Static and dynamic correspondences in Markov sampling

Joao Basso

February 5th, 2026*

Contents

1 Introduction

Let Ω be a space with a potential $H\colon \Omega \to \mathbb{R}$ along with a parameter $\beta > 0$. We define the associated Gibbs measure as $\mu\colon \Omega \to [0, 1]$ as

$$\mu(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z}, \qquad \text{where} \qquad Z = \sum_{\omega \in \Omega} e^{-\beta H(\sigma)}. \tag{1.1}$$

A widespread technique to sample from μ is to design a Markov chain with transition matrix P whose stationary distribution is π. A long line of research has to been to elucidate the connections between properties of μ to those of P. In this note we will discuss in particular the connection between uniqueness of the Gibbs measure in the infinite volume, decay of correlations and mixing times.

As our running example, we will consider the ferromagnetic 2D Ising model. That is, let $G(V, E)$ be a finite subgraph of $\mathbb{Z}^2$ and $\Omega = \{\pm 1\}^n$. For $\sigma \in \Omega$, the potential is then

$$H(\sigma) = -\sum_{(i,j) \in E} \sigma_i \sigma_j. \tag{1.2}$$

Note that this has two minimum energy states: $(1, \dots, 1)$ and $(-1, \dots, -1)$.

When talking about mixing, we will refer to the *mixing time*, defined as the minimum time such that a Markov chain started in any state is δ away from the stationary distribution in TV distance.

*These notes accompanied a talk given at the Quantum Many-Body seminar at the math department at UC Berkeley.

2 Nonuniqueness of Gibbs states

We will keep the mathematical formalism to a minimum. For a completely rigorous treatment, see e.g. [FV17].

We start by defining the notion of the infinite volume Gibbs measure. For ease, we specialize it to the 2D Ising model, but this may be defined in larger generality [Dob68c].

Definition 2.1 (Gibbsian distribution). *Let $\{\xi(t) \colon t \in \mathbb{Z}^2\}$ be a random field, where $\xi \colon \mathbb{Z}^2 \to \{\pm 1\}$. Let $V = \{t_1, \dots, t_{|V|}\} \subset \mathbb{Z}^2$ and $x \colon \mathbb{Z}^2 \setminus V \to \{\pm 1\}$ be a set of boundary conditions. Denote by E the set of edges of $\mathbb{Z}^2$. We say that ξ is a* Gibbsian distribution *if*

$$\mathbb{P}\left[\xi(t_1) = x_1, \dots, \xi(t_{|V|}) = x_{|V|}\right] = q_V(x_1, \dots, x_{|V|} \mid x(t)), \quad \text{and} \tag{2.1}$$

$$\mathbb{P}\left[\xi(t) = x(t)\right] = 1, \quad t \in \mathbb{Z}^d \setminus V, \tag{2.2}$$

where

$$q_V(x_1, \dots, x_{|V|} \mid x(t)) := \frac{e^{-\beta H_V(x_1, \dots, x_{|V|} | x(t))}}{\sum_{x_1, \dots, x_{|V|} \in X} e^{-\beta H_V(x_1, \dots, x_{|V|} | x(t))}}, \tag{2.3}$$

$$H_V(x_1, \dots, x_{|V|} \mid x(t)) := -\sum_{1 \le i < j \le |V|} x_i x_j \mathbf{1}\{(t_i, t_j \in E)\} - \sum_{i=1}^{|V|} \sum_{t \in \mathbb{Z}^2 \setminus V} x_i x(t) \mathbf{1}\{(t_i, t) \in E\}. \tag{2.4}$$

In words, q_V is the Gibbs distribution of V with $x(t)$ as the boundary condition, and H_V is the potential for V along with its boundary. The Gibbsian field is a distribution on $\mathbb{Z}^2$ that has the correct conditional probabilities when restricted to V. One may wonder if such an extension always exists. Existence can be proven by taking V larger and larger. The next natural question is:

Question 2.2. *Are Gibbsians distributions unique for all $V, x(t)$?*

This turns out to not always be the case. For an example, take V to be a $\sqrt{n} \times \sqrt{n}$ sublattice. The intuitive idea is that, at low enough temperatures, there are long range correlations since neighboring sites want to be aligned due to the strong ferromanetic bonds. Hence a boundary condition of all $+1$ will force a large fraction of V to be $+1$, and there will only be scattered islands of -1. The opposite happens if we flip the boundary. Hence these two configurations lead to different Gibbsian distributions.

To formalize this idea, we use Peierl's argument [Pei36, Gri64, Bon14]. Our strategy will be to consider the average magnetization. For $\sigma \in V$, let

$$m(\sigma) = \frac{1}{n} \sum_{i=1}^{n} \sigma(i) = 1 - \frac{2N_-(\sigma)}{n}, \tag{2.5}$$

where N_- counts the number of -1's. So it suffices to compute $\langle N_- \rangle$. First fix all spins on ∂V to $+1$. Our strategy will be to represent each possible configuration of V by contours. Enclose each spin by a square, creating a dual lattice. Then we draw a contour around every square with a -1. We delete two drawn edges that overlap and, if they meet at corners, shave them a bit so they don't touch. See Fig. 1 for an illustration.

The set of all such drawings $\mathcal{C}$ is in bijection with the set of configurations of V given an assignment of ∂V, since every contour edge indicates different bits on each side of it. Hence, we will count contours instead of assignments directly. This is easier than counting the signs directly, as it's hard to know the energy of an arbitrary configuration, whereas the contours keep track of disagreements – their appear exactly on the perimeter. Let Γ_ℓ be the set of all possible contours with perimeter ℓ and $A(\gamma)$ be the area enclosed by $\gamma \in \bigcup_\ell \Gamma_\ell$. Then

$$\langle N_- \rangle = \left\langle \sum_{\gamma \in \bigcup_\ell \Gamma_\ell} A(\gamma) \mathbf{1}\{\gamma \text{ occurs}\} \right\rangle = \sum_\ell \sum_{\gamma \in \Gamma_\ell} A(\gamma) \langle \mathbf{1}\{\gamma \text{ occurs}\} \rangle = \sum_\ell \sum_{\gamma \in \Gamma_\ell} A(\gamma)\, \mathbb{P}[\gamma \text{ occurs}]. \tag{2.6}$$

So we must compute the area and the probability of each possible contour.

