}(1/n^{1+1/\sqrt{q}})$ and mixing time $O_{p,q}(n^{1+1/\sqrt{q}} \log n)$.*

Remark 76. As was previously mentioned, in an independent work, Guo–Mousa [Mou22] have shown $O_{p,q}(n^2 \log n)$-mixing of the Glauber dynamics, where only the constant depends on the parameters p, q. In particular, they showed that the modified log-Sobolev constant of $\mathcal{P}^2_\mu$ satisfies the lower bound $\varrho(\mathcal{P}^2_\mu) \geq \frac{1}{\left(1+\frac{p}{q(1-p)}\right)\cdot n^2}$. Thus, their mixing time bound supersedes ours. An interesting direction is to see if the mixing time can be further improved, perhaps by combining their methods with the spectral independence we obtain.

The proof uses the following fact concerning the marginals of the random cluster measure, which is simple to prove.

Fact F.3.3 (Marginal Bounds for Random Cluster Measures; see e.g. [Gri09])**.** *Let $\mathcal{M} = (\mathscr{U}, \mathcal{X})$ be any matroid, and let $0 \leq p \leq 1$ and $q > 0$. Then for every element $i \in \mathscr{U}$, the marginal probability of i under the random cluster measure $\mu = \mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ satisfies the bounds*

$$\min\left\{p, \frac{p}{p+q(1-p)}\right\} \leq \Pr_{\sigma\sim\mu}[i \in \sigma] \leq \max\left\{p, \frac{p}{p+q(1-p)}\right\}.$$

Proof. By the Law of Total Probability, the marginal probability $\Pr_{\sigma\sim\mu}[\sigma_i = 1]$ is a convex combination of conditional marginal probabilities $\Pr_{\sigma\sim\mu}[\sigma_i = 1 \mid \sigma_{-i} = \tau]$ over all $\tau : \mathscr{U} \setminus \{i\} \to \{0,1\}$. Hence, it suffices to show that for every such τ, $\Pr_{\sigma\sim\mu}[\sigma_i = 1 \mid \sigma_{-i} = \tau] \in \left\{p, \frac{p}{p+q(1-p)}\right\}$. This holds simply because conditioning on τ corresponds to contracting the elements in $\tau^{-1}(1)$ and deleting the elements in $\tau^{-1}(0)$. In the resulting matroid, i is either a loop, which gives probability p, or a nonloop, which gives probability $\frac{p}{p+q(1-p)}$. □

Note that the $\frac{p(1-p)(1-q)}{p+q(1-p)}(n-k-1)$ bound in Theorem F.3.1 immediately follows from Fact F.3.3 just by controlling each entry of Ψ_μ and bounding $\|\Psi_\mu\|_\infty$. All that remains is to establish the $\frac{1}{\sqrt{q}}$-bound.

Proof of Theorem F.3.1. Given what we saw in Chapter 5, one natural strategy for establishing spectral independence for $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ would be to establish strong log-concavity of its multivariate generating polynomial, which is equal to the following (up to normalization by a constant):

$$\sum_{\sigma\subseteq[n]} q^{-\operatorname{rank}(\sigma)} \prod_{i\in\sigma} x_i \prod_{i\in[n]\setminus\sigma} y_i.$$

Unfortunately, this clearly cannot be the case, since this would imply that $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ is in fact $(0,\ldots,0)$-spectrally independent, which is false. A simple counterexample would be the matroid consisting simply of two parallel elements; see also Theorem F.1.1.

Instead, we will use Theorem F.2.1, since log-concavity of $h_{\mathcal{M},q}$ (see Eq. (5.7) and Theorem 5.4.3) implies that

$$\sum_{\sigma\subseteq[n]} q^{-\operatorname{rank}(\sigma)} \prod_{i\in\sigma} x_i$$

is completely log-concave on all of $\mathbb{R}^n_{\geq 0}$. Combined with Fact F.3.3, this already establishes the bound $\frac{p}{q(1-p)}$. To finish the proof, we will use a matroid duality trick to establish the bound $\frac{1-p}{p}$. Once we have these two bounds, the desired $\frac{1}{\sqrt{q}}$-bound immediately follows by maximizing $\min\left\{\frac{p}{q(1-p)}, \frac{1-p}{p}\right\}$ over p. This is achieved at the well-known *self-dual point* $p_{\mathsf{sd}}(q) \stackrel{\mathsf{def}}{=} \frac{\sqrt{q}}{1+\sqrt{q}}$, which can be shown by observing $\frac{p}{q(1-p)}$ (resp. $\frac{1-p}{p}$) is monotone decreasing (resp. increasing) in p, setting them equal, and solving for p.

