Combinatorics, Probability and Computing (2008) 17, 761–779. © 2008 Cambridge University Press
doi:10.1017/S0963548308009437 Printed in the United Kingdom

Dobrushin Conditions and Systematic Scan

MARTIN DYER[1], LESLIE ANN GOLDBERG[2]
and MARK JERRUM[3†]

[1]School of Computing, University of Leeds, Leeds LS2 9JT, UK

[2]Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK

[3]School of Mathematical Sciences, Queen Mary, University of London
Mile End Road, London E1 4NS, UK

Received 12 October 2006; revised 23 July 2008; first published online 3 October 2008

We consider Glauber dynamics on finite spin systems. The mixing time of Glauber dynamics can be bounded in terms of the influences of sites on each other. We consider three parameters bounding these influences: α, the total influence on a site, as studied by Dobrushin; α', the total influence of a site, as studied by Dobrushin and Shlosman; and α'', the total influence of a site in any given context, which is related to the path-coupling method of Bubley and Dyer. It is known that if any of these parameters is less than 1 then random-update Glauber dynamics (in which a randomly chosen site is updated at each step) is rapidly mixing. It is also known that the Dobrushin condition $\alpha < 1$ implies that systematic-scan Glauber dynamics (in which sites are updated in a deterministic order) is rapidly mixing. This paper studies two related issues, primarily in the context of systematic scan: (1) the relationship between the parameters α, α' and α'', and (2) the relationship between proofs of rapid mixing using Dobrushin uniqueness (which typically use analysis techniques) and proofs of rapid mixing using path coupling. We use matrix balancing to show that the Dobrushin–Shlosman condition $\alpha' < 1$ implies rapid mixing of systematic scan. An interesting question is whether the rapid mixing results for scan can be extended to the $\alpha = 1$ or $\alpha' = 1$ case. We give positive results for the rapid mixing of systematic scan for certain $\alpha = 1$ cases. As an application, we show rapid mixing of systematic scan (for any scan order) for heat-bath Glauber dynamics for proper q-colourings of a degree-Δ graph G when $q \geqslant 2\Delta$.

1. Introduction

A *spin system* consists of a set of *sites* and a set of spins. In this paper, both sets will be finite. We use $[n] = \{1, \ldots, n\}$ to denote the set of sites, and C to denote the set of spins. A *configuration* is an assignment of spins to sites, and Ω^+ denotes the set of all configurations.

† Partially supported by the EPSRC grant Discontinuous Behaviour in the Complexity of Randomized Algorithms. Some of the work was done while the authors were visiting the Mathematical Sciences Research Institute in Berkeley. A preliminary version of this paper appeared in the *Proceedings of the 10th International Workshop on Randomization and Computation* (RANDOM 2006).

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

Sites interact locally, and these interactions specify the relative likelihood of possible (local) sub-configurations. Taken together, these give a well-defined probability distribution, π, on the set of configurations Ω^+. *Glauber dynamics* is a random walk on configurations that updates spins one site at a time, and converges to π. Before giving some examples, we formalize these concepts in a way that will be useful for this paper.

We use the following notation. If x is a configuration and j is a site then x_j denotes the spin at site j in x. For each site j, S_j denotes the set of pairs of configurations that agree off of site j. That is, S_j is the set of pairs $(x, y) \in \Omega^+ \times \Omega^+$ such that, for all $i \neq j$, $x_i = y_i$. For each site j, we will have a transition matrix $P^{[j]}$ on the state space Ω^+ which satisfies two properties.

(1) $P^{[j]}$ moves from one configuration to another by updating site j. That is, if $P^{[j]}(x, y) > 0$, then $(x, y) \in S_j$.
(2) Also, π is invariant with respect to $P^{[j]}$.

Most theoretical results about Glauber dynamics consider *random updates,* in which the random walk on configurations proceeds as follows. At each step, a site j is chosen uniformly at random. The configuration is then updated according to transition matrix $P^{[j]}$. Formally, random-update Glauber dynamics corresponds to a Markov chain $\mathcal{M}$ with state space Ω^+ and transition matrix $P = (1/n)\sum_{j=1}^{n} P^{[j]}$.

For example, consider the spin system corresponding to proper q-colourings of an n-vertex graph G with maximum degree $\Delta \leqslant q - 2$. The sites are the vertices $1, \ldots, n$. C is the set of colours $C = \{1, \ldots, q\}$. The distribution π assigns equal probability to all proper colourings (colourings with no monochromatic edges) and it assigns zero probability to all improper colourings. In so-called 'heat-bath' Glauber dynamics, the transition matrix $P^{[j]}$ makes the following transition from a configuration x. Let C_x denote the set of colours that are not assigned to neighbours of site j in x. Let $x \to^j c$ denote the configuration obtained from x by changing the spin at site j to c. $P^{[j]}$ makes a transition to a uniformly chosen configuration in $\{x \to^j c \mid c \in C_x\}$. Another example is the 'Metropolis' Glauber dynamics for proper q-colourings of G. In this case, the transition matrix $P^{[j]}$ makes the following transition from x. A colour $c \in C$ is chosen uniformly at random. If $c \in C_x$ then a transition is made to $x \to^j c$; otherwise the new configuration is x. Further examples include corresponding dynamics for the Potts model (see Section 2.3) and for the hard-core lattice gas model.

There has been much work on analysing the *mixing time* of random-update Glauber dynamics. The mixing time from a specified initial configuration x (as a function of the deviation ε from stationarity) is

$$\tau_x(\mathcal{M}, \varepsilon) = \min\{t > 0 : \mathrm{d_{TV}}(P^t(x, \cdot), \pi(\cdot)) \leqslant \varepsilon\},$$

where $\mathrm{d_{TV}}$ denotes total variation distance. The mixing time of $\mathcal{M}$ is given by $\tau(\mathcal{M}, \varepsilon) = \max_{x \in \Omega^+} \tau_x(\mathcal{M}, \varepsilon)$. $\mathcal{M}$ is said to be 'rapidly mixing' if $\tau(\mathcal{M}, \varepsilon)$ is at most a polynomial in n and $\log(\varepsilon^{-1})$.

It is well known that the mixing time can be bounded in terms of the influences of sites on each other. To be more precise, let $\mu_j(x, \cdot)$ be the distribution on spins at site j induced by $P^{[j]}(x, \cdot)$ and let $\rho_{i,j}$ be the influence of site i on site j, which is given by

$$\rho_{i,j} = \max_{(x,y) \in S_i} \mathrm{d_{TV}}(\mu_j(x, \cdot), \mu_j(y, \cdot)).$$

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

We will be interested in three quantities. Let α be the total influence *on* a site, defined by

$$\alpha = \max_{j\in[n]} \sum_{i\in[n]} \rho_{i,j}.$$

Let α' be the total influence *of* a site, defined by

$$\alpha' = \max_{i\in[n]} \sum_{j\in[n]} \rho_{i,j}.$$

Finally, let α'' be the total influence of a site *in any given context,* defined by

$$\alpha'' = \max_{i\in[n]} \max_{(x,y)\in S_i} \sum_{j\in[n]} \mathrm{d_{TV}}(\mu_j(x,\cdot),\mu_j(y,\cdot)).$$

The *Dobrushin condition* $\alpha < 1$, which says that the total influence on a site is small, implies that $\mathcal{M}$ is rapidly mixing. In particular, $\tau(\mathcal{M},\varepsilon) = O(\frac{n}{1-\alpha}\log(n\varepsilon^{-1}))$. Dobrushin's original result [7] was not stated in terms of rapid mixing: instead, he was concerned with a closely related issue, uniqueness of the Gibbs measure for countable (not finite) spin systems. For a proof that the condition implies rapid mixing see, for example, Weitz's paper [21].

An easy application of the *path-coupling* method of Bubley and Dyer [2] shows that $\alpha'' < 1$ implies rapid mixing. In particular, $\tau(\mathcal{M},\varepsilon) = O(\frac{n}{1-\alpha''}\log(n\varepsilon^{-1}))$. Path coupling can also be used to show rapid mixing for the case $\alpha'' = 1$ provided the change in path length has enough variance. For details, see Dyer and Greenhill's survey paper [11].