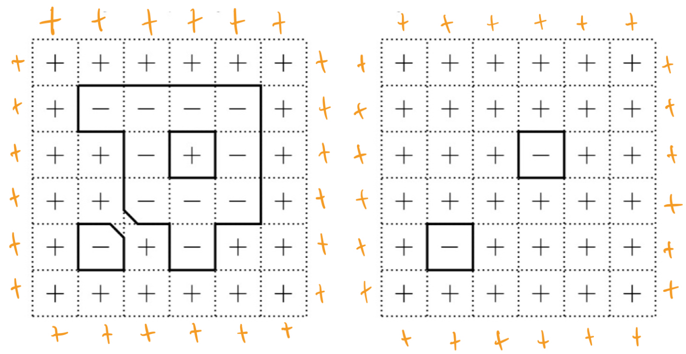

Figure 1: Left: Illustration of drawing the contours in Peierl's argument. The boundary is all +1s. Right: If C is the larger contour on the left, this is $\bar{C}$. Image partly borrowed from [Bon14].

Lemma 2.3 (Area of contours)**.** *Suppose $\gamma \in \Gamma_\ell$. Then*

$$A(\gamma) \leq \frac{\ell^2}{16}. \tag{2.7}$$

Proof. We can surround the contour with a minimal rectangle with dimensions $a \times b$. Note that the perimeter of γ can only be larger than the perimeter of the rectangle: $2(a+b) \leq \ell$. It is a standard optimization problem that the square is the shape that minimizes area for a fixed perimeter. So we can take $a = b = \ell/4$. Squaring finishes the proof. □

Lemma 2.4 (Probability of contour)**.** *Suppose that $\gamma \in \Gamma_\ell$. Then*

$$\mathbb{P}[\gamma \text{ occurs}] \leq e^{-2\beta\ell}. \tag{2.8}$$

This captures the intuition that there should not be many islands with spins opposite to the boundary: the probability of such islands is exponentially small on their size and β.

Proof. We can write

$$\mathbb{P}[\gamma \text{ occurs}] = \frac{\sum_{C \in \mathcal{C}:\gamma \in C} e^{-\beta H(C)}}{\sum_{C \in \mathcal{C}} e^{-\beta H(C)}}, \tag{2.9}$$

where $H(C)$ is the energy of the drawing C and we wrote $\gamma \in C$ if the contour γ appears in C. If $\gamma \in C$, let $\bar{C}$ be the same drawing, except all signs inside γ are flipped; that is, the contour γ disappears since now we have agreement across its perimeter; see Fig. 1 for an illustration. Since the only bonds that change are across the perimeter of γ, it follows that

$$H(\bar{C}) = H(C) - 2L. \tag{2.10}$$

Now we remove some of the drawings in the denominator:

$$\mathbb{P}[\gamma \text{ occurs}] \leq \frac{\sum_{C \in \mathcal{C}:\gamma \in C} e^{-\beta H(C)}}{\sum_{\bar{C}:\, C \in \mathcal{C}, \gamma \in C} e^{-\beta H(\bar{C})}} = \frac{\sum_{C \in \mathcal{C}:\gamma \in C} e^{-\beta H(C)}}{e^{2\beta\ell} \sum_{\bar{C}:\, C \in \mathcal{C}, \gamma \in C} e^{-\beta H(C)}} = e^{-2\beta\ell}. \tag{2.11}$$

□

Lemma 2.5 (Number of contours with perimeter)**.** *It holds that*

$$|\Gamma_\ell| \leq \frac{2n}{\ell} 3^\ell. \tag{2.12}$$

Proof. The lattice with n vertices has fewer edges than a 4-regular graph with n vertices, which in turn has $2n$ edges. Hence there are at most $2n$ initial locations for edges, and afterwards we can pick one out of 3, yielding $2n3^\ell$. We overcounted since we can start at any edge, so we divide by ℓ. □

Returning to Eq. (2.6), we can combine these lemmas to see that

$$\langle N_- \rangle \leq \sum_\ell |\Gamma_\ell| \frac{\ell^2}{16} e^{-2\beta\ell} = \frac{n}{8} \sum_\ell \ell (3e^{-2\beta})^\ell. \tag{2.13}$$

For large enough β, this sequence is converging to less than 1/2 let's say, and returning to Eq. (2.5), we see that we get positive magnetization. Inverting the boundary, we get the opposite scenario. This shows that there cannot be a unique infinite-volume Gibbsian distribution.

Question 2.6. *When are Gibbsians distributions unique for all $V, x(t)$?*

3 Dobrushin's conditions for uniqueness

Dobrushin (later joined by Shlosman) developed a long line of work seeking to answer Question 2.6 [Dob68a, Dob68b, Dob70, DS85, DS87], which culminated in 12 equivalent properties which guarantee the uniqueness of the Gibbsian distribution. We consider a particular one related to correlations.

Definition 3.1 (Influence matrix)**.** *Let $x \in \Omega$, $i \in [n]$ and π the stationary distribution. Define, for $b \in \{\pm 1\}$,*

$$\pi_i(x, b) := \mathop{\mathbb{P}}_{z \sim \pi}[z_i = b \mid z_j = x_j, j \neq i]. \tag{3.1}$$

Also define

$$S_j := \{(x, y) \in \Omega^2 \colon x_k = y_k, k \neq j\}, \tag{3.2}$$

that is, all bitstrings agreeing outside j. Then we define the influence matrix $R \in \mathbb{R}^{n \times n}$ *as*

$$R_{ij} = \max_{(x,y) \in S_j} ||\pi_i(x, \cdot) - \pi_i(y, \cdot)||_{\mathrm{TV}}. \tag{3.3}$$

Intuitively, this captures how much the update of bit i is affected by flipping bit j, that is, the *influence of j on i*. Naturally, if column j has many entries, j affects many bits; if row i has many entries, i is affected by many bits. Hence, we can understand correlations by looking at norms of R.

The literature usually refers to the following two conditions:

$$\textit{Dobrushin condition:} \quad \|R\|_1 < 1 \quad \text{(max row sum, most affected bit)} \tag{3.4}$$

$$\textit{Dobrushin-Shlosman condition:} \quad \|R\|_\infty < 1 \quad \text{(max column sum, most affecting bit)} \tag{3.5}$$

This condition is enough to ensure the uniqueness of the infinite volume Gibbsian distribution. Initially it was necessary to assume that the underlying graph was a sublattice of $\mathbb{Z}^d$, but later [Wei05] showed how to drop that assumption.