For the $\frac{1-p}{p}$-bound, observe that by Theorem F.2.1 and Fact F.3.3, it suffices to show that the polynomial

$$\sum_{\sigma\subseteq[n]} q^{-\operatorname{rank}(\sigma)} \prod_{i\notin\sigma} y_i$$

is also completely log-concave on all of $\mathbb{R}^n_{\geq 0}$. We use matroid duality. Since $\operatorname{rank}_{\mathcal{M}^*}(S) = \operatorname{rank}_{\mathcal{M}}([n] \setminus S) + |S| - \operatorname{rank}_{\mathcal{M}}([n])$ (see Eq. (1.5)), we have that

$$h_{\mathcal{M}^*,q}(1, y_1, \ldots, y_n) = q^r \sum_{\sigma \subseteq [n]} q^{-\operatorname{rank}(\sigma)} \prod_{i \notin \sigma} \frac{y_i}{q},$$

from which the desired log-concavity follows again from Theorem 5.4.3. □

Proof Sketch of Corollary F.3.2. The $\frac{1}{\sqrt{q}}$-bound on the spectral independence of $\mu = \mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ from Theorem F.3.1 implies that the homogeneous multivariate generating polynomial of μ is $\frac{\sqrt{q}}{1+\sqrt{q}}$-fractionally log-concave in the sense of [Ali+21]. It follows from [Ana+21a] that μ is $\left(1 + \frac{1}{\sqrt{q}}\right)$-entropically independent, so $\mu^{\mathsf{RC}}_{\mathcal{M},p,q}$ satisfies $\left\lceil 1 + \frac{1}{\sqrt{q}} \right\rceil$-uniform block factorization of entropy with constant $\Omega(n^{1+1/\sqrt{q}})$. One can then compare the $\left\lceil 1 + \frac{1}{\sqrt{q}} \right\rceil$-uniform block dynamics with the Glauber dynamics by further factorizing the entropy. This can be done using the standard log-Sobolev constant, which can be bounded by comparison with the spectral gap. This takes advantage of the marginal boundedness of μ as well as the second spectral independence bound in Theorem F.3.1. □

F.4 Approximation Guarantees for Local Search Algorithms

In this section, we show how variant(s) of the approximate exchange properties used to give tight mixing time analyses in Chapter 11 can be used to give approximation ratio guarantees for simple local search algorithms. For instance, we show that a simple local search algorithm gives a $(k!)^2$-approximation for the problem of maximizing $\mu(S)$ for $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ with a log-concave generating polynomial (Lemma F.4.2). If the generating polynomial of $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is moreover real stable, then the slightly stronger exchange property from Lemma 11.2.2 implies the greedy algorithm gives a $(k!)^2$-approximation of $\max_S \mu(S)$ (see Lemma F.4.1). This is a generalization of a classical result of Khachiyan [Kha95], which says that greedy produces a $k^{O(k)}$-approximation for the (sub)determinant maximization problem [Kha95; Di +14; Nik15]; [see also KD16]. This also generalizes more recent results which show that greedy gives a $k^{O(k)}$-approximation for the largest j-dimensional simplex problem [Pac04; ÇM10]. The best result on the largest j-dimensional simplex problem is a $2^{O(k)}$-approximation by [Nik15], matching the lower bound given by [Di +14; ÇM10]. These results give further evidence for the efficacy of local search and greedy methods; see also [Fed72].

We first precisely state the greedy algorithm and analyze it in Lemma F.4.1. We then precisely state the simple local search algorithm and analyze it in Lemma F.4.2. For convenience, for a subset T of $[n]$ of size $\leq k$, let $\mu(T) = \sum_{S \in \binom{[n]}{k} : S \supseteq T} \mu(S)$.