An inspection of the definition of α' reveals that

$$\alpha' = \max_{i} \sum_{j} \max_{(x,y)\in S_i} \mathrm{d_{TV}}(\mu_j(x,\cdot),\mu_j(y,\cdot)) \geqslant \alpha''.$$

Therefore, $\alpha' < 1$ implies $\tau(\mathcal{M},\varepsilon) = O(\frac{n}{1-\alpha'}\log(n\varepsilon^{-1}))$. Dobrushin and Shlosman [8] were the first to derive uniqueness from the condition $\alpha' < 1$, which says that the total influence of a site is small. (As Weitz points out [21], Dobrushin and Shlosman stated their result in terms of the total influence *on* a site but they worked in a translation-invariant setting and what they used is that the total influence *of* a site is small. In fact, Dobrushin and Shlosman worked in a more general block-dynamics setting. This will be discussed below.)

While theoretical results about Glauber dynamics typically consider random updates, experimental work is often carried out using systematic strategies that cycle through sites in a deterministic manner, a dynamics we refer to as *systematic scan.* Formally, systematic scan corresponds to a Markov chain $\mathcal{M}_{\rightarrow}$ with state space Ω^+ and transition matrix $P_{\rightarrow} = \prod_{j=1}^{n} P^{[j]}$.

The Dobrushin condition implies that systematic scan is rapidly mixing. In particular,

$$\tau(\mathcal{M}_{\rightarrow},\varepsilon) = O\left(\frac{1}{1-\alpha}\log(n\varepsilon^{-1})\right).$$

A proof follows easily from the account of Dobrushin uniqueness in Simon's book [18], some of which is derived from the account of Föllmer [12].

This paper explores two related issues, primarily in the context of systematic scan.

(1) What is the relationship between conditions bounding the influence *on* a site (upper bounds on α) and conditions bounding the influence *of* a site (upper bounds on α' or α'')? Do they imply rapid mixing in the same circumstances?

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

(2) What is the relationship between proofs of rapid mixing using Dobrushin uniqueness, which typically use analysis techniques and proofs of rapid mixing using path coupling? Can they be used to prove rapid mixing in the same circumstances?

The second of these issues was first raised by Sokal [20], who shows how to translate a proof based on a Markovian coupling to analytic language. The first of these issues was highlighted by Weitz [21] who elucidated the dual nature of the Dobrushin condition $\alpha < 1$ and the Dobrushin–Shlosman condition $\alpha' < 1$.

A preliminary issue is the relationship between the parameters α, α' and α''. We have observed above that the two 'influence of a site' parameters, α' and α'', are related by $\alpha'' \leqslant \alpha'$. Weitz [21, Section 5.3] gives an example in a block-dynamics context where the parameter analogous to α'' is less than one (allowing one to infer rapid mixing), but the other parameters are greater than one. In Section 2.1 (Observation 2.1) we give a similar example in our (single-site) context. In this example, α'' can be made arbitrarily smaller than the other parameters. In Section 2.2 (Observation 2.2) we give an example where the 'influence of a site' parameters α' and α'' are less than one but the 'influence on a site' parameter α is greater. In Section 2.3 (Observation 2.3) we give an example where the 'influence on a site' parameter α is less than one but the other parameters are greater.

The primary motivation for our work was the observation that Dobrushin's condition (the influence on a site is small) implies rapid mixing for systematic scan. There are some known coupling results for systematic scan using the fact that the influence *of* a site is small, but only in very special circumstances, for example, proper colourings of a path, where the scan is left-to-right along the path [9]. When Dobrushin's condition is satisfied then systematic scan is rapidly mixing regardless of the order in which vertices are scanned. In this paper we show that rapid mixing also occurs (for every scan order) if the influence *of* a site is small. In particular, we prove the following theorem.

Theorem 1.1. *Suppose* $\alpha' < 1$. *Then*

$$\tau(\mathcal{M}_{\rightarrow}, \varepsilon) \leqslant \frac{2}{1-\alpha'} \log(4n^2(1-\alpha')^{-1}\varepsilon^{-1}).$$

The main ingredient in the proof is matrix balancing. This enables us to translate the Dobrushin–Shlosman condition (the influence of a site is small) into the Dobrushin condition (the influence on a site is small). The matrix balancing can be viewed as a generic way of deriving weights. Simon's version of Dobrushin uniqueness corresponds to showing convergence in the L_1-norm of a particular vector. This convergence occurs for $\alpha < 1$. The matrix-balancing approach corresponds to showing convergence in a weighted L_1-norm. Weights have been used previously in mixing results, for example [2, 9, 21]. The point here is that the weights are derived automatically by the balancing. Also, the contribution to the mixing-time bound arising from the ratios between the weights is accounted for in the balancing, so does not need to appear explicitly in the statement of the theorem. So rapid mixing holds for $\alpha' < 1$. For a discussion of weighted Dobrushin conditions, see Section 3.6.

The statement of Theorem 1.1 can be generalized. In particular, the proof does not require each of the sites to be updated the same number of times. All that is really required is that the sequence of updates contains a *subsequence* of at least $\frac{2}{1-\alpha'} \log(4n^2(1-\alpha')^{-1}\varepsilon^{-1})$ scans.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

The same generalization holds for the Dobrushin case. It will be useful to record the following corollary. Let $P_{\leftarrow} = \prod_{k=1}^{n} P^{[n-k+1]}$ and let $P_{\rightarrow\leftarrow} = P_{\rightarrow}P_{\leftarrow}$. Let $\mathcal{M}_{\rightarrow\leftarrow}$ be the Markov chain with state space Ω^+ and transition matrix $P_{\rightarrow\leftarrow}$.

Corollary 1.2. *Suppose* $\alpha' < 1$. *Then*

$$\tau(\mathcal{M}_{\rightarrow\leftarrow}, \varepsilon) \leqslant \frac{2}{1-\alpha'} \log(4n^2(1-\alpha')^{-1}\varepsilon^{-1}).$$

The mixing time of a Markov chain is closely related to its spectral gap: the gap between the largest eigenvalue of its transition matrix and the second-largest eigenvalue. We prove the following theorem, which applies when the eigenvalues of the transition matrix are real.

Theorem 1.3. *Suppose* $\alpha' < 1$. *Let* β_1 *be the second-largest eigenvalue of* $P_{\rightarrow}$. *If* β_1 *and its associated eigenvector are real then* $1 - \beta_1 \geqslant (1-\alpha')/2$.

Theorems 1.1 and 1.3 are proved in Section 3.

Having seen that the Dobrushin condition and the Dobrushin–Shlosman condition imply rapid mixing for systematic scan, an interesting question is whether these results can be extended to the $\alpha = 1$ or $\alpha' = 1$ case. Section 4 (Observation 4.1) provides an example where $\alpha = 1$ and random-update Glauber dynamics is rapidly mixing, but systematic scan is not even ergodic. Nevertheless, we give positive results for the rapid mixing of systematic scan for certain $\alpha = 1$ cases, particularly some cases for which there are *symmetric* upper bounds on the dependencies corresponding to a bound of 1, both on the total influence of a site, and on the total influence on a site.

A *dependency matrix* for a spin system is an $n \times n$ matrix R in which $R_{i,j} \geqslant \rho_{i,j}$ (so $R_{i,j}$ is an upper bound on the influence of site i on site j. We will be particularly interested in the case in which R is symmetric. In this case we can view R as a weighted adjacency matrix, and we refer to the resulting (undirected) graph on sites as the *dependency graph* of R. We say that R is *connected* if the resulting dependency graph is connected in the sense that there is a positive-weight path from every site to every other. In Section 5.1 we prove the following result, which says that systematic scan is rapidly mixing if there is a dependency matrix which (1) is symmetric, (2) has row and column sums at most 1 (corresponding to total influence at most 1 for every site), and (3) every connected component has a site with a row sum less than 1 (corresponding to total influence less than 1).

The mixing bound given in the theorem is a function of n, the number of sites, and also of N, the 'precision' of R. We say that a dependency matrix R has precision N (for a positive integer N) if every entry $R_{i,j}$ can be expressed as a fraction of integers with denominator N.

Theorem 1.4. *Suppose that a spin system has a precision-*N *symmetric dependency matrix* R *with row sums and column sums at most* 1. *Suppose that every connected component has a site whose row sum is less than* 1. *Then* $\tau(\mathcal{M}_{\rightarrow}, \varepsilon) = O(n^3 N \log(n^2\varepsilon^{-1}))$ *and* $\tau(\mathcal{M}_{\rightarrow\leftarrow}, \varepsilon) = O(n^3 N \log(n^2\varepsilon^{-1}))$.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

We now apply Theorem 1.4 to an example considered previously. Suppose that G is an n-vertex connected graph with maximum degree $\Delta \geqslant 2$. Let $q = 2\Delta$ and consider the spin system corresponding to heat-bath Glauber dynamics for proper q-colourings of G. If G is not Δ-regular then there is a connected symmetric dependency matrix R in which some vertex s of degree less than Δ has small total influence so its row sum is less than 1. As we will see in Section 5, Theorem 1.4 implies that systematic scan is rapidly mixing. Using the decomposition method of Martin and Randall, we can extend the result to the case in which G is Δ-regular (and every vertex has total influence 1), proving the following.

