CS294 Partition Functions **Fall 2020**

Lecture 9: September 29

Instructor: Alistair Sinclair

Introduction

In this lecture we look at two further applications of coupling of Markov chains: versions of the so-called "Dobrushin conditions", which give non-trivial (but usually not tight) conditions for rapid mixing of Glauber dynamics on bounded-degree graphs; and linear extensions of a partial order, which involves a more delicate use of path coupling.

9.1 Dobrushin conditions

The "Dobrushin condition", and the related "Dobrushin-Shlosman condition", were developed many years ago as sufficient conditions under which the Gibbs measure of a spin system is unique. We shall have more to say about uniqueness of Gibbs measures later in the class, but for now we can think of this as meaning that correlations between spins at vertices (in a countably infinite graph) decay to zero with distance. What we shall show here, using a simple coupling argument, is that these same conditions actually also imply that the associated Glauber dynamics has mixing time $O(n \log n)$ on any connected n-vertex graph. In fact we'll show that a weaker condition, which is implied by both the Dobrushin and the Dobrushin-Shlosman conditions, suffices for this. We note also that $O(n \log n)$ mixing time is the best one can hope for for *any* spin system on a bounded degree graph [HS05].

All of these conditions are properties of the *influence matrix* of the spin system. Consider a spin system on the finite graph $G = (V, E)$ with n vertices. The *influence matrix* $R = (\rho_{ij})$ is an $n \times n$ matrix indexed by the vertices (sites), with entries defined by

$$\rho_{ij} := \max_{(\sigma,\tau)\in S_j} \|\pi_i(\sigma,\cdot) - \pi_i(\tau,\cdot)\|_{\mathrm{TV}}, \tag{9.1}$$

where S_j is the set of all pairs of configurations $(\sigma, \tau) \in \Omega^2$ such that σ, τ agree on all sites except j, and $\pi_i(\sigma, \cdot)$ is the distribution on the spin at site i in configuration σ (i.e., the distribution used by the Glauber dynamics when updating the spin at site i in configuration σ). Thus ρ_{ij} is a natural measure of the influence of spin j on its neighbor i. (Note that, by the spatial Markov property, the influence defined in this way is zero when i, j are not neighbors.)

Recall that the *spectral norm* of an $n \times n$ square matrix A is the ℓ_2 operator norm defined as $\|A\| := \sup_{v\neq 0} \frac{\|Av\|_2}{\|v\|_2}$. Equivalently, this is the largest singular value of A (the singular values are the square roots of the eigenvalues of the symmetric matrix AA^*, where A^* is the transpose of A). When A is symmetric it is just the largest eigenvalue of A.

We will prove the following result, due to Hayes [Hay06]. For a more systematic treatment, and sharper results (which replace the spectral norm by *any* matrix norm), see Dyer, Goldberg and Jerrum [DGJ09].

Theorem 9.1. *If the spectral norm of the influence matrix R is at most $1-\delta$ for some $\delta > 0$, then the mixing time of the (heat-bath) Glauber dynamics is $\tau_{\text{mix}} \leq \frac{c}{\delta} n \ln n$ for a universal constant c.*

Proof. We will use coupling. Let $(X_t), (Y_t)$ be two copies of the Glauber dynamics, started in arbitrary configurations X_0, Y_0. We couple their evolutions as follows: at each time step, X_t, Y_t both choose the same site i to update, and then we *maximally couple* their choices of a new spin at i, i.e., we ensure that

$$\Pr[X_{t+1}(i) \neq Y_{t+1}(i) | X_t, Y_t] = \|\pi_i(X_t, \cdot) - \pi_i(Y_t, \cdot)\|_{\text{TV}}.$$

Writing $p_t(i)$ for $\Pr[X_{t+1}(i) \neq Y_{t+1}(i) | X_0, Y_0]$, we can then see that

$$p_{t+1}(i) \leq \left(1 - \frac{1}{n}\right) p_t(i) + \frac{1}{n} \sum_{j \in N(i)} \rho_{ij} p_t(j), \tag{9.2}$$

where $N(i)$ is the neighborhood of i in G. The first term comes from choosing any vertex other than i (in which case $p_t(i)$ is unchanged), and the second from choosing i itself. The fact that the total contribution in this case is bounded by the sum over the influences ρ_{ij} follows from the definition of the maximal coupling and the triangle inequality; more formally, it can be justified by path coupling, which expresses $p_{t+1}(i)$ in terms of the same probabilities just for adjacent pairs of configurations (i.e., those in S_j for each j). [**Exercise:** Convince yourself of the validity of (9.2).]