Algorithm 1 Greedy

Initialize $S \leftarrow \emptyset$
While $|S| < k$: Pick $i \notin S$ that maximizes $\mu(S \cup \{i\})$, and update $S \leftarrow S \cup \{i\}$

Lemma F.4.1. *Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be such that its multivariate generating polynomial g_μ is real stable. Then the output $S_{\mathsf{Greedy}} \in \binom{[n]}{k}$ of the greedy algorithm (see Algorithm 1) is a $(k!)^2$-approximation of $\max_{T \in \binom{[n]}{k}} \mu(T)$.*

Proof. Without loss of generality, assume μ is not identically zero. For $j \in [k]$, let i_j be the element added to S_{Greedy} at iteration j of the while loop. Let $S_0 = \emptyset$, $S_j = S_{j-1} \cup \{i_j\}$ and $\mathsf{OPT}_j \overset{\mathsf{def}}{=} \arg\max_{T \in \binom{[n]}{k} : T \supseteq S_j} \mu(T)$. Note that $S_k = S_{\mathsf{Greedy}}$, $\mu(\mathsf{OPT}_0) = \max_{T \in \binom{[n]}{k}} \mu(T)$, and $\mathsf{OPT}_k = S_{\mathsf{Greedy}}$. We show by induction on $j \in [k]$ that $\mu(\mathsf{OPT}_{j-1}) \leq (k - j + 1)^2 \mu(\mathsf{OPT}_j)$, from which $\mu(\mathsf{OPT}) \leq (k!)^2 \mu(S_{\mathsf{Greedy}})$ immediately follows.

First, observe that $\mu(S_j) > 0$ for all $j \in \{0, \ldots, k\}$. For $j = 0$, this is trivially true since $\mu(\emptyset) = \sum_S \mu(S)$. For $j \geq 1$, this follows by induction, since

$$\mu(S_j) = \max_{i \notin S_{j-1}} \mu(S_{j-1} \cup \{i\}) \geq \frac{1}{k} \sum_{i \notin S_{j-1}} \mu(S_{j-1} \cup \{i\}) \geq \frac{1}{kn} \mu(S_{j-1}) > 0.$$

Consider $\mu_{j-1} : \binom{[n]\setminus S_{j-1}}{k+1-j} \to \mathbb{R}_{\geq 0}$ defined by $\mu_{j-1}(T) = \mu(T \cup S_{j-1})$. Observe that μ_{j-1} is generated by the real stable polynomial $\partial_{i_{j-1}} \cdots \partial_{i_1} g_\mu$ (see Theorem 5.7.3). Let $X = \mathsf{OPT}_{j-1} \setminus S_{j-1}$. Applying (11.5) in Lemma 11.2.2 to X and $i_j \notin S_{j-1}$, we have

$$\mu_{j-1}(X)\mu_{j-1}(i_j) \leq (k+1-j) \sum_{e \in X} \mu_{j-1}(X + i_j - e)\mu_{j-1}(e) \leq (k+1-j)^2 \mu(\mathsf{OPT}_j)\mu_{j-1}(i_j)$$

where the last inequality follows from $i_j = \arg\max_{i \notin S_{j-1}} \mu(S_{j-1} \cup \{i\}) = \arg\max_{i \notin S_{j-1}} \mu_{j-1}(i)$, $\mathsf{OPT}_j = \arg\max_{T \in \binom{[n]}{k} : T \supseteq S_j} \mu(T)$ and the fact that $S_j \subseteq (X \cup S_{j-1} \cup \{i_j\} \setminus \{e\})$ for $e \in X$. Dividing both sides by $\mu_{j-1}(i_j) = \mu(S_{j-1} \cup \{i_j\}) > 0$ gives

$$\mu(\mathsf{OPT}_{j-1}) = \mu_{j-1}(X) \leq (k+1-j)^2 \mu(\mathsf{OPT}_j).$$

□

We remark that similar guarantees can be obtained for a closely related local search algorithm, which moves between sets of size k, each time replacing one element by another. Note that our improved exchange property for strongly Rayleigh distributions is crucial in obtaining $k^{O(k)}$-approximation. For arbitrary log-concave distributions, we can show the α-approximate exchange property in Lemma 11.2.2 with $\alpha \leq 2^{O(k)}$ instead of $\alpha \leq k^2$, thus proving a $2^{O(k^2)}$-approximation guarantee for greedy. Furthermore, we show that local search yields $k^{O(k)}$-approximation.

Algorithm 2 α-local search ($\alpha \leq 1$)

Initialize $S \leftarrow S_0$ for some S_0 with $\mu(S_0) \neq 0$
While $\mu(S) < \alpha \cdot \mu(S - i + j)$ for some $j \notin S$ and $i \in S$ do: Update $S \leftarrow \arg\max_{S' \in \{S-i+j : j \notin S, i \in S\}} \mu(S')$

Lemma F.4.2. *Let $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be such that its multivariate generating polynomial g_μ is log-concave. Then the output $S_{\mathsf{LS}} \in \binom{[n]}{k}$ of α-local search (see Algorithm 2) is a $(k!)^2/\alpha^k$-approximation of $\max_{T \in \binom{[n]}{k}} \mu(T)$.*

