Systematic scan for sampling colourings *

Martin Dyer †
School of Computing
University of Leeds

Leslie Ann Goldberg‡
Department of Computer Science
University of Warwick

Mark Jerrum§
School of Informatics
University of Edinburgh

June 21, 2004

Abstract

We address the problem of sampling colourings of a graph G by Markov chain simulation. For most of the article we restrict attention to proper q-colourings of a path on n vertices (in statistical physics terms, the 1-dimensional q-state Potts model at zero temperature), though in later sections we widen our scope to general "H-colourings" of arbitrary graphs G. Existing theoretical analyses of the mixing time of such simulations relate mainly to a dynamics in which a random vertex is selected for updating at each step. However, experimental work is often carried out using systematic strategies that cycle through coordinates in a deterministic manner, a dynamics sometimes known as *systematic scan*. The mixing time of systematic scan seems more difficult to analyse than that of random updates, and little is currently known. In this article, we go some way towards correcting this imbalance. By adapting a variety of techniques, we derive upper and lower bounds (often tight) on the mixing time of systematic scan. An unusual feature of systematic scan as far as the analysis is concerned is that it fails to be time reversible.

*This work was partially supported by the IST Programme of the EU under contract number IST-1999-14036 (RAND-APX), and by the UK EPSRC through the research grant *Discontinuous Behaviour in the Complexity of Randomized Algorithms*.

†`http://www.comp.leeds.ac.uk/~dyer/`, School of Computing, University of Leeds, Leeds LS2 9JT, United Kingdom.

‡`http://www.dcs.warwick.ac.uk/~leslie/`, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom.

§`http://www.dcs.ed.ac.uk/home/mrj/`, School of Informatics, University of Edinburgh, JCMB, The King's Buildings, Edinburgh EH9 3JZ, United Kingdom.

1 Introduction

Many models in statistical physics come under the heading of "spin systems". Such a system is specified by a graph G, in our case finite. Configurations of the system are assignments of "spins" to the vertices of G. There are assumed to be q possible spins, and hence potentially q^n configurations, where n is the number of vertices of G, though some of these configurations may be illegal. Each configuration has an energy that comes from summing, over all edges of G, the interaction energies between adjacent spins. These energies specify a probability distribution, called the *Boltzmann distribution*, on configurations. The Potts model and the hard-core lattice gas model are examples of spin systems.

In this paper, for consistency with previous literature, we shall refer to spins as *colours* and to configurations as *states*. Sampling from the Boltzmann distribution is a challenging computational task. Often, the only feasible way of going about it is to simulate a suitable random "dynamics" on configurations. The dynamics has the property of converging to a stationary distribution which is the Boltzmann distribution. This is usually straightforward to arrange. The hard part is proving that the dynamics is "rapidly mixing", i.e., converges rapidly to stationarity.

Identifying the vertices of G with the integers $\{1, 2, \ldots, n\}$, we may think of the state space as having coordinates. There is a substantial body of literature concerned with bounding *mixing time* (i.e., time to convergence to near-stationarity) of systems such as those described above. Almost all this theoretical work relates to random single-site updates, which choose a random coordinate for updating at each transition. We shall refer to this strategy as *Glauber dynamics*.[1] However, experimental work is often carried out using systematic strategies that cycle through coordinates in a deterministic manner, a dynamics we refer to as *systematic scan* (or just "scan" for short). The mixing time of systematic scan seems more difficult to analyse that that of Glauber, and little is currently known.

In this paper, we take some first steps in analysing systematic scan for spin systems. Our setting will be very simple; indeed, for the most part, we will restrict attention to proper q-colourings of a path of n vertices (in statistical physics terms, the 1-dimensional q-state Potts model at zero temperature). To compensate for the simple setting, we provide tight (i.e., matching within a constant factor) upper and lower bounds on mixing time. Measuring mixing time in terms of the number of updates of individual vertices (so that one scan equates to n updates) we show that: when $q = 3$, mixing occurs in $\Theta(n^3 \log n)$ updates, whether Glauber dynamics or systematic scan is used; while when $q \geq 4$, mixing occurs in $\Theta(n \log n)$ updates, again independently of whether Glauber or scan is used. Our main tools are: harmonic analysis (Wilson [27]), path coupling (Bubley and Dyer [5]) and disagreement percolation (van den Berg [26]).

Later in the paper, we considerably widen the setting from usual proper colourings to general H-colourings (also known a graph homomorphisms), but staying at first with the path as the underlying graph. H-colourings model arbitrary spin systems with symmetric "hard" constraints. We show that, for any H, Glauber mixes in $O(n^5)$ updates and scan in $O(n^6)$ updates. The former bound is unlikely to be tight, and the latter even less so. The method here is that of canonical paths (Sinclair and Jerrum [25] and Diaconis and Stroock [9]).

Finally, we consider H-colourings of a general graph G, and compare the mixing times of

[1] The term "Glauber Dynamics" appears not to have a precise agreed meaning. Here we are using the term to signify *single site* updates performed in a *random sequence*. These are certainly aspects of the dynamics first considered by Glauber [17].

scan and Glauber. We show that for any H these are within a polynomial factor of each other (in terms of total number of individual updates performed), at least when G is of bounded degree. The question of whether scan can ever be faster than Glauber, or vice versa, remains a tantalising open problem. The only situation where a gap is known is the rather uninteresting one that arises when G is the empty graph, where Glauber requires $\Theta(n \log n)$ updates (Dyer et al. [12]) while scan clearly mixes in one sweep.

1.1 Previous Work

Amit [3] has investigated systematic scan in the context of sampling from multivariate Gaussian distributions. In this instance, one iteration of systematic scan applies a "heat-bath" update to each coordinate axis in turn. Amit precisely calculates the spectral gap of the scan operator, and hence bounds the mixing time. He also estimates the spectral gap of a similar process on perturbed Gaussian distributions.

In another application of systematic scan — this time more combinatorial in nature and slightly closer to the one studied here — Diaconis and Ram [7] consider the problem of generating random elements of a finite group. The *systematic scan* Metropolis algorithm cycles through the generators in order, and flips coins to decide whether or not to multiply by each generator in turn. The *random update* algorithm chooses a generator uniformly at random at each step. For the symmetric group, they show that the systematic scan algorithm mixes in $\Theta(n)$ scans, so consideration of $\Theta(n^2)$ generators is necessary and sufficient for mixing. They consider two different scanning strategies from [16] — the same results hold for both strategies. Matching results (in terms of the number of generators considered) are given by Benjamini et al. [4] for the random update strategy. Diaconis and Ram also consider the hypercube and the dihedral group. For the hypercube, they show that $\Theta(n \log n)$ updates are necessary and sufficient, whether one is doing random updates or systematic scan. For the dihedral group, both strategies take $\Theta(n)$ updates. Diaconis and Ram point out that careful analysis of rates of convergence for the Metropolis algorithm is completely open in non-group cases.

For a brief review of other work on systematic scan, consult Diaconis and Ram [7, §2b].

2 Definitions and notation

The *variation distance* between distributions θ_1 and θ_2 on Ω is

$$\mathrm{d_{TV}}(\theta_1, \theta_2) = \frac{1}{2} \sum_i |\theta_1(i) - \theta_2(i)| = \max_{A \subseteq \Omega} |\theta_1(A) - \theta_2(A)|.$$

For a discrete ergodic Markov chain $\mathcal{M}$ with transition matrix P and stationary distribution π, and a distinguished state x, the mixing time (as a function of the deviation ε from stationarity) is

$$\mathrm{Mix}_x(\mathcal{M}, \varepsilon) = \min \left\{ t > 0 : \mathrm{d_{TV}}(P^t(x, \cdot), \pi(\cdot)) \leq \varepsilon \right\}.$$

The mixing time of $\mathcal{M}$ is $\mathrm{Mix}(\mathcal{M}, \varepsilon) = \max_x \mathrm{Mix}_x(\mathcal{M}, \varepsilon)$.

Suppose G is an undirected graph with vertex set $\{1, \ldots, n\}$. To avoid trivialities, we assume $n > 3$. We consider q-colourings of G, where $q \geq 3$. Formally, a colouring σ is a vector $\sigma = (\sigma_1, \ldots, \sigma_n)$ in which $\sigma_i \in \{0, \ldots, q-1\}$ denotes the colour of vertex i. A colouring is *proper* if adjacent vertices receive different colours. $\Omega^+ = \{0, \ldots, q-1\}^n$ is the set of all colourings (proper and improper), while Ω is the set of all proper colourings.

A Markov chain with state space Ω starts at a colouring $\sigma(0)$ and visits a sequence of colourings $\sigma(0), \sigma(1), \ldots$. We often use τ to denote a colouring (when we need two names). The two Markov chains that we study are as follows.

- $\mathcal{M}_{\mathrm{Gl}}$ (Glauber): Choose vertex v uniformly at random; do Metropolis(v).
- $\mathcal{M}_{\rightarrow}$ (Systematic scan): For $v := 1$ to n do Metropolis(v).

The procedure Metropolis(v) used in both of the above dynamics performs as follows: A colour c is chosen uniformly at random. A proposed new colouring is formed by recolouring vertex v with colour c. This proposed move is accepted if and only if colour c is not used at any neighbour of v.

Let P_{Gl} be the transition matrix of $\mathcal{M}_{\mathrm{Gl}}$ and Let $P_{\rightarrow}$ be the transition matrix of $\mathcal{M}_{\rightarrow}$. It will be convenient in our proofs to consider reverse systematic scan:

- $\mathcal{M}_{\leftarrow}$ (Reverse scan): For $v := n$ down to 1 do Metropolis(v).

Let $P_{\leftarrow}$ be the transition matrix of $\mathcal{M}_{\leftarrow}$. Observe that $\mathcal{M}_{\leftarrow}$ is the time reversal of $\mathcal{M}_{\rightarrow}$, since $P_{\rightarrow}(\sigma, \sigma') = P_{\leftarrow}(\sigma', \sigma)$ for all $\sigma, \sigma' \in \Omega$.

Let $\mathcal{M}$ be any discrete Markov chain with transition matrix P, stationary distribution π and state space Ω. Define the *optimal Poincaré constant* of $\mathcal{M}$ by

$$\lambda(\mathcal{M}) = \inf_{f:\Omega \to \mathbb{R}} \frac{\mathcal{E}_{\mathcal{M}}(f,f)}{\mathrm{var}_\pi(f)},$$

where the inf is over all non-constant functions from Ω to $\mathbb{R}$ and the *Dirichlet form* is given by

$$\mathcal{E}_{\mathcal{M}}(f,f) = \frac{1}{2} \sum_{x,y \in \Omega} \pi(x) P(x,y) (f(x) - f(y))^2$$

and

$$\mathrm{var}_\pi(f) = \sum_{x \in \Omega} \pi(x)(f(x) - \mathbf{E}_\pi\, f)^2 = \frac{1}{2} \sum_{x,y \in \Omega} \pi(x)\pi(y)(f(x) - f(y))^2.$$

If $\mathcal{M}$ is time-reversible with respect to π (i.e., $\pi(\sigma)P(\sigma,\tau) = \pi(\tau)P(\tau,\sigma)$) then the eigenvalues of P are real and can be written $1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{|\Omega'|-1} > -1$. Then $\lambda(\mathcal{M})$ is is equal to $1 - \beta_1$.

Some of our rapid-mixing proofs will use the method of *path coupling* [5]. In our path-coupling proofs, we will define partial couplings on the set S, which will always be the set of pairs of colourings that differ on a single vertex.

For most of the paper we consider the case in which G is a path going left-to-right from vertex 1 to vertex n. Kenyon and Randall [22] have shown that, for every q, the *block dynamics,* which updates a sufficiently large constant-length path at each step, mixes in time $O(n \log n)$. Our results show that this upper bound holds for single-site dynamics for $q \geq 4$, but not for $q = 3$.

In our analysis for $q = 3$ we will study the auxiliary Markov chains $\mathcal{M}^{\pm}_{\mathrm{Gl}}$ and $\mathcal{M}^{\pm}_{\rightarrow}$ on state space $\Upsilon = \{-1, 1\}^{n-1}$. A configuration $X \in \Upsilon$ is a vector $X = (X_1, \ldots, X_{n-1})$. The chain evolves as $X(0), X(1), \ldots$. We use the notation $\mathrm{Ham}(X, Y)$ to denote the Hamming distance between configurations X and Y.

3 The analysis technique for $q = 3$

The following develops an idea of Wilson [27] for lower bounding the convergence rate of certain types of Markov chains.