Defining the matrix $A := \frac{n-1}{n} I + \frac{1}{n} R$, where I is the $n \times n$ identity matrix, we can rewrite (9.2) as $p_{t+1} \leq A p_t$ (componentwise), and hence by induction $p_t \leq A^t p_0$ for any t. Note that the assumption $\|R\| \leq 1 - \delta$ and subadditivity of matrix norms implies that the operator norm of A satisfies

$$\|A\| \leq \frac{n-1}{n} \|I\| + \frac{1}{n} \|R\| \leq 1 - \frac{\delta}{n}.$$

Now, we get

$$\begin{aligned}
\Pr[X_t \neq Y_t] &\leq \|p_t\|_1 && \text{by a union bound} \\
&\leq \sqrt{n} \|p_t\|_2 && \text{by Cauchy-Schwartz} \\
&\leq \sqrt{n} \|A^t p_0\|_2 && \text{since } p_t \leq A^t p_0 \text{ componentwise} \\
&\leq \sqrt{n} \|A\|^t \|p_0\|_2 && \text{by defn of spectral norm} \\
&\leq n \left(1 - \frac{\delta}{n}\right)^t && \text{by spectral norm of } A.
\end{aligned}$$

This immmediately implies that $\tau(\varepsilon) \leq \frac{n}{\delta} \ln(\frac{n}{\varepsilon})$, which yields the claimed bound on mixing time. □

Here are a few simple applications of Theorem 9.1.

1. **Hard-core model.** Recall that the configurations are independent sets of G, and the weight of independent set σ is $\lambda^{|\sigma|}$, where σ is the number of (occupied) vertices in σ. The Glauber dynamics picks a vertex v u.a.r. and, if v has no occupied neighbors, makes v occupied with probability $\frac{\lambda}{1+\lambda}$. [**Exercise:** Check that this is the correct definition of the heat-bath Glauber dynamics in this example.] What are the influences ρ_{ij} for this model? Fix any vertex i. If any neighbor of i is occupied then i is unoccupied with probability 1; otherwise, i is unoccupied with probability $\frac{1}{1+\lambda}$. Therefore, the maximum in (9.1) is attained when the configurations σ, τ make all neighbors of i except for j unoccupied; the resulting variation distance is then $\frac{\lambda}{1+\lambda}$. So in this case we have

$$R = \frac{\lambda}{1+\lambda} A_G,$$

where A_G is the adjacency matrix of G. (Note that R is symmetric in this case, though this is not generally true.) As a consequence,

$$\|R\| = \frac{\lambda}{1+\lambda}\|A_G\| = \frac{\lambda}{1+\lambda}\lambda_0(G),$$

where $\lambda_0(G)$ is the principal eigenvalue of A_G.[1] Now it is well known that $\lambda_0(G) \le \Delta$ (the maximum degree of G), so if we take $\lambda \le \frac{1-\varepsilon}{\Delta}$ then $\|R\| \le \frac{\Delta(1-\varepsilon)}{\Delta+1-\varepsilon} \le 1-\varepsilon$. Hence Theorem 9.1 implies that the mixing time of the Glauber dynamics on any graph of maximum degree Δ is $O(n \log n)$ as long as the fugacity parameter λ satisfies $\lambda \le \frac{1-\varepsilon}{\Delta}$ for some fixed constant $\varepsilon > 0$. As we shall see later in the course, the dependency $\frac{1}{\Delta}$ on the maximum degree is optimal, though the constant is not.

2. **Ferromagnetic Ising model at zero field.** Recall that the weight of a configuration is $w(\sigma) = \lambda^k$, where k is the number of disagreeing edges and $\lambda = \exp(-2\beta)$ with $\beta \ge 0$. Recall also the Glauber dynamics for this model from an earlier lecture. To compute the influences ρ_{ij} here, fix a vertex i. If the numbers of neighbors of i with spins $+1$ and -1 are d^+ and d^- respectively, then the spin of i is $+1$ with probability

$$\frac{\lambda^{d^-}}{\lambda^{d^+} + \lambda^{d^-}}.$$

Suppose w.l.o.g. we change the spin at one of the neighbors j from $+1$ to -1. Then after a bit of algebra [**exercise!**] we see that the variation distance between the above distribution and the new distribution of the spin at i is

$$\frac{\lambda^{-1} - \lambda}{(z + z^{-1})(\lambda^{-1}z + \lambda z^{-1})},$$

where $z := \lambda^{(d^+ - d^-)/2}$. Elementary calculus [**exercise!**] shows that this function attains its maximum value when $d^+ + d^-$ (the degree of i) is odd and $d^+ = d^- + 1$, and that value is

$$\frac{\lambda^{-1/2} - \lambda^{1/2}}{\lambda^{-1/2} + \lambda^{1/2}} = \tanh(\beta).$$