Proof. We sketch a proof for the case $\alpha = 1$. The proof for general $\alpha \leq 1$ is entirely analogous. The main ingredient in our proof is an inequality of the form

$$\mu(S)\mu(T) \leq \frac{1}{|S \setminus T|^2} \left(\sum_{i \in S \setminus T, j \in T \setminus S} \mu(S - i + j) \right) \left(\sum_{i \in S \setminus T, j \in T \setminus S} \mu(T + i - j) \right),$$

which is a variant of the approximate exchange property from Definition 48. For $S \cap T = \emptyset$, this inequality follows from Proposition 5.5.3 and complete log-concavity of the polynomial

$$\tilde{f}_\mu(y, x) = \mu(S)y^k + \left(\sum_{i \in S, j \in T} \mu(S - i + j) \right) y^{k-1}x + \cdots + \left(\sum_{i \in S, j \in T} \mu(T + i - j) \right) yx^{k-1} + \mu(T)x^k$$

obtained by setting $z_i = y$ for $i \in S$, $z_j = x$ for $j \in T$, and $z_i = 0$ otherwise, in the generating polynomial g_μ. We can reduce the remaining cases to the case $S \cap T = \emptyset$ by taking the derivative $\prod_{i \in S \cap T} \partial_i$ of g_μ.

When S is a local maxima, we get

$$\mu(S)\mu(T) \leq k^2 \mu(S) \max_{i \in S \setminus T, j \in T \setminus S} \mu(T + i - j).$$

Successively apply this inequality, first with $T = T_0 \stackrel{\mathsf{def}}{=} \arg\max_{T'} \mu(T')$, then with

$$T_\ell \stackrel{\mathsf{def}}{=} \arg\max_{i \in S \setminus T_{\ell-1}, j \in T_{\ell-1} \setminus S} \mu(T_{\ell-1} + i - j) \text{ for } \ell \geq 1.$$

Note that $|T_\ell \setminus S_{\mathsf{LS}}|$ strictly decreases in each iteration, so we get $T_k = S_{\mathsf{LS}}$, and

$$\mu(T_0) \leq k^2 \mu(T_1) \leq k^2(k-1)^2 \mu(T_2) \leq \cdots \leq (k!)^2 \mu(S_{\mathsf{LS}})$$

□

F.5 "Lossless Function-Wise" Local-to-Global Entropy Contraction

Our goal in this section is to try to give an exact formula for the rate of entropy contraction for a global function f_n under a single application of the down operator. We will arrive at a "kind of" average-case local-to-global theorem, but one in which the distribution we are averaging over depends on the local entropies of f_n. This hasn't found any applications (yet), but we record it here in case it can be useful for something.

Fix a global function $f_n : \binom{\mathscr{U}}{n} \to \mathbb{R}_{\geq 0}$, which we again think of as the density of some measure μ on $\binom{\mathscr{U}}{n}$ w.r.t. μ. For a feasible $\tau \subseteq \mathscr{U}$ with $|\tau| = k \leq n-2$, we define the *local contraction coefficient* $\alpha^\tau(f_n)$ of a function f_n as the largest number in $[0,1]$ satisfying

$$\mathrm{Ent}_1^\tau\left(f_1^\tau\right) \leq (1-\alpha^\tau(f_n)) \cdot \mathrm{Ent}_2^\tau\left(f_2^\tau\right).$$

When these local entropies are nonzero, we simply have

$$\alpha^\tau(f_n) = 1 - \frac{\mathrm{Ent}_1^\tau\left(f_1^\tau\right)}{\mathrm{Ent}_2^\tau\left(f_2^\tau\right)}.$$

Similarly, the *global contraction coefficient* $C(f_n)$ of f_n is defined as the largest number in $[0,1]$ satisfying

$$\mathrm{Ent}_{n-1}\left(f_{n-1}\right) \leq (1-C(f_n)) \cdot \mathrm{Ent}_n\left(f_n\right).$$

We are interested in understanding $C(f_n)$. Towards this, the following reformulations are useful.