Theorem 3.2 (Uniqueness from Dobrushin's condition)**.** *If $\|R\|_1 < 1$, then the Gibbsian distribution is unique for all $V, x(t)$.*

4 Dobrushin's conditions and mixing

Now we turn to the dynamics. We will consider the canonical Glauber dynamics, where we pick a bit a random, and refresh it according to $\pi_i(x, \cdot)$. For more background on Glauber dynamics, see Section A.

It turns out that both conditions above, Eqs. (3.4) and (3.5), imply $O(n \log n)$ mixing [AH87], and in fact so does $\|R\| \leq 1 - \epsilon$ for any norm $\|\cdot\|$ [DGJ09]. For spin systems, this is optimal [HS05]. We will prove the version with the 2-norm studied in [Hay06].

Theorem 4.1 (Dobrushin condition implies fast mixing)**.** *Suppose that* $\|R\|_2 \le 1 - \epsilon$*. Then*

$$t_{\mathrm{mix}}(\delta) \le \frac{n}{\epsilon} \log \frac{n}{\delta}. \tag{4.1}$$

Proof. We will a coupling argument. For background on couplings, see Section B. For $x, y \in \Omega$, let two initial states be $X_0 = x_0, Y_o = y_0$. For $t \ge 0$, will run the optimal coupling on (X_t, Y_t), that is, at all steps, we pick the same site i at random to update, and update both copies based on the optimal coupling. Define the vector $\vec{p}_t \in \mathbb{R}^n$ so that

$$p_t(i) = \mathbb{P}[X_t(i) \ne Y_t(i)]. \tag{4.2}$$

$\vec{p}$ will be our way of tracking coalescence of the two copies. That is, by an union bound,

$$\mathbb{P}[X_t \ne Y_t] \le \sum_{i=1}^{n} p_t(i) = \|\vec{p}\|_1. \tag{4.3}$$

Let's see how $p(i)$ evolves in one step. If some $j \ne i$ is chosen to be updated, which happens with probability $\frac{n-1}{n}$, $p(i)$ does not change. If i is chosen, which happens with probability $\frac{1}{n}$, then we must compute this change.

$$\begin{aligned} p_{t+1}(i) &= \left(1 - \frac{1}{n}\right) p_t(i) + \frac{1}{n} \mathbb{P}[X_{t+1}(i) \ne Y_{t+1}(i) \mid i \text{ is chosen}] && (4.4)\\ &= \left(1 - \frac{1}{n}\right) p_t(i) + \frac{1}{n} \mathop{\mathbb{E}}_{X_t, Y_t} [\mathbb{P}[X_{t+1}(i) \ne Y_{t+1}(i) \mid X_t, Y_t, i \text{ is chosen}]] && (4.5)\\ &= \left(1 - \frac{1}{n}\right) p_t(i) + \frac{1}{n} \mathop{\mathbb{E}}_{X_t, Y_t} ||\pi_i(X_t, \cdot) - \pi_i(Y_t, \cdot)||_{\mathrm{TV}}, && (4.6) \end{aligned}$$

by the law of total expectation and optimality of the coupling (see Proposition B.6.) Now we perform a hybrid argument. Let $Z^{(0)}, \ldots, Z^{(m)} \in \Omega$ where $Z^{(0)} = X_t, Z^{(m)} = Y_t$, $|Z^{(k)} - Z^{(k+1)}| = 1$ and $m = |X_t = Y_t|$, where $|\cdot|$ is the Hamming weight. Then

$$\begin{aligned} p_{t+1}(i) &\le \left(1 - \frac{1}{n}\right) p_t(i) + \frac{1}{n} \mathop{\mathbb{E}}_{X_t, Y_t} \sum_{k=0}^{m-1} \left|\left|\pi_i(Z^{(k)}, \cdot) - \pi_i(Z^{(k+1)}, \cdot)\right|\right|_{\mathrm{TV}} && (4.7)\\ &\le \left(1 - \frac{1}{n}\right) p_t(i) + \frac{1}{n} \sum_{j=1}^{n} R_{ij} \mathop{\mathbb{E}}_{X_t, Y_t} [\mathbf{1}\{X_t(j) \ne Y_t(j)\}] && (4.8)\\ &= \left(1 - \frac{1}{n}\right) p_t(i) + \frac{1}{n} \sum_{j=1}^{n} R_{ij} p_t(j), && (4.9) \end{aligned}$$

by triangle inequality and Eq. (3.3). We can rewrite this as a vector equation,

$$\vec{p}_{t+1} \le A\vec{p}_t \quad \text{with} \quad A = \left(1 - \frac{1}{n}\right) I + \frac{1}{n} R, \tag{4.10}$$

and by induction, we have $\vec{p}_t = A^t \vec{p}_0$. Returning to Eq. (4.3), we get

$$\begin{aligned} \mathbb{P}[X_t \ne Y_t] &\le \|\vec{p}_t\|_1 && (4.11)\\ &= \left\|A^t \vec{p}_0\right\|_1 && (4.12)\\ &\le \sqrt{n} \left\|A^t \vec{p}_0\right\|_2 && \text{(Cauchy-Schwarz)} \quad (4.13)\\ &\le \sqrt{n} \|A\|_2^t \|\vec{p}_0\|_2 && \text{(submultiplicativity)} \quad (4.14)\\ &\le n\|A\|_2^t. && (0 \le p_0(i) \le 1) \quad (4.15) \end{aligned}$$

Finally, by the Dobrushin condition,

$$\|A\|_2 \leq 1 - \frac{1}{n} + \frac{1}{n}\|R\|_2 \leq 1 - \frac{1}{n} + \frac{1-\epsilon}{n} = 1 - \frac{\epsilon}{n}, \tag{4.16}$$

which implies the bound

$$\mathbb{P}[X_t \neq Y_t] \leq \left(1 - \frac{\epsilon}{n}\right)^t n \leq ne^{-\epsilon t/n}. \tag{4.17}$$

A coalescence time immediately implies a mixing time bound. That is, again by Proposition B.6,

$$||X_t - Y_t||_{\mathrm{TV}} \leq \mathbb{P}[X_t \neq Y_t] \leq ne^{-\epsilon t/n}, \tag{4.18}$$

and choosing $t = \frac{n}{\epsilon}\log\frac{n}{\delta}$ makes this quantity $\leq \delta$. □

To see this in practice, let's apply it to the Ising model.