Theorem 1.5. *Let G be an n-vertex connected graph with maximum degree $\Delta \geqslant 2$ and let $q = 2\Delta$. Consider the spin system corresponding to heat-bath updates on proper colourings and let x be any proper colouring. Then $\tau_x(\mathcal{M}_{\rightarrow\leftarrow}, \varepsilon) = O(n^3 \log(n) \log((\varepsilon\pi(x))^{-1}))$.*

Theorem 1.5 is proved in Section 5.2 and a generalization (to more general spin systems) is stated and proved in Section 6.

It would be nice to have a full characterization of the situations in which $\alpha = 1$ (or $\alpha' = 1$) implies rapid mixing for systematic scan. Another interesting open question is whether bounds on the path-coupling parameter α'' imply rapid mixing for systematic scan.

All of our mixing results build upon the methods used in Simon's account of Dobrushin uniqueness (using analysis techniques). Thus, we might ask the question whether path coupling is a less useful technique for proving rapid mixing for systematic scan. In Section 7 we show that this may not be the case. In particular, we provide an alternative proof that the Dobrushin condition implies rapid mixing of systematic scan. This proof uses path coupling.

An issue that is not treated in this paper is how these methods generalize to dynamics other than Glauber. Dobrushin and Shlosman's result [8] actually applied to block dynamics rather than just to single-site dynamics, though only when the underlying dependency graph is Z^d. For random-update Glauber dynamics, Weitz [21] shows rapid mixing both when the influence of a site is small and when the influence on a site is small. Weitz's results apply to 'block dynamics' and to an arbitrary dependency graph. Both results are proved using coupling. Our results about systematic scan can be similarly generalized. This work, by Pedersen [15], is in preparation.

2. Relationships between the influence parameters

This section gives some examples illustrating the relationships between the influence parameters. The 'influence of a site' parameters α' and α'' are related by $\alpha'' \leqslant \alpha'$ and we give an example in which this inequality is strict. We give examples showing that the influence on a site can be either smaller than, or greater than, the influence of a site.

2.1. Example where the path-coupling parameter is smallest

Observation 2.1. *There is a spin system for which $\alpha'' < 1$ but $\alpha > 1$ and $\alpha' > 1$.*

Let H be a uniform hypergraph with n vertices, which correspond to the sites of our spin system. Let m be the size of the hyperedges and suppose that each vertex is contained in Δ edges.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

We will have two spins, so a configuration x will correspond to a subset of sites, namely the subset of sites with spin 1. The subset is said to be *independent* if it does not contain any hyperedge. Let π be the distribution in which the probability of an independent set of size s is proportional to λ^s and the probability of a non-independent set is zero. Let $P^{[j]}$ be the 'Metropolis' Glauber dynamics, which makes the following transition from a configuration x. If site j is in the set corresponding to x, then move to a new configuration by deleting j with probability $1/(1+\lambda)$ and stay at configuration x with the remaining probability. If j is not in the set corresponding to x, and the number of hyperedges contained in $x \cup \{j\}$ is at most the number of hyperedges contained in x, then move to the configuration corresponding to $x \cup \{j\}$ with probability $\lambda/(1+\lambda)$ and stay at x with the remaining probability. If j is not in the set corresponding to x but the number of hyperedges contained in $x \cup \{j\}$ exceeds the number of hyperedges contained in x, then stay at x.

Now consider $(x, y) \in S_i$. Suppose for $j \neq i$ that vertex j has $\mathrm{d}_{\mathrm{TV}}(\mu_j(x,\cdot), \mu_j(y,\cdot)) > 0$. First note that j is not in the set corresponding to x or y. So there is a hyperedge containing both i and j such that every other vertex in the hyperedge (other than i and j) is contained in x (and therefore in y). Note that there is at most one such j for this edge, since j is not in the set corresponding to x or y. Also, $\mathrm{d}_{\mathrm{TV}}(\mu_j(x,\cdot), \mu_j(y,\cdot)) \leqslant \lambda/(1+\lambda)$. Since $\mathrm{d}_{\mathrm{TV}}(\mu_i(x,\cdot), \mu_i(y,\cdot)) \leqslant \lambda/(1+\lambda)$, $\alpha'' \leqslant (\Delta+1)\lambda/(1+\lambda)$. Thus $\alpha'' < 1$ if λ is sufficiently small with respect to Δ. However, α is larger than α'' by (approximately) a factor of m, which can be arbitrarily large. In particular, $\alpha \geqslant \Delta(m-1)\lambda/(1+\lambda)$. To see this, consider a site j. For each of the Δ hyperedges containing j and for each of the $m-1$ other vertices i in the hyperedge, we can consider $(x, y) \in S_i$ such that j is not in the set corresponding to x (or y) but every other vertex in the hyper-edge (other than i or j) is in x (hence in y). $\mathrm{d}_{\mathrm{TV}}(\mu_j(x,\cdot), \mu_j(y,\cdot)) = \lambda/(1+\lambda)$ so $\alpha \geqslant \Delta(m-1)\lambda/(1+\lambda)$. The same analysis applies to α'. Thus, we have an example where the path-coupling constant α'' is less than one, but where α and α' are arbitrarily large.

2.2. Example where the influence of a site is small but the influence on a site is large

Observation 2.2. *There is a spin system for which $\alpha' < 1$ and $\alpha'' < 1$ but $\alpha > 1$.*

Suppose that G is an n-vertex star, *i.e.*, a graph with $n-1$ vertices adjacent to a central one. Let $q = n+1$. Consider heat-bath Glauber dynamics on proper colourings. If v is the centre vertex and w is another vertex then $\rho_{v,w} = 1/(q-1)$. Now to calculate $\rho_{w,v}$, suppose $(x, y) \in S_w$. Suppose $x_w = 1$ and $y_w = 2$. The simplest thing is to look at three cases, depending on whether none, one, or both, of 1 and 2 appear in the set of colours assigned to the other leaves by x (or equivalently, y). In the first case, suppose there are q' distinct colours at the other leaves. Then the number of available colours for v in both copies is $q-q'-1$ and the variation distance is $1/(q-q'-1)$ which is maximized by setting $q' = q-3$. The other cases are better, so $\rho_{w,v} = 1/2$. Now $\alpha \geqslant \sum_w \rho_{w,v} = (q-2)/2 > 1$ but

$$\alpha' \leqslant \max\left(\sum_{w \neq v} \rho_{v,w}, \rho_{w,v}\right) = \max\left(\frac{q-2}{q-1}, \frac{1}{2}\right) < 1.$$

Theorem 1.1 implies that systematic scan is rapidly mixing for this spin system.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

2.3. Example where the influence on a site is small but the influence of a site is large

Observation 2.3. *There is a spin system for which* $\alpha < 1$ *but* $\alpha'' > 1$ *and* $\alpha' > 1$

The spin system corresponds to the Ising model (2-state Potts model) on a length-2 path. There are just 3 sites, w, v and w'. There are two spins, 1 and 2, and the weight of an assignment x is λ^ℓ, where $\lambda < 1$ is a parameter of the spin system and $\ell \in \{0, 1, 2\}$ is the number of bichromatic edges in the colouring that x assigns to the path w, v, w'. π is the distribution in which the probability of a configuration is proportional to its weight. The dynamics is heat-bath: from a configuration x, $P^{[j]}$ transitions to a new configuration which is either $x \to^j 1$ or $x \to^j 2$. The relative probabilities of these are proportional to the weights of the two configurations.

First, note that $\rho_{v,w} = (1 - \lambda)/(1 + \lambda)$. (To see this, note that if $x_v = 1$ then the probability that colour 1 is chosen for w is $1/(1 + \lambda)$.) $\rho_{v,w'}$ is the same. Next, note that $\rho_{w,v} = 1/(1 + \lambda^2) - 1/2$. (To see this, assume without loss of generality that $x_{w'} = 1$. If $x_w = 1$ then the probability that 1 is chosen for v is $1/(1 + \lambda^2)$. If $x_w = 2$ then the probability is $1/2$.)