Thus the influence matrix in this case satisfies

$$R \le \tanh(\beta) A_G,$$

and hence by monotonicity of the spectral norm in the matrix entries of a non-negative matrix,

$$\|R\| \le \tanh(\beta)\|A_G\| = \tanh(\beta)\lambda_0(G).$$

So as in the previous example, we get $O(n \log n)$ mixing time on all graphs of maximum degree Δ provided $\tanh(\beta) \le \frac{1-\varepsilon}{\Delta}$ for some fixed $\varepsilon > 0$. This almost matches the so-called "tree uniqueness threshold" for the Ising model, which is $\tanh(\beta) = \frac{1}{\Delta-1}$.

We note that the same calculations hold for the antiferromagnetic Ising model with the roles of λ and λ^{-1} reversed.

3. **q-colorings.** Finally, let's see what the Dobrushin condition gives us for coloring. Fix a vertex i. After some thought [**exercise!**] one can verify that the maximum in (9.1) is achieved when the colorings σ, τ assign *distinct* colors to all neighbors of i, and differ only at j. In this case, the variation distance on the color at i will be $\frac{1}{q-\deg(i)} \le \frac{1}{q-\Delta}$. So once again the influence matrix is bounded by a constant multiple of the adjacency matrix, i.e., $R \le \frac{1}{q-\Delta} A_G$, and we get

$$\|R\| \le \frac{1}{q-\Delta}\lambda_0(G). \tag{9.3}$$

[1] Apologies for the overloading of the letter λ here!

Plugging in $\lambda_0(G) \le \Delta$, we see that the mixing time is $O(n \log n)$ as long as $q \ge (2+\varepsilon)\Delta$ for any fixed $\varepsilon > 0$. Actually, for any fixed Δ, the mixing time is $O(n \log n)$ for $q \ge 2\Delta + 1$, which matches the argument we saw earlier via path coupling (and indeed in some sense the two analyses are at root equivalent).

Remarks

1. Note that while we have focused in this note on random sampling via MCMC, all our positive results lead directly to efficient approximate counting algorithms (fpras) via the reductions we saw earlier in the class.
2. The above examples capture in an elegant way "vanilla" applications of (path) coupling to Glauber dynamics on spin systems on bounded degree graphs, which typically work over a non-trivial range of parameter values that however stops short of the "optimal" range determined by some phase transition. To get optimal algorithms, we will need either more powerful methods or restrictions on the class of graphs. We'll have more to say about this later in the course.
3. The original Dobrushin and Dobrushin-Shlosman conditions [Dob70, DS85a, DS85b] supplied the weaker criterion that the row sums (respectively, the column sums) of the influence matrix are bounded by $1-\varepsilon$; this means that the total influence *on* the spin at site i (respectively, *of* the spin at site i) is bounded away from 1. These criteria correspond to bounds on the ℓ_1 and ℓ_∞ operator norms of R, respectively. Since the spectral norm is bounded above by both of these norms, the criterion in Theorem 9.1 is more powerful. (In fact the bounds we gave in the examples above also follow from the original D-S criteria, though the derivation of mixing time bounds via norms is cleaner.) Indeed, Dobrushin and Shlosman were motivated by conditions for uniqueness of the Gibbs measure on infinite graphs, rather than by rapid mixing of the dynamics, and it turns out that the weaker spectral norm condition in Theorem 9.1, while sufficient to ensure rapid mixing, does *not* in general imply uniqueness of the Gibbs measure. An example is provided by the Ising model on a tree, where Theorem 9.1 can be used to show rapid mixing for $\beta = O(1/\sqrt{\Delta})$, while uniqueness is known to hold only up to $\beta = \Theta(1/\Delta)$.
4. To get a sense of the added power of Theorem 9.1 over the basic Dobrushin-Shlosman framework, consider the class of planar graphs of maximum degree Δ. In contrast to general graphs, the maximum eigenvalue of planar graphs of maximum degree $\Delta \ge 6$ is known to be at most $c\sqrt{\Delta}$ rather than Δ for a constant c. ($c = 2\sqrt{3}$ certainly suffices [Hay06]. For $\delta \le 5$, Δ-regular planar graphs exist and the bound $\lambda_0(G) \le \Delta$ is best possible.) Plugging in this bound for $\lambda_0(G)$ in place of Δ in (9.3) gives us $O(n \log n)$ mixing time for $q \ge \Delta + \frac{c}{1-\varepsilon}\sqrt{\Delta} = (1 + o_\Delta(1))\Delta$, a significant improvement over the bounds for general graphs. Analogous improvements for planar graphs hold for the hard-core and Ising model examples discussed above. Other classes of sparse graphs, for which the largest eigenvalue can be effectively bounded, yield similar improvements. See [DGJ09] for details.