Let $\mathcal{M}$ be a finite ergodic Markov chain with transition matrix P and state space $\Upsilon \subseteq \mathbb{Z}^m$. Suppose there exists a matrix A such that $\mathbf{E}[X(1) \mid X(0)] = AX(0)$ for all $X(0) \in \Upsilon$.[2] We will assume that A has real eigenvalues, though it is possible to extend the method to complex eigenvalues. We may further assume that A has only non-negative eigenvalues, since otherwise we can consider the two-step chain $\mathcal{M}^2 = (\Upsilon, P^2)$ which converges exactly twice as fast. Now let λ be any eigenvalue of A, with left eigenvector w. Then,

$$\mathbf{E}[wX(t) \mid X(0)] = wA^t X(0) = \lambda^t wX(0). \tag{1}$$

Let $\Phi_t = wX(t)$. To obtain the strongest lower bound, we choose λ to be the largest eigenvalue such that there exist $x, y \in \Upsilon$ with $wx \neq wy$. Then we choose $X(0) = \mathrm{argmax}_x |wx|$. Since w is defined only to scalar multiples, we may assume $wX(0) > 0$. It follows from (1) that $\lambda \leq 1$. Otherwise $\limsup_{t\to\infty} \mathbf{E}[\Phi_t] = \infty$, contradicting the finiteness of $\mathcal{M}$. If $\lambda = 1$, we have $\mathbf{E}[\Phi_t] = \Phi_0$ for all t. But $\Phi_t \leq \Phi_0$, so we must have $\Phi_t = \Phi_0$ for all t. Using ergodicity of $\mathcal{M}$, this implies $wx = wy$ for all $x, y \in \Upsilon$, contradicting our choice of λ. Thus $\lambda < 1$, and hence $\lim_{t\to\infty} \mathbf{E}[wX(t)] = 0$. If $X(\infty)$ denotes the equilibrium distribution, it follows that $\mathbf{E}[wX(\infty)] = 0$.

We will now consider the quantities $\mathbf{E}[\Phi_t \mid \Phi_{t-1}]$ and $\mathrm{var}(\Phi_t \mid \Phi_{t-1})$. Definitions of conditional expectations and variances can be found in [10] (pages 190–198). We will use the fact that $\mathrm{var}(Y) = \mathbf{E}[\mathrm{var}(Y \mid X)] + \mathrm{var}(\mathbf{E}[Y \mid X])$ (page 198). Suppose that $\mathbf{E}[\mathrm{var}(\Phi_t \mid \Phi_{t-1})] \leq \rho$ for all $t > 0$, and let $\nu = \rho/(1-\lambda^2)$. Now using $\mathbf{E}[\Phi_t \mid \Phi_{t-1}] = \lambda\Phi_{t-1}$ and $\mathrm{var}(\Phi_0) = 0$,

$$\begin{aligned}
\mathrm{var}(\Phi_t) &= \mathbf{E}[\mathrm{var}(\Phi_t \mid \Phi_{t-1})] + \mathrm{var}(\mathbf{E}[\Phi_t \mid \Phi_{t-1}]) \\
&= \mathbf{E}[\mathrm{var}(\Phi_t \mid \Phi_{t-1})] + \mathrm{var}(\lambda\Phi_{t-1}) \\
&= \mathbf{E}[\mathrm{var}(\Phi_t \mid \Phi_{t-1})] + \lambda^2 \mathrm{var}(\Phi_{t-1}) \ \leq\ \rho + \lambda^2 \mathrm{var}(\Phi_{t-1}) \\
&\leq \sum_{i=0}^{t-1} \lambda^{2i}\rho \ <\ \rho/(1-\lambda^2) \ =\ \nu.
\end{aligned} \tag{2}$$

Instead of (2), Wilson uses $\nu = R/2\gamma$, where $\gamma = 1 - \lambda$ and $\mathbf{E}[(\Phi_t - \Phi_{t-1})^2 \mid \Phi_{t-1}] < R$. The calculation to justify this is longer, and the conclusion is not valid for all λ. However, since $\rho < R$ and usually $\lambda = o(1)$, (2) implies Wilson's bound asymptotically, but in general they are incomparable. Now, using Chebyshev's inequality,

$$\Pr\left(\Phi_t < \lambda^t\Phi_0 - \sqrt{2\nu/\varepsilon}\right) < \tfrac{1}{2}\varepsilon, \quad \text{and} \quad \Pr\left(\Phi_\infty > \sqrt{2\nu/\varepsilon}\right) < \tfrac{1}{2}\varepsilon.$$

Thus $\mathrm{d_{TV}}(\Phi_t, \Phi_\infty) \leq 1 - \varepsilon$ only if $\lambda^t\Phi_0 < 2\sqrt{2\nu/\varepsilon}$. (We will abuse the notation $\mathrm{d_{TV}}(\cdot,\cdot)$ for variation distance by extending it to random variables.) For $\varepsilon \leq \frac{1}{8}$, this holds only if

$$t > \frac{\ln\left(\sqrt{\varepsilon/8}\,\Phi_0/\sqrt{\nu}\right)}{\ln(1/\lambda)} \geq \frac{\lambda\ln(\varepsilon\Phi_0/\sqrt{\nu}\,)}{1-\lambda}.$$

[2] If $A - I$ is invertible then an affine dependence $\mathbf{E}[X(1)] = AX(0) + b$ can be reduced to this by moving the origin in Υ. In particular, $\mathbf{E}[X(1)] = AX(0) + b$ is the same as $\mathbf{E}[X(1) + c] = A(X(0) + c)$ for $b = (A - I)c$.

Setting $\varepsilon' = 1 - \varepsilon$, we find that for $\varepsilon' \geq \frac{7}{8}$,

$$\mathrm{Mix}(\mathcal{M}, \varepsilon') \geq \frac{\lambda \ln((1-\varepsilon')\Phi_0/\sqrt{\nu}\,)}{1-\lambda}. \tag{3}$$

We say that a Markov chain is *monotone* with respect to a partial order $\leq$ on its state space if two realisations $X(t)$ and $Y(t)$ of it may be coupled so that $X(0) \geq Y(0)$ entails $X(t) \geq Y(t)$ for all $t \in \mathbb{N}$. Suppose $\mathcal{M}$ is monotone with respect to the product partial order $\leq$ on $\mathbb{R}^m$, i.e., the partial order defined by $x \leq y$ if and only if $x_i \leq y_i$ for all $i \in \{1, \ldots, m\}$. If the weight vector $w > 0$ (in the product order), we can use it to bound the mixing time from above. Let us re-scale w so that $\min_i w_i = 1$. Let $d(x, y) = \sum_{i=1}^m w_i|x_i - y_i|$ for $x, y \in \Upsilon$. Then d is a metric, since $w > 0$. The mixing time of $\mathcal{M}$ is bounded by the monotone coupling time for two copies $X(t)$, $Y(t)$, with $X(0) = \top$ and $Y(0) = \bot$, where $\top$ and $\bot$ denote the maximum and minimum states in the partial order, assumed unique. Then, since $X(t) \geq Y(t)$,

$$\begin{aligned}\mathbf{E}[d(X(t), Y(t))] &= \mathbf{E}[w(X(t) - Y(t))] = wA(X(t-1) - Y(t-1)) \\ &= \lambda\, w(X(t-1) - Y(t-1)) = \lambda\, d(X(t-1), Y(t-1)).\end{aligned}$$

So

$$\mathrm{d_{TV}}(X(t), Y(t)) \leq \Pr[X(t) \neq Y(t)] \leq \mathbf{E}[d(X(t), Y(t))] \leq \lambda^t d(X(0), Y(0)) \leq 2\lambda^t \Phi_0 \leq \varepsilon \tag{4}$$

when $t \geq \ln(2\Phi_0/\varepsilon)/\ln(1/\lambda)$, which holds if

$$t \geq \ln(2\Phi_0/\varepsilon)/(1-\lambda).$$

Thus,

$$\mathrm{Mix}(\mathcal{M}, \varepsilon) \leq \ln(2\Phi_0/\varepsilon)/(1-\lambda). \tag{5}$$

When ν is sufficiently small with respect to Φ_0, the upper bound (5) and the lower bound (3) agree to within a constant factor for the time to reach variation distance 7/8, say.

3.1 Bounding $\mathbf{E}[\mathrm{var}(\Phi_t \mid \Phi_{t-1})]$

In order to use the technique in Section 3, we have to find a ρ such that $\mathbf{E}[\mathrm{var}(\Phi_t \mid \Phi_{t-1})] \leq \rho$, where the expectation is over Φ_{t-1}.

For this, let $Z_t = \Phi_t - \Phi_{t-1}$. Then

$$\begin{aligned}\mathbf{E}[\mathrm{var}(\Phi_t \mid \Phi_{t-1})] &= \mathbf{E}[\mathrm{var}(\Phi_{t-1} + Z_t \mid \Phi_{t-1})] \\ &= \mathbf{E}[\mathrm{var}(Z_t \mid \Phi_{t-1})] \\ &= \mathbf{E}\left[\mathbf{E}[Z_t^2 \mid \Phi_{t-1}] - (\mathbf{E}[Z_t \mid \Phi_{t-1}])^2\right] \\ &\leq \mathbf{E}\left[\mathbf{E}[(\Phi_t - \Phi_{t-1})^2 \mid \Phi_{t-1}]\right] \\ &\leq \max_{\Phi_{t-1}} \mathbf{E}[(\Phi_t - \Phi_{t-1})^2 \mid \Phi_{t-1}].\end{aligned}$$

We will use the above sequence of inequalities to find a suitable ρ.

3.2 A benchmark example

We illustrate the technique by applying it to a simple example whose analysis was also given in the Introduction to [7]. Consider mixing on the cube $\{-1,+1\}^m$ of the chain which changes the sign of a uniform random coordinate with probability $\frac{1}{2}$. Then

$$\mathbf{E}[X_i(t+1)] = (1-1/m)X_i(t),$$

so $A = (1-1/m)I$, and all its eigenvalues are equal to $(1-1/m)$. We may choose an arbitrary w, say the vector $(1,1,\ldots,1)$. Then we can take $\rho = 2$, so $\nu = 2m/(2-1/m) \leq 2m$. Taking $X(0) = w$, $\Phi_0 = m$, and for constant ε, the lower bound (3) for mixing time is $\frac{1}{2}m\ln m + O(m)$. This chain is monotone, and the upper bound (5) is $m\ln m + O(m)$.

4 Glauber dynamics for $q = 3$ mixes in $\Theta(n^3 \log n)$ updates

Let G be a path going left-to-right from vertex 1 to vertex n. Recall that Ω is the set of all proper colourings of G.

4.1 Analysis of a related Markov chain

Let σ be a colouring in Ω. Note that for every $i \in \{1,\ldots,n-1\}$ we either have $\sigma_{i+1} = \sigma_i + 1 \pmod 3$ or $\sigma_{i+1} = \sigma_i - 1 \pmod 3$. We can associate σ with a vector $X \in \Upsilon = \{-1,1\}^{n-1}$. X_i is 1 if $\sigma_{i+1} = \sigma_i + 1 \pmod 3$ and $X_i = -1$ otherwise. (Note that three colourings are mapped to the same configuration X — given σ_1 and X, the colouring σ can be recovered.)

The Markov chain $\mathcal{M}_{\mathrm{Gl}}$ can be associated with a Markov chain $\mathcal{M}_{\mathrm{Gl}}^{\pm}$ on Υ. The moves of $\mathcal{M}_{\mathrm{Gl}}^{\pm}$ (from configuration X) are as follows. Choose $r \in \{1,\ldots,n\}$ uniformly at random. If $r = 1$ or $r = n$ then, with probability $\frac{1}{3}$, change the sign of X_1 or X_{n-1} respectively. Otherwise exchange X_{r-1} and X_r with probability $\frac{1}{3}$. In all other cases, do nothing.

