Lecture 6: Phase Transitions, Correlation Decay, and the Hardcore Gas Model

September 26, 2023

In the previous lecture, we saw the Jerrum–Sinclair FPRAS for the ferromagnetic Ising model. This algorithmic result is remarkable for a number of reasons, one of which is that it overcomes a natural *phase transition* which occurs in the model. For most interesting classes of graphs (e.g. complete graphs, bounded-degree graphs, tori, etc.) there exists a critical temperature β_c such that

- if $\beta < \beta_c$, then Glauber dynamics mixes in polynomial-time (typically $O(n \log n)$ steps), and
- if $\beta > \beta_c$, then Glauber dynamics requires $\exp(\Omega(n))$ steps to mix.[1]

This algorithmic phase transition stems from an underlying phase transition in the structural properties of the Gibbs distribution, namely the presence or absence of a precise notion of *correlation decay* between the spins assigned to distant vertices. In the next two lectures, we will build up to a beautiful *deterministic* algorithm directly based on correlation decay for approximate counting which is very different in nature from Markov chain Monte Carlo, due independently to Weitz [Wei06] and Bandyopadhyay–Gamarnik [BG08] (with numerous additional follow-ups e.g. [Bay+07; GK07]). We will also see sharper *complexity* phase transition phenomena in some models (albeit without proofs for the hardness side).

1 Spin Systems and Spatial Mixing

How does one "detect" a phase transition? There are many ways to do this. For the ferromagnetic Ising model, we can look at the *magnetization* $\sum_{v \in V} \sigma_v$. When $\beta < \beta_c$, this statistic concentrates around 0, and when $\beta > \beta_c$, its distribution becomes bimodal. But this statistic is a bit tailored to the specific features of the ferromagnetic Ising model; we'd like a theory which works much more broadly. Later on in the course, we'll see a method based on studying where discontinuities in the *free energy* $\log Z$ "cluster". Here, we look at the presence or absence of a structural property of the Gibbs distribution known as *correlation decay*. To state it, we first define a large class of probability measures for which this notion will be useful.

Definition 1 (q-Spin System)**.** *Let $q \in \mathbb{N}$ be an integer satisfying $q \geq 2$. A q-*spin system *is specified by a graph $G = (V, E)$, a nonnegative symmetric* interaction matrix $\boldsymbol{A} \in \mathbb{R}_{\geq 0}^{q \times q}$*, and a nonnegative vector of* external fields $\boldsymbol{\lambda} \in \mathbb{R}_{\geq 0}^{V \times [q]}$*. These data give rise to a* Gibbs distribution *$\mu = \mu_{G,\boldsymbol{A},\boldsymbol{\lambda}}$ on $[q]^V$ given by*

$$\mu(\sigma) \propto \prod_{uv \in E} \boldsymbol{A}_{\sigma(u),\sigma(v)} \prod_{v \in V} \boldsymbol{\lambda}_{v,\sigma(v)}, \qquad \forall \sigma : V \to [q],$$

with corresponding partition function

$$Z = Z_{G,\boldsymbol{A}}(\boldsymbol{\lambda}) \overset{\mathsf{def}}{=} \sum_{\sigma:V \to [q]} \prod_{uv \in E} \boldsymbol{A}_{\sigma(u),\sigma(v)} \prod_{v \in V} \boldsymbol{\lambda}_{v,\sigma(v)}.$$

The elements of $[q]$ are often called *spins* or *colors*. Often when $q = 2$, we will instead take the space of spins to be either $\{0, 1\}$ or $\{\pm 1\}$ depending on context.

[1] We already saw an example in the first problem set, namely the *Curie–Weiss model* where $G = K_n$.

Spin systems are sometimes also known as *Markov random fields*, and are a special class of *probabilistic graphical models* [KF09]. Even though the "description" of such a system is extremely compact, the resulting Gibbs distribution is extremely complex, and displays a rich set of behaviors. Their study follows one of the central themes of statistical mechanics, where we study "macroscopic" (or "global") properties of large-scale systems given only the "microscopic" (or "local") interactions between nearby components.

Here are a few common examples of spin systems arising in statistical physics, combinatorics, theoretical computer science, etc.

- Suppose $q = 2$ and $\boldsymbol{A} = \begin{bmatrix} e^\beta & 1 \\ 1 & e^\beta \end{bmatrix}$ where $\beta \geq 0$. Then this 2-spin system exactly recovers the ferromagnetic Ising model.[2] A natural generalization to larger $q \geq 2$, where $\boldsymbol{A} = (e^\beta - 1) \cdot I + \mathbf{1}_q \mathbf{1}_q^\top$ gives rise to the *ferromagnetic Potts model.*

- Suppose $q = 2$, $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$, and $\boldsymbol{\lambda} = (\lambda \mathbf{1}_V, \mathbf{1}_V)$ for some $\lambda \in \mathbb{R}_{\geq 0}$. Then we can view the assignments $\sigma : V \to [q]$ as indicators of subsets of vertices. Furthermore, μ is supported on *independent sets* of G, i.e. subsets $I \subseteq V$ such that no pair of vertices in I are connected by an edge. For such sets of vertices, we have $\mu(I) \propto \lambda^{|I|}$. This is called the *hardcore (gas) model.*[3]

- Suppose $\boldsymbol{A} = \mathbf{1}_q \mathbf{1}_q^\top - I$ and $\boldsymbol{\lambda} = \mathbf{1}$. Then μ is uniform over the *(proper) q-colorings* of G, i.e. assignments $\chi : V \to [q]$ such that $\chi(u) \neq \chi(v)$ for all $uv \in E$.[4]

Given such a Gibbs measure, we can talk about its conditional distributions. Some natural events to condition on are, for instance, that a given subset of vertices have certain prescribed assignments. This is known as pinning.