Lemma F.5.1. *1. Local Contraction* $\iff$

$$\mathrm{Ent}_1^\tau\left(f_1^\tau\right) \leq \left(\frac{1}{\alpha^\tau(f_n)} - 1\right) \cdot \mathbb{E}_{v\sim\mu_1^\tau}\left[\mathrm{Ent}_1^{\tau\cup\{v\}}\left(f_1^{\tau\cup\{v\}}\right)\right].$$

2. Global Contraction $\iff$

$$\mathrm{Ent}_n\left(f_n\right) \leq \frac{1}{C(f_n)} \cdot \mathbb{E}_{\tau\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\tau\left(f_1^\tau\right)\right].$$

Proof. Apply the Law of Total Entropy (see Lemma 9.2.3) and rearrange as necessary. □

Proposition F.5.2 (Exact Local-to-Global)**.** *Fix a global function f_n with local contraction rates $\alpha^\tau(f_n)$. For each feasible $\sigma \subseteq \mathscr{U}$ with $|\sigma| = n-1$, define the* ***flag factorization coefficient*** *of f_n w.r.t. σ as*

$$\mathscr{F}_\sigma\left(f_n\right) = \mathbb{E}_{\substack{\text{orderings} \\ u_1,\ldots,u_{n-1} \\ \text{of } \sigma}} \sum_{j=0}^{n-2} \prod_{i=0}^{n-j-2} \left(\frac{1}{\alpha^{\{u_1,\ldots,u_i\}}(f_n)} - 1\right)$$

Then the global contraction coefficient $C(f_n)$ satisfies the identity

$$\frac{1}{C(f_n)} = \frac{\mathbb{E}_{\sigma\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\sigma\left(f_1^\sigma\right) \cdot \mathscr{F}_\sigma\left(f_n\right)\right]}{\mathbb{E}_{\sigma\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\sigma\left(f_1^\sigma\right)\right]}.$$

Proof. We apply the Law of Total Entropy (see Lemma 9.2.3) repeatedly to obtain

$$\mathrm{Ent}_n\left(f_n\right) = \sum_{j=0}^{n-1} \mathbb{E}_{\tau\sim\mu_j}\left[\mathrm{Ent}_1^\tau\left(f_1^\tau\right)\right].$$

Now, we repeatedly apply Lemma F.5.1 Item 1 to obtain the inequality

$$\begin{aligned}
\mathrm{Ent}_n\left(f_n\right) &= \mathbb{E}_{\sigma\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\sigma\left(f_1^\sigma\right) \cdot \mathbb{E}_{\substack{\text{orderings} \\ u_1,\ldots,u_{n-1} \\ \text{of } \sigma}} \sum_{j=0}^{n-2} \prod_{i=0}^{n-j-2} \left(\frac{1}{\alpha^{\{u_1,\ldots,u_i\}}(f_n)} - 1\right)\right] \\
&= \mathbb{E}_{\sigma\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\sigma\left(f_1^\sigma\right) \cdot \mathscr{F}_\sigma\left(f_n\right)\right] \\
&= \frac{\mathbb{E}_{\sigma\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\sigma\left(f_1^\sigma\right) \cdot \mathscr{F}_\sigma\left(f_n\right)\right]}{\mathbb{E}_{\sigma\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\sigma\left(f_1^\sigma\right)\right]} \cdot \mathbb{E}_{\sigma\sim\mu_{n-1}}\left[\mathrm{Ent}_1^\sigma\left(f_1^\sigma\right)\right].
\end{aligned}$$

Applying Lemma F.5.1 Item 2 yields the claim. □

F.6 Trickle-Down for Two-Sided ℓ_∞-Independence

In this section, we focus on distributions μ over $2^{[n]}$ for simplicity. We prove an analog of Oppenheim's vanilla Trickle-Down Theorem (see Theorem 3.2.1) for $\|\Psi_\mu - \mathsf{Id}\|_\infty$ using just the Law of Total Probability. The argument does not have any spectral analysis in it. The hope is that this helps give some more insight/intuition into trickle-down methods.

Theorem F.6.1. *For every $0 \le d \le n-2$, let $\gamma_d \ge 0$ be the smallest real number satisfying the following condition: For every $S \subseteq [n]$ satisfying $|S| = d$ and every $i \notin S$, we have the inequality*

$$\sum_{j \notin S \cup \{i\}} \max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(i \to j)| \le \gamma_d.$$

Then for all $0 \le d \le n-3$, we have the recursive inequality

$$\gamma_d \le \frac{\frac{n-d-1}{n-d-2}\gamma_{d+1}}{1 - \frac{1}{n-d-2}\gamma_{d+1}}.$$

In particular, if $\gamma_{n-2} \le \alpha$, then

$$\gamma_d \le \frac{(n-d-1)\alpha}{1-(n-d-2)\alpha}$$

for all $0 \le d \le n-2$.