In this section, we analyse the mixing rate of $\mathcal{M}_{\mathrm{Gl}}^{\pm}$. Note that this chain is monotone with respect to the usual partial order on Υ. Then straightforward calculations give

$$\mathbf{E}[X(t+1)] = AX(t), \quad \text{where } A = I - \frac{1}{3n}B,$$

and

$$B = \begin{bmatrix} 3 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & & \ddots & & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 3 \end{bmatrix}.$$

Note that A is symmetric so has all real eigenvalues. Moreover A is non-negative and irreducible, so its largest eigenvalue λ has a positive eigenvector w. The eigenvectors are identical to those of B, and $\lambda = (1-\lambda'/3n)$, where λ' is the smallest eigenvalue of B. The "generic" row gives the equation

$$-w_{i-1} + 2w_i - w_{i+1} = \lambda' w_i, \tag{6}$$

the form of which suggests a simple harmonic oscillation. So we will try the solution $w_i = \sin(\alpha i + \beta)$. Substituting in (6) gives $\lambda' = 2(1 - \cos\alpha) = 4\sin^2(\alpha/2)$. We also have the two "boundary conditions"

$$3w_1 - w_2 = \lambda' w_1, \quad -w_{n-2} + 3w_{n-1} = \lambda' w_{n-1}. \tag{7}$$

The first equation in (7) gives $\sin(\alpha + \beta) = -\sin\beta$, i.e. $\beta = -\alpha/2$. The second then gives $\sin((n - \frac{1}{2})\alpha) = -\sin((n - \frac{3}{2})\alpha)$, so $(n - \frac{1}{2})\alpha = 2\pi - (n - \frac{3}{2})\alpha$, i.e. $\alpha = \pi/(n-1)$. Thus

$$w_i = \sin\left(\frac{\pi(i - \frac{1}{2})}{n - 1}\right) > 0 \qquad (i = 1, \ldots, n-1),$$

so w is the eigenvector corresponding to the largest eigenvalue. Now, if we let $w_0 = w_n = 0$, we may take

$$\rho = 2 \max_{1 \le i \le n} (w_i - w_{i-1})^2 = 2(w_2 - w_1)^2 \sim 2\pi^2/n^2.$$

Also $\lambda = 1 - 4\sin^2(\pi/(2n-2))/3n$, so $1 - \lambda \sim \pi^2/3n^3$. Hence $\nu \sim 3n$. Taking $X(0)$ to be the all 1's vector,

$$\Phi_0 = \sum_{i=1}^{n-1} \sin\left(\frac{\pi(i - \frac{1}{2})}{n-1}\right) = \operatorname{Im}\left[\sum_{j=1}^{n-1} \exp\left(\frac{\mathbf{i}\pi(j - \frac{1}{2})}{n-1}\right)\right] = \operatorname{cosec}\left(\frac{\pi}{2(n-1)}\right) \sim \frac{2n}{\pi},$$

where $\mathbf{i} = \sqrt{-1}$ in the second equality and the final equality follows from simplifying the geometric series as follows. For $\ell = i\pi/(n-1)$, the sum is equal to

$$e^{-\ell/2}\left(\frac{e^{\ell} - e^{\ell n}}{1 - e^{\ell}}\right) = \frac{-2}{\exp\left(\frac{i\pi}{2(n-1)}\right) - \exp\left(\frac{-i\pi}{2(n-1)}\right)} = \frac{i}{\sin\left(\frac{\pi}{2(n-1)}\right)}.$$

Now for any fixed $\varepsilon \ge 7/8$, the mixing time lower bound (3) is $\mathrm{Mix}(\mathcal{M}^{\pm}_{\mathrm{Gl}}, \varepsilon) \ge \frac{3n^3}{2\pi^2}\ln n + O(n^3)$. Also (for any positive ε) the upper bound (5) is $\mathrm{Mix}(\mathcal{M}^{\pm}_{\mathrm{Gl}}, \varepsilon) \le 3\pi^{-2} n^3 \ln n + O(n^3)$. Thus the mixing time of $\mathcal{M}^{\pm}_{\mathrm{Gl}}$ is $\Theta(n^3 \log n)$.

4.2 Distance measures and a lower bound for Glauber dynamics

We will use two distance measures to analyse the Glauber-dynamics Markov chain $\mathcal{M}_{\mathrm{Gl}}$. First, we define the distance $d_1(\sigma, \tau)$ for $\sigma \in \Omega$ and $\tau \in \Omega$ is follows. Let X be the member of Υ associated with σ and Y be the member of Υ associated with τ. Let $d_1(\sigma, \tau) = \mathrm{Ham}(X, Y)$, where $\mathrm{Ham}(X, Y)$ is Hamming distance between X and Y as in Section 2.

Using distance measure d_1, the lower bound from Section 4.1 applies directly to $\mathcal{M}_{\mathrm{Gl}}$. Thus, we obtain the following theorem.

Theorem 1. *Let G be the n-vertex path. Let $q = 3$. Consider the Markov chain $\mathcal{M}_{\mathrm{Gl}}$ on the state space Ω. For any fixed $\varepsilon \ge 7/8$, $\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}}, \varepsilon) \ge \frac{3n^3}{2\pi^2}\ln n + O(n^3)$.*

In order to upper-bound the mixing time of $\mathcal{M}_{\mathrm{Gl}}$, we will also define a second distance measure.

Give the vertices $1, \dots, n$ weights $\lambda_1, \dots, \lambda_n$ respectively. These weights are positive rationals. If $(\sigma, \tau) \in S$ differs on the colour of vertex i then let $\phi(\sigma, \tau) = \lambda_i$. Define the function d_2 on $\Omega \times \Omega$ as follows. For each pair $(\sigma, \tau) \in \Omega \times \Omega$, let

$$d_2(\sigma, \tau) = \min_{\omega(0), \dots, \omega(k)} \sum_{j=0}^{k-1} \phi(\omega(j), \omega(j+1)), \tag{8}$$

where the minimum is over all paths $\sigma = \omega(0), \dots, \omega(k) = \tau$ such that each $\omega_j \in \Omega$ and each pair $(\omega(j), \omega(j+1)) \in S$. A path $\omega(0), \dots, \omega(k)$ satisfying (8) is referred to as a *geodesic* path from σ to τ.

In our couplings, we will want to be able to bound the expected change in the distance d_2. In order to do this, we use height functions. A *height function* h corresponding to a proper colouring σ is a vector in $\mathbb{Z}^n$ satisfying the following properties.

1. For every vertex i, $h_i \equiv i \pmod 2$.
2. For every vertex i, $h_i \equiv \sigma_i \pmod 3$.
3. For every edge $(i, i+1)$, $|h_i - h_{i+1}| = 1$.

Let $\mathcal{H}(\sigma)$ denote the set of height functions corresponding to colouring σ. The distance between two height functions, h and h^*, is given by

$$d(h, h^*) = \sum_{i \in \{1, \dots, n\}} \frac{|h_i - h_i^*| \, \lambda_i}{2}.$$

Lemma 2. *For any pair of colourings $(\sigma, \tau) \in \Omega \times \Omega$,*

$$d_2(\sigma, \tau) = \min_{\substack{h \in \mathcal{H}(\sigma), \\ h^* \in \mathcal{H}(\tau)}} d(h, h^*).$$

Proof. To show that

$$d_2(\sigma, \tau) \geq \min_{\substack{h \in \mathcal{H}(\sigma), \\ h^* \in \mathcal{H}(\tau)}} d(h, h^*),$$

consider a geodesic path from σ to τ. Let $h'(0)$ be any height function in $\mathcal{H}(\sigma)$ and let $h'(0), \dots, h'(k)$ be the sequence of height functions corresponding to the geodesic path. Now

$$\min_{\substack{h \in \mathcal{H}(\sigma), \\ h^* \in \mathcal{H}(\tau)}} d(h, h^*) \leq d(h'(0), h'(k)) \leq \sum_{i=0}^{k-1} d(h'(i), h'(i+1)) = d_2(\sigma, \tau).$$

To show that the

$$d_2(\sigma, \tau) \leq \min_{\substack{h \in \mathcal{H}(\sigma), \\ h^* \in \mathcal{H}(\tau)}} d(h, h^*),$$

consider any $h \in \mathcal{H}(\sigma)$ and $h^* \in \mathcal{H}(\tau)$. A "height-function transformation" (see [19]) either takes a local maximum of a height function and pushes it down by two or takes a local minimum and pushes it up by two. We can show that there is a sequence $h = h(0), \dots, h(k) =$

h^* of height-function transformations transforming h into h^* that chooses each vertex v only $|h_v - h_v^*|/2$ times. (This can be proved by induction on $\sum_v |h_v - h_v^*|$. See Lemma 4.3 of [19].) Now let $\omega(0), \ldots, \omega(k)$ be the sequence of colourings corresponding to $h(0), \ldots, h(k)$. Note that

$$\sum_{j=0}^{k-1} \phi(\omega(j), \omega(j+1)) \leq d(h, h^*).$$

Thus, $d_2(\sigma, \tau) \leq d(h, h^*)$. □

4.3 An upper bound for Glauber dynamics

Suppose $(\sigma(0), \tau(0)) \in \Omega \times \Omega$. Then

$$\Pr(d_1(\sigma(t), \tau(t)) \geq 1) \leq \mathbf{E}(d_1(\sigma(t), \tau(t))).$$

Applying Equation (4) to the analysis in Section 4.1, this is at most $2\lambda^t \Phi_0$ where $\Phi_0 \in \Theta(n)$ and $1 - \lambda \sim \pi^2/2n^3$. So for some $t' = O(n^3 \log n)$, we will have $d_1(\sigma(t'), \tau(t')) = 0$ with probability at least $\frac{39}{40}$. By Lemma 2, $d_1(\sigma(t'), \tau(t')) = 0$ implies that $d_2(\sigma(t'), \tau(t')) = 0$ or $d_2(\sigma(t'), \tau(t')) = \sum_{i \in \{1,\ldots,n\}} \lambda_i$.

Now choose weights $\lambda_1 = \lambda_n = 1/2$ and $\lambda_2 = \cdots = \lambda_{n-1} = 1$. We use path coupling on pairs $(\sigma(0), \tau(0)) \in S$. Start with such a pair, run t' steps to get $(\sigma(t'), \tau(t'))$. With probability at least 39/40, $d_1(\sigma(t'), \tau(t')) = 0$. Either $\sigma(t') = \tau(t')$ or $d_2(\sigma(t'), \tau(t')) = n - 1$. Suppose the latter. We now carry on from $(\sigma(t'), \tau(t'))$ using the identity coupling. We will show in Section 4.3.1 below that if we take any $(\sigma, \tau) \in \Omega \times \Omega$ and produce (σ', τ') by one step of the identity coupling then $\mathbf{E}[d_2(\sigma', \tau')] \leq d_2(\sigma, \tau)$. Thus $D_t = d_2(\sigma(t' + t), \tau(t' + t))$ is a super-martingale with $D_0 = n - 1$. In Section 4.3.2 below, we will define a quantity $V \in \Omega(1/n)$ and we will show that for all t, $\mathbf{E}[(D_{t+1} - D_t)^2 | D_t] \geq V$. Let $B = 10n$. Consider the process which is D_t until either (a) $D_t = 0$ (i.e. coupling occurs) or (b) $D_t = B$. Let T be the stopping time at which (a) or (b) first occurs, and suppose that D_t is constant for $t \geq T$. It is not difficult to show (see [23]) that $(B - D_t)^2 - V \min(t, T)$ is a sub-martingale. Let p be the probability that (a) occurs. Then, using the optional stopping theorem twice,

$$(1 - p)B \leq D_0 \quad \text{and} \quad pB^2 - V\,\mathbf{E}[T] \geq (B - D_0)^2 > 0.$$

From the first we get $p \geq 1 - D_0/B \geq \frac{9}{10}$. From the second we get $\mathbf{E}[T] \leq pB^2/V$. If T_1 is the stopping time conditional on (a) occurring, it follows that $\mathbf{E}[T_1] \leq B^2/V$. Hence $\Pr(T_1 > 10B^2/V) \leq \frac{1}{10}$. So, if we now run the d_2 coupling for (at most) $10B^2/V = O(n^3)$ steps, then σ and τ will fail to couple with probability at most $\frac{1}{40} + 2 \times \frac{1}{10} < \frac{1}{4}$. Thus we have shown the following.

Theorem 3. *Let G be the n-vertex path. Let $q = 3$. Consider the Markov chain $\mathcal{M}_{\mathrm{Gl}}$ on the state space Ω.* $\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}}, \frac{1}{4}) \leq O(n^3 \log n)$.

We can boost the coupling probability in the usual way to bound $\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}}, \varepsilon)$ for $\varepsilon \leq 1/4$.

4.3.1 The coupling breaks even

Recall from Section 4.3 that $\lambda_1 = \lambda_n = \frac{1}{2}$ and $\lambda_2 = \cdots = \lambda_{n-1} = 1$.

Lemma 4. *Suppose* $(\sigma,\tau)\in S$ *differs at vertex* i. *Obtain* (σ',τ') *by one step of the identity coupling. Then* $\mathbf{E}[d_2(\sigma',\tau')]\le d_2(\sigma,\tau)$.

Proof. Recall that $n>3$. There are three cases.

Suppose $i\in\{1,n\}$. Then $\mathbf{E}[d_2(\sigma',\tau')]-\frac{1}{2}$ is equal to

$$-\frac{2}{3n}\lambda_1+\frac{1}{3n}\lambda_2=0.$$

The first term in the sum comes from the two colours which could be chosen at vertex i, causing coupling. The second term comes from the one bad colour which could be chosen at i's neighbour, causing one of the height functions to change by 2.

Suppose $i\in\{2,n-1\}$. Then $\mathbf{E}[d_2(\sigma',\tau')]-1$ is equal to

$$\frac{2}{3n}\lambda_1-\frac{2}{3n}\lambda_2+\frac{1}{3n}\lambda_3=0.$$

Suppose $i\in\{3,\ldots,n-2\}$. Then $\mathbf{E}[d_2(\sigma',\tau')]-1$ is equal to

$$\frac{1}{3n}\lambda_{i-1}-\frac{2}{3n}\lambda_i+\frac{1}{3n}\lambda_{i+1}=0.$$

□

We can conclude from Lemma 4 by path-coupling that if we take any $(\sigma,\tau)\in\Omega\times\Omega$ and produce (σ',τ') by one step of the identity coupling then $\mathbf{E}[d_2(\sigma',\tau')]\le d_2(\sigma,\tau)$.