9.2 Linear Extensions of a Partial Order

In this section we give one more example of path coupling, to the combinatorial problem of counting (or randomly sampling) linear extensions of a partial order.

Given a partial order[2] $\preceq$ on $V = \{1, 2, \ldots, n\}$, a *linear extension* of $\preceq$ is a total order $\sqsubseteq$ on V which respects $\preceq$, i.e., for all $x, y \in V$ $x \preceq y$ implies $x \sqsubseteq y$. Hence a linear extension of $\preceq$ can be written as a permutation $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ of $\{1, 2, \ldots, n\}$ such that $\sigma(i) \le \sigma(j)$ if $v_i \preceq v_j$.

[2] A *partial order* $(V, \preceq)$ is a binary relation $\preceq$ over a set V which is reflexive, antisymmetric, and transitive, *i.e.*, for all a, b, and $c \in V$, we have, (i) $a \preceq a$, (ii) $a \preceq b$ and $b \preceq a \implies a = b$, (iii) $a \preceq b$ and $b \preceq c \implies a \preceq c$.

Note that, given a partial order $\preceq$ on $V = \{1, 2, \ldots, n\}$, one can easily construct a linear extension of $\preceq$ (**exercise**). For example, in Figure 9.1, the right hand picture represents a linear extension of the partial order shown on the left.

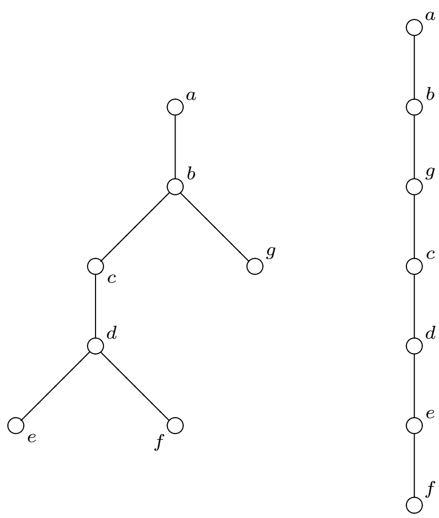

Figure 9.1: A partial order and one of its linear extensions

Let $\Omega = \Omega(\preceq)$ denote the set of all linear extensions of $\preceq$. Being able to count the elements of Ω, or equivalently sample u.a.r. from Ω, has a variety of applications in combinatorics, near-optimal sorting and decision theory. Brightwell and Winkler [BW91] showed that counting linear extensions is $\#P$-complete via a clever reduction. Here we present a MCMC algorithm that samples linear extensions, and thus in turn also yields an fpras for $|\Omega|$.

The Markov Chain on Ω is defined as follows:

1. With probability 1/2 do nothing.
2. Else pick a position $p \in \{1, 2, \ldots, n-1\}$ uniformly at random, and exchange the elements at positions p and $p+1$ if "legal", i.e., if the resulting total order is a linear extension of $\preceq$.

It is easy to verify that this Markov Chain is symmetric and aperiodic. Irreducibility of the chain follows from the following exercise.

Exercise: Prove that it is always possible to reach a given linear extension from any another using at most $\binom{n}{2}$ legal exchanges of consecutive elements.

Hence the above chain is ergodic and converges to the uniform distribution π on Ω. We will use path coupling, for which we need to define a pre-metric on Ω.

Pre-metric:
Two states X and Y in Ω are *adjacent* iff they differ at exactly two positions, say, i and j, $1 \leq i < j \leq n$. We denote this by $X = Y \circ (i, j)$. In this case, the distance $d(X, Y)$ between X and Y is defined to be $j - i$ (see Figure 9.2).