Remark 77. Note that this result gives $\left(\frac{(n-1)\alpha}{1-(n-2)\alpha}, \ldots, \frac{\alpha}{1-\alpha}, \alpha\right)$-spectral independence, where $\alpha \ge \gamma_{n-2}$. In particular, $\gamma_d/(n-d-1)$ gives a bound on the second largest eigenvalue of all local random walks, and so this exactly recovers Oppenheim's vanilla Trickle-Down Theorem (see Theorem 3.2.1, stated in the language of local random walks). The only difference is that we obtain a stronger conclusion via a stronger assumption, namely we are working with bounds on $\|\Psi_\mu - \mathsf{Id}\|_\infty$ instead of $\lambda_{\max}(\Psi_\mu) - 1$.

The proofs hinge on the following key technical lemma.

Lemma F.6.2 (Recursive Influence Bound)**.** *For all distinct $i, j, k \in [n]$, we have the inequality*

$$|\Psi_\mu(i \to j)| \le \max_{s \in \{0,1\}} \left\{\left|\Psi_{\mu^{k \leftarrow s}}(i \to j)\right|\right\} + |\Psi_\mu(i \to k)| \cdot \max_{s \in \{0,1\}} \left\{\left|\Psi_{\mu^{i \leftarrow s}}(k \to j)\right|\right\}.$$

Remark 78. If one isn't careful when applying the Law of Total Probability, one might become convinced that the inequality $|\Psi_\mu(i \to j)| \le \max_{s \in \{0,1\}} \left\{\left|\Psi_{\mu^{k \leftarrow s}}(i \to j)\right|\right\}$ (or similar such inequalities) holds. Such an inequality is too good to be true in general. However, Lemma F.6.2 shows that this is "almost" true, with some additional loss.

We prove this lemma in Appendix F.6.1 below. We first use it to prove Theorem F.6.1.

Proof of Theorem F.6.1. Fix a set $S \subset [n]$, $i \notin S$, $j \notin S \cup \{i\}$ and a feasible assignment $\sigma : S \to \{0,1\}$. By averaging over $k \notin S \cup \{i,j\}$ and applying Lemma F.6.2, we obtain

$$\begin{aligned}|\Psi_{\mu^\sigma}(i \to j)| &\le \frac{1}{n-d-2} \sum_{k \notin S \cup \{i,j\}} \max_{s \in \{0,1\}} \left\{\left|\Psi_{\mu^{\sigma, k \leftarrow s}}(i \to j)\right|\right\} \\ &\quad + \frac{1}{n-d-2} \sum_{k \notin S \cup \{i,j\}} |\Psi_{\mu^\sigma}(i \to k)| \cdot \max_{s \in \{0,1\}} \left\{\left|\Psi_{\mu^{\sigma, i \leftarrow s}}(k \to j)\right|\right\}.\end{aligned}$$

Maximizing over all such feasible assignments $\sigma : S \to \{0,1\}$, we have that

$$\begin{aligned}\max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(i \to j)| &\le \frac{1}{n-d-2} \sum_{k \notin S \cup \{i,j\}} \max_{\tau : S \cup \{k\} \to \{0,1\}} |\Psi_{\mu^\tau}(i \to j)| \\ &\quad + \frac{1}{n-d-2} \sum_{k \notin S \cup \{i,j\}} \max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(i \to k)| \cdot \max_{\tau : S \cup \{i\} \to \{0,1\}} |\Psi_{\mu^\tau}(k \to j)|.\end{aligned}$$

Summing over all $j \notin S \cup \{i\}$, we obtain

$$\begin{aligned}
&\sum_{j \notin S \cup \{i\}} \max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(i \to j)| \\
&\leq \frac{1}{n-d-2} \sum_{j \notin S \cup \{i\}} \sum_{k \notin S \cup \{i,j\}} \max_{\tau : S \cup \{k\} \to \{0,1\}} |\Psi_{\mu^\tau}(i \to j)| \\
&+ \frac{1}{n-d-2} \sum_{j \notin S \cup \{i\}} \sum_{k \notin S \cup \{i,j\}} \max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(i \to k)| \cdot \max_{\tau : S \cup \{i\} \to \{0,1\}} |\Psi_{\mu^\tau}(k \to j)| \\
&= \frac{1}{n-d-2} \sum_{k \notin S \cup \{i\}} \underbrace{\sum_{j \notin S \cup \{i,k\}} \max_{\tau : S \cup \{k\} \to \{0,1\}} |\Psi_{\mu^\tau}(i \to j)|}_{\leq \gamma_{d+1}} \\
&+ \frac{1}{n-d-2} \sum_{k \notin S \cup \{i\}} \max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(i \to k)| \underbrace{\sum_{j \notin S \cup \{i,k\}} \max_{\tau : S \cup \{i\} \to \{0,1\}} |\Psi_{\mu^\tau}(k \to j)|}_{\leq \gamma_{d+1}} \\
&\leq \frac{n-d-1}{n-d-2} \gamma_{d+1} + \frac{1}{n-d-2} \gamma_{d+1} \underbrace{\sum_{k \notin S \cup \{i\}} \max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(i \to k)|}_{(*)}.
\end{aligned}$$