Definition 2 (Pinning)**.** *A* pinning *is a partial assignment $\tau : \Lambda \to [q]$ where $\Lambda \subseteq V$ is a subset of vertices. Given such a pinning, we write μ^τ for the induced* conditional (Gibbs) distribution*, given by*

$$\mu^\tau(\sigma) \propto \begin{cases} \mu(\sigma), & \text{if } \sigma(v) = \tau(v), \forall v \in \Lambda \\ 0, & \text{otherwise} \end{cases}.$$

We can also "restrict attention" for a subset of vertices by marginalizing out all other vertices.

Definition 3 (Marginal Distribution)**.** *Let $A \subseteq V$ be a subset of vertices. We write μ_A for the induced* marginal distribution *over partial assignments $\eta : A \to [q]$ given by*

$$\mu_A(\eta) = \sum_{\substack{\sigma : V \to [q] \\ \sigma|_A = \eta}} \mu(\sigma).$$

Note we can combine both notations and look at the marginal distribution μ_A^τ of $A \subseteq V$ conditioned on a pinning of some other vertices $\Lambda \subseteq V \setminus A$.

An appealing property of graphical models is that they satisfy the *(global) Markov property* (or *conditional independence* property).

Fact 1.1. *Let $\boldsymbol{A}, \boldsymbol{\lambda}$ be the parameters of a q-spin system on a graph $G = (V, E)$. Then for every partition $A \sqcup S \sqcup B$ of V such that every path from a vertex in A to a vertex in B must pass through a vertex in S and every pinning $\tau : S \to [q]$, the conditional measure μ^τ factorizes as $\mu_A^\tau \otimes \mu_B^\tau$.*[5]

With the formalism of spin systems in hand, we can now define what exponential decay of correlations means.

[2] One can also view this combinatorially as a weighted distribution over *cuts* in the graph.

[3] This is a discretization of the original *hard spheres* model, where gas particles are viewed as solid spheres in $\mathbb{R}^3$ satisfying the constraint that no pair of gas particles can intersect.

[4] In statistical physics lingo, this is the Gibbs distribution of the *zero-temperature antiferromagnetic q-state Potts model.*

[5] Such a set S is called a *separator* between A and B.

Definition 4 (Spatial Mixing)**.** *We say μ exhibits* Weak Spatial Mixing (WSM) *if there exist constants $C > 0$ and $0 < \delta < 1$ such that for every $r \in V$, every boundary set of vertices $\Lambda \subseteq V \setminus \{r\}$, and every pair of pinnings $\tau, \sigma : \Lambda \to [q]$,*

$$\|\mu_r^\tau - \mu_r^\sigma\|_{\mathsf{TV}} \le C \cdot (1-\delta)^{\mathrm{dist}(r,\Lambda)}. \tag{1}$$

Similarly, we say μ exhibits Strong Spatial Mixing (SSM) *if we can upgrade Eq. (1) to*

$$\|\mu_r^\tau - \mu_r^\sigma\|_{\mathsf{TV}} \le C \cdot (1-\delta)^{\mathrm{dist}(r,\Lambda_{\tau,\sigma})}, \tag{2}$$

where $\Lambda_{\tau,\sigma} = \{v \in \Lambda : \tau(v) \neq \sigma(v)\}$ denotes the set of vertices of disagreement between τ, σ.

Another way to think of strong spatial mixing is that we demand weak spatial mixing holds for all conditional distribution of μ, not only μ itself. While there are many other notions of phase transition and correlation decay, for the purposes of this and the next lecture, correlation decay will mean one of weak/strong spatial mixing, and phase transition means a sharp threshold for some natural underlying parameter which delineates whether or not this form of correlation decay holds.

Finally, we mention that SSM remains open for many models of interest.

Conjecture 1. *The uniform distribution over proper q-colorings of an arbitrary graph of maximum degree Δ exhibits* SSM *as long as $q \ge \Delta + 1$.*

Surprisingly, this wasn't even known for bounded-degree trees until recently [Che+23].

2 The Hardcore Model

We now specialize our discussion to the hardcore model, where we will see a sharp *complexity-theoretic* phase transition. Recall that in the hardcore model, we have a graph $G = (V, E)$ and a parameter[6] $\lambda \ge 0$; we typically view λ as a constant independent of n. The corresponding Gibbs distribution $\mu = \mu_{G,\lambda}$ is a distribution over independent sets $I \subseteq V$ of G given by

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

with partition function

$$Z_G(\lambda) \overset{\mathsf{def}}{=} \sum_{I \subseteq V \text{ independent}} \lambda^{|I|}.$$

The latter is sometimes called the *(univariate) independence polynomial* of G. Throughout, we slightly abuse notation, and identify v with the event that v is contained in a random independent set drawn from the Gibbs distribution. At the same time, we write $\overline{v}$ for the complementary event.