Lemma 4.2 (Influence matrix bound for the Ising model)**.** *Consider the (ferro or antiferromagnetic) Ising model with inverse temperature β on a graph $G = (V, E)$ with adjacency matrix A. That is, the energy is $H(\sigma) = \pm\sum_{(i,j)\in E}\sigma_i\sigma_j$. Then standard Glauber dynamics (c.f. Section A) has an influence matrix R such that*

$$\|R\|_2 \leq \tanh(\beta)\|A\|_2. \tag{4.19}$$

Proof. We compute for the ferromagnetic case but the proof is the same for the ferromagnetic case with some sign flips. Our strategy will be to upper-bound R_{ij} for all i, j. Fix $i \neq j \in [n]$ and $\sigma \in \Omega$. Suppose i has d neighbors, r of which are set to 1 (and $d - r$ to -1). By Eq. (A.6),

$$\begin{aligned}\pi_i(\sigma, \pm 1) &= \frac{1 \pm \tanh\left(\beta\sum_{j\in N(i)}\sigma(j)\right)}{2} && (4.20)\\ &= \frac{1 \pm \tanh\left(\beta(2r-d)\right)}{2}. && (4.21)\end{aligned}$$

To calculate R_{ij}, we consider another bitstring $\rho \in \Omega$ which only differs from σ at k. If $k \notin N(i)$, the two marginals are the same so the variation distance is 0. If $k \in N(i)$, $\sum_{j\in N(i)}\rho_k = 2r - d + 2$ or $= 2r - d - 2$.

Starting with the first case, we compute the TV distance with Lemma B.2:

$$\begin{aligned}||\pi_i(\sigma,\cdot) - \pi_i(\rho,\cdot)||_{\mathrm{TV}} &= \frac{1}{2}\sum_{b\in\{\pm 1\}}|\pi_i(\sigma,b) - \pi_i(\rho,b)| && (4.22)\\ &= \frac{1}{2}\cdot 2\left|\frac{\tanh(\beta(2r-d)) - \tanh(\beta(2r-d+2))}{2}\right| && (4.23)\\ &= \frac{\tanh(\beta(2r-d+2)) - \tanh(\beta(2r-d))}{2}, && (4.24)\end{aligned}$$

since $\tanh(\cdot)$ is increasing. Maximizing $\tanh(x + r) - \tanh(x)$ and equating the derivative to zero gives

$$\operatorname{sech}^2(x + r) = \operatorname{sech}^2(x). \tag{4.25}$$

Since $\operatorname{sech}^2(\cdot)$ is even and decreasing on $[0, \infty)$, we must have $x + r = -x$, or in our current symbols,

$$\beta(2r - d + 2) = -\beta(2r - d) \implies r = \frac{d-1}{2}, \tag{4.26}$$

which attains the value $\tanh(\beta(2r - d + 2)) - \tanh(\beta(2r - d)) = 2\tanh(\beta)$. The second case is similar, also with a maximum of $2\tanh(\beta)$ when $r = \frac{d+1}{2}$. All in all, we get $||\pi_i(\sigma,\cdot) - \pi_i(\rho,\cdot)||_{\mathrm{TV}} \leq \tanh(\beta)$ and so

$$R_{ij} \leq \tanh(\beta) \text{ for all } i, j. \tag{4.27}$$

The next observation is that, by the spatial Markov property, if $(i,j) \notin E$, that is, $A_{ij} = 0$, then R_{ij}. This can be seen from Eq. (A.6), where $\pi_i(\sigma, \cdot)$ is not affected by vertices not neighbors of i, so $||\pi_i(x,\cdot) - \pi_i(y,\cdot)||_{\mathrm{TV}}$ in Eq. (3.3) is zero. From these observations, we get

$$R_{ij} \le \tanh(\beta) A_{ij}. \tag{4.28}$$

An application of Lemma D.1 completes the proof. □

Lemma 4.3 (Influence matrix for 2D Ising model)**.** *Consider the Ising model with inverse temperature β on an $n \times n$ lattice. Then standard Glauber dynamics (c.f. Section A) has an influence matrix R such that*

$$\|R\|_2 \le 4 \tanh(\beta) \tag{4.29}$$

Proof. Let A be the adjacency matrix of the 2D grid. Since A is Hermitian, $\|A\|_2 = \max \operatorname{spec}(|A|)$. The spectrum of A is studied in Section C. By Corollary C.6, the largest eigenvalue in magnitude is given by

$$\|A\|_2 = 4\cos\left(\frac{\pi}{n+1}\right) \le 4. \tag{4.30}$$

Applying Lemma 4.2 finishes the proof. □

Remark 4.4 (Generalizing to any bounded degree graph)**.** Although computing the spectrum of A is entertaining, we can avoid it by employing a more general argument. Since the spectrum of the adjacency matrix of a graph is always upper-bounded by its maximum degree (see Lemma D.2 for a proof), we immediately deduce that $\|A\|_2 \le 4$ in the 2D Ising setting.

Corollary 4.5 (Fast mixing for the 2D Ising model)**.** *Let $0 < \epsilon = O(1)$ and suppose that*

$$\beta \le \operatorname{arctanh}\left(\frac{1-\epsilon}{4}\right). \tag{4.31}$$

Then standard Glauber dynamics has a mixing time of

$$t_{\mathrm{mix}}(\delta) \le \frac{n}{\epsilon} \log \frac{n}{\delta}. \tag{4.32}$$

Proof. This follows from combining Theorem 4.1 and Lemma 4.3. □

That is, Dobrushin's criteria guarantees fast mixing beyond the threshold of $\beta = \operatorname{arctanh}(1/4) \approx 0.255413$. Note that this is still away from the known critical threshold of $\beta_c = \log(1+\sqrt{2})/2 \approx 0.440687$ established by Onsager [Ons44, LS12]. Fast mixing all the way to β_c was proved by sharper techniques in [MO94a, MO94b].