4.3.2 Lower bounding V

Let $w=\min_i\lambda_i$. Start with σ and τ such that $\sigma\ne\tau$. We will identify a vertex z and a colour C such that, if we obtain σ' from σ by trying C at z and we obtain τ' from τ by trying C at z then $d_2(\sigma',\tau')\le d_2(\sigma,\tau)-w$. Since (z,C) is chosen with probability $1/(3n)$, we get $V=w^2/(3n)$.

Our method is this. Given σ and τ, choose $h\in\mathcal{H}(\sigma)$ and $h^*\in\mathcal{H}(\tau)$ such that $d_2(\sigma,\tau)=d(h,h^*)$. Construct h' from h by applying the choice (z,C) in h and construct h'^* from h^* by applying the choice (z,C) in h^*. We will show that $d(h',h'^*)\le d(h,h^*)-w$ so $d_2(\sigma',\tau')\le d(h',h'^*)\le d_2(\sigma,\tau)-w$.

Without loss of generality, assume that there is a vertex v such that $h_v>h^*_v$. Let $m=\max_v h_v-h^*_v$ and let $R=\{v\mid h_v-h^*_v=m\}$. By construction, R is non-empty.

Case 1: R **is the whole line**

Let z be any local maximum in h and let C be the colour that is not used at z or at its neighbours in h. z is also a local maximum in h^* (since R is the whole line) but C is used either at z or at its neighbours in h^* (since $\sigma\ne\tau$). Choose (z,C). Then $h'_z=h_z-2$. But $h'^*_z=h^*_z$. So $d(h,h^*)-d(h',h'^*)=\lambda_z$.

Case 2: There is a vertex $z\in R$**, all of whose neighbours are in** $\overline{R}$**.**

Note that all edges from z to $\overline{R}$ in h go down (that is, height decreases along these edges). Also, all edges from z to $\overline{R}$ in h^* go up. Thus, z is a local maximum in h and a local minimum in h^*. Let C be the colour that is not used at z or at its neighbours in h. Choose (z,C). Then $h'_z=h_z-2$. Since z is a local minimum in h^*, $h'^*_z\ge h^*_z$. Also, $h'_z\ge h'^*_z$ since we choose the same colour in both copies. Thus, $d(h,h^*)-d(h',h'^*)\ge\lambda_z$.

Case 3: There is a vertex $z \in R$ which has a neighbour $w \in \overline{R}$ and a neighbour $r \in R$.
Note that the edge from z to r goes the same direction (up or down) in h as in h^*. Suppose first that it goes down. Then z is a local maximum in h. Let C be the colour that is not used at z or at its neighbours in h. Choose (z, C). Then $h'_z = h_z - 2$. Also, $h'^*_z = h^*_z$ (since z has a neighbour below and a neighbour above, and won't be recoloured in h^*). Thus, $d(h, h^*) - d(h', h'^*) \geq \lambda_z$.

Suppose instead that the edge from z to r goes up. Then z is a local minimum in h^*. Let C be the colour that is not used at z or at its neighbours in h^*. Choose (z, C). Then $h'^*_z = h^*_z + 2$ and $h'_z = h_z$ so $d(h, h^*) - d(h', h'^*) \geq \lambda_z$.

5 Systematic scan for $q = 3$ mixes in $\Theta(n^2 \log n)$ sweeps

As in Section 4, we consider the path G with vertices 1 through n with $q = 3$ colours. We consider the dynamics $\mathcal{M}_{\rightarrow}$.

5.1 Analysis of a related Markov chain

As in Section 4.1, the Markov chain $\mathcal{M}_{\rightarrow}$ can be associated with a Markov chain $\mathcal{M}^{\pm}_{\rightarrow}$ on Υ. Each move of $\mathcal{M}^{\pm}_{\rightarrow}$ starts with a configuration $X \in \Upsilon$ and makes n moves of the chain $\mathcal{M}^{\pm}_{\text{Gl}}$ from Section 4.1 corresponding to the choices $r = 1, r = 2, \ldots, r = n$ (in order).

Consider the transition from configuration X to configuration X' corresponding to one step of $\mathcal{M}^{\pm}_{\rightarrow}$. Let $\tilde{X}_i$ denote the colour of vertex i in the intermediate configuration which is obtained after the choices $r = 1, r = 2, \ldots, r = i$. Then

$$\begin{aligned}
\mathbf{E}[\,\tilde{X}_1\,] &= \tfrac{1}{3} X_1, \\
\mathbf{E}[\,\tilde{X}_i\,] &= \tfrac{2}{3} X_i + \tfrac{1}{3}\,\mathbf{E}[\,\tilde{X}_{i-1}] \qquad (i = 2, \ldots, n-1), \\
\mathbf{E}[\,X'_i\,] &= \tfrac{2}{3}\,\mathbf{E}[\tilde{X}_i] + \tfrac{1}{3} X_{i+1} \qquad (i = 1, \ldots, n-2), \\
\mathbf{E}[X'_{n-1}] &= \tfrac{1}{3}\,\mathbf{E}[\tilde{X}_{n-1}].
\end{aligned}$$

Solving these gives

$$\begin{aligned}
\mathbf{E}[\,X'_1\,] &= \tfrac{2}{9} X_1 + \tfrac{1}{3} X_2, \\
\mathbf{E}[\,X'_i\,] &= \tfrac{2}{3^{i+1}} X_1 + \sum_{j=2}^{i} \tfrac{4}{3^{i+2-j}} X_j + \tfrac{1}{3} X_{i+1} \qquad (i = 2, \ldots, n-2), \\
\mathbf{E}[X'_{n-1}] &= \tfrac{1}{3^n} X_1 + \sum_{j=2}^{n-2} \tfrac{2}{3^{n+1-j}} X_j + \tfrac{2}{9} X_{n-1}.
\end{aligned}$$

So the matrix

$$A = \begin{bmatrix} \frac{2}{9} & \frac{1}{3} & 0 & & & & \\ \frac{2}{27} & \frac{4}{9} & \frac{1}{3} & 0 & & & \\ \frac{2}{81} & \frac{4}{27} & \frac{4}{9} & \frac{1}{3} & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \\ \vdots & \vdots & \vdots & & \ddots & \ddots & 0 \\ \frac{2}{3^{n-1}} & \frac{4}{3^{n-2}} & \frac{4}{3^{n-3}} & \frac{4}{3^{n-4}} & \cdots & \frac{4}{9} & \frac{1}{3} \\ \frac{1}{3^n} & \frac{2}{3^{n-1}} & \frac{2}{3^{n-2}} & \frac{2}{3^{n-3}} & \cdots & \frac{2}{27} & \frac{2}{9} \end{bmatrix}$$

Here A is not symmetric, but is non-negative and irreducible, so has a positive eigenvector w corresponding to its (real) largest eigenvalue λ. Now w, λ satisfy the equations

$$\begin{aligned} \lambda w_1 &= \tfrac{2}{9} w_1 + \sum_{j=2}^{n-2} \tfrac{2}{3^{j+1}} w_j + \tfrac{1}{3^n} w_{n-1} \\ \lambda w_i &= \tfrac{1}{3} w_{i-1} + \sum_{j=i}^{n-2} \tfrac{4}{3^{j-i+2}} w_j + \tfrac{2}{3^{n-i+1}} w_{n-1} \qquad (i = 2, \ldots, n-2), \\ \lambda w_{n-1} &= \tfrac{1}{3} w_{n-2} + \tfrac{2}{9} w_{n-1}. \end{aligned}$$

These can be simplified by subtracting one-third of the $(i+1)$th equation from the ith for $i = 2, \ldots, n-2$, and one-sixth the second from the first, giving

$$\lambda w_2 - (6\lambda - 1) w_1 = 0, \tag{9}$$

$$\lambda w_{i+1} - (3\lambda - 1) w_i + w_{i-1} = 0 \qquad (i = 2, \ldots, n-2), \tag{10}$$

$$-3 w_{n-2} + (9\lambda - 2) w_{n-1} = 0. \tag{11}$$

If λ is close to 1, the form of (10) suggests a slightly damped harmonic oscillation, so we will try a solution of the form $e^{\gamma i} \sin(\alpha i + \beta)$. Substituting this in (10) and equating coefficients of $\sin(\alpha i + \beta)$, $\cos(\alpha i + \beta)$ gives

$$\lambda = e^{-2\gamma} \quad \text{and} \quad \cos\alpha = (3e^{-\gamma} - e^{\gamma})/2, \quad \text{i.e. } e^{\gamma} = \sqrt{3 + \cos^2\alpha} - \cos\alpha. \tag{12}$$

(The second of these follows from $\sin(x+y) = \sin x \cos y + \cos x \sin y$ and $\sin(x-y) = \sin x \cos y - \cos x \sin y$ and the third used the quadratic formula with the choice $\cos\alpha \geq 0$.) Then (9) and (11) give

$$\frac{\sin(2\alpha + \beta)}{\sin(\alpha + \beta)} = \cos\alpha + \cot(\alpha + \beta) \sin\alpha = 6e^{-\gamma} - e^{\gamma} \tag{13}$$

$$\frac{\sin((n-2)\alpha + \beta)}{\sin((n-1)\alpha + \beta)} = \cos\alpha - \cot((n-1)\alpha + \beta) \sin\alpha = \frac{9e^{-\gamma} - 2e^{\gamma}}{3}. \tag{14}$$

Using (12) to eliminate γ in (13) and (14) gives

$$\tan(\alpha + \beta) = \frac{\sin\alpha}{2\cos\alpha + \sqrt{3 + \cos^2\alpha}}, \quad \tan((n-1)\alpha + \beta) = \frac{-3\sin\alpha}{2\cos\alpha + \sqrt{3 + \cos^2\alpha}},$$

implying

$$\tan((n-1)\alpha + \beta) = -3\tan(\alpha + \beta), \quad \tan\alpha = 4\tan(\alpha + \beta)/(1 - 3\tan^2(\alpha + \beta)). \tag{15}$$

To see the second of these equalities, note that the left-hand equation on the previous line is equivalent to

$$\tan(\alpha+\beta) = \frac{\sin\alpha}{2\cos\alpha + \sqrt{4\cos^2\alpha + 3\sin^2\alpha}}$$

using $cos^2\alpha + sin^2\alpha = 1$. But this is equal to

$$\frac{\tan\alpha}{2+\sqrt{4+3\tan^2\alpha}}.$$

Now solve this for $\tan\alpha$. The equalities in (15) imply

$$\pi - (n-1)\alpha - \beta = \arctan(3\tan(\alpha+\beta)), \alpha = \alpha + \beta + \arctan(3\tan(\alpha+\beta)).$$

The first of these uses $\tan(\pi - x) = -\tan(x)$ and the second uses $\tan(x+y) = (\tan x + \tan y)/(1 - \tan x \tan y)$ with $x = \alpha+\beta$ and $y = \arctan(3\tan(\alpha+\beta))$. So finally we have

$$\alpha = \frac{\pi}{n-1}, \quad \tan\beta = -3\tan\left(\beta + \frac{\pi}{n-1}\right) \quad e^\gamma = \sqrt{3+\cos^2\left(\frac{\pi}{n-1}\right)} - \cos\left(\frac{\pi}{n-1}\right).$$

Note that β is the solution of a trigonometric equation, but it is easily checked that $-\pi/(n-1) < \beta < 0$. Hence $w > 0$, corresponding to the largest eigenvalue λ of A. Asymptotically we have

$$\alpha \sim \pi/n, \quad \beta \sim -3\pi/4n, \quad \gamma \sim \pi^2/4n^2, \quad \text{so } 1-\lambda \sim \pi^2/2n^2. \tag{16}$$

If we take X_0 to be the all 1's vector, then it is easy to check that $\Phi_0 \sim 2n/\pi$. Clearly we can take ρ to be n times the quantity used for $\mathcal{M}^{\pm}_{\mathrm{Gl}}$, i.e., $2\pi^2/n$. Hence we have $\nu \sim 2n$. Now, from (3), for fixed $\varepsilon \geq 7/8$, the lower bound is $\mathrm{Mix}(\mathcal{M}^{\pm}_{\rightarrow}, \varepsilon) \geq \pi^{-2}n^2 \ln n + O(n)$. Since the sweep is also monotone, (5) gives the upper bound $\mathrm{Mix}(\mathcal{M}^{\pm}_{\rightarrow}, \varepsilon) \leq 2\pi^{-2}n^2 \ln n + O(n)$. It may be observed that the bounds bounds for $\mathcal{M}^{\pm}_{\mathrm{Gl}}$ are both about n times these quantities, so there is no evidence that the scan gives a significant speed-up. However, there will be a considerable saving in random number generation.

5.2 A lower bound for systematic scan

We will use distance measures d_1 and d_2 from Section 4.2. Using distance measure d_1, the lower bound from Section 5.1 applies directly to $\mathcal{M}_{\rightarrow}$. Thus, we have

Theorem 5. *Let G be the n-vertex path. Let $q = 3$. Consider the Markov chain $\mathcal{M}_{\rightarrow}$ on the state space Ω. For any fixed $\varepsilon \geq 7/8$, $\mathrm{Mix}(\mathcal{M}_{\rightarrow}, \varepsilon) \geq \pi^{-2}n^2 \ln n + O(n)$.*

5.3 An upper bound for systematic scan

As in Section 4.3, we find that for some $t' = O(n^2 \log n)$, we will have $d_1(\sigma(t'), \tau(t')) = 0$ with probability at least $\frac{39}{40}$.