Why are we considering independent sets and not some other model? When we discuss the cluster expansion, we'll see that this model is in some sense "universal" in that many other models can be "embedded" as hardcore models. Another is a matter of convenience. Most of our discussion will extend well beyond the hardcore model, but independent sets possess a nice self-reducibility feature.

Fact 2.1. *For any vertex $v \in V$, the measure $\mu_{G,\lambda}^v$ obtained by conditioning on v being in a random draw $I \sim \mu_{G,\lambda}$ is simply $\mu_{G-N[v],\lambda}$. Similarly, the conditional measure $\mu_{G,\lambda}^{\overline{v}}$ is simply $\mu_{G-v,\lambda}$.*

So, our goal is now to (approximately) sample a random independent set according to the Gibbs distribution $\mu = \mu_{G,\lambda}$. Intuitively, this should be easy if λ is small, since the distribution concentrates mass on small independent sets like the empty set. On the other hand, if λ large, then the distribution concentrates mass on the maximum independent sets in G, which are in general hard to find; in fact, there are dramatic *hardness of approximation* results for finding maximum independent sets [Hås99; Hås01].

[6] This is the analog of the "inverse temperature" from the ferromagnetic Ising model. It is sometimes called the *fugacity*.

It turns out, there is a sharp threshold separating these two regimes. Throughout, let $\Delta \geq 1$ denote the maximum degree of the graph $G = (V, E)$; we also think of Δ as an absolute constant independent of n. Define

$$\lambda_c(\Delta) \stackrel{\text{def}}{=} \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} \approx \frac{e}{\Delta-1}.$$

This is the *uniqueness threshold* for the hardcore model on the infinite Δ-regular tree $\mathbb{T}_\Delta$ [Kel85]. It delineates the presence or absence of correlation decay on $\mathbb{T}_\Delta$; we'll spell out this connection in further detail in a moment. For now, we state more precisely the sharp complexity-theoretic phase transition in the hardcore model. This is based on the following two seminal results.

Theorem 2.2 (Weitz [Wei06]). *If $\lambda < \lambda_c(\Delta)$, then* SSM *holds for the hardcore model on any graph of maximum degree Δ. Furthermore, there exists an* FPTAS *for estimating $Z_G(\lambda)$ for every graph $G = (V, E)$ of maximum degree Δ and every $\lambda < \lambda_c(\Delta)$. If $\lambda \leq (1-\delta)\lambda_c(\Delta)$ for a constant $0 < \delta < 1$, the running time of this algorithm scales as $(n/\epsilon)^{O\left(\frac{1}{\delta}\log\Delta\right)}$, where $0 < \epsilon < 1$ is the estimation error.*

Theorem 2.3 (Sly [Sly10]; see also [SS14; Gal+14; GŠV15; GŠV16]). *Suppose there exists $\lambda > \lambda_c(\Delta)$ and an* FPRAS *for estimating $Z_G(\lambda)$ on all graphs of maximum degree Δ. Then* NP = RP.

Both of these results were breakthroughs in the field of approximate counting and sampling. Unlike the ferromagnetic Ising model, where above criticality, we could only rule out fast mixing of Glauber dynamics, Theorem 2.3 rules out the existence of *any* efficient algorithm (assuming NP $\neq$ RP). Theorem 2.2 is remarkable because its proof shows that strong spatial mixing on the infinite Δ-regular tree implies strong spatial mixing for every graph of maximum degree Δ. Thus, in a very precise sense, the infinite Δ-regular tree is the "worst case". Theorem 2.2 is also interesting because the algorithm is fully deterministic.

Remark 1. While the algorithm in Theorem 2.2 is deterministic, its running time isn't so favorable, and one can ask if faster algorithms exist. Later in the course, we'll use the proof techniques for Theorem 2.2 to establish $O(n \log n)$ mixing of Glauber dynamics for all $\lambda < \lambda_c(\Delta)$.

The goal of this and the next lecture is to prove Theorem 2.2. Unfortunately, we won't have time to prove Theorem 2.3, but we'll give some small indications of how the proof goes. We conclude this section with an open problem.

Question 1. *Does there exist an* FPRAS *for estimating $Z_G(\lambda)$ at $\lambda = \lambda_c(\Delta)$ on any graph $G = (V, E)$ of maximum degree Δ?*

The original paper of Weitz establishes that the correlations decay *polynomially fast*, but it is unclear how to convert this to an algorithm. Nonetheless, this still establishes uniqueness of the Gibbs measure on the infinite Δ-regular tree even when $\lambda = \lambda_c(\Delta)$.

3 Correlation Decay on Trees

To build up towards Theorem 2.2, let us start by studying correlation decay on finite trees of maximum degree Δ. We will first prove the following very special case of the strong spatial mixing claim in Theorem 2.2.

Theorem 3.1. *Suppose $\lambda < \lambda_c(\Delta)$. Then* SSM *holds for the hardcore Gibbs measure on all trees of maximum degree Δ.*

Even though as stated this result only holds for trees, this will turn out to be the "worst case" out of all graphs with maximum degree Δ. In particular, even if we don't use Theorem 3.1 directly towards proving Theorem 2.2, we will use the ingredients in the proof of Theorem 3.1.