A Glauber dynamics

Suppose that the sample space is $\Omega = \Sigma^n$. The generic Glauber dynamics is that, at each step, we pick $i \in [n]$ at random, and refresh it according to the marginal from the stationary distribution π. That is, if the current state is $\sigma_t \in \Omega$, then, for $v \in \Sigma$,

$$\mathbb{P}[\sigma_{t+1}(i) = v] = \mathbb{P}_{\sigma\sim\pi}[\sigma(i) = v \mid \sigma(j) = \sigma_t(j), j \ne i]. \tag{A.1}$$

For concreteness, suppose we have a graph $G = (V, E)$ and a bitstring $\sigma \in \{\pm 1\}^{|V|}$ has energy given by

$$H(\sigma) = -\sum_{(i,j)\in E} J_{ij}\sigma_i\sigma_j, \tag{A.2}$$

for some coupling matrix J. Suppose the stationary distribution is, for some $\beta \ge 0$,

$$\pi(\sigma) \propto e^{-\beta H(\sigma)}. \tag{A.3}$$

Then we can compute the updates explicitly. Let $N(i)$ be the neighbors of i in G. Eq. (A.1) becomes

$$\mathbb{P}[\sigma_{t+1}(i) = \pm 1] = \frac{e^{-\beta\left(\mp \sum_{j\in N(i)} J_{ij}\sigma_t(j) + \text{rest}\right)}}{e^{-\beta\left(\mp \sum_{j\in N(i)} J_{ij}\sigma_t(j) + \text{rest}\right)} + e^{-\beta\left(\pm \sum_{j\in N(i)} J_{ij}\sigma_t(j) + \text{rest}\right)}} \tag{A.4}$$

$$= \frac{e^{\pm\beta\left(\sum_{j\in N(i)} J_{ij}\sigma_t(j)\right)}}{e^{\pm\beta\left(\sum_{j\in N(i)} J_{ij}\sigma_t(j)\right)} + e^{\mp\beta\left(\sum_{j\in N(i)} J_{ij}\sigma_t(j)\right)}} \tag{A.5}$$

$$= \frac{1 \pm \tanh\left(\beta \sum_{j\in N(i)} J_{ij}\sigma_t(j)\right)}{2}. \tag{A.6}$$

From an algorithmic perspective, this is indeed implementable since we only need the current bits and the couplings J.

B TV distance and couplings

Here we elaborate on the key connection, used in the proof of Theorem 4.1, between a distance of distributions and couplings of random variables. A great source for this is [LP17]; [Woo21, Col24] are also nice videos.

We start with a notion of distance between two probability distributions.

Definition B.1 (Total Variance (TV) distance)**.** *Let $\mu, \nu \colon \Omega \to [0,1]$ be two measures. The we define their* TV distance *as*

$$||\mu - \nu||_{\text{TV}} = \sup_{A\subseteq\Omega} |\mu(A) - \nu(A)|. \tag{B.1}$$

That is, the TV distance is the largest separation that an event may witness. There's an alternative characterization of this distance that is often easier to compute.

Lemma B.2 (Another TV distance characterization)**.** *Let $\mu, \nu \colon \Omega \to [0,1]$ be two measures. Then*

$$||\mu - \nu||_{\text{TV}} = \frac{1}{2} \sum_{\omega\in\Omega} |\mu(\omega) - \nu(\omega)|. \tag{B.2}$$

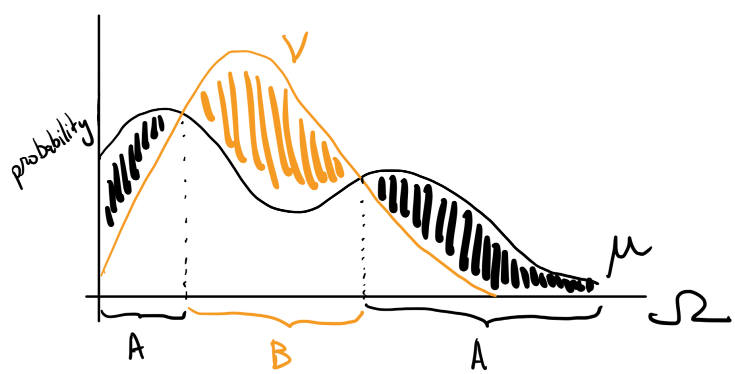

Figure 2: Illustration for the proof of Lemma B.2. The orange and black regions have the same area, which equals $||\mu - \nu||_{\text{TV}}$.

Proof. It is helpful to consider Fig. 2. Let

$$A := \{\omega \in \Omega \colon \mu(\omega) \geq \nu(\omega)\}, \qquad B := \{\omega \in \Omega \colon \nu(\omega) > \mu(\omega)\}. \tag{B.3}$$

Note that this is the greedy choice that achieves the sup in Eq. (B.1) (see orange region in Fig. 2),

$$||\mu - \nu||_{\mathrm{TV}} = \mu(A) - \nu(A). \tag{B.4}$$

because if any other elements are included, they will reduce the difference. To be precise, let $S \subseteq \Omega$ and note that

$$\mu(S) - \nu(S) \le \mu(S \cap A) - \nu(S \cap A). \qquad \text{(removed some elements in } B\text{; difference cannot decrease)} \tag{B.5}$$

$$\le \mu(A) - \mu(A) \qquad \text{(added some elements in } A\text{; difference cannot decrease)} \tag{B.6}$$

A parallel argument can be made to show that $\nu(B) - \mu(B)$ also achieves the sup, as B is the greedy choice to make this difference as large as possible (see black region in Fig. 2):

$$||\mu - \nu||_{\mathrm{TV}} = \nu(B) - \mu(B). \tag{B.7}$$

This is all consistent: since $A \cup B = \Omega$, we have

$$\mu(A) + \mu(B) = 1 = \nu(A) + \nu(B) \implies \mu(A) - \nu(A) = \nu(B) - \nu(A). \tag{B.8}$$

Combining Eqs. (B.4) and (B.7), we get

$$||\mu - \nu||_{\mathrm{TV}} = \frac{1}{2}\left(\mu(A) - \nu(A) + \nu(B) - \nu(B)\right) \tag{B.9}$$

$$= \frac{1}{2}\left(|\nu(A) - \mu(A)| + |\nu(B) - \nu(B)|\right) \tag{B.10}$$

$$= \frac{1}{2}\sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|. \tag{B.11}$$

□

Example B.3 (Disjoint distributions)**.** *Let* $\Omega = \{1, 2, 3, 4\}$ *and* $\mu(\{1\}) = \mu(\{2\}) = \frac{1}{2} = \nu(\{3\}) = \nu(\{4\})$*. Then, using Lemma B.2,*

$$||\mu - \nu||_{\mathrm{TV}} = \frac{1}{2}\left(\left|\frac{1}{2} - 0\right| + \left|\frac{1}{2} - 0\right| + \left|0 - \frac{1}{2}\right| + \left|0 - \frac{1}{2}\right|\right) = 1. \tag{B.12}$$

And in general, if μ, ν *have disjoint supports, we have*

$$||\mu - \nu||_{\mathrm{TV}} = |\mu(\Omega) - \nu(\Omega)| = 1. \tag{B.13}$$

Example B.4 (Two Bernoullis)**.** *Let* μ, ν *be Bernoulli distributions with parameters* $p \ge q$ *respectively. We may think of them as coins. By Lemma B.2,*

$$||\mu - \nu||_{\mathrm{TV}} = \frac{1}{2}\left(|p - q| + |(1 - p) - (1 - q)|\right) = p - q. \tag{B.14}$$

We now describe couplings.