Now choose the following weights. Let $\lambda_1 = \frac{1}{4}$, $\lambda_2 = \cdots = \lambda_{n-1} = 1$ and $\lambda_n = \frac{3}{4}$.

We now use path coupling on pairs $(\sigma(0), \tau(0)) \in S$. Start with such a pair, run t' steps to get $(\sigma(t'), \tau(t'))$. With probability at least 39/40, $d_1(\sigma(t'), \tau(t')) = 0$. Either $\sigma(t') = \tau(t')$ or $d_2(\sigma(t'), \tau(t')) = n-1$. Suppose the latter. We now carry on from $(\sigma(t'), \tau(t'))$ using the identity coupling. We show in Section 5.3.1 that if we take any $(\sigma, \tau) \in \Omega \times \Omega$ and produce (σ', τ') by one scan using the identity coupling then $\mathbf{E}[d_2(\sigma', \tau')] \leq d_2(\sigma, \tau)$. Thus

$D_t = d_2(\sigma(t'+t), \tau(t'+t))$ is a super-martingale with $D_0 = n-1$. In Section 5.3.2, we define $V = 1/27$ and we show that for all t, $\mathbf{E}[(D_{t+1} - D_t)^2 | D_t] \geq V$. Let $B = 10n$. Consider the process which is D_t until either (a) $D_t = 0$ (i.e. coupling occurs) or (b) $D_t = B$. Let T be the stopping time at which (a) or (b) first occurs, and suppose that D_t is constant for $t \geq T$. It is not difficult to show (see [23]) that $(B - D_t)^2 - V \min(t, T)$ is a sub-martingale. Let p be the probability that (a) occurs. Then, using the optional stopping theorem twice,

$$(1-p)B \leq D_0 \quad \text{and} \quad pB^2 - V\,\mathbf{E}[T] \geq (B - D_0)^2 > 0.$$

From the first we get $p \geq 1 - D_0/B \geq \frac{9}{10}$. From the second we get $\mathbf{E}[T] \leq pB^2/V$. If T_1 is the stopping time conditional on (a) occurring, it follows that $\mathbf{E}[T_1] \leq B^2/V$. Hence $\Pr(T_1 > 10B^2/V) \leq \frac{1}{10}$. So, if we now run the d_2 coupling for (at most) $10B^2/V = O(n^2)$ steps, then σ and τ will fail to couple with probability at most $\frac{1}{40} + 2 \times \frac{1}{10} < \frac{1}{4}$. Thus we have shown the following.

Theorem 6. *Let G be the n-vertex path. Let $q = 3$. Consider the Markov chain $\mathcal{M}_{\rightarrow}$ on the state space Ω.* $\mathrm{Mix}(\mathcal{M}_{\rightarrow}, \frac{1}{4}) \leq O(n^2 \log n)$.

We can boost the coupling probability in the usual way to bound $\mathrm{Mix}(\mathcal{M}_{\rightarrow}, \varepsilon)$ for $\varepsilon \leq 1/4$.

5.3.1 The coupling breaks even

Recall that the vertices of the path G are labelled $1, \ldots, n$ going from left to right.

Lemma 7. *Suppose that σ and τ differ at vertex $i < n$ and agree to the right of vertex i. Obtain σ' and τ' by scanning left-to-right, starting at vertex $i+1$, doing the identity coupling. Then*

$$\mathbf{E}[d_2(\sigma', \tau')] - d_2(\sigma, \tau) \leq \frac{1}{2}.$$

Proof. Choose $h \in \mathcal{H}(\sigma)$ and $h^* \in \mathcal{H}(\tau)$ such that $d_2(\sigma, \tau) = d(h, h^*)$. Let h', h'^* be the transformed height functions produced by the scan. If $v = i + \ell$ for $\ell \in \{1, \ldots, n-i-1\}$, then the probability that vertex v is changed by the coupling is $(\frac{1}{3})^\ell$. If there is a change, then one of the height functions changes by 2, so the change in $d(h', h'^*)$ is 1. For $v = n$, the probability that v changes is $(\frac{1}{3})^{n-i-1}\frac{2}{3}$. The change to $d(h', h'^*)$ in this case is $\frac{3}{4}$. Thus,

$$\begin{aligned}
\mathbf{E}[d_2(\sigma', \tau')] &\leq \mathbf{E}[d(h', h'^*)] \\
&= d(h, h^*) + \sum_{\ell=1}^{n-i-1} \left(\frac{1}{3}\right)^\ell + \frac{2}{3}\frac{3}{4}\left(\frac{1}{3}\right)^{n-i-1} \\
&= d(h, h^*) + \frac{\frac{1}{3} - \left(\frac{1}{3}\right)^{n-i}}{1 - \frac{1}{3}} + \frac{1}{2}\left(\frac{1}{3}\right)^{n-i-1} \\
&= d(h, h^*) + \frac{1}{2}
\end{aligned}$$

□

Lemma 8. *Suppose that σ and τ differ only at vertex 1. Obtain σ' and τ' by scanning left-to-right, starting at vertex 1, doing the identity coupling. Then*

$$\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{4}.$$

Proof. With probability $\frac{2}{3}$, the first vertex agrees, so $\sigma' = \tau'$. With probability $\frac{1}{3}$, the first vertex is left unchanged. Thus (using Lemma 7), $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{3}(\frac{1}{4} + \frac{1}{2})$. □

Lemma 9. *Suppose that σ and τ differ only at vertex 2. Obtain σ' and τ' by scanning left-to-right, starting at vertex 1, doing the identity coupling. Then*

$$\mathbf{E}[d_2(\sigma', \tau')] \leq 1.$$

Proof. Say that σ starts 2 0 2 and τ starts 2 1 2. Consider the coupling of first vertex.

- With probability $\frac{2}{3}$:
  The first vertex is made to disagree, for example σ now starts 2 0 2 but τ starts 0 1 2.

  The coupling of the second vertex in this case is as follows.

  - With probability $\frac{1}{3}$:
    The second vertex is made to agree, for example both become 1. In this case, $\mathbf{E}[d_2(\sigma', \tau')] = \frac{1}{4}$.
  - With probability $\frac{2}{3}$:
    The second vertex is unchanged. In this case, $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{4} + 1 + \frac{1}{2} = \frac{7}{4}$.

- With probability $\frac{1}{3}$:
  The first vertex is unchanged. By analogy to the proof of Lemma 8, $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{3}(1 + \frac{1}{2}) = \frac{1}{2}$.

Adding it all up, $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{2}{3}\left(\frac{1}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{7}{4}\right) + \frac{1}{3} \cdot \frac{1}{2} = 1$. □

Lemma 10. *Let $2 < i < n$. Suppose that σ and τ differ only at vertex i. Obtain σ' and τ' by scanning left-to-right, starting at vertex 1, doing the identity coupling. Then*

$$\mathbf{E}[d_2(\sigma', \tau')] \leq 1.$$

Proof. Say that vertices $i - 2 \ldots i + 1$ of σ are 1 2 0 2 and of τ are 1 2 1 2.
Consider the coupling of vertex $i - 1$.

- With probability $\frac{1}{3}$:
  Vertex $i - 1$ is made to disagree, so σ now starts 1 2 0 2 but τ starts 1 0 1 2. By analogy with the proof of Lemma 9, $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{3} \cdot 1 + \frac{2}{3}(1 + 1 + \frac{1}{2}) = 2$.

- With probability $\frac{2}{3}$:
  Vertex $i - 1$ is unchanged. By analogy with the proof of Lemma 8, $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{3}(1 + \frac{1}{2}) = \frac{1}{2}$.

Adding it all up, $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{3} \cdot 2 + \frac{2}{3} \cdot \frac{1}{2} = 1$. □

Lemma 11. *Suppose that σ and τ differ only at vertex n. Obtain σ' and τ' by scanning left-to-right, starting at vertex 1, doing the identity coupling. Then*

$$\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{3}{4}.$$

Proof. Say that vertices $n - 2 \ldots n$ of σ are 0 2 0 and of τ are 0 2 1.
Consider the coupling of vertex $n - 1$:

- With probability $\frac{1}{3}$:
  Vertex $n-1$ is made to disagree, so σ now ends with 010 and τ with 021.

  The coupling of vertex n (the last) in this case is as follows.

  - With probability $\frac{1}{3}$:
    The last vertex is made 0, with resulting cost 1.
  - With probability $\frac{1}{3}$:
    The last vertex is unchanged, with resulting cost $1 + \frac{3}{4} = \frac{7}{4}$.
  - With probability $\frac{1}{3}$:
    The last vertex becomes 2 in σ, with resulting cost $1 + 2 \cdot \frac{3}{4} = \frac{5}{2}$. (The claimed final cost is witnessed by the sequence of transitions $0\,1\,2 \to 0\,1\,0 \to 0\,2\,0 \to 0\,2\,1$.)

- With probability $\frac{2}{3}$:
  Vertex $n-1$ is unchanged. Now with probability $\frac{2}{3}$, vertex n will agree and with probability $\frac{1}{3}$ it will be unchanged. Thus $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$.

Adding it all up, $\mathbf{E}[d_2(\sigma', \tau')] \leq \frac{1}{3}(\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{7}{4} + \frac{1}{3} \cdot \frac{5}{2}) + \frac{2}{3} \cdot \frac{1}{4} = \frac{3}{4}$. □

Lemma 8, 9, 10 and 11 show that if $(\sigma, \tau) \in S$ and we obtain σ' and τ' by scanning left-to-right, starting at vertex 1, doing the identity coupling. Then $\mathbf{E}[d_2(\sigma', \tau')] \leq d_2(\sigma, \tau)$. By path coupling, we find that if we take any $(\sigma, \tau) \in \Omega \times \Omega$ and we produce (σ', τ') by one scan using the identity coupling then $\mathbf{E}[d_2(\sigma', \tau')] \leq d_2(\sigma, \tau)$.

5.3.2 The coupling has enough variance (Lower bounding V)

Recall that $w = \min_i \lambda_i$. Suppose $\sigma \neq \tau$. In Section 4.3.2, we considered several cases. For each case, we identified a vertex z and a colour C such that, if we obtain $\sigma^{'}$ from σ by trying C at z and we obtain $\tau^{'}$ from τ by trying C at z then $d_2(\sigma', \tau') \leq d_2(\sigma, \tau) - w$.

In this section, we reconsider each case. Obtain σ^* and τ^* by scanning σ and τ left to right, using the identity coupling. For each case in Section 4.3.2, we prove the following.

- If $z > 1$ then there is a colour c_ℓ (depending only on σ_{z-1}, σ_z, τ_{z-1} and τ_z) such that choosing colour c_ℓ for vertex $z-1$ ensures $\sigma^*_{z-1} = \sigma_{z-1}$ and $\tau^*_{z-1} = \tau_{z-1}$. Actually, it is easy to see that c_ℓ exists – just take any colour in $\{\sigma_{z-1}, \sigma_z\} \cap \{\tau_{z-1}, \tau_z\}$.

- Suppose $z < n$. For any colour c there is a colour $c_r(c)$ (depending only on σ_{z-1}, σ_z, σ_{z+1}, τ_{z-1}, τ_z, τ_{z+1} and c) such that if we choose c_ℓ for vertex $z-1$, c for vertex z, and $c_r(c)$ for vertex $z+1$, then $\sigma^*_{z+1} = \sigma_{z+1}$ and $\tau^*_{z+1} = \tau_{z+1}$.

- There is a colour C' such that, if we obtain $\sigma^{'}$ from σ by trying C' at z and we obtain $\tau^{'}$ from τ by trying C' at z then $\sigma^{'}_z = \sigma_z$ and $\tau^{'}_z = \tau_z$.

This is enough to establish $V = 1/27$. We will consider the event that c_ℓ is chosen for $z-1$ and, whatever colour, c, is chosen for z, $c_r(c)$ is chosen for $z+1$. This event occurs with probability $1/9$. Conditioned on the fact that this event occurs, we can choose the colour c for vertex z *after* choosing all other colours. That is, the choice of c is independent of the rest of the scan. Let $\sigma^\dagger$ and $\tau^\dagger$ be random variables defined by a left-to-right scan of σ and τ which uses c_ℓ at $z-1$ and $c_r(c)$ at $z+1$ and misses out the re-colouring at z.

If $|d_2(\sigma^\dagger,\tau^\dagger) - d_2(\sigma,\tau)| \geq w/2$ then we choose colour C' for vertex z so $\sigma^* = \sigma^\dagger$ and $\tau^* = \tau^\dagger$. Otherwise, we choose colour C for vertex v so $d_2(\sigma^*,\tau^*) \leq d_2(\sigma^\dagger,\tau^\dagger) - w$. Either way, we get $|d_2(\sigma^*,\tau^*) - d_2(\sigma,\tau)| \geq w/2$.