3.1 The Tree Recursion

To prove Theorem 3.1, let us arbitrarily fix some *root* vertex $r \in V$ in the tree $T = (V, E)$. We wish to show that the marginal distribution of r is stable w.r.t. small perturbations in the boundary condition imposed on far away vertices. The key tool that will give us a handle on this marginal distribution is the *tree recursion*.

For an arbitrary tree T rooted at a vertex r, write p_r for the probability that r is contained in $I \sim \mu_{T,\lambda}$. If u is any other vertex in T, we write T_u for the unique subtree of T rooted at u which is "below" u. By a slight abuse of notation, we write p_u for the probability that u is contained in $I \sim \mu_{T_u,\lambda}$; note that this distribution is w.r.t. the subtree T_u, not the full tree T.

Lemma 3.2. *Let T be an arbitrary tree rooted at r, and suppose r has children $u_1, \dots, u_d$ with corresponding rooted subtrees $T_1, \dots, T_d$. Then writing $p_i = p_{u_i}$,*

$$p_r = F_d(p_1, \dots, p_d) \stackrel{\text{def}}{=} \frac{\lambda \prod_{i=1}^d (1 - p_i)}{1 + \lambda \prod_{i=1}^d (1 - p_i)}. \tag{3}$$

In the univariate case where $p_1, \dots, p_d$ are all equal to some p, we write

$$f_d(p) \stackrel{\text{def}}{=} \frac{\lambda(1-p)^d}{1 + \lambda(1-p)^d}$$

for the univariate recursion.

Note that this recursion immediately yields an efficient dynamic programming algorithm for *exactly* computing the marginals, as well as the partition function, for the hardcore model on any tree. A more general recursion which works for any spin system is stated in Lemma A.1 and proved in Appendix A. This recursion can be derived a number of ways, but the key property of spin systems which enables such recursions is conditional independence (see Fact 1.1).

Since $\|\mu_r^\tau - \mu_r^\sigma\|_{\mathsf{TV}} = |p_r^\tau - p_r^\sigma|$, the high-level strategy to proving Theorem 3.1 given Eq. (3) is to show that for an appropriate measure of "distance" ψ on $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$, we have that for all $\boldsymbol{p}, \boldsymbol{q} \in [0,1]_{\geq 0}^d$,

$$\psi(F_d(\boldsymbol{p}), F_d(\boldsymbol{q})) \leq (1 - \delta) \cdot \max_{1 \leq i \leq d} \psi(\boldsymbol{p}_i, \boldsymbol{q}_i).$$

If this *contraction* property holds, then by translating between $\psi(\cdot,\cdot)$ and total variation distance (with whatever losses incurred going into the constant C), we can obtain Theorem 3.1. This is formalized in Lemma 3.4 below. This proof strategy follows a number of follow-up works (see e.g. [Res+13; LLY13]) to Weitz's original paper, which had a different style of argument. However, before we prove Theorem 3.1 in full, we first look at the univariate case and how the threshold $\lambda_c(\Delta)$ is derived.

3.2 The Univariate Case: Where $\lambda_c(\Delta)$ Comes From

Before we prove contraction in full, let us consider the special case of $(\Delta - 1)$-ary trees of depth-L, with L tending to $+\infty$. We use $\widehat{\mathbb{T}}_{\Delta,L}$ to denote this tree, with distinguished root vertex r. Certainly, as a prerequisite to Theorem 3.1, it must be that when the boundary Λ consists of all leaves of $\widehat{\mathbb{T}}_{\Delta,L}$ and $\tau \equiv 0, \sigma \equiv 1$ on Λ, we have

$$|p_r^\tau - p_r^\sigma| \lesssim (1 - \delta)^L.$$

Because of the symmetry of $\widehat{\mathbb{T}}_{\Delta,L}$, this weaker claim is equivalent to showing that for all $L \in \mathbb{N}$,

$$\left| f_{\Delta-1}^{\circ L}(0) - f_{\Delta-1}^{\circ L}(1) \right| \lesssim (1 - \delta)^L. \tag{4}$$

Since these two τ, σ are the "extremal" cases, i.e. they minimize/maximize the marginal of the root respectively (simply by monotonicity of $f_{\Delta-1}$), Eq. (4) says that regardless of the choice of the initializing marginal probability p, the sequence $\{f_{\Delta-1}^{\circ L}(p)\}_{L=0}^\infty$ always converges to the same *fixed point* of $f_{\Delta-1}(\cdot)$. The threshold $\lambda_c(\Delta)$ precisely delineates between when this convergence happens and when it fails.