Definition B.5 (Couplings)**.** *Let* μ, ν *be two probability measures. A* coupling *of* μ, ν *is a pair* (X, Y) *of two jointly distributed random variables whose marginals agree with* μ, ν*:*

$$\mathbb{P}[X = x] = \mu(x) \qquad \textit{and} \qquad \mathbb{P}[Y = y] = \nu(y). \tag{B.15}$$

Couplings are intimately related to the TV distance.

Proposition B.6 (TV distance and couplings)**.** *Let* μ, ν *be two probability distributions on the same space* Ω*. Then*

$$||\mu - \nu||_{\mathrm{TV}} = \inf\{\mathbb{P}[X \neq Y] \colon (X, Y) \textit{ is a coupling of } \mu, \nu\}. \tag{B.16}$$

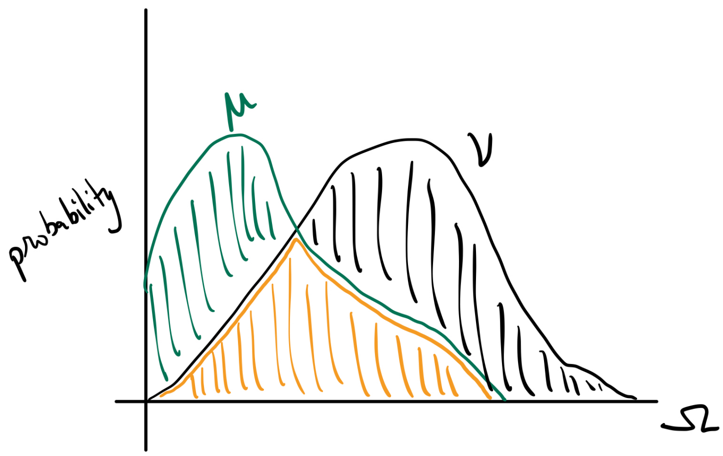

Figure 3: Illustration used in the proof of Proposition B.6

Proof. For one direction, let $A \subseteq \Omega$. Then

$$\begin{aligned}
\mu(A) - \nu(A) &= \mathbb{P}[X \in A] - \mathbb{P}[Y \in B] && \text{(B.17)}\\
&= \big(\mathbb{P}[X \in A, Y \in A] + \mathbb{P}[X \in A, Y \notin A]\big) - \big(\mathbb{P}[Y \in A, X \in A] - \mathbb{P}[Y \in A, X \notin A]\big) && \text{(B.18)}\\
&= \mathbb{P}[X \in A, Y \notin A]\big) - \mathbb{P}[Y \in A, X \notin A]\big) && \text{(B.19)}\\
&\le \mathbb{P}[X \in A, Y \notin A]\big) && \text{(B.20)}\\
&\le \mathbb{P}[X \neq Y]. && \text{(B.21)}
\end{aligned}$$

Taking the supremum over A, we conclude that

$$||\mu - \nu||_{\mathrm{TV}} \le \mathbb{P}[X \neq Y]. \tag{B.22}$$

Conversely, to show that there is a coupling that achieves the infimum, we explicitly construct it. First let

$$p = \sum_{\omega \in \Omega} \min\{\mu(x), \nu(x)\}. \tag{B.23}$$

This adds the probabilities in the orange curve in Fig. 3. Note that

$$\begin{aligned}
p &= \sum_{\omega:\, \mu(\omega) \le \nu(\omega)} \mu(\omega) + \sum_{\omega:\, \nu(\omega) < \mu(\omega)} \nu(\omega) && \text{(B.24)}\\
&= \left(\sum_{\omega:\, \mu(\omega) \le \nu(\omega)} \mu(\omega) + \sum_{\omega:\, \mu(\omega) > \nu(\omega)} \mu(\omega) \right) + \left(\sum_{\omega:\, \nu(\omega) < \mu(\omega)} \nu(\omega) - \sum_{\omega:\, \mu(\omega) > \nu(\omega)} \mu(\omega) \right) && \text{(B.25)}\\
&= 1 - ||\mu - \nu||_{\mathrm{TV}}, && \text{(B.26)}
\end{aligned}$$

by $\sum_\omega \mu(\omega) = 1$ and Eq. (B.4). We will make X, Y agree as much as possible, which is when we are below the orange curve. Then

- with probability p, set

$$X = Y = \omega \qquad \text{with probability} \qquad \frac{\min\{\mu(\omega), \nu(\omega)\}}{p}. \tag{B.27}$$

This samples from the orange region in Fig. 3.

- otherwise, set

$$X = \begin{cases} \frac{\mu(\omega)-\nu(\omega)}{||\mu-\nu||_{\mathrm{TV}}} & \text{if} \quad \mu(\omega) > \nu(\omega), \\ 0 & \text{otherwise,} \end{cases} \tag{B.28}$$

which samples from the green region in Fig. 3,and

$$Y = \begin{cases} \frac{\nu(\omega)-\mu(\omega)}{||\mu-\nu||_{\mathrm{TV}}} & \text{if} \quad \nu(\omega) > \mu(\omega), \\ 0 & \text{otherwise,} \end{cases} \tag{B.29}$$

which samples from the black region in Fig. 3.