Here are cases 1 and 2 from Section 4.3.2:

Case	$\sigma_{z-1}\cdots\sigma_{z+1}$	$\tau_{z-1}\cdots\tau_{z+1}$	C'	$c_r(0)$	$c_r(1)$	$c_r(2)$
1	010	121	1	0	1	2
1	010	202	0	0	1	2
2	010	101	0	0	0	2
2	010	212	1	0	1	2
2	010	020	0	0	0	2

In Case 3, say $\sigma_{z-1}\cdots\sigma_{z+1} = 010$ and $\tau_{z-1}\cdots\tau_{z+1}$ is monotonic. Then C' is any colour in $\{0,1\}$. $c_r(2)$ is any colour in $\{0,2\}\cap\{\tau_z,\tau_{z+1}\}$ and, for any $i \neq 2$, $c_r(i)$ is any colour in $\{0,1\}\cap\{\tau_z,\tau_{z+1}\}$.

The case where $\tau_{z-1}\cdots\tau_{z+1} = 101$ and $\sigma_{z-1}\cdots\sigma_{z+1}$ is monotonic is similar.

6 Optimal mixing of Glauber and scan when $q = 4$

6.1 Distance measures

In this section, G is the n-vertex path. We take the state space to be Ω^+ (i.e., all colourings, whether proper or not). The results that we get by analysing our Markov chains on state space Ω^+ also apply to the same chains with state space Ω — this is because the chains do not make transitions from states in Ω to states outside of Ω. (Thus, the stationary distribution is uniform on Ω – states in $\Omega^+ \setminus \Omega$ have zero measure.) We ought to note that, on the extra states in $\Omega^+ \setminus \Omega$, what we are calling a "Metropolis" update does not strictly fit the official definition. For example, with a natural definition of the "energy" of a colouring, and using the usual Metropolis filter, the transition $\ldots 0\,0\,1 \ldots \to \ldots 0\,1\,1 \ldots$ would occur with positive probability. Nevertheless, we disallow this transition on account of the adjacent colour 1 vertices in the final state. However, our version of Metropolis agrees with the usual one on the significant part of the state space, namely Ω.

6.2 Glauber with $q = 4$: $O(n \log n)$ updates suffice

We'll use Theorem 2.2 of [12].

Suppose $(\sigma,\tau) \in S$ differ on vertex i. Construct (σ',τ') from (σ,τ) by using the following coupling. Choose the same vertex v to recolour in σ and in τ. Choose the same colour in both copies unless $v \in \{i-1, i+1\}$. In that case, choose colour σ_i in one copy while choosing τ_i in the other (and choose the same colour otherwise).

We first show that the value β in Theorem 2.2 of [12] is 1. That is, we show that $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')] \leq 1$. Consider the choices made in σ. If we choose vertex $i-1$ and colour τ_i then the Hamming distance might go up by 1. Similarly, if we choose vertex $i+1$ and colour τ_i then the Hamming distance might go up by 1. If we choose vertex i and any of the (at least two) colours not in $\{\sigma_{i-1},\sigma_{i+1}\}$ then the Hamming distance goes down by 1. These are the only choices which can cause the distance to change.

Now consider a multi-step coupling from (σ,τ). Assume for now that $i \in \{3,\ldots,n-2\}$, so there are at least 2 vertices to the left of vertex i and at least 2 vertices to the right of vertex i. The other cases are easier and we will consider them later. Let c be a colour which is not in $\{\sigma_i, \tau_i\}$ (there are two such colours, but choose an arbitrary one and call it c). Let Ψ be the set containing the following 6 choices (in σ): choose i with any colour, choose $i-1$ with τ_i, or choose $i+1$ with τ_i. Let the stopping time T be the first time a choice from Ψ is made. (That is, a choice from Ψ is made in the transition from $(\sigma(T-1), \tau(T-1))$ to $(\sigma(T), \tau(T))$ where $(\sigma(0), \tau(0)) = (\sigma, \tau)$.)

Let Ξ be the set containing the following 14 choices (in σ): Choose $i-1$ with any colour besides τ_i. Choose $i+1$ with any colour besides τ_i. Choose $i-2$ with any colour. Choose $i+2$ with any colour. Let C be the set containing all $4n-20$ choices that are not in Ψ or Ξ. Let $z_1, \ldots, z_t$ denote the choices made (in σ) in the transitions $(\sigma(0), \tau(0)), (\sigma(1), \tau(1)), \ldots, (\sigma(t), \tau(t))$. We will say that the sequence $z_1, \ldots, z_t$ is *good* if the only choices in $\Xi \cup \Psi$ are

- For some $t_1 \in [1, t]$, z_{t_1} consists of vertex $i-2$ along with the "smallest" colour that is not in $\{c, \sigma_{i-3}(t_1 - 1), \sigma_{i-1}\}$, and
- for some $t_2 \in [t_1 + 1, t]$, z_{t_2} consists of vertex $i+2$ along with the "smallest" colour that is not in $\{c, \sigma_{i+3}(t_1 - 1), \sigma_{i+1}\}$, and
- for some $t_3 \in [t_2 + 1, t]$, $z_{t_3} = (i-1, c)$, and
- for some $t_4 \in [t_3 + 1, t]$, $z_{t_4} = (i+1, c)$.

Now,

$$\Pr(z_1, \ldots, z_{t-1} \text{ is good} \mid T = t) = \binom{t-1}{4} \left(\frac{1}{4n-6}\right)^4 \left(\frac{4n-20}{4n-6}\right)^{t-5}. \tag{17}$$

Let α be any positive constant which is at most $1/6$. Let δ be a positive constant, independent of n, such that, for all $t \in [\alpha n, n]$, the expression in Equation (17) is at least δ. Now

$$\mathbf{E}\left[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T = t \text{ and } z_1, \ldots, z_{t-1} \text{ is good}\right] \leq \frac{3 \times 0 + 1 \times 1 + 2 \times 2}{6} = \frac{5}{6}.$$

In particular, $\mathrm{Ham}(\sigma(T), \tau(T)) = 1$ if $z_t = (i, c)$ and $\mathrm{Ham}(\sigma(T), \tau(T)) = 0$ if z_t consists of vertex i with some other colour. Otherwise, $\mathrm{Ham}(\sigma(T), \tau(T)) \leq 2$. Similarly,

$$\mathbf{E}\left[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T = t \text{ and } z_1, \ldots, z_{t-1} \text{ is not good}\right] \leq \frac{2 \times 0 + 2 \times 1 + 2 \times 2}{6} = 1,$$

so

$$\begin{aligned}
\mathbf{E}[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T = t] &= \Pr(z_1, \ldots, z_{t-1} \text{ is good} \mid T = t) \\
&\quad \times \mathbf{E}\left[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T = t \text{ and } z_1, \ldots, z_{t-1} \text{ is good}\right] \\
&\quad + \Pr(z_1, \ldots, z_{t-1} \text{ is not good} \mid T = t) \\
&\quad\quad \times \mathbf{E}\left[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T = t \text{ and } z_1, \ldots, z_{t-1} \text{ is not good}\right] \\
&\leq \Pr(z_1, \ldots, z_{t-1} \text{ is good} \mid T = t) \left(1 - \frac{1}{6}\right) \\
&\quad + (1 - \Pr(z_1, \ldots, z_{t-1} \text{ is good} \mid T = t)) \\
&= 1 - \frac{1}{6} \Pr(z_1, \ldots, z_{t-1} \text{ is good} \mid T = t).
\end{aligned}$$

Thus if $t \in [\alpha n, n]$,

$$\mathbf{E}[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T = t] \le 1 - \frac{\delta}{6}.$$

Finally,

$$\begin{aligned}
\mathbf{E}[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T \le n] &= \sum_{t=1}^{n} \mathbf{E}[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T = t] \Pr(T = t \mid T \le n) \\
&\le 1 - \frac{\delta}{6} \sum_{t=\alpha n}^{n} \Pr(T = t \mid T \le n) \\
&\le 1 - \frac{\delta}{6} \sum_{t=\alpha n}^{n} \Pr(T = t). \qquad (18)
\end{aligned}$$

Since $\alpha \le 1/6$, we have

$$\Pr(T < \alpha n) = 1 - \left(\frac{4n-6}{4n}\right)^{\alpha n} = 1 - \left(1 - \frac{6}{4n}\right)^{\alpha n} \le \frac{6\alpha}{4} \le \frac{1}{4}.$$

Also,

$$\Pr(T > n) = \left(\frac{4n-6}{4n}\right)^{n} = \left(1 - \frac{6}{4n}\right)^{n} \le \exp(-6/4) \le \frac{1}{4}.$$

Thus, Equation (18) gives

$$\mathbf{E}\left[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T \le n\right] \le 1 - \frac{\delta}{12}. \qquad (19)$$

Now Theorem 2.2 of [12] tells us that

$$\mathbf{E}[\mathrm{Ham}(\sigma(n), \tau(n)) - 1] \le \Pr(T \le n)\big(\mathbf{E}[\mathrm{Ham}(\sigma(T), \tau(T)) \mid T \le n] - 1\big),$$

and by Equation (19), this is at most $-\delta/12$ so

$$\mathbf{E}[\mathrm{Ham}(\sigma(n), \tau(n))] \le 1 - \frac{\delta}{12}.$$

By the "delayed path coupling lemma" of Czumaj et al. (Lemma 2.1 of [12]), the mixing time satisfies

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}}, \varepsilon) \le \frac{12 \log(n\varepsilon^{-1})}{\delta} n.$$

In the preceding argument, we assumed that $i \in \{3, \ldots, n-2\}$ so that vertices $i-1, i-2$ and $i+1, i+2$ all exist. The argument still goes through if i has fewer neighbours to the left (or right). In that case, we just modify the argument by changing the definition of "good" so that it doesn't mention vertices that don't exist.

Thus, we have proved the following.

Theorem 12. *Let G be the n-vertex path. Let $q = 4$. Consider the Markov chain $\mathcal{M}_{\mathrm{Gl}}$ on the state space Ω^+.* $\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}}, \varepsilon) \le \frac{12}{\delta} n \log(n\varepsilon^{-1})$, *where δ is the constant mentioned above.*

6.3 Systematic scan for $q = 4$: $O(\log n)$ sweeps suffice

We will only define the coupling for pairs $(\sigma, \tau) \in S$. Each such pair disagrees at a single vertex i. Thus, when we come to re-colour a vertex j during the scan, at most one of $\{j-1, j+1\}$ has a disagreement. The coupling that we will use is as follows. If vertex j is not adjacent to a disagreement, then we use the same colours in both copies. On, the other hand, if (say) vertex $j-1$ has a disagreement, then we couple the choice of σ_{j-1} for σ_j and τ_{j-1} for τ_j and we couple the choice of τ_{j-1} for σ_j and σ_{j-1} for τ_j. Otherwise we choose the same colour in both copies. The coupling if $j+1$ has a disagreement is similar.

In the following sequence of lemmas, we let i denote the rightmost vertex where there is a disagreement between the colourings σ and τ. Lemmas 13 through 16 are valid for any $q \geq 4$, and we state them in terms of q so that we can re-use them later for $q > 4$. The first lemma, Lemma 13, analyses a scan starting from vertex $i+1$.

Lemma 13. *Suppose that σ and τ differ at vertex $i < n$ and agree to the right of vertex i. Obtain σ' and τ' by scanning left-to-right, starting at vertex $i+1$. Then*

$$\mathbf{E}[\mathrm{Ham}(\sigma', \tau')] - \mathrm{Ham}(\sigma, \tau) \leq \frac{1}{q-1}.$$

Proof. If $z = i+\ell$ for $\ell \in \{1, \ldots, n-i\}$, then the probability that vertex z becomes a disagreement after the recolouring is $(1/q)^{\ell}$. Thus, the expected number of additional disagreements is

$$\left(\frac{1}{q}\right)^{1} + \left(\frac{1}{q}\right)^{2} + \cdots + \left(\frac{1}{q}\right)^{n-i} \leq \frac{1}{q} \times \frac{1}{1-1/q} = \frac{1}{q-1}.$$

□

The next two lemmas analyse a scan starting from vertex i.