Now $\alpha = \max(\rho_{v,w}, 2\rho_{w,v})$, which is less than one for any positive λ. Also

$$\alpha' = \max(2\rho_{v,w}, \rho_{w,v}) = 2\rho_{v,w},$$

which is greater than 1 for $\lambda < 1/3$. Note that $\alpha'' = \alpha'$.

3. Rapid mixing for $\alpha' < 1$

In this section, we prove rapid mixing of systematic scan for the case $\alpha' < 1$. We prove Theorem 1.1, Corollary 1.2, and Theorem 1.3. Consider a spin system with dependency matrix R. Suppose that some entry of R is non-zero and that the row sums of R are less than 1. Let $\gamma = 1 - \max_{i \in [n]} \sum_{j \in [n]} R_{i,j}$. Note that $\gamma \in (0, 1]$. Also, if a spin system is non-trivial in the sense that it has sites i and j with $\rho_{i,j} > 0$ and it has $\alpha' < 1$ then it has such a dependency matrix with $\gamma = 1 - \alpha'$.

3.1. Matrix balancing

In this section we prove the following lemma.

Lemma 3.1. *There is an* $n \times n$ *diagonal matrix* W *such that, for every* $j \in [n]$, $\gamma/(2n) < W_{j,j} < 1$ *and every column sum of* WRW^{-1} *is less than* $1 - \gamma/2$.

Proof. Let R' be the matrix which is the same as R, except that we add a new column $n + 1$ and a new row $n + 1$. For $i \in [n]$, let $R'_{i,n+1} = (1 - r_i)/2$, where r_i is the row sum of R. Note that $R'_{i,n+1} \geqslant \gamma/2$ and that the row sums of R' are at most $1 - \gamma/2$. For $j \in [n]$, let $R'_{n+1,j}$ be $(1 - \gamma/2)/n$, and $R'_{n+1,n+1} = 0$, so $\sum_{j=1}^{n+1} R'_{n+1,j} = 1 - \gamma/2$. Thus R' is an irreducible non-negative matrix with largest eigenvalue at most $1 - \gamma/2$ [17, Theorem 1.5, Corollary 1]. It follows from the Perron–Frobenius theorem [17, Theorem 1.6] that there exists $\pi' \in \mathbb{R}^{n+1}$, with $\pi' > 0$ and $\sum_{j=1}^{n} \pi'_j = 1$, such that

$$\sum_{i=1}^{n+1} \pi'_i R'_{i,j} \leqslant (1 - \gamma/2)\pi'_j \qquad (j \in [n + 1]).$$

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

Now let $W = \mathrm{diag}(\pi'_1, \ldots, \pi'_{n+1})$ and let $R'' = WR'W^{-1}$. This multiplies row i by π'_i and divides column j by π'_j. Now

$$\sum_{i=1}^{n} R''_{i,j} \;=\; \sum_{i=1}^{n} R'_{i,j}\frac{\pi'_i}{\pi'_j} \;<\; \frac{1}{\pi'_j}\sum_{i=1}^{n+1}\pi'_i R'_{i,j} \;\leqslant\; 1-\gamma/2.$$

Thus (if we move the extra row and column from W), the column sums of WRW^{-1} are less than $1-\gamma/2$. Also $\pi'_{n+1} > \gamma/2$, since

$$\pi'_{n+1} \;>\; (1-\gamma/2)\pi'_{n+1} \geqslant \sum_{i=1}^{n}\pi'_i(1-r_i)/2 \;\geqslant\; (\gamma/2)\sum_{i=1}^{n}\pi'_i \;=\; \gamma/2.$$

Since $\pi'_j > 0$ ($j \in [n]$) and $\sum_{j=1}^{n}\pi'_j = 1$, we have $W_{j,j} = \pi'_j < 1$ ($j \in [n]$). Finally, for $j \in [n]$, $W_{j,j} \;=\; \pi'_j \;\geqslant\; \frac{1}{1-\gamma/2}\,\pi'_{n+1}\frac{1-\gamma/2}{n} \;>\; \frac{\gamma}{2n}$. □

3.2. The effect of a applying $P^{[j]}$ or $P_{\rightarrow}$.

Let $M^{[j]}$ be the matrix constructed from the identity matrix by replacing column j with the jth column of R. For any function f from Ω^+ to $\mathbb{R}$, let $\delta_i(f) = \max_{(x,y)\in S_i}|f(x)-f(y)|$ and let $\delta(f)$ be the column vector $\delta(f) = (\delta_1(f), \ldots, \delta_n(f))$. Let $P^{[j]}f$ be the function from Ω^+ to $\mathbb{R}$ given by $P^{[j]}f(x) = \sum_y P^{[j]}(x,y)f(y)$. The following lemma is (a slight generalization of one) in Simon's book [18]. The proof is included for completeness.

Lemma 3.2. *The vector $\delta(P^{[j]}f)$ is componentwise less than or equal to the vector $M^{[j]}\delta(f)$.*

Proof. Suppose $(x,y) \in S_i$. Then

$$\begin{aligned}
&|P^{[j]}f(x) - P^{[j]}f(y)| \\
&= \left|\sum_c f(x\rightarrow^j c)P^{[j]}(x, x\rightarrow^j c) - \sum_c f(y\rightarrow^j c)P^{[j]}(y, y\rightarrow^j c)\right| \\
&= \left|\sum_c f(x\rightarrow^j c)\mu_j(x,c) - \sum_c f(y\rightarrow^j c)\mu_j(x,c) + \sum_c f(y\rightarrow^j c)(\mu_j(x,c)-\mu_j(y,c))\right| \\
&\leqslant \sum_c \left|f(x\rightarrow^j c) - f(y\rightarrow^j c)|\mu_j(x,c) + \left|\sum_c f(y\rightarrow^j c)(\mu_j(x,c)-\mu_j(y,c))\right|\right. \\
&\leqslant \max_c |f(x\rightarrow^j c) - f(y\rightarrow^j c)| \\
&\qquad + (\max_c f(y\rightarrow^j c) - \min_c f(y\rightarrow^j c))\mathrm{d_{TV}}(\mu_j(x,\cdot),\mu_j(y,\cdot)) \\
&\leqslant 1_{i\neq j}\delta_i(f) + R_{i,j}\delta_j(f).
\end{aligned}$$

□

Now let W be the matrix from Lemma 3.1. Let $Q^{[j]} = WM^{[j]}W^{-1}$ and note that $Q^{[j]}$ is the matrix constructed from the identity matrix by replacing column j with the vector $(q_{1,j}, \ldots, q_{n,j})$, where $q_{i,j} = R_{i,j}w_i/w_j$. From Lemma 3.1, $\sum_{i=1}^{n} q_{i,j} \leqslant 1-\gamma/2$. Now let $k_i(f) = w_i\delta_i(f)$ and let $k(f)$ be the column vector $k(f) = (k_1(f), \ldots, k_n(f))$. Let $K(f) = \sum_{i=1}^{n} k_i(f)$. The following

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

lemma is a weighted version of a lemma in Simon's book (using the weights from the matrix balancing).

Lemma 3.3. *For any $m \in \{0, \ldots, n\}$,*

$$K(P^{[1]} \cdots P^{[m]} f) \leqslant \sum_{i=1}^{m} (1 - \gamma/2) k_i(f) + \sum_{i=m+1}^{n} k_i(f).$$

Proof. First note that lemma 3.2 implies that the vector $k(P^{[j]} f)$ is componentwise less than or equal to the vector $Q^{[j]} k(f)$. Using this fact, the proof is by induction on m with base case $m = 0$. The proof now follows Simon's. By the inductive hypothesis, $K(P^{[1]} \cdots P^{[m]} f)$ is at most $\sum_{i=1}^{m-1} (1 - \gamma/2) k_i(P^{[m]} f) + \sum_{i=m}^{n} k_i(P^{[m]} f)$. Applying our fact, this is at most

$$\sum_{i=1}^{m-1} (1 - \gamma/2) [Q^{[m]} k(f)]_i + \sum_{i=m}^{n} [Q^{[m]} k(f)]_i.$$

Then using the observation about the structure of $Q^{[m]}$ from above, we can rewrite this as

$$\sum_{i=1}^{m-1} (1 - \gamma/2)(k_i(f) + q_{i,m} k_m(f)) + q_{m,m} k_m(f) + \sum_{i=m+1}^{n} k_i(f) + q_{i,m} k_m(f).$$