Proposition 3.3 ([Kel85]; see also [Bar16])**.** *For every λ and every Δ, the univariate function $f_{\Delta-1}$ has a* unique *fixed point $\hat{p} = \hat{p}(\lambda, \Delta)$. Furthermore, the following hold:*

- *Suppose $\lambda \leq \lambda_c(\Delta)$. Then for every $p \in [0,1]$, the sequence $\{f_{\Delta-1}^{\circ L}(p)\}_{L=0}^\infty$ converges to $\hat{p}$. If $\lambda \leq (1 - \delta) \cdot \lambda_c(\Delta)$, then $\left| f_{\Delta-1}'(\hat{p}) \right| \leq 1 - O(\delta)$ and*

$$\left| \hat{p} - f_{\Delta-1}^{\circ L}(p) \right| \lesssim (1 - O(\delta))^L, \qquad \forall L \in \mathbb{N}.$$

- *Suppose $\lambda > \lambda_c(\Delta)$. Then there exist two additional fixed points $\hat{p}_{\mathsf{odd}} < \hat{p} < \hat{p}_{\mathsf{even}}$ of $f_{\Delta-1}^{\circ 2}$ such that*

$$\lim_{L\to\infty} f_{\Delta-1}^{\circ 2L}(p) = \begin{cases} \hat{p}_{\mathsf{odd}}, & \text{if } p < \hat{p} \\ \hat{p}_{\mathsf{even}}, & \text{if } p > \hat{p} \\ \hat{p}, & \text{if } p = \hat{p} \end{cases}.$$

Proof Sketch. For simplicity, we'll only prove existence of the requisite fixed points and establish the local behavior of f_d around its unique fixed point; we refer the reader to [Bar16] for a full proof. For convenience, write $d = \Delta - 1$. That f_d has a unique fixed point $\hat{p} = \hat{p}(\lambda, d)$ for all λ and all d follows from the fact that f_d is monotone decreasing, with $f_d(0) = \frac{\lambda}{1+\lambda} > 0$ and $f_d(1) = 0 < 1$. Furthermore, if $p < \hat{p}$, then $f_d(p) > \hat{p}$, and vice versa. Since f_d is increasing in λ, it also follows that $\hat{p}$ is continuous and increasing as a function of λ.

We now establish that $\lambda \leq \lambda_c(d+1)$ if and only if locally around this fixed point $\hat{p}$, we have contraction, i.e. $|f_d'(\hat{p})| \leq 1$. Furthermore, if $\lambda \leq (1-\delta)\cdot\lambda_c(d+1)$, then we actually get a gap $|f_d'(\hat{p})| \leq 1 - O(\delta)$. A calculation reveals that

$$f_d'(p) = -d \cdot \frac{1 - f_d(p)}{1-p} \cdot f_d(p), \qquad \forall p \in [0,1],$$

and so at the fixed point $\hat{p}$, we have $|f_d'(p)| = d\cdot\hat{p}$. We claim that $\hat{p} \leq \frac{1}{d}$ if and only if $\lambda \leq \lambda_c(d+1)$, and furthermore, if $\lambda \leq (1-\delta)\cdot\lambda_c(d+1)$, then $\hat{p} \leq (1-O(\delta))\cdot\frac{1}{d}$. To see this, observe that $\hat{p} = f_d(\hat{p})$ is equivalent to $\hat{p} = \lambda(1-\hat{p})^{d+1}$. If $\lambda \leq \lambda_c(d+1) = \frac{d^d}{(d-1)^d}$ and $\hat{p} > 1/d$, then

$$\lambda(1-\hat{p})^{d+1} < \lambda_c(d+1)\cdot\left(1 - \frac{1}{d}\right)^{d+1} = \frac{1}{d} < \hat{p},$$

contradicting $\hat{p}$ being a fixed point. It follows that $\lambda \leq \lambda_c(d+1)$ forces $\hat{p} \leq \frac{1}{d}$, and by making this argument quantitative, we get that $\lambda \leq \lambda_c(d+1)$ forces $\hat{p} \leq (1-O(\delta))\cdot\frac{1}{d}$. A nearly-identical argument shows that $\lambda > \lambda_c(d+1)$ forces $\hat{p} > 1/d$.

At this juncture, we already see that if $\lambda \leq (1-\delta)\cdot\lambda_c(\Delta)$, then at least for p close to $\hat{p}$, the sequence $\{f_d^{\circ L}(p)\}_{L=0}^{\infty}$ converges to $\hat{p}$. This is because $|f_d'(\hat{p})| \leq 1 - O(\delta)$ implies by continuity that $|f_d'(q)| \leq 1 - O(\delta)$ for all q in a small neighborhood of $\hat{p}$, whence for p in this small neighborhood of $\hat{p}$, we have $|\hat{p} - f_d(p)| \leq |f_d'(q)|\cdot|\hat{p} - p| \leq (1-O(\delta))\cdot|\hat{p}-p|$ by the Mean Value Theorem (where q is some convex combination of p and $\hat{p}$).