Lemma 14. *Suppose $(\sigma, \tau) \in S$ differ on vertex i. Let $C = |\{\sigma_{i-1}, \sigma_{i+1}\}|$. ($C$ is the number of colours that are used at neighbours of i in colouring σ.) Obtain σ' and τ' by scanning left-to-right, starting at vertex i. Then $\mathbf{E}[\mathrm{Ham}(\sigma', \tau')] \leq C/(q-1)$.*

Proof. Consider the recolouring of vertex i in copy σ. With probability $1 - C/q$, the chosen colour is not in $\{\sigma_{i-1}, \sigma_{i+1}\}$ so $\mathrm{Ham}(\sigma', \tau') = 0$. On the other hand, no matter what colour is chosen for vertex i, Lemma 13 guarantees that (conditioned on this choice) $\mathbf{E}[\mathrm{Ham}(\sigma', \tau')] \leq 1 + 1/(q-1)$. Thus, we have

$$\mathbf{E}[\mathrm{Ham}(\sigma', \tau')] \leq \frac{C}{q}\left(1 + \frac{1}{q-1}\right) = \frac{C}{q-1}.$$

□

Lemma 15. *Suppose colourings σ and τ differ just on vertices $i-1$ and i. Obtain σ' and τ' by scanning left-to-right, starting at vertex i. Then*

$$\mathbf{E}[\mathrm{Ham}(\sigma', \tau')] \leq 1 + \frac{3}{q-1}.$$

Proof. Consider the recolouring of vertex i in copy σ. With probability at least $1-3/q$, the chosen colour is not in $\{\sigma_{i-1},\tau_{i-1},\sigma_{i+1}\}$ so $\mathrm{Ham}(\sigma',\tau')=1$. On the other hand, no matter what colour is chosen for vertex i, Lemma 13 guarantees that $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le 2+1/(q-1)$. Thus, we have

$$\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\left(1-\frac{3}{q}\right)\cdot 1+\frac{3}{q}\left(2+\frac{1}{q-1}\right),$$

which simplifies to the claimed upper bound. □

The next three lemmas analyse a scan starting from vertex $\max\{1,i-1\}$.

Lemma 16. *Suppose $(\sigma,\tau)\in S$ differ on vertex i. Obtain σ' and τ' by scanning left-to-right, starting at vertex $\max\{1,i-1\}$. Then*

$$\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\frac{3}{q-1}.$$

Proof. If $i=1$, then Lemma 14 with $C=1$ shows $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le 1/(q-1)$, which is at most the expression given in the statement of the lemma. Suppose $i>1$. Consider the recolouring of vertex $i-1$ in copy σ. With probability $1/q$ colour τ_i is chosen. By Lemma 15, $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le 1+3/(q-1)$. Otherwise, $\sigma'_{i-1}=\tau'_{i-1}$, so Lemma 14 guarantees that $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le 2/(q-1)$. Hence,

$$\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\frac{1}{q}\left(1+\frac{3}{q-1}\right)+\left(1-\frac{1}{q}\right)\frac{2}{q-1}=\frac{3}{q-1},$$

as claimed. □

For the rest of this section, we restrict attention to the case $q=4$, which corresponds to the "break even" situation in Lemma 16.

Lemma 17. *Suppose $(\sigma,\tau)\in S$ differ on vertex $i<n$. Suppose that $\sigma_{i+1}\notin\{\sigma_i,\tau_i,\sigma_{i-2}\}$. Obtain σ' and τ' by scanning left-to-right, starting at vertex $\max\{1,i-1\}$. Then $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\frac{11}{12}$.*

Proof. If $i=1$, then the lemma follows from Lemma 14 with $C=1$. Suppose $i>1$. Consider the recolouring of vertex $i-1$ in copy σ. With probability $\frac{1}{4}$, colour σ_{i+1} is chosen. The same colour is chosen in copy τ, and Lemma 14 with $C=1$ guarantees that $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\frac{1}{3}$. With probability $\frac{1}{2}$, the colour chosen for vertex $i-1$ is not in $\{\sigma_{i+1},\tau_i\}$, so $\sigma'_{i-1}=\tau'_{i-1}$. By Lemma 14 with $C=2$, $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\frac{2}{3}$. Otherwise, Lemma 15 guarantees that $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le 2$. Thus, we have $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{2}{3}+\frac{1}{4}\cdot 2=\frac{11}{12}$. □

Lemma 18. *Suppose $(\sigma,\tau)\in S$ differ on vertex $i<n$. Suppose that $\sigma_{i+1}=\sigma_i$. Suppose that $\sigma_i\neq\sigma_{i-2}$ and $\tau_i\neq\sigma_{i-2}$. Obtain σ' and τ' by scanning left-to-right, starting at vertex $\max\{1,i-1\}$. Then $\mathbf{E}[\mathrm{Ham}(\sigma',\tau')]\le\frac{11}{12}$.*

Proof. If $i=1$, then the lemma follows from Lemma 14 with $C=1$.

Suppose $i>1$. Consider the recolouring of vertex $i-1$. With probability $\frac{1}{4}$ colour τ_i is chosen in copy σ and σ_i is chosen in copy τ. Both of these choices are accepted. In this case, consider the recolouring of vertex i. With probability $\frac{1}{2}$, the colour chosen is not in $\{\sigma_i, \tau_i\}$ and is accepted in both copies, leaving $\text{Ham}(\sigma', \tau') = 1$. Otherwise, by Lemma 13, $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{7}{3}$. Thus, conditioned on this colour choice for vertex $i-1$, we have $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{7}{3} = \frac{5}{3}$. For any other choice at vertex $i-1$, Lemma 14 guarantees that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{2}{3}$. We conclude that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{1}{4} \cdot \frac{5}{3} + \frac{3}{4} \cdot \frac{2}{3} = \frac{11}{12}$. □

For the remaining lemmas we analyse a scan starting from vertex 1. These three lemmas imply the result.

Lemma 19. *Suppose $(\sigma, \tau) \in S$ differ on vertex n. Obtain σ' and τ' by scanning left-to-right, starting at vertex 1. Then $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{11}{16}$.*

Proof. Consider the recolouring of vertex $n-1$ in colouring σ. With probability $\frac{1}{4}$, colour τ_n is chosen. In this case, $\text{Ham}(\sigma', \tau') \leq 2$. Otherwise, $\sigma'_{n-1} = \tau'_{n-1}$ so $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{1}{4}$. Thus, $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{1}{4} \cdot 2 + \frac{3}{4} \cdot \frac{1}{4} = \frac{11}{16}$. □

Lemma 20. *Suppose $(\sigma, \tau) \in S$ differ on vertex $i < n$. Suppose that $\sigma_{i+1} \notin \{\sigma_i, \tau_i\}$. Obtain σ' and τ' by scanning left-to-right, starting at vertex 1. Then $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq 47/48$.*

Proof. If $i \leq 2$ then the lemma follows from Lemma 17. Suppose $i > 2$ and consider the recolouring of vertex $i-2$. With probability $\frac{1}{4}$, the colour that is chosen is the first colour that is not in $\{\sigma'_{i-3}, \sigma_{i-1}, \sigma_{i+1}\}$. This is accepted so Lemma 17 guarantees that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{11}{12}$. Otherwise, Lemma 16 guarantees that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq 1$. Putting this together, $\mathbf{E}[\text{Ham}(\sigma'\tau')] \leq \frac{1}{4} \cdot \frac{11}{12} + \frac{3}{4} \cdot 1 = 47/48$. □

Lemma 21. *Suppose $(\sigma, \tau) \in S$ differ on vertex $i < n$. Suppose that $\sigma_{i+1} = \sigma_i$. Obtain σ' and τ' by scanning left-to-right, starting at vertex 1. Then $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq 191/192$.*

Proof. If $i \leq 2$ then the lemma follows from Lemma 18. Next suppose $i = 3$. Consider the recolouring of vertex $i-2$. With probability $\frac{1}{4}$, it is recoloured with the first colour that is not in $\{\sigma_{i-1}, \sigma_i, \tau_i\}$. Now Lemma 18 guarantees that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{11}{12}$. Otherwise, Lemma 16 guarantees that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq 1$. Our conclusion for $i = 3$ is that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{1}{4} \cdot \frac{11}{12} + \frac{3}{4} \cdot 1 = 47/48$. Finally, suppose $i > 3$. Consider the recolouring of vertex $i-3$. Let c be the first colour that is not in $\{\sigma_{i-1}, \sigma_i, \tau_i\}$. With probability $\frac{1}{4}$, the colour that is chosen for vertex $i-3$ is the first colour that is not in $\{\sigma_{i-4}, \sigma_{i-2}, c\}$. Suppose this happens. Then with probability $\frac{1}{4}$, c is chosen for vertex $i-2$. Then Lemma 18 guarantees that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{11}{12}$. Otherwise, Lemma 16 guarantees that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq 1$. We conclude that $\mathbf{E}[\text{Ham}(\sigma', \tau')] \leq \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{11}{12} + \frac{15}{16} \cdot 1 = 191/192$. □

Lemma 19, 20 and 21 imply the following result (by path coupling).

Theorem 22. *Let G be the n-vertex path. Let $q = 4$. Consider the Markov chain $\mathcal{M}_{\rightarrow}$ on the state space Ω^+.* $\text{Mix}(\mathcal{M}_{\rightarrow}, \varepsilon) \leq 192 \log(n\varepsilon^{-1})$.

6.4 Lower bounds for $q \geq 4$

In this section we prove that Glauber requires $\Omega(n \log n)$ updates and scan requires $\Omega(\log n)$ sweeps. We use the "disagreement percolation" method of van den Berg [26].

6.4.1 Calculating the stationary distribution for bounded line segments.

Consider an s-edge path (for any s). Consider the $q \times q$ "transfer matrix"

$$A = \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 0 & & 1 & 1 \\ \vdots & \vdots & & \ddots & & \vdots \\ 1 & 1 & 1 & & 0 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{pmatrix}.$$

Note that $A^s[i,j]$ is the number of colourings of the path in which the right vertex is coloured with colour i and the left vertex is coloured with colour j. We will write e_i to denote the row vector with a 1 in column i and zeros elsewhere. Write f to denote the row vector $(1,1,\ldots,1)$. Write v_i to denote the row vector with $q-1$ in column i and -1 elsewhere. Let e'_i, f' and v'_i be the corresponding column vectors. Thus, $e_i A^s e'_j$ is the number of colourings from colour i on the right to colour j on the left.

Now the (right) eigenvectors of A are f' with eigenvalue $q-1$ and, for every j, v'_j with eigenvalue -1. Since $e_j = q^{-1} f + q^{-1} v_j$ we have

$$A^s e'_j = q^{-1}(A^s f' + A^s v'_j) = q^{-1}((q-1)^s f' + (-1)^s v'_j).$$

Thus, if s is even we have

$$A^s e'_j = q^{-1}((q-1)^s f' + v'_j).$$

So, for $i \neq j$, the number of paths from colour i to colour j is

$$e_i A^s e'_j = q^{-1}((q-1)^s - 1). \tag{20}$$

Also, the number of paths from colour j to colour j is

$$e_j A^s e'_j = q^{-1}((q-1)^s + q - 1) = q^{-1}(q-1)^s(1 + (q-1)^{-(s-1)}). \tag{21}$$

6.4.2 Calculating the induced distribution on the colour of an internal vertex.

Suppose that ℓ and r are positive even integers and let $k = \ell + r$. Consider a path on vertices $1, \ldots, 1+k$. Consider the uniform distribution on colourings in which vertices 1 and $1+k$ are both coloured with colour j.

We wish to bound the probability that vertex $1+\ell$ is coloured with colour j. From Equation (21), this is

$$\frac{q^{-1}(q-1)^r\big(1+(q-1)^{-(r-1)}\big) \times q^{-1}(q-1)^\ell\big(1+(q-1)^{-(\ell-1)}\big)}{q^{-1}(q-1)^k(1+(q-1)^{-(k-1)})} \tag{22}$$

$$= q^{-1}\big(1+(q-1)^{-(r-1)}\big)\frac{1+(q-1)^{-(\ell-1)}}{1+(q-1)^{-(k-1)}}$$

$$\geq q^{-1}\big(1+(q-1)^{-(r-1)}\big). \tag{23}$$

6.4.3 Dividing the line into segments

Let r be the largest even number not exceeding $\frac{1}{3}\log_{q-1} n$. Let ℓ be the smallest even number that is at least $48\log_e n$. Let $k = \ell + r$ and $m = \lfloor (n-1)/k \rfloor$. For $i \in \{0, \ldots, m-1\}$, let L_i be the vertex $1 + ik$ and M_i be the vertex $1 + ik + \ell$. Finally, let $L_m = 1 + mk$ and R_i be L_{i+1}. The idea is to divide the line into line segments. Segment i has left endpoint L_i and "middle" point M_i (which is not quite in the middle!) and right endpoint R_i.

Let Z_i be the indicator for the event that vertex M_i is coloured with colour 0. Let $Z = \sum_{i=0}^{m-1} Z_i$. Of course, the expectations of Z_i and Z are only well-defined if we focus attention on a particular distribution over Ω. We will use the notation $\mathbf{E}_\pi(Z_i)$ to refer to the expectation of Z_i in distribution π.

6.4.4 Calculating the distribution of Z in stationarity

Let π be the uniform distribution on Ω. We can sample from π by filling in the colours from left to right. There are $q(q-1)^{n-1}$ possible colourings in Ω. Given the colours of vertices $1, \ldots, n-v$ (for $v < n$), there are $(q-1)^v$ ways to finish the colouring, and these are chosen uniformly. The probability that $Z_1 = 1$ is $1/q$. For any $i > 1$, we can use Equation (21) (observing that the quantity in Equation (20) is smaller) to see that the probability that $Z_i = 1$, conditioned on colours $\sigma_1, \ldots, \sigma_{M_{i-1}}$, is at most

$$\frac{q^{-1}(q-1)^k(1 + (q-1)^{-(k-1)})}{(q-1)^k} = q^{-1}(1 + (q-1)^{-(k-1)}).$$

Thus, Z is dominated from above by the sum of m independent Bernoulli random variables with success probability $p = q^{-1}(1 + (q-1)^{-(k-1)})$.