Rearranging, this is

$$\sum_{i=1}^{m-1} (1 - \gamma/2) k_i(f) + \sum_{i=m+1}^{n} k_i(f) + k_m(f) \left(\sum_{i=1}^{m-1} (1 - \gamma/2) q_{i,m} + \sum_{i=m}^{n} q_{i,m} \right),$$

and the result now follows since $\sum_{i=1}^{n} q_{i,j} \leqslant 1 - \gamma/2$, for all j. □

Corollary 3.4.

$$K(P_{\rightarrow} f) \leqslant (1 - \gamma/2) K(f).$$

3.3. The bound on the mixing time: Proof of Theorem 1.1 and Corollary 1.2

For a test function f, let $f_t(x) = \sum_z P_{\rightarrow}^t(x, z) f(z)$. Thus, Corollary 3.4 gives

$$K(f_t) \leqslant (1 - \gamma/2) K(f_{t-1}).$$

Now let f be the indicator variable for being in some subset A of Ω^+. Then

$$\begin{aligned} \max_x f_t(x) - \min_y f_t(y) &\leqslant \sum_i \frac{k_i(f_t)}{w_i} \leqslant \left(\frac{1}{\min_i W_{i,i}} \right) K(f_t) \leqslant \frac{4n}{\gamma} K(f_t) \\ &\leqslant \frac{4n}{\gamma} (1 - \gamma/2)^t K(f_0) \leqslant \frac{4n^2}{\gamma} (1 - \gamma/2)^t, \end{aligned}$$

which is at most ε for $t \geqslant (2/\gamma) \log(4n^2 \gamma^{-1} \varepsilon^{-1})$. Also, $\min_y f_t(y) \leqslant E_\pi f_t \leqslant \max_x f_t(x)$ and $E_\pi f_t = \pi(A)$, which gives us

$$\tau(\mathcal{M}_{\rightarrow}, \varepsilon) \leqslant \frac{2}{\gamma} \log(4n^2 \gamma^{-1} \varepsilon^{-1}).$$

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

This proves Theorem 1.1. Corollary 1.2 comes from the fact that the right-hand side of the expression in Lemma 3.3 is at most $K(f)$. Thus, extra updates do no harm.

3.4. The bound on the spectral gap: Proof of Theorem 1.3

Let f be the eigenvector associated with the second-largest eigenvalue β_1 of $P_{\rightarrow}$. Suppose β_1 and f are real.

$$K(P_{\rightarrow}f) = \sum_{i=1}^{n} k_i(P_{\rightarrow}f) = \beta_1 \sum_{i=1}^{n} k_i(f) = \beta_1 K(f).$$

By Corollary 3.4, $1 - \beta_1 \geqslant \gamma/2$. This proves Theorem 1.3.

3.5. Remarks: contraction in various norms

Lemma 3.2 tells us that $\delta(P^{[j]}f) \leqslant M^{[j]}\delta(f)$. We are interested in the effect of M on the vector $\delta(f)$ where $M = \prod_{j=1}^{n} M^{[j]}$. Lemma 3.3 shows that $\delta(f)$ is contracting in the weighted L_1-distance $K(f)$. The contraction comes from the fact that the column sums of the weighted matrix $Q^{[j]}$ are less than 1. Simon's proof of Dobrushin's result for $\alpha < 1$ corresponds to taking W to be the identity matrix (so it is L_1-contraction). Our proof uses contraction in weighted L_1-distance, but there are other possibilities. For example, in Section 3.7 we revisit the random-updates Markov chain $\mathcal{M}$ and we prove mixing by observing a contraction in the L_∞-norm. This is possible when the row sums of the appropriate weighted matrix are less than 1.

3.6. Remark: weighted Dobrushin conditions

As mentioned in the Introduction, the matrix balancing in the proof of Theorem 1.1 can be viewed as a construction of weights under which the Markov chain exhibits convergence in a weighted L_1-norm. Another way to formalize this notion is to define *weighted* influences.

Thus, if $w_1, \ldots, w_n$ are weights of the n sites, we can define the weighted influence parameters

$$\alpha_w = \max_{j \in [n]} \sum_{i \in [n]} \frac{w_i}{w_j} \rho_{i,j}$$

and

$$\alpha'_w = \max_{i \in [n]} \sum_{j \in [n]} \frac{w_i}{w_j} \rho_{i,j}.$$

Lemma 3.1 can then be viewed as the statement that, if $\alpha' < 1$ (specifically, $\alpha' \leqslant 1 - \gamma$), then there is a weight vector w such that $\alpha_w < 1$ (specifically, $\alpha_w < 1 - \gamma/2$). This statement does not in itself imply rapid mixing since the mixing rate depends upon the ratio of the different weights w_j and $w_{j'}$. However, crucially, the lemma ensures that this ratio is bounded, since all weights are between $\gamma/2n$ and 1. Note that, by transposition, the lemma also allows the conclusion $\alpha'_w < 1$ from $\alpha < 1$.

3.7. A similar proof for random updates

It is known that $\alpha' < 1$ implies a bound on the spectral gap of the random-updates Markov chain $\mathcal{M}$. For example, this can be obtained by path coupling. For completeness, here is a proof in the Dobrushin framework.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

Let

$$M' = \frac{1}{n}\sum_j M^{[j]} = \frac{n-1}{n}I + \frac{1}{n}R.$$

Now note that every row of M' has a row sum which is at most $1-\gamma/n$. Suppose that P is reversible and let $\lambda = 1-\beta_1$ be its spectral gap. Recall that $Pf(x) = \sum_z P(x,z)f(z)$. Choose f to be the second-largest eigenvalue of P so $Pf = (1-\lambda)f$. Then, since the maximum component of $\delta(f)$ is shrunk by a factor of $1-\gamma/n$ when we multiply by M', we have

$$(1-\lambda)\max_i \delta_i(f) = \max_i \delta_i(Pf) \leqslant (1-\gamma/n)\max_i \delta_i(f),$$

so $(1-\lambda) \leqslant 1-\gamma/n$ and therefore $\lambda \geqslant \gamma/n$.

4. An $\alpha = 1$ case where systematic scan does not mix

Having seen that the Dobrushin condition and the Dobrushin–Shlosman condition imply rapid mixing for systematic scan, an interesting question is whether these results can be extended to the $\alpha = 1$ or $\alpha' = 1$ case. In this section we provide an example where $\alpha = 1$ and random-update Glauber dynamics is rapidly mixing, but systematic scan is not rapidly mixing.

Observation 4.1. *There is a spin system with $\alpha = 1$ for which random-update Glauber dynamics is rapidly mixing, but systematic scan is not ergodic.*

To establish Observation 4.1, consider a spin system with $n = 2m$ sites and spins 1 and 2. For every odd number j, the transition matrix $P^{[j]}$ deterministically flips the spin at site j (either from spin 1 to spin 2 or from spin 2 to spin 1). For every even number j, the transition matrix $P^{[j]}$ moves from configuration x to configuration $x_1 \cdots x_{j-1}x_{j-1}x_{j+1} \cdots x_n$.

Note that for an odd number i, $\rho_{i,i} = 1$ and $\rho_{i,i+1} = 1$. Every other $\rho_{i,j}$ is 0. Thus, $\alpha = 1$ but both α' and α'' are 2.

To see that systematic scan is not ergodic, note that it is periodic. For an odd number i, sites i and $i+1$ cycle through configurations $00, 10, 11, 01$.

To see that the random-update dynamics is rapidly mixing, use path coupling. The transitions on sites i and $i+1$ (for an odd number i) are the cycles above plus self-loops on 00 and 11 plus transitions from 10 to 00 and from 01 to 11.

Remarks. **(1)** Observation 4.1 indicates that perhaps Dobrushin-like arguments cannot be extended to the $\alpha = 1$ case. The reason for this is that Dobrushin-like arguments imply rapid mixing for systematic scan, as well as for random updates. But here is an example where systematic scan is clearly not rapidly mixing.

(2) The reader might argue that it is too much to expect a Dobrushin-like argument to apply when $\alpha = 1$. After all, path coupling does not apply to the $\alpha' = 1$ case unless there is sufficient variance – a side-condition that has not been carried over to the Dobrushin setting. This may be true, and a proof of rapid mixing must require additional considerations. It seems unlikely that such conditions can be Dobrushin-like, *i.e.*, based purely on the dependency matrix. We

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

have no example which directly supports this view, but the following shows something weaker. Combined with Observation 4.1, it implies that any such conditions could not be translated from α to α', and *vice versa*, as they are in the contractive case.