We must verify a few things. First, Eq. (B.27) is a valid distribution because of Eq. (B.23). Similarly, Eqs. (B.28) and (B.29) are also valid distributions because of Eqs. (B.4) and (B.7). Additionally, X has the correct marginal since $X = \omega$ with probability

$$p \cdot \frac{\min\{\mu(\omega), \nu(\omega)\}}{p} + (1-p)\frac{\mu(\omega)-\nu(\omega)}{||\mu-\nu||_{\mathrm{TV}}}\mathbf{1}\{\mu(\omega) > \nu(\omega)\} = \mu(\omega). \tag{B.30}$$

This is effectively adding the green and orange regions in Fig. 3, recovering μ. By the same computation, Y has the correct marginal, which amounts to adding the orange and black regions, recovering ν. And finally, $X \neq Y$ exactly in the second case, which happens with probability $1 - p = ||\nu - \mu||_{\mathrm{TV}}$ by Eq. (B.26). ☐

The intuition is that TVD is measuring how much the distributions are distinguishable. The optimal coupling can make them agree as much as they are undistinguishable; however, they have to disagree however much they are distinguishable.

Example B.7 (Independent coupling)**.** *Consider two Bernoulli coins X, Y with probabilities $p \geq q$. If we use an independent coupling, where the joint distribution is just the product distribution. In this case,*

$$\mathbb{P}[X \neq Y] = p(1-q) + q(1-p) > p - q, \tag{B.31}$$

the latter being the TV distance of their marginals Example B.4.

Example B.8 (Optimal coupling)**.** *We want to make X, Y agree as much as possible. This is usually accomplished by sharing the same randomness. Let's pick a uniform random number r in $[0, 1]$ X, Y and make X heads if $r < p$ and Y heads if $r < q$. Note that they have the correct marginals. Now*

$$\mathbb{P}[X \neq Y] = \mathbb{P}[q < r < p] = p - q. \tag{B.32}$$

Hence this is an optimal coupling by Example B.4.

C Eigenvalues of the 2D grid

Let G_n be the $n \times n$ grid graph. Here we compute its eigenvalues. We will combine a few lemmas.

Definition C.1 (Graph cartesian product)**.** *Let $G = (V_1, E_1), H = (V_2, E_2)$ be two graphs. We define the* graph cartesian product *$G \square H := (V, E)$ to be the graph where $V := V_1 \times V_2$ and, for $v_1, u_1 \in V_1, v_2, u_2 \in V_2$,*

$$((v_1, v_2), (u_1, u_2)) \in E \iff (v_1 = u_1 \text{ and } (v_2, u_2) \in E_2) \text{ or } ((v_1, u_1) \in E_1 \text{ and } v_2 = u_2). \tag{C.1}$$

This is useful since G_n is a cartesian product.

Lemma C.2 (G is a cartesian product)**.** *Let G_n be the graph of a $n \times n$ grid, and let P_n be the path graph of length n. Then*

$$G_n = P_n \square P_n. \tag{C.2}$$

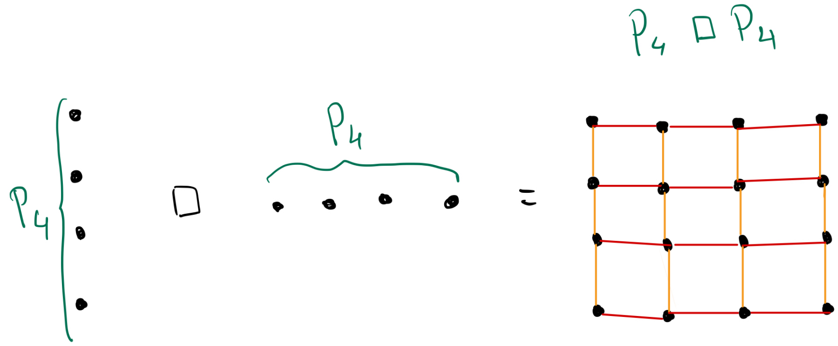

Figure 4: The 4×4 grid is given by $P_4 \square P_4$. Orange edges are those where $v_1 = u_1$ and $(v_2, u_2) \in E_2$, and red those where $(v_1, u_1) \in E_1$ and $v_2 = u_2$.

This can be seen from Fig. 4. That is, line up one P_n vertically and the other horizontally. The vertex set will be the cartesian pairs, and the edges connect nearest neighbors.

Furthermore, we can characterize the adjacency matrix of a cartesian product from the adjacency matrices of its components.

Lemma C.3 ([KR05])**.** *Let G, H be graphs with adjacency matrices A, B respectively, where $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{m\times m}$. Then the adjacency matrix of $G \square H$ is*

$$A \otimes I_m + I_n \otimes B, \tag{C.3}$$

where I_k is the identity matrix of dimension k.

Corollary C.4 (Eigenvalues of G_n from those of P_n)**.** *If P_n has eigenvalues $\{\lambda_i\}_{i=1}^n$, then G_n has eigenvalues*

$$\{\lambda_i + \lambda_j\}_{i,j=1}^n. \tag{C.4}$$

Proof. This can be seen by multiplying any eigenvector of A, B by Eq. (C.3). □

All that remains then is to compute the spectrum of the path graph.

Lemma C.5 (Eigenvalues of P_n)**.** *Let P_n be the path graph on n vertices. Then its adjancency matrix has eigenvalues*

$$\left\{-2\cos\left(\frac{k\pi}{n+1}\right) : 1 \le k \le n\right\}. \tag{C.5}$$

Proof. We can see that the adjacency matrix A_n is a Toeplitz matrix:

$$A_n = \begin{bmatrix} 0 & 1 & 0 & 0 & \\ 1 & 0 & 1 & 0 & \ddots \\ 0 & 1 & 0 & 1 & \ddots \\ 0 & 0 & 1 & 0 & \ddots \\ & & & \ddots & \end{bmatrix}. \tag{C.6}$$

Let the characteristic polynomial be $p_n(\lambda) = \det(A_n - \lambda I)$. Doing the cofactor expansion, we get

$$p_n(\lambda) = \det \begin{bmatrix} -\lambda & 1 & 0 & 0 & \\ 1 & -\lambda & 1 & 0 & \ddots \\ 0 & 1 & -\lambda & 1 & \ddots \\ 0 & 0 & 1 & -\lambda & \ddots \\ & & & \ddots & \end{bmatrix} = -\lambda p_{n-1}(\lambda) - \det \begin{bmatrix} 1 & 1 & 0 & \ddots \\ 0 & -\lambda & 1 & \ddots \\ 0 & 1 & -\lambda & \ddots \\ & & \ddots & \end{bmatrix}, \tag{C.7}$$

which yields the recursion

$$p_n(\lambda) = -\lambda p_{n-1}(\lambda) - p_{n-2}(\lambda) \qquad \text{with} \qquad p_1(\lambda) = -\lambda, p_0(\lambda) = 1. \tag{C.8}$$