Let $\varepsilon = m^{-3/8}$. By a Chernoff bound,

$$\Pr_\pi(Z \geq (1+\varepsilon)mp) \leq \exp(-\varepsilon^2 mp/3). \tag{24}$$

Note that $(q-1)^{k-1} = \omega(n^{1/3})$ and that $m = \Theta(n/\log n)$. Thus, Equation (24) implies

$$\Pr_\pi(Z \geq q^{-1}m + \frac{1}{2}mn^{-1/3}) \leq \Pr_\pi(Z \geq (1+\varepsilon)mp) \leq \exp(-\varepsilon^2 mp/3) = o(1). \tag{25}$$

6.4.5 An initial distribution for the Markov chains

Equation (25) shows that, in the stationary distribution π, Z is unlikely to exceed $q^{-1}m + \frac{1}{2}mn^{-1/3}$. In this section, we will define an initial distribution π_0 on colourings in Ω.

The idea will be to show that if $\sigma(0)$ is chosen from π_0 and $\sigma(0), \sigma(1), \ldots, \sigma(t)$ evolves according to the dynamics (either Glauber or scan) and t is too small, then, in the distribution of $\sigma(t)$, Z is likely to exceed $q^{-1}m + \frac{1}{2}mn^{-1/3}$, allowing us to conclude that the chain does not mix by step t.

Let π_0 be the uniform distribution on colourings in which vertices $L_0, \ldots, L_m$ are coloured 0. By Equation (23), $\mathbf{E}_{\pi_0} Z \geq mq^{-1}(1 + (q-1)^{-(r-1)}) \geq mq^{-1}(1 + (q-1)n^{-1/3})$. By a Chernoff bound,

$$\Pr_{\pi_0}\left(Z \geq q^{-1}m + \tfrac{1}{2}mn^{-1/3}\right) \geq 1 - o(1). \tag{26}$$

6.4.6 The t-step distribution for systematic scan

Suppose $\sigma(0)$ is chosen from π_0. Let $\sigma(0), \sigma(1), \ldots$ evolve according to the dynamics of $\mathcal{M}_{\rightarrow}$. Let $\mathcal{L}_{\rightarrow,t}(Z)$ denote the distribution of the random variable Z in the colouring $\sigma(t)$.

Suppose $\tau(0)$ is chosen from π_0. Let $\tau(0), \tau(1), \ldots$ evolve according to the "clamped dynamics" $\mathcal{M}^{c}_{\rightarrow}$, which is the same as $\mathcal{M}_{\rightarrow}$ except that all moves involving vertices $\{L_0, \ldots, L_m\}$ are rejected (so the colour of these vertices cannot change). Let $\mathcal{L}^{c}_{\rightarrow,t}(Z)$ denote the distribution of the random variable Z in the colouring $\tau(t)$. By construction the distribution $\mathcal{L}^{c}_{\rightarrow,t}(Z)$ is same as the distribution of Z in π_0.[3]

To upper bound $d_{TV}(\mathcal{L}_{\rightarrow,t}(Z), \mathcal{L}^{c}_{\rightarrow,t}(Z))$, we will will consider a joint process $(\sigma(t), \tau(t))$ in which the first component has the same distribution as $(\sigma(t))$ and the second component has the same distribution as $(\tau(t))$. The total variation distance $d_{TV}(\mathcal{L}_{\rightarrow,t}(Z), \mathcal{L}^{c}_{\rightarrow,t}(Z))$ is upper-bounded by the probability that some vertex M_i gets different colours in $\sigma(t)$ and $\tau(t)$.

The particular joint process that we will consider starts with $\sigma(0) = \tau(0)$. To move from $(\sigma(t-1), \tau(t-1))$ to $(\sigma(t), \tau(t))$, we use the "switch coupling". When we consider vertex v for recolouring, we will couple the colour choices as follows.

(A) If we consider colour $\sigma(v-1)$ for v in σ then consider colour $\tau(v-1)$ for v in τ,

(B) if we consider colour $\tau(v-1)$ for v in σ then consider colour $\sigma(v-1)$ for v in τ,

(C) otherwise consider the same colour for v in τ as in σ.

We will be particularly interested in $t < r$. For such a t, and for any $i \in \{0, \ldots, m-1\}$, the probability that vertex M_i gets different colours in $\sigma(t)$ and $\tau(t)$ is at most the probability that we chose option (B) in order for vertices $L_i + 1, \ldots, L_i + \ell$ over the t scans.

Say that vertex $L_i + v$ is "interrupting" (that is, it interrupts the disagreement percolation) if, the first time that we consider this vertex when we have a disagreement at vertex $L_i + v - 1$, we choose some option other than option (B) for vertex $L_i + v$.

The probability that vertex M_i gets different colours in $\sigma(t)$ and $\tau(t)$ is at most the probability that we have fewer than t interrupting vertices in $L_i + 1 \ldots, L_i + \ell$, this is dominated (from above) by the probability of having fewer than t successes in ℓ Bernoulli trials with success probability $(q-1)/q$. So if we take any $t \le r/2 \le (2/3)\ell(q-1)/q$, a Chernoff bound says that the probability of having fewer than t interrupting vertices is at most

$$\exp(-(1/3)^2 \ell (q-1)/(2q)),$$

which is at most n^{-2} by the definition of ℓ. Thus, the probability that there exists an i such that vertex M_i gets different colours in $\sigma(t)$ and $\tau(t)$ is at most $mn^{-2} = o(1)$.

Thus, for any $t \le r/2$, $d_{TV}(\mathcal{L}_{\rightarrow,t}(Z), \mathcal{L}^{c}_{\rightarrow,t}(Z)) = o(1)$ so by Equation (26),

$$\Pr_{\mathcal{L}_{\rightarrow,t}(Z)}\left(Z \ge q^{-1}m + \tfrac{1}{2}mn^{-1/3}\right) \ge 1 - o(1). \tag{27}$$

[3] This follows because π_0 is the stationary distribution of $\mathcal{M}^{c}_{\rightarrow}$. This can be proved as follows, where Ω_0 is the set of colourings in which vertices $L_0, \ldots, L_m$ are coloured 0. Let $P^{c}_{\rightarrow}$ by the transition matrix of $\mathcal{M}^{c}_{\rightarrow}$ and let $P^{c}_{\leftarrow}$ be the transition matrix of the reversal. Then any stationary distribution π' of $\mathcal{M}^{c}_{\rightarrow}$ satisfies

$$\pi'(\sigma') = \sum_{\sigma \in \Omega_0} \pi'(\sigma) P^{c}_{\rightarrow}(\sigma, \sigma') = \sum_{\sigma \in \Omega_0} \pi'(\sigma) P^{c}_{\leftarrow}(\sigma', \sigma),$$

but the latter equation is satisfied by the uniform distribution $\pi' = \pi_0$. (Also, the chain is ergodic so has a unique stationary distribution.)

Combining this with Equation (25), we find that $d_{TV}(\mathcal{L}_{\to,t}, \pi) \geq 1 - o(1)$ so systematic scan does not mix in t steps. Thus, we obtain the following theorem.

Theorem 23. *Let G be the n-vertex path. Let $q \geq 4$. Consider the Markov chain $\mathcal{M}_{\to}$ on the state space Ω. For any fixed $\varepsilon < 1$,*

$$\mathrm{Mix}(\mathcal{M}_{\to}, \varepsilon) \geq \frac{1}{2}\left(\frac{1}{3}\log_{q-1} n - 2\right).$$

6.4.7 The t-step distribution for Glauber

A similar argument to that of Section 6.4.6 can be used to show that Glauber dynamics does not mix in t steps for some $t = \Omega(n \log_{q-1} n)$. The particular value of t for which the straightforward argument works is around $nr/(q^2 e)$. We prefer to give a stronger argument which gives a better bound as a function of q. The idea for the stronger argument is as follows. In Section 6.4.6, we showed that the distribution of $\mathcal{L}_{\to,t}(Z)$ and $\mathcal{L}^{c}_{\to,t}(Z)$ were close by showing that, with high probability, there was no $i \in \{0, \ldots, m-1\}$ for which a disagreement at vertex L_i or R_i could percolate to vertex M_i. Here we observe that the distributions $\mathcal{L}_{\to,t}(Z)$ and $\mathcal{L}^{c}_{\to,t}(Z)$ would be close even if some, but not many, of the percolations occur.

We start with some notation. It will be helpful to keep track of the nearest endpoint to an arbitrary vertex v. For this purpose, if v is in the range $L_i + 1, \ldots, M_i - 1$, its "*important neighbour*" will be vertex $v - 1$. On the other hand, if v is in the range $M_i, \ldots, R_i - 1$, its important neighbour will be vertex $v + 1$.

As in Section 6.4.6, we will consider a process $\sigma(0), \sigma(1), \ldots$ in which $\sigma(0)$ is drawn from π_0 and $\sigma(t)$ evolves according to $\mathcal{M}_{\mathrm{Gl}}$. We will also consider the process $\tau(0), \tau(1), \ldots$ in which $\tau(0) = \sigma(0)$ and $\tau(t)$ evolves according to a *clamped dynamics* $\mathcal{M}^{c}_{\mathrm{Gl}}$ in which moves involving $L_0, L_1, \ldots, L_m$ are rejected. We will construct a joint process $(\sigma(t), \tau(t))$ with $\sigma(0) = \tau(0)$. To move from $(\sigma(t-1), \tau(t-1))$ to $(\sigma(t), \tau(t))$ we choose the same vertex v in both copies. If $v = L_i$ for some i then only σ is changed. If $v > L_m$ then we use the same colour in both copies. Otherwise, we do a switch coupling, based on the important neighbour, w, of v. In particular, we couple the colour choices as follows.

(A) If we consider colour $\sigma(w)$ for v in σ then consider colour $\tau(w)$ for v in τ,

(B) if we consider colour $\tau(w)$ for v in σ then consider colour $\sigma(w)$ for v in τ,

(C) otherwise consider the same colour for v in τ as in σ.

Suppose that $t \leq (qnr)/(2e(q-1))$. The probability that M_i gets different colours in $\sigma(t)$ and $\tau(t)$ is at most the probability that (at least) one of the following occurs during the t steps.

- During some ordered sequence of $\ell - 1$ steps, the process recolours vertices $L_i + 1, L_i + 2, \ldots, L_i + \ell - 1 = M_i - 1$ using option (B).
- During some ordered sequence of r steps, the process recolours vertices $R_i - 1, R_i - 2, \ldots, R_i - r = M_i$ using option (B).

The probability that one of these occurs is at most

$$2\binom{t}{r}\left(\frac{1}{qn}\right)^r \leq 2\left(\frac{te}{rqn}\right)^r \leq \frac{1}{8}\left(\frac{2te}{rqn}\right)^r \leq \frac{1}{8}\left(\frac{1}{q-1}\right)^r,$$

where we have crudely used $r \geq 4$ in the second inequality. Thus,

$$d_{\mathrm{TV}}(\mathcal{L}_{\mathrm{Gl},t}(Z_i), \mathcal{L}^{\mathrm{c}}_{\mathrm{Gl},t}(Z_i)) \leq \frac{1}{8}\left(\frac{1}{q-1}\right)^r. \tag{28}$$

Now combining Equation (23) and Equation (28), we have

$$\begin{aligned}
\Pr_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i = 1) &\geq \Pr_{\mathcal{L}^{\mathrm{c}}_{\mathrm{Gl},t}}(Z_i = 1) - d_{\mathrm{TV}}(\mathcal{L}_{\mathrm{Gl},t}(Z_i), \mathcal{L}^{\mathrm{c}}_{\mathrm{Gl},t}(Z_i)) \\
&\geq q^{-1}(1 + (q-1)^{-(r-1)}) - \frac{1}{8}\left(\frac{1}{q-1}\right)^r \\
&\geq q^{-1} + \frac{5}{8}(q-1)^{-r} \\
&\geq q^{-1} + \frac{5}{8}n^{-1/3}.
\end{aligned}$$

So

$$\mathbf{E}_{\mathcal{L}_{\mathrm{Gl},t}}(Z) \geq q^{-1}m + \frac{5}{8}mn^{-1/3}.$$

Also,

$$\mathrm{var}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i) = \Pr_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i = 1)\Pr_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i = 0) \leq 1.$$

We will show in Lemma 25 (below) that for $i \neq j$ $\mathrm{cov}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i, Z_j) \leq m^{-1}$. So

$$\begin{aligned}
\mathrm{var}_{\mathcal{L}_{\mathrm{Gl},t}}(Z) &= \sum_i \mathrm{var}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i) + \sum_{i \neq j} \mathrm{cov}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i, Z_j) \\
&\leq m + \sum_{i \neq j} \mathrm{cov}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i, Z_j) \\
&\leq 2m.
\end{aligned}$$