Conversely, if $\lambda > \lambda_c(\Delta)$, then the fixed point $\hat{p}$ is repulsive, and so for small $\epsilon > 0$, we have $f_d^{\circ 2}(p) > p$ for $p \in (\hat{p}, \hat{p}+\epsilon)$ and $f_d^{\circ 2}(p) < p$ for $p \in (\hat{p}-\epsilon, \hat{p})$. Since $f_d^{\circ 2}(\hat{p}) = \hat{p}$, $f_d^{\circ 2}(0) = f_d\left(\frac{\lambda}{1+\lambda}\right) > 0$ and $f_d^{\circ 2}(1) = f_d(0) = \frac{\lambda}{1+\lambda} < 1$, it follows by continuity and the Intermediate Value Theorem that $f_d^{\circ 2}$ must have (at least) two other fixed points. These fixed points must come in pairs, in the sense that if $\hat{q} < \hat{p}$ is a fixed point of $f_d^{\circ 2}$ then so is $f_d(\hat{q}) > \hat{p}$. That there are exactly two other fixed points follows by showing that there exists p^* such that $f_d^{\circ 2}$ is convex in the interval $[0, p^*]$ and concave in the interval $[p^*, 1]$. For the details of the remainder of the argument, we refer to [Bar16]. □

What Proposition 3.3 says is that if $\lambda > \lambda_c(\Delta)$, then macroscopic *oscillations* start appearing in the marginals of the vertices of $\widehat{\mathbb{T}}_{\Delta,L}$ which persist as one goes up the tree. Each application of $f_{\Delta-1}$ just bounces back and forth between $\hat{p}_{\mathsf{odd}}$ and $\hat{p}_{\mathsf{even}}$. On the other hand, when $\lambda \leq \lambda_c(\Delta)$, all marginals contract towards the unique fixed point $\hat{p}$. This is analogous to how we need aperiodicity when studying mixing times of Markov chains.

The term "mixing" is also analogous in meaning. Here, spatial mixing refers to decorrelation (w.r.t. space) of the marginal of r from other vertices which are far away. In the setting of Markov chains, "mixing" (or "temporal mixing") refers to decorrelation (w.r.t. time) of the current state X_t from, say, the initial one X_0. Later in the course, we will study the relationship between these two notions of mixing. In particular, we will prove that for the hardcore model with $\lambda < \lambda_c(\Delta)$, Glauber dynamics mixes in $O(n \log n)$ steps.

Remark 2 (The Uniqueness Problem for Infinite-Volume Gibbs Measure). A mathematical problem of interest is rigorously defining a Gibbs distribution on an infinite graph like the lattice $\mathbb{Z}^d$ or the Δ-regular tree $\mathbb{T}_\Delta$. One natural approach would be to take appropriate "limits", e.g. for $\mathbb{T}_\Delta$, we could look at "the limit" of the distributions $\mu_{\widehat{\mathbb{T}}_{\Delta,L}}$ for $(\Delta-1)$-ary trees of increasing larger depth L. Since $\widehat{\mathbb{T}}_{\Delta,L}$ "converges" to $\mathbb{T}_\Delta$ as $L \to \infty$, one might hope that $\mu_{\widehat{\mathbb{T}}_{\Delta,L}}$ also "converges" to some unique probability measure.

What Proposition 3.3 essentially says is that this is possible if $\lambda \le \lambda_c(\Delta)$. On the other hand, when $\lambda > \lambda_c(\Delta)$, the parity of the depth L actually matters. In particular, the distributions $\{\mu_{\widehat{\mathbb{T}}_{\Delta,2L}}\}_{L=0}^\infty$ must have a different limit compared to the distributions $\{\mu_{\widehat{\mathbb{T}}_{\Delta,2L+1}}\}_{L=0}^\infty$, simply by virtue of the discrepancy between marginals they induce at the root vertex r. For this reason, the threshold $\lambda_c(\Delta)$ is sometimes called the *uniqueness threshold* for the hardcore model on the infinite Δ-regular tree, and the interval $[0, \lambda_c(\Delta)]$ is sometimes called the *uniqueness regime*.

3.3 Correlation Decay via Contraction

It turns out a good choice for the distance is given by composition with a suitable *potential function*. More specifically, we $\psi(p,q) \stackrel{\text{def}}{=} |\varphi(p) - \varphi(q)|$, where $\varphi : [0,1] \to \mathbb{R}_{\ge 0} \cup \{+\infty\}$ is a smooth strictly monotone function; note that strict monotonicity guarantees that φ admits a smooth and strictly monotone inverse map $\varphi^{-1} : \mathbb{R}_{\ge 0} \cup \{\infty\} \to [0,1]$. For the present discussion, we won't instantiate φ. We typically write $\Phi \stackrel{\text{def}}{=} \varphi'$.

Since our goal is now to study $|\varphi(p_r^\tau) - \varphi(p_r^\sigma)|$, it is natural to consider the induced tree recursion for the new variables $m = \varphi(p)$ given by

$$G_d(m_1, \dots, m_d) \stackrel{\text{def}}{=} \varphi\left(F_d\left(\varphi^{-1}(m_1), \dots, \varphi^{-1}(m_d)\right)\right).$$

We also write $g_d \stackrel{\text{def}}{=} \varphi \circ f_d \circ \varphi^{-1}$ for the analogous univariate recursion.

Lemma 3.4. *Assume $\varphi : [0,1] \to \mathbb{R}_{\ge 0} \cup \{+\infty\}$ is a smooth strictly monotone potential function. For convenience, further assume there exists some positive universal constants $A, B > 0$ such that $|\varphi(p)| \le A$ and $|\varphi'(p)| \ge B$ for all $p \in [0,1]$. Suppose for some constant $0 < \delta < 1$, we have that $\|\nabla G_d(\boldsymbol{m})\|_1 \le 1 - O(\delta)$ for every $\boldsymbol{m} \in \mathbb{R}_{\ge 0}^d$ and every $1 \le d \le \Delta - 1$. Then the conclusion of Theorem 3.1 holds.*

This is essentially a consequence of the Mean Value Theorem, which allows one to translate a bound on the gradient norm to direct contraction of G_d, plus a simple inductive argument. Lemma 3.4 reduces the task to finding a good φ satisfying $\|\nabla G_d(\boldsymbol{m})\|_1 \le 1 - O(\delta)$. We will see one such choice in the next lecture.