Observation 4.2. *Let R be the dependency matrix corresponding to Observation 4.1. Then there is a spin system with $\alpha' = 1$ and dependency matrix R^T such that systematic scan is rapidly mixing.*

To establish Observation 4.2, consider a spin system with $n = 2m$ sites and spins 1, 2 and 3. For every odd number j, the transition matrix $P^{[j]}$ deterministically moves from a configuration x with $x_j x_{j+1} = 11$ to $x \to^j 2$. If $x_j x_{j+1}$ is anything other than 11 then $P^{[j]}$ moves to either $x \to^j 1$ or $x \to^j 3$ with equal probability. For every even number j, the transition matrix $P^{[j]}$ moves to either $x \to^j 1$ or $x \to^j 2$ with equal probability. It can be checked that this example has dependency matrix R^T with $\alpha' = 1$. Coupling can be used to show that systematic scan is rapidly mixing: each pair of sites $(j, j+1)$ (where j is odd) can be coupled independently.

Remark. Note that the (directed) dependency graph corresponding to R is not strongly connected in the sense that there is not a positive-weight path from every site to every other. It is worth considering whether Dobrushin-like arguments do apply when $\alpha = 1$ and R is strongly connected. Connectivity turns out to be useful in the proof of Theorem 1.4 below, though it is not essential for the setting of that theorem (as is shown below).

5. Positive results for $\alpha = 1$

5.1. Symmetric dependency matrices with a row sum less than 1

In this section we prove Theorem 1.4, which says that systematic scan is rapidly mixing if there is a dependency matrix which (1) is symmetric, (2) has row and column sums at most 1 (corresponding to total influence at most 1 for every site), and (3) every connected component has a site with a row sum less than 1 (corresponding to total influence less than 1). We start with the connected case. Recall that a matrix has precision N if every entry in it can be expressed as a fraction of integers with denominator N.

Theorem 5.1. *Suppose that a spin system has a precision-N symmetric connected dependency matrix R with row sums and column sums at most 1. Suppose that there is a site s with row sum less than 1. Then $\tau(\mathcal{M}_{\rightarrow}, \varepsilon) = O(n^3 N \log(n^2 \varepsilon^{-1}))$ and $\tau(\mathcal{M}_{\rightarrow\leftarrow}, \varepsilon) = O(n^3 N \log(n^2 \varepsilon^{-1}))$.*

Remark. The dependence of the running time on the precision N is one way to express the condition that the dependency matrix R needs to be sufficiently 'mixing'. There are other possible choices. See [4].

Proof. Without loss of generality assume that for $i \neq s$ we have $\sum_j R_{i,j} = 1$. If this is not the case then $R_{i,i}$ can be increased until it is. Note that increasing $R_{i,i}$ to make the row sum 1 does not increase the precision of the matrix.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

Let $\gamma = 1 - \sum_{j\in[n]} R_{s,j}$. Since the sum of row s is less than 1, γ is positive. Also, γ is less than 1 (otherwise sites in $[n]$ do not depend upon s, contradicting connectivity).

Construct $R^{[2]}$ from R by adding an extra row and column. Set $R^{[2]}_{s,n+1} = R^{[2]}_{n+1,s} = \gamma$ and $R^{[2]}_{n+1,n+1} = 1-\gamma$ and make the rest of the entries in the new row and column equal to zero. Note that the row and column sums of $R^{[2]}$ are 1. It is the transition matrix of an ergodic Markov chain. Its stationary distribution, $\pi^{[2]}$, is uniform (by symmetry). Let $H(i,j)$ denote the hitting time of j from i and let H denote the maximum hitting time. Note that $H \leqslant (n+1)^2 N$. (To see this, view the Markov chain with transition matrix $R^{[2]}$ as a random walk on a N-regular undirected graph, with multiple edges and loops allowed. Then $H(G)$ is at most the number of vertices times the number of (directed) edges [13].) Note also that $H(i,i) = 1/\pi^{[2]}_i = (n+1)$.

Now let $R^{[3]} = R^{[2]} - E$, where E is the all-zero matrix except that row $n+1$ of E is $(-\xi,\ldots,-\xi,n\xi)$, where

$$\xi = \frac{1}{n(n+1)^2 N}.$$

Note that γ can be expressed as a fraction of integers with denominator N. Since it is less than 1, $\gamma \leqslant \frac{N-1}{N} < 1 - n\xi$, so $R^{[3]}_{n+1,n+1}$ is non-negative. It is clear that the other entries of $R^{[3]}$ are also non-negative. Furthermore, $R^{[3]}$ is the transition matrix of an ergodic Markov chain. Let $\pi^{[3]}$ be its stationary distribution. We now use Theorem 2.1 of [3]. This says that for all j,

$$|\pi^{[2]}_j - \pi^{[3]}_j| \leqslant \frac{||E||_\infty}{2} \frac{\max_{i\neq j} H(i,j)}{H_{j,j}} \leqslant \frac{1}{2(n+1)}.$$

Now let $W = \mathrm{diag}(\pi^{[3]}{}_1,\ldots,\pi^{[3]}{}_{n+1})$ and let $R^{[4]} = WR^{[3]}W^{-1}$ (this is multiplying row i by $\pi^{[3]}_i$ and dividing column j by $\pi^{[3]}_j$). Note by stationarity of $\pi^{[3]}$ that the columns of $R^{[4]}$ have sum 1. That is,

$$\sum_{i=1}^{n+1} R^{[3]}_{i,j} \frac{\pi^{[3]}_i}{\pi^{[3]}_j} = \frac{1}{\pi^{[3]}_j} \sum_{i=1}^{n+1} \pi^{[3]}_i R^{[3]}_{i,j} = 1.$$

Also for $j \in [n]$, we have

$$R^{[4]}_{n+1,j} = R^{[3]}_{n+1,j} \frac{\pi^{[3]}_{n+1}}{\pi^{[3]}_j} \geqslant \frac{1}{3} R^{[3]}_{n+1,j} \geqslant \frac{\xi}{3},$$

so $\sum_{i=1}^{n} R^{[4]}_{i,j} \leqslant 1 - \xi/3$. Finally, observe that for i and j in $[n]$,

$$R^{[4]}_{i,j} = R^{[3]}_{i,j} \frac{\pi^{[3]}_i}{\pi^{[3]}_j} = R_{i,j} \frac{\pi^{[3]}_i}{\pi^{[3]}_j},$$

so W (with row and column $n+1$ deleted) can be used as a weight matrix for R. Also $W_{j,j} = \pi^{[3]}_j \geqslant 1/(2(n+1))$.

Now we proceed as in Section 3, and to get the variation distance down to ε takes

$$\frac{3}{\xi} \log(2n(n+1)\varepsilon^{-1}) = 3n(n+1)^2 N \log(2n(n+1)\varepsilon^{-1})$$

scans. This proves the bound on $\tau(\mathcal{M}_{\rightarrow},\varepsilon)$ in the theorem. The bound on $\tau(\mathcal{M}_{\rightarrow\leftarrow},\varepsilon)$ is established in the same way as Corollary 1.2. □

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

To prove Theorem 1.4, apply Theorem 5.1 to each connected component individually. Suppose that t is the maximum, over the components, of $\tau(\mathcal{M}_{\rightarrow}, \varepsilon/n)$. Then $\mathrm{d}_{\mathrm{TV}}(P_{\rightarrow}^{t}(x,\cdot),\pi) \leqslant \varepsilon$. The same argument applies to $\tau(\mathcal{M}_{\rightarrow\leftarrow}, \varepsilon)$.

5.2. Heat-bath updates and proper colourings with 2Δ colours

Let G be a connected graph with maximum degree $\Delta \geqslant 2$ and let $q = |C| = 2\Delta$. Let Ω be the set of proper q-colourings of G. Let π be the uniform distribution on Ω (so $\pi(x) = 0$ for all configurations $x \in \Omega^{+} - \Omega$) . Let $P^{[j]}$ be the transition matrix for a 'heat-bath' update on site j. To be precise, $\mu_j(x,\cdot)$ is the uniform distribution on colours that are not used at neighbours of site j in configuration x. For all edges (i,j) of G, $\rho_{i,j} = 1/(q - \mathrm{d}(j))$, where $\mathrm{d}(j)$ is the degree of site j, so we take $R_{i,j} = 1/(q-\Delta) = 1/\Delta$. If (i,j) is not an edge then $R_{i,j} = 0$ ([16]).