The key observation is that $(*)$ is precisely the expression we wanted to bound in the beginning. Rearranging, we obtain

$$\sum_{i \notin S \cup \{j\}} \max_{\sigma : S \to \{0,1\}} |\Psi_{\mu^\sigma}(j \to i)| \leq \frac{\frac{n-d-1}{n-d-2} \gamma_{d+1}}{1 - \frac{1}{n-d-2} \gamma_{d+1}}.$$

As this holds for all choices of $S \subset [n]$ with $|S| = d$ and all $i \notin S$, we obtain the desired inequality

$$\gamma_d \leq \frac{\frac{n-d-1}{n-d-2} \gamma_{d+1}}{1 - \frac{1}{n-d-2} \gamma_{d+1}}.$$

Finally, we show that $\gamma_d \leq \frac{(n-d-1)\alpha}{1-(n-d-2)\alpha}$ if $\gamma_{n-2} \leq \alpha$. We go by descending induction beginning with $d = n-2$. The base case $d = n-2$ trivially holds. Suppose we know $\gamma_{d+1} \leq \frac{(n-d-2)\alpha}{1-(n-d-3)\alpha}$. Then

$$\gamma_d \leq \frac{\frac{n-d-1}{n-d-2} \gamma_{d+1}}{1 - \frac{1}{n-d-2} \gamma_{d+1}} \leq \frac{\frac{n-d-1}{n-d-2} \cdot \frac{(n-d-2)\alpha}{1-(n-d-3)\alpha}}{1 - \frac{1}{n-d-2} \cdot \frac{(n-d-2)\alpha}{1-(n-d-3)\alpha}} = \frac{\frac{(n-d-1)\alpha}{1-(n-d-3)\alpha}}{1 - \frac{\alpha}{1-(n-d-3)\alpha}} = \frac{(n-d-1)\alpha}{1-(n-d-2)\alpha}.$$

□

F.6.1 Proof of Lemma F.6.2

Recall that

$$\Psi_\mu(i \to j) = \Pr_{\sigma \sim \mu}[\sigma(j) = 1 \mid \sigma(i) = 1] - \Pr_{\sigma \sim \mu}[\sigma(j) = 1 \mid \sigma(i) = 0].$$

To keep the notation succinct in this proof, we will abuse notation and write i for the event that $\sigma(i) = 1$ and $\bar{i}$ for the even that $\sigma(i) = 0$. We will drop the subscripts on all probabilities. Thus, for instance, we may compactly write $\Psi_\mu(i \to j)$ as $\Pr[j \mid i] - \Pr[j \mid \bar{i}]$.

Fix i, j, k distinct. We have by the Law of Total Probability that

$$\begin{aligned}
&|\Psi_\mu(i \to j)| = \left|\Pr[j \mid i] - \Pr[j \mid \bar{i}]\right| \\
&= \left|(\Pr[j \mid i,k] \Pr[k \mid i] - \Pr[j \mid \bar{i},k] \Pr[k \mid \bar{i}]) + (\Pr[j \mid i,\bar{k}] \Pr[\bar{k} \mid i] - \Pr[j \mid \bar{i},\bar{k}] \Pr[\bar{k} \mid \bar{i}]\right|.
\end{aligned}$$

Now, we insert some additional terms by adding and subtracting the same term. The main idea is that these terms use $\Pr[k]$ to interpolate between $\Pr[k \mid i]$ and $\Pr[k \mid \overline{i}]$. We obtain