Remark 3. Typically, it is enough to impose the boundedness assumptions on φ and Φ to only hold in some interval $[a,b] \subseteq [0,1]$ for constants $0 < a, b < 1$, so long as, say, a constant number of applications of F_d to any collection of marginal vectors lands in $[a,b]$. For instance, for the hardcore model, by monotonicity of F_d, we get that

$$F_d(\boldsymbol{p}) \le F_d(\boldsymbol{0}) = \frac{\lambda}{1+\lambda}, \qquad \forall \boldsymbol{p} \in [0,1]^d.$$

Similarly,

$$F_d(F_d(\boldsymbol{p}_1), \dots, F_d(\boldsymbol{p}_d)) \ge F_d\left(\frac{\lambda}{1+\lambda} \cdot \boldsymbol{1}\right) = \frac{\lambda}{\lambda + (1+\lambda)^d}, \qquad \forall \boldsymbol{p}_1, \dots, \boldsymbol{p}_d \in [0,1]^d.$$

Such boundedness requirements are very easy to satisfy; the main challenge is ensuring the gradient norm bound.

Proof of Lemma 3.4. We first prove that $|\varphi(p_r^\tau) - \varphi(p_r^\sigma)| \le (1-\delta)^{\mathrm{dist}(r, \Lambda_{\tau,\sigma})} \cdot A$. We will then convert this back to total variation distance at some loss in the constant. We write $m = \varphi(p)$ for the new variables after applying the transformation φ. Observe that by the Mean Value Theorem, if $\boldsymbol{m}, \boldsymbol{m}' \in \mathbb{R}_{\ge 0}^d$ are arbitrary, then

$$\begin{aligned}
|G_d(\boldsymbol{m}) - G_d(\boldsymbol{m})| &= |\langle \nabla G_d(\boldsymbol{m}''), \boldsymbol{m} - \boldsymbol{m}' \rangle| && \text{(Mean Value Theorem)} \\
&\le \|\nabla G_d(\boldsymbol{m}'')\|_1 \cdot \|\boldsymbol{m} - \boldsymbol{m}'\|_\infty && \text{(Hölder's Inequality)} \\
&\le (1 - O(\delta)) \cdot \max_{1 \le i \le d} |m_i - m_i'|, && \text{(Gradient Assumption)}
\end{aligned}$$

where $\boldsymbol{m}''$ is some convex combination of $\boldsymbol{m}, \boldsymbol{m}'$. Applying this inequality inductively and writing $L_r(\ell)$ for the set of vertices at distance exactly ℓ from the root r, we see that

$$\begin{aligned}
|\varphi(p_r^\tau) - \varphi(p_r^\sigma)| &\le (1 - O(\delta)) \cdot \max_{u \in L_r(1)} |\varphi(p_u^\tau) - \varphi(p_u^\sigma)| \\
&\le \cdots \\
&\le (1-\delta)^{\mathrm{dist}(r, \Lambda_{\tau,\sigma})} \cdot \max_{u \in L_r(\mathrm{dist}(r,\Lambda_{\tau,\sigma}))} |\varphi(p_u^\tau) - \varphi(p_u^\sigma)| \\
&\le (1-\delta)^{\mathrm{dist}(r, \Lambda_{\tau,\sigma})} \cdot A. && (\varphi \text{ is } A\text{-bounded})
\end{aligned}$$

Note that during this induction, we could have hit a vertex in $\Lambda \setminus \Lambda_{\tau,\sigma}$, but this doesn't matter since for any such vertex u, we automatically have $|\varphi(p_u^\tau) - \varphi(p_u^\sigma)| = 0$. To complete the proof, we just convert back to total variation distance. We have

$$\begin{aligned}
\|\mu_r^\tau - \mu_r^\sigma\|_{\mathsf{TV}} &= |p_r^\tau - p_r^\sigma| \\
&= \left|(\varphi^{-1})'(\varphi(q))\right| \cdot |\varphi(p_r^\tau) - \varphi(p_r^\sigma)| && \text{(Mean Value Theorem)} \\
&\le \frac{1}{B} \cdot A \cdot (1-\delta)^{\mathrm{dist}(r,\Lambda_{\tau,\sigma})}. && \text{(Boundedness Assumptions)}
\end{aligned}$$

Setting $C = A/B$ completes the proof. □

References

[Bar16] Alexander Barvinok. *Combinatorics and Complexity of Partition Functions.* Vol. 30. Springer Algorithms and Combinatorics, 2016 (cit. on pp. 5, 6).

[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 p. 1).

[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 p. 1).

[Che+23] Zongchen Chen, Kuikui Liu, Nitya Mani, and Ankur Moitra. "Strong spatial mixing for colorings on trees and its algorithmic applications". In: *arXiv preprint arXiv:2304.01954* (2023). To Appear in FOCS 2022 (cit. on p. 3).