Unlike the rest of the paper we now use the notation $\mathcal{M}_{\rightarrow\leftarrow}$ to refer to the Markov chain with transition matrix $P_{\rightarrow\leftarrow}$ but restricted to state space Ω.

We now prove Theorem 1.5. If G is not Δ-regular, then the theorem follows from Theorem 5.1. Some vertex has degree less than Δ, so has row sum less than 1. So assume from now on that G is Δ-regular.

Note that on Ω, $P^{[j]}$ is reversible with respect to the uniform distribution, so $P_{\rightarrow}(x,y) = P_{\leftarrow}(y,x)$ which implies that $\mathcal{M}_{\rightarrow\leftarrow}$ is reversible (with respect to the uniform distribution).

We now use the decomposition theorem of Martin and Randall [14]. For every spin $c \in C$, let $\Omega_c = \{x \in \Omega \mid x_n = c\}$. Let $P_{\rightarrow\leftarrow c}$ be the restriction Markov chain on Ω_c defined as follows. If $x \neq y$ and x and y are in Ω_c then $P_{\rightarrow\leftarrow c}(x,y) = P_{\rightarrow\leftarrow}(x,y)$. Finally, if $x \in \Omega_c$ then

$$P_{\rightarrow\leftarrow c}(x,x) = 1 - \sum_{y \in \Omega_c : y \neq x} P_{\rightarrow\leftarrow c}(x,y).$$

Note by symmetry that $\pi(\Omega_c) = 1/q$. Let $\overline{P}$ be the Markov chain on state space C whose transition matrix is defined by

$$\overline{P}(c,c') = q \sum_{x \in \Omega_c, y \in \Omega_{c'}} \pi(x) P_{\rightarrow\leftarrow}(x,y).$$

Note that $\overline{P}$ is reversible with respect to the uniform distribution on C.

Theorem 4.2 of [14] gives us

$$\mathrm{Gap}(P_{\rightarrow\leftarrow}) \geqslant \frac{1}{2}\mathrm{Gap}(\overline{P}) \min_{c \in C} \mathrm{Gap}(P_{\rightarrow\leftarrow c}), \tag{5.1}$$

where $\mathrm{Gap}(P)$ denotes the spectral gap of a reversible ergodic Markov chain with transition matrix P.

5.3. A lower bound on $\mathrm{Gap}(P_{\rightarrow\leftarrow c})$.

Let P_c^{*} be the Markov chain on Ω_c with transition matrix $P_1 \cdots P_{n-1} P_{n-1} \cdots P_1$. Note that both P_c^{*} and $P_{\rightarrow\leftarrow c}$ are reversible with respect to the uniform distribution on Ω_c. Furthermore, for any pair of distinct states x and y in Ω_c,

$$\frac{1}{q^2} P_c^{*}(x,y) \leqslant P_{\rightarrow\leftarrow c}(x,y).$$

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

This inequality implies that there is a very straightforward comparison between P_c^* and $P_{\rightarrow\leftarrow c}$ in which all of the flow corresponding to a transition of P_c^* uses the transition itself. The same inequality holds in the case $x = y$ (since every self-loop transition in $P_{\rightarrow\leftarrow c}$ occurs with probability at least q^{-2}). Thus, the flow is an odd flow, and $\mathrm{Gap}(P_{\rightarrow\leftarrow c}) \geqslant q^{-2}\mathrm{Gap}(P_c^*)$. (See Lemma 4.3 of [14]. The requirement that the flow be an odd flow is necessary to bound the gap between the smallest eigenvalue and -1. See [5, 10].)

The analysis in Section 5.1 can be used to upper-bound the mixing time of P_c^*, which is equivalent to lower-bounding its spectral gap. We will use the following inequality to convert the mixing-time bound to a bound on spectral gap:

$$\mathrm{Gap}(P_c^*) \geqslant \frac{1}{\tau(P_c^*, \varepsilon) + \ln(1/(2\varepsilon))} \ln\left(\frac{1}{2\varepsilon}\right).$$

This inequality is based on Proposition 1(ii) of [19] by Sinclair, which is based on a continuous version by Aldous [1]. A proof is given (see Theorem 5) in [10]. The inequality holds for any $\varepsilon > 0$. For example, we will take $\varepsilon = 1/4$.

We will now give an upper bound on $\tau(P_c^*, \varepsilon)$. The upper bound will actually be for the state space Ω_c^+ consisting of all colourings (proper and improper) with $x_n = c$. However, this implies an upper bound for Ω_c. Recall that our $O()$ notation suppresses terms that depend upon q and Δ, but not terms that depend upon n. The dependency matrix R^* for this chain can be obtained from R by deleting row n and column n. Every connected component of the corresponding dependency graph contains a site s which is adjacent to n in G. Then

$$\sum_{j=1}^{n-1} R_{s,j} = (\Delta - 1)(1/\Delta) = 1 - 1/\Delta.$$

We can apply Theorem 1.4 to get the bound $\tau(P_c^*, \varepsilon) = O(n^3 \log(n\varepsilon^{-1}))$. Taking $\varepsilon = 1/4$, we get $\mathrm{Gap}(P_{\rightarrow\leftarrow c}) = \Omega(1/(n^3 \log(n)))$.

5.3.1. A lower bound on $\mathrm{Gap}(\overline{P})$**.** Let Q be the Markov chain on $\{1, \ldots, q\}$ such that for all c, $Q(c, \cdot)$ is the uniform distribution. We will compare $\overline{P}$ to Q. First note that

$$\sum_{c \neq c'} \overline{P}(c, c') = q \sum_{x,y : x_n \neq y_n} \pi(x) P_{\rightarrow\leftarrow}(x, y) \geqslant q \sum_x \pi(x) \frac{1}{4} \geqslant \frac{q}{4}.$$

By symmetry, $\overline{P}(c, c')$ is the same for all pairs of distinct colours (c, c') so

$$P_{\rightarrow\leftarrow}(c, c') \geqslant \frac{1}{4(q-1)} = \frac{q}{4(q-1)} Q(c, c').$$

Similarly,

$$\sum_c \overline{P}(c, c) = q \sum_{x,y : x_n = y_n} \pi(x) P_{\rightarrow\leftarrow}(x, y) \geqslant q \sum_x \pi(x) \frac{1}{q} \geqslant 1,$$

so $\overline{P}(c, c) \geqslant Q(c, c)$ by symmetry. This gives

$$\mathrm{Gap}(\overline{P}) \geqslant \frac{q}{4(q-1)} \mathrm{Gap}(Q) = \frac{q}{4(q-1)}.$$

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

5.3.2. Putting it all together. By decomposition, we now have

$$\mathrm{Gap}(P_{\to\leftarrow}) = \Omega(1/(n^3 \log(n))).$$

Applying an inequality of Diaconis and Stroock [6, Proposition 3] and Sinclair [19, Proposition 1(i)], we get

$$\tau_x(\mathcal{M}_{\to\leftarrow}, \varepsilon) \leqslant O(n^3 \log(n)) \ln \frac{1}{\varepsilon\pi(x)}.$$

Note that $\pi(x) \geqslant q^{-n}$, so

$$\tau_x(\mathcal{M}_{\to\leftarrow}, 1/(2e)) \leqslant O(n^4 \log(n)).$$

6. A generalized $\alpha = 1$ case

The following is a generalization of Theorem 1.5. Consider a general spin system. Let $\Omega = \{x \in \Omega^+ \mid \pi(x) > 0\}$. As in Section 5.2, let $\Omega_c = \{x \in \Omega \mid x_n = c\}$ and let $P_{\to\leftarrow c}$ be the Markov chain on Ω_c defined by $P_{\to\leftarrow c}(x, y) = P_{\to\leftarrow}(x, y)$ for distinct x and y in Ω_c with $P_{\to\leftarrow c}(x, x) = 1 - \sum_{y\in\Omega_c : y\neq x} P_{\to\leftarrow c}(x, y)$.

Theorem 6.1. *Suppose that a spin system has a precision-N symmetric connected dependency matrix R with row sums equal to* 1. *Suppose there is a positive real ξ such that:*

(1) *for every site j, $P^{[j]}$ is reversible with respect to π,*
(2) *for every spin c, $P_{\to\leftarrow c}$ is irreducible,*
(3) *for every configuration x and every site j, $P^{[j]}(x, x_j) \geqslant \xi$,*
(4) *for every configuration x and every colour c, $\Pr(\tau_n = c) \geqslant \xi$ when τ is drawn from $P_{\to\leftarrow}(x, \cdot)$.*

Then $\tau_x(\mathcal{M}_{\to\leftarrow}, \varepsilon) \leqslant O(\xi^{-3} n^3 N \log(n)) \log(1/(\varepsilon\pi(x)))$.