With the change of variables $\lambda = -2\cos\theta$, we recognize that this is the recurrence relation obeyed by the Chebyshev polynomials of the second kind. Hence

$$p_n(\lambda) = U_n(-\lambda/2). \tag{C.9}$$

Since

$$U_n(\cos\theta) = \frac{\sin((n+1)\theta)}{\sin\theta}, \tag{C.10}$$

the roots of U_n satisfy

$$\sin((n+1)\theta) = 0 \text{ but } \sin\theta \neq 0, \text{ so } \theta = \frac{k\pi}{n+1} \text{ with } 1 \le k \le n. \tag{C.11}$$

Hence the roots of p_n are

$$\lambda_k = -2\cos\left(\frac{k\pi}{n+1}\right) \text{ for } 1 \le k \le n. \tag{C.12}$$

□

Applying Corollary C.4, we conclude:

Corollary C.6 (Spectrum of 2D grid)**.** *Let G_n be the graph of an $n \times n$ grid. Then its adjancency matrix has eigenvalues*

$$\left\{-2\cos\left(\frac{j\pi}{n+1}\right) - 2\cos\left(\frac{k\pi}{n+1}\right) : 1 \le j, k \le n\right\}. \tag{C.13}$$

D Miscellaneous lemmas

Lemma D.1 (Spectral norm preserves inequalities of nonnegative matrices)**.** *Suppose that $c \ge 0$, A, b have nonnegative entries and*

$$A_{ij} \le cB_{ij}. \tag{D.1}$$

Then

$$\|A\|_2 \le c\|B\|_2. \tag{D.2}$$

Proof. First note that $\|A\|_2$ is achieved by a vector with nonnegative entries. Indeed, note that

$$\|Ax\|_2^2 = \sum_i \left(\sum_j A_{ij}x_j\right)^2 \le \sum_i \left(\sum_j |A_{ij}x_j|\right)^2 = \sum_i \left(\sum_j A_{ij}|x_j|\right)^2 = \|A|x|\|_2^2. \tag{D.3}$$

So let x be the unit-norm nonnegative vector which achieves the $\|A\|_2$. Then

$$\|A\|_2^2 = \|Ax\|_2^2 = \sum_i \left(\sum_j A_{ij} x_j \right)^2 \leq \sum_i \left(\sum_j c B_{ij} x_j \right)^2 = c^2 \|Bx\|_2 \leq c^2 \left(\max_{\|y\|_2 = 1} \|By\|_2 \right)^2 = c^2 \|B\|_2^2, \tag{D.4}$$

since every term appearing is positive. Taking square roots finishes the proof. □

Lemma D.2 (General bound on spectrum of a graph)**.** *Let A be the adjacency matrix of a graph with maximum degree Δ. Then*

$$\|A\|_2 \leq \Delta. \tag{D.5}$$

Proof. Since A is Hermitian and $A_{ii} = 0$ for all i, by Gershgorin's theorem we have

$$\|A\|_2 \leq \max_j \sum_{i \neq j} |A_{i,j}| = \Delta. \tag{D.6}$$

□

References

[AH87] Michael Aizenman and Richard Holley. Rapid convergence to equilibrium of stochastic ising models in the dobrushin shlosman regime. In *Percolation theory and ergodic theory of infinite particle systems*, pages 1–11. Springer, 1987. 4

[Bon14] Claudio Bonati. The peierls argument for higher dimensional ising models. *European Journal of Physics*, 35(3):035002, 2014. 2, 3

[Col24] Spectral Collective. How many times do you ACTUALLY need to shuffle? YouTube, 2024. URL https://www.youtube.com/watch?v=5txcQy3rGwo&t=1203s. 8

[DGJ09] Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. Matrix norms and rapid mixing for spin systems. 2009. 4

[Dob68a] PL Dobruschin. The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory of Probability & Its Applications*, 13(2):197–224, 1968. 4

[Dob68b] Roland L'vovich Dobrushin. Gibbsian random fields for lattice systems with pairwise interactions. *Functional Analysis and its applications*, 2(4):292–301, 1968. 4

[Dob68c] Roland L'vovich Dobrushin. The problem of uniqueness of a gibbsian random field and the problem of phase transitions. *Functional analysis and its applications*, 2(4):302–312, 1968. 2

[Dob70] Roland L Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970. 4

[DS85] Roland Lvovich Dobrushin and Senya B Shlosman. Constructive criterion for the uniqueness of gibbs field. In *Statistical Physics and Dynamical Systems: Rigorous Results*, pages 347–370. Springer, 1985. 4

[DS87] Roland L Dobrushin and Senya B Shlosman. Completely analytical interactions: constructive description. *Journal of Statistical Physics*, 46(5):983–1014, 1987. 4

[FV17] Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017. 2

[Gri64] Robert B Griffiths. Peierls proof of spontaneous magnetization in a two-dimensional ising ferromagnet. *Physical Review*, 136(2A):A437, 1964. 2

[Hay06] Thomas P Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 39–46. IEEE, 2006. 4

[HS05] Thomas P Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 511–520. IEEE, 2005. 4

[KR05] A Kaveh and H Rahami. A unified method for eigendecomposition of graph products. *Communications in numerical methods in engineering*, 21(7):377–388, 2005. 12

[LP17] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017. 8

[LS12] Eyal Lubetzky and Allan Sly. Critical ising on the square lattice mixes in polynomial time. *Communications in Mathematical Physics*, 313(3):815–836, 2012. 7

[MO94a] Fabio Martinelli and Enzo Olivieri. Approach to equilibrium of glauber dynamics in the one phase region: I. the attractive case. *Communications in Mathematical Physics*, 161(3):447–486, 1994. 7

[MO94b] Fabio Martinelli and Enzo Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region: II. The general case. *Communications in Mathematical Physics*, 161(3):487–514, 1994. 7

[Ons44] Lars Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Physical review*, 65(3-4):117, 1944. 7

[Pei36] Rudolf Peierls. On ising's model of ferromagnetism. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 32, pages 477–481. Cambridge University Press, 1936. 2

[Wei05] Dror Weitz. Combinatorial criteria for uniqueness of gibbs measures. *Random Structures & Algorithms*, 27(4):445–475, 2005. 4

[Woo21] Mary Wootters. Class 15, Video 1: Total Variation Distance. YouTube, 2021. URL `https://www.youtube.com/watch?v=aJP0Bj6jVkI`. 8