[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 p. 4).

[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 p. 1).

[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 p. 4).

[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 p. 4).

[Hås01] Johan Håstad. "Some Optimal Inapproximability Results". In: *J. ACM* 48.4 (July 2001), pp. 798–859. ISSN: 0004-5411 (cit. on p. 3).

[Hås99] Johan Håstad. "Clique is hard to approximate within $n^{1-\epsilon}$". In: *Acta Mathematica* 182 (1999), pp. 105–142 (cit. on p. 3).

[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 pp. 4, 5).

[KF09] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009 (cit. on p. 2).

[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 p. 5).

[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 p. 5).

[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 p. 4).

[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 p. 4).

[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. 1, 4).

A Unfinished Proofs

Lemma A.1 (General Tree Recursion). *Let $\boldsymbol{A} \in \mathbb{R}_{\geq}^{q\times q}, \boldsymbol{\lambda} \in \mathbb{R}_{\geq 0}^{q}$ be the parameters of a q-spin system. Let T be a tree, and let $r \in T$ be an arbitrarily chosen root vertex, with children $u_1, \ldots, u_d$ and corresponding subtrees $T_1, \ldots, T_d$. Then for every color $\mathfrak{c} \in [q]$,*

$$\mu_{T,r}(\mathfrak{c}) = \frac{\boldsymbol{\lambda}_{\mathfrak{c}} \prod_{i=1}^{d} \sum_{\mathfrak{b}\in[q]} A_{\mathfrak{b},\mathfrak{c}} \cdot \mu_{T_i,u_i}(\mathfrak{b})}{\sum_{\mathfrak{a}\in[q]} \boldsymbol{\lambda}_{\mathfrak{a}} \prod_{i=1}^{d} \sum_{\mathfrak{b}\in[q]} A_{\mathfrak{b},\mathfrak{a}} \cdot \mu_{T_i,u_i}(\mathfrak{a})}.$$

Proof. We use a simple combinatorial argument. Imagine we split r into d distinct vertices $r_1, \ldots, r_d$. The resulting graph $\tilde{T}$ now consists of d trees $\tilde{T}_1, \ldots, \tilde{T}_d$ which are disconnected from each other, where $\tilde{T}_i$ is formed by taking T_i and adding a new vertex r_i with a single edge connecting r_i to u_i; in particular, r_i has degree-1 in $\tilde{T}_i$. For each such copy r_i, we reset its external field vector to $\{\boldsymbol{\lambda}_{\mathfrak{c}}^{1/d}\}_{\mathfrak{c}\in[q]}$. Then

$$\begin{aligned}
\mu_{T,r}(\mathfrak{c}) &= \frac{Z_T(r \leftarrow \mathfrak{c})}{\sum_{\mathfrak{a}\in[q]} Z_T(r \leftarrow \mathfrak{a})} = \frac{Z_{\tilde{T}}(r_1, \ldots, r_d \leftarrow \mathfrak{c})}{\sum_{\mathfrak{a}\in[q]} Z_{\tilde{T}}(r_1, \ldots, r_d \leftarrow \mathfrak{a})} \\
&= \frac{\prod_{i=1}^{d} Z_{\tilde{T}_i}(r_i \leftarrow \mathfrak{c})}{\sum_{\mathfrak{a}\in[q]} \prod_{i=1}^{d} Z_{\tilde{T}_i}(r_i \leftarrow \mathfrak{a})}, && (\tilde{T}_i \text{ are disconnected from each other})
\end{aligned}$$

where we write $Z_T(r \leftarrow \mathfrak{c})$ for the partition function of the spin system on T restricted to terms $\sigma : V \to [q]$ satisfying $\sigma(r) = \mathfrak{c}$. Since r_i has degree-1 with u_i as its unique neighbor,

$$Z_{\tilde{T}_i}(r_i \leftarrow \mathfrak{a}) = \boldsymbol{\lambda}_{\mathfrak{a}}^{1/d} \sum_{\mathfrak{b}\in[q]} A_{\mathfrak{b},\mathfrak{a}} \cdot Z_{T_i}(u_i \leftarrow \mathfrak{b}), \qquad \forall \mathfrak{a} \in [q],$$

and so

$$\mu_{T,r}(\mathfrak{c}) = \frac{\boldsymbol{\lambda}_{\mathfrak{c}} \prod_{i=1}^{d} \sum_{\mathfrak{b}\in[q]} A_{\mathfrak{b},\mathfrak{c}} \cdot Z_{T_i}(u_i \leftarrow \mathfrak{b})}{\sum_{\mathfrak{a}\in[q]} \boldsymbol{\lambda}_{\mathfrak{a}} \prod_{i=1}^{d} \sum_{\mathfrak{b}\in[q]} A_{\mathfrak{b},\mathfrak{a}} \cdot Z_{T_i}(u_i \leftarrow \mathfrak{a})}.$$

Since $\mu_{T_i,u_i}(\mathfrak{b}) = \frac{Z_{T_i}(u_i \leftarrow \mathfrak{b})}{Z_{T_i}}$ by definition, dividing all terms by Z_{T_i} in both the numerator and denominator yields the claim. $\square$