Proof. The proof follows the proof of Theorem 1.5. Condition (1) implies that $\mathcal{M}_{\to\leftarrow}$ is reversible with respect to π. The projection chain $\overline{P}$ is defined by

$$\overline{P}(c, c') = \frac{1}{\pi(\Omega_c)} \sum_{x\in\Omega_c, y\in\Omega_{c'}} \pi(x) P_{\to\leftarrow}(x, y).$$

Note that $\overline{P}$ is reversible with respect to the distribution that chooses spin c with probability $\pi(\Omega_c)$. Condition (4) implies that $\overline{P}$ is irreducible and aperiodic. $P_{\to\leftarrow c}$ is reversible by construction. It is aperiodic by condition (3) and irreducible by condition (2). Similarly, (3) implies that $P_{\to\leftarrow}$ is aperiodic and (2) and (4) imply that it is irreducible. Since $\overline{P}$, $P_{\to\leftarrow}$ and $P_{\to\leftarrow c}$ are reversible and ergodic, we can apply the decomposition theorem.

By condition (3), $P_{\to\leftarrow c}(x, y) \geqslant \xi^2 P_c^*(x, y)$, so $\mathrm{Gap}(P_{\to\leftarrow c}) \geqslant \xi^2 \mathrm{Gap}(P_c^*)$. Applying Theorem 1.4, we get $\mathrm{Gap}(P_c^*) = \Omega(1/(n^3 N \log(n)))$.

Condition (4) gives $\overline{P}(c, c') \geqslant \xi q Q(c, c')$, so $\mathrm{Gap}(\overline{P}) \geqslant \xi q \mathrm{Gap}(Q) = \xi q$. The remainder of the proof is unchanged. □

Remarks. **(1)** It is easy to see that Theorem 6.1 implies Theorem 1.5, though the implicit constants are slightly worse since condition (4) gives a slightly worse analysis than the analysis

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

given in Section 5.3.1. The theorem also applies to Glauber dynamics on spin systems such as the Potts model or the hard-core lattice gas model.

(2) The connectivity requirement in Theorem 6.1 can be removed by considering the connected components separately as in the proof of Theorem 1.4.

7. Proving rapid mixing for systematic scan using path coupling

All of the mixing results so far have built upon Dobrushin uniqueness. We conclude by sketching an alternate proof, based on path coupling [2], that the Dobrushin condition $\alpha < 1$ implies rapid mixing of systematic scan. Similar ideas are implicit in [21], though this does not explicitly consider systematic scan.

We consider coupled chains X_t, Y_t. Let the (path) coupling be given by choosing the same vertex w_t in both chains, and then coupling the choice of spin maximally. Suppose the initial states X_0, Y_0 have shortest path P_0. The length of P_0 is the Hamming distance $H(X_0, Y_0)$. Consider the evolution of this path at time t to $P = (Z_0, Z_1, \ldots, Z_{\ell-1}, Z_\ell)$, with length $\ell \geqslant H(X_t, Y_t)$. (Note that we do not optimize the path length after each step, but assume instead that the path evolves naturally.) We will call any edge of P (Z_{r-1}, Z_r) ($r \in [\ell]$) an edge in S_i if $(Z_{r-1}, Z_r) \in S_i$. Note that P_0 has at most one edge in S_i for each $i = 1, 2, \ldots, n$. Suppose v_i is the total number of edges of P in S_i. Clearly $\ell = \sum_{i=1}^n v_i$, so $\mathrm{E}[\ell] \leqslant n \max_i \mathrm{E}[v_i]$.

Suppose $w_t = j$. For every edge in S_i ($i \neq j$), an edge in S_i will persist, and a new edge in S_j will appear with probability at most $\rho_{i,j}$. Every edge in S_j will either disappear, or persist with probability at most $\rho_{j,j}$. Thus, denoting the quantities at time $t+1$ with primes, $\mathrm{E}[v_i'] = \mathrm{E}[v_i]$ ($i \neq j$) and $\mathrm{E}[v_j'] \leqslant \sum_{i=1}^n \rho_{i,j} \mathrm{E}[v_i] \leqslant \alpha \max_i \mathrm{E}[v_i]$. In a complete scan we have $w_t = j$ for every j and some t. Hence the resulting values $\overrightarrow{v}_i$ after the scan will satisfy $\max_i \mathrm{E}[\overrightarrow{v}_i] \leqslant \alpha \max_i \mathrm{E}[v_i]$. Since $\max_i \mathrm{E}[v_i] \leqslant 1$ initially, after s complete scans $\max_i \mathrm{E}[v_i] \leqslant \alpha^s$, and thus $\mathrm{d_{TV}}(X_{ns}, Y_{ns}) \leqslant \Pr(X_{ns} \neq Y_{ns}) \leqslant \mathrm{E}[H(X_{ns}, Y_{ns})] \leqslant \mathrm{E}[\ell] \leqslant n\alpha^s$.

Acknowledgements

We are very grateful to Alan Sokal for pointing out to us that Dobrushin's condition implies mixing of systematic scan, and for referring us to [18], and to Christian Borgs for bringing Föllmer's work to our attention.

References

[1] Aldous, D. (1982) Some inequalities for reversible Markov chains. *J. London Math. Soc.* (2) **25** 564–576.

[2] Bubley, R. and Dyer, M. E. (1997) Path coupling: A technique for proving rapid mixing in Markov chains. In *Proc. 38th Annual Symposium on Foundations of Computer Science* (FOCS '97), pp. 223–231.

[3] Cho, G. E. and Meyer, C. D. (2000) Markov chain sensitivity measured by mean first passage times. *Linear Algebra Appl.* **316** 21–28.

[4] Cho, G. E. and Meyer, C. D. (2001) Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra Appl.* **335** 137–150.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press

[5] Diaconis, P. and Saloff-Coste, L. (1993) Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730.
[6] Diaconis, P. and Stroock, D. (1991) Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.
[7] Dobrushin, R. L. (1970) Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl.* **15** 458–486.
[8] Dobrushin, R. L. and Shlosman, S. B. (1985) Constructive criterion for the uniqueness of a Biggs field. In *Statistical Mechanics and Dynamical Systems* (J. Fritz, A. Jaffe and D. Szasz, eds), Birkhäuser, Boston, pp. 347–370.
[9] Dyer, M., Goldberg, L. A. and Jerrum, M. (2006) Systematic scan for sampling colourings. *Ann. Appl. Probab.*, **16** 185–230.
[10] Dyer, M., Goldberg, L. A., Jerrum, M. and Martin, R. (2006) Markov chain comparison. *Probability Surveys* **3** 89–111.
[11] Dyer, M. and Greenhill, C. (1999) Random walks on combinatorial objects. In *Surveys in Combinatorics* (J. D. Lamb and D. A. Preece, eds), Vol. 267 of *London Mathematical Society Lecture Notes*, Cambridge University Press, pp. 101–136.
[12] Föllmer, H. (1982) A covariance estimate for Gibbs measures. *J. Funct. Anal.* **46** 387–395.
[13] Lovász, L. and Winkler, P. (1995) Mixing of random walks and other diffusions on a graph. In *Surveys in Combinatorics* (P. Rowlinson, ed.), Vol. 218 of *London Mathematical Society Lecture Notes*, pp. 119–154.
[14] Martin, R. and Randall, D. (2000) Sampling adsorbing staircase walks using a new Markov chain decomposition method. In *Proc. 41st Annual IEEE Symposium on Foundations of Computer Science* (FOCS 2000), pp. 492–502.
[15] Pedersen, K. Personal communication.
[16] Salas, J. and Sokal, A. D. (1997) Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem. *J. Statist. Phys.* **86** 551–579.
[17] Seneta, E. (2006) *Non-Negative Matrices and Markov Chains*, 2nd edn, Springer, New York.
[18] Simon, B. (1993) *The Statistical Mechanics of Lattice Gases*, Princeton University Press.
[19] Sinclair, A. (1992) Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* **1** 351–370.
[20] Sokal, A. (2001) A personal list of unsolved problems concerning lattice gases and antiferromagnetic Potts models. *Markov Processes and Related Fields*, **7** 21–38.
[21] Weitz, D. (2005) Combinatorial criteria for uniqueness of Gibbs measures. *Random Struct. Alg.* **27** 445–475.

https://doi.org/10.1017/S0963548308009437 Published online by Cambridge University Press