Exercise: Check that the above is indeed a pre-metric.

This pre-metric extends to a metric on Ω, denoted again by d. Note that this metric may be rather complicated to describe, but the power of path coupling lies in the fact that we never need to do so. Next we

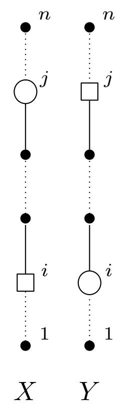

Figure 9.2: Two adjacent states

define the coupling for adjacent states.

9.2.1 Coupling for adjacent pairs

Let (X, Y) be a pair of adjacent states. Let i and j be the positions where they differ, $1 \le i < j \le n$. Then the transition $(X, Y) \mapsto (X', Y')$ is defined as follows.

Case 1: If $j \neq i + 1$:

- X, Y both pick the same $p \in \{1, \dots, n-1\}$
- X, Y then both make the same move (i.e., both do nothing with probability $\frac{1}{2}$, else both attempt to exchange the elements at positions $p, p+1$).

Case 2: If $j = i + 1$:

- X, Y both pick the same $p \in \{1, \dots, n-1\}$
- If $p \neq i$, then both make the same move as above. If $p = i$ then either X does nothing while Y exchanges the elements at positions i and $i + 1 = j$, or vice versa with the roles of X, Y reversed, each with probability $\frac{1}{2}$. (This ensures that X, Y become equal after the move.)

(Note that X' and Y' may no longer be adjacent after a move under this coupling.)

Exercise: Check that this is indeed a valid coupling.

9.2.2 Analysis

To analyze the coupling we need to consider three cases, based on the position p chosen by X and Y.

Case I: $p \notin \{i-1, i, j-1, j\}$.
In this case the exchanges in X and Y are either both legal or both not legal (since X, Y agree everywhere except at i, j). Thus X' and Y' again differ exactly at i and j and $d(X', Y') = j - i = d(X, Y)$.

Case II: $p = i - 1$ or $p = j$.

These two cases are symmetric, so we consider only the case $p = i - 1$. If the exchanges are both legal in X and Y, then we have $d(X', Y') = j - (i - 1) = d(X, Y) + 1$. If exactly one of them is legal (say in X) then $Y' = Y = X \circ (i, j) = X' \circ (i - 1, i) \circ (i, j)$ and $d(X', Y') = (j - i) + 1 = d(X, Y) + 1$. If the exchanges are not legal in both X and Y then of course $d(X', Y') = d(X, Y)$. Hence d increases by at most 1 in this case. The probability of an increase is at most $2 \times \frac{1}{2(n-1)} = \frac{1}{n-1}$.

Case III: $p = i$ or $p = j - 1$.

There are two subcases, depending on the value of $j - i$.

- $j - i > 1$.
  Again, by symmetry we need only consider the case $p = i$. But the crucial observation is that the elements at positions i, $i + 1$ and j are incomparable in $\preceq$. Hence the exchanges in X and Y are both legal and $d(X', Y') = j - (i + 1) = d(X, Y) - 1$. The probability of this decrease is $2 \times \frac{1}{2(n-1)} = \frac{1}{n-1}$.
- $j - i = 1$.
  Notice that in this case we make an exchange between the elements at positions i and $j = i + 1$ in exactly one of X and Y keeping the other undisturbed and the exchange is always legal. So, $d(X', Y') = 0 = d(X, Y) - 1$. This case happens only for $p = i$, so the probability is $\frac{1}{n-1}$.

Hence $d(X', Y') = d(X, Y) + 1$ with probability *at most* $1/(n - 1)$, and $d(X', Y') = d(X, Y) - 1$ with probability exactly $1/(n - 1)$. Otherwise $d(X', Y') = d(X, Y)$. Therefore we have

$$\mathrm{E}\left[d(X', Y') | X, Y\right] \leq d(X, Y). \tag{9.4}$$

Note that the Path Coupling Lemma doesn't apply directly here as we have only proved that $d(X_t, Y_t)$ is non-increasing in expectation, rather than that it contracts. However, (9.4) does imply that $d(X_t, Y_t)$ is dominated by a symmetric random walk on the integer interval $[0, D]$, where $D = \max_{X,Y \in \Omega} d(X, Y) \leq \binom{n}{2}$. By a standard martingale argument, the expected time until this walk hits zero from any starting point is $O(\beta^{-1} D^2)$, where $\beta = \min_{X,Y \in \Omega} \Pr[|d(X', Y') - d(X, Y)| = 1] \geq \frac{1}{n-1}$. Hence the mixing time is bounded as $\tau_{\text{mix}} = O(n^5)$.