$$\begin{aligned}
&\Big|(\Pr[j \mid i,k]\Pr[k \mid i] - \Pr[j \mid i,k]\Pr[k]) + (\Pr[j \mid i,k]\Pr[k] - \Pr[j \mid \overline{i},k]\Pr[k]) \\
&+ (\Pr[j \mid \overline{i},k]\Pr[k] - \Pr[j \mid \overline{i},k]\Pr[k \mid \overline{i}]) + (\Pr[j \mid i,\overline{k}]\Pr[\overline{k} \mid i] - \Pr[j \mid i,\overline{k}]\Pr[\overline{k}]) \\
&+ (\Pr[j \mid i,\overline{k}]\Pr[\overline{k}] - \Pr[j \mid \overline{i},\overline{k}]\Pr[\overline{k}]) + (\Pr[j \mid \overline{i},\overline{k}]\Pr[\overline{k}] - \Pr[j \mid \overline{i},\overline{k}]\Pr[\overline{k} \mid \overline{i}])\Big| \\
&= \Big|\Pr[j \mid i,k] \cdot (\Pr[k \mid i] - \Pr[k]) + \Pr[j \mid \overline{i},k] \cdot (\Pr[k] - \Pr[k \mid \overline{i}]) \\
&+ \Pr[j \mid i,\overline{k}] \cdot (\Pr[\overline{k} \mid i] - \Pr[\overline{k}]) + \Pr[j \mid \overline{i},\overline{k}] \cdot (\Pr[\overline{k}] - \Pr[\overline{k} \mid \overline{i}]) \\
&+ \Pr[k] \cdot (\Pr[j \mid i,k] - \Pr[j \mid \overline{i},k]) + \Pr[\overline{k}] \cdot (\Pr[j \mid i,\overline{k}] - \Pr[j \mid \overline{i},\overline{k}])\Big|.
\end{aligned}$$

Now, we observe that $\Pr[k \mid i] - \Pr[k] = \Pr[\overline{i}] \cdot (\Pr[k \mid i] - \Pr[k \mid \overline{i}]) = -(\Pr[\overline{k} \mid i] - \Pr[\overline{k}])$ and $\Pr[k] - \Pr[k \mid \overline{i}] = \Pr[i] \cdot (\Pr[k \mid i] - \Pr[k \mid \overline{i}]) = -(\Pr[\overline{k}] - \Pr[\overline{k} \mid \overline{i}])$. Hence, the above further simplifies to

$$\begin{aligned}
&\Big|(\Pr[j \mid i,k] - \Pr[j \mid i,\overline{k}])(\Pr[k \mid i] - \Pr[k]) + (\Pr[j \mid \overline{i},k] - \Pr[j \mid \overline{i},\overline{k}]) \cdot (\Pr[k] - \Pr[k \mid \overline{i}]) \\
&+ \Pr[k] \cdot (\Pr[j \mid i,k] - \Pr[j \mid \overline{i},k]) + \Pr[\overline{k}] \cdot (\Pr[j \mid i,\overline{k}] - \Pr[j \mid \overline{i},\overline{k}])\Big| \\
&= \Big|(\Pr[k \mid i] - \Pr[k \mid \overline{i}]) \cdot (\Pr[\overline{i}] \cdot (\Pr[j \mid i,k] - \Pr[j \mid i,\overline{k}]) + \Pr[i] \cdot (\Pr[j \mid \overline{i},k] - \Pr[j \mid \overline{i},\overline{k}])) \\
&+ \Pr[k] \cdot (\Pr[j \mid i,k] - \Pr[j \mid \overline{i},k]) + \Pr[\overline{k}] \cdot (\Pr[j \mid i,\overline{k}] - \Pr[j \mid \overline{i},\overline{k}])\Big|.
\end{aligned}$$

At the moment, this looks completely unusable, although up until this point, everything has been an equality. However, a simple application of the Triangle Inequality and the observation that $\Pr[k] + \Pr[\overline{k}] = \Pr[j] + \Pr[\overline{j}] = 1$ gives

$$\begin{aligned}
&\max\{\left|\Pr[j \mid i,k] - \Pr[j \mid \overline{i},k]\right|, \left|\Pr[j \mid i,\overline{k}] - \Pr[j \mid \overline{i},\overline{k}]\right|\} \\
&+ \left|\Pr[k \mid i] - \Pr[k \mid \overline{i}]\right| \cdot \max\{\left|\Pr[j \mid i,k] - \Pr[j \mid i,\overline{k}]\right|, \left|\Pr[j \mid \overline{i},k] - \Pr[j \mid \overline{i},\overline{k}]\right|\} \\
&= \max_{s \in \{0,1\}} \left\{\left|\Psi_{\mu^{k \leftarrow s}}(i \to j)\right|\right\} + |\Psi_\mu(i \to k)| \cdot \max_{s \in \{0,1\}} \left\{\left|\Psi_{\mu^{i \leftarrow s}}(k \to j)\right|\right\}
\end{aligned}$$

as desired.