9.2.3 A modified Markov chain

Bubley and Dyer [BD99] showed that the convergence can be accelerated by picking position p according to some other distribution instead of choosing it uniformly at random. Here is their modified chain:

1. With probability 1/2 do nothing.
2. Else pick a position $p \in \{1, 2, \ldots, n - 1\}$ with probability $\frac{Q(p)}{Z}$, where $Z = \sum_{i=1}^{n-1} Q(i)$, and exchange the elements at positions p and $p + 1$ if legal.

Note that this Markov chain is still symmetric and aperiodic. If $Q(p) > 0$ for all $1 \leq p \leq n - 1$ then it is also irreducible and hence ergodic with uniform stationary distribution. Note that the same coupling stated before can be defined for this Markov chain except that the position p is now chosen with probability $Q(p)/Z$. Exactly the same analysis holds here too, and we have that $d(X', Y') = d(X, Y) + 1$ with probability *at most* $[Q(i - 1) + Q(j)]/2Z$ and $d(X', Y') = d(X, Y) - 1$ with probability $[Q(i) + Q(j - 1)]/2Z$. Otherwise $d(X', Y') = d(X, Y)$. Therefore,

$$\mathrm{E}\left[d(X', Y') - d(X, Y) | X, Y\right] \leq \frac{1}{2Z}\left[Q(i - 1) + Q(j) - Q(i) - Q(j - 1)\right].$$

We want the R.H.S. to be $\leq -\alpha \cdot d(X,Y) = -\alpha(j-i)$. Observe that the R.H.S. is the difference between the discrete derivatives of the function Q at j and at i, which will be linear if we take $Q(p)$ to be a quadratic polynomial, say $Q(p) = ap^2 + bp + c$. We need to choose the parameters a, b and c in such a way that the value of α is maximized. The optimal choice of Q is given by $Q(p) = p(n-p)$ (**Exercise:** check this!). This gives $Z = \sum_p Q(p) = (n^3 - n)/6$ and $\alpha = 1/Z \geq 6/n^3$.

Plugging in, we get the expected change in distance

$$\mathrm{E}\left[d(X', Y') - d(X,Y) | X, Y\right] \leq -\frac{6}{n^3} d(X,Y).$$

Hence, $\tau_{\text{mix}} = O(\alpha^{-1} \log D) = O(n^3 \log n)$.

9.2.4 Remarks

1. Wilson [Wil04] improved the upper bound on the mixing time for the original Markov chain to $O(n^3 \log n)$ and showed that this is tight. Interestingly, this is the correct order for the mixing time even in the case when $\preceq$ is empty (so that Ω includes all permutations of the elements); in this case the Markov chain is equivalent to shuffling cards by random adjacent transpositions.
2. As far as I know, it remains open whether the mixing time of the modified Markov chain can be strictly better than that of the original.

References

[BD99] R. Bubley and M.E. Dyer. Faster random generation of linear extensions. *Discrete Mathematics*, 201:81–88, 1999.

[BW91] G. Brightwell and P. Winkler. Counting linear extensions. *Order*, 8:225–242, 1991.

[DGJ09] M. Dyer, L. Goldberg, and M. Jerrum. Matrix norms and rapid mixing for spin systems. *Annals of Applied Probability*, 19:71–107, 2009.

[Dob70] R.L. Dobrushin. Prescribing a system of random variables by the help of conditional distributions. *Theory of Probability and its Applications*, 15:469–497, 1970.

[DS85a] R.L. Dobrushin and S.B. Shlosman. Completely analytical Gibbs fields. In J. Fritz, A. Jaffe, and D. Szasz, editors, *Statistical Mechanics and Dynamical Systems*, pages 371–403. Birkhäuser, Boston, 1985.

[DS85b] R.L. Dobrushin and S.B. Shlosman. Constructive criterion for the uniqueness of a Gibbs field. In J. Fritz, A. Jaffe, and D. Szasz, editors, *Statistical Mechanics and Dynamical Systems*, pages 347–370. Birkhäuser, Boston, 1985.

[Hay06] T. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 39–46, 2006.

[HS05] T. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 511–520, 2005.

[Wil04] D.B. Wilson. Mixing times of lozenge tiling and card shuffling Markov chains. *Annals of Applied Probability*, 14:274–325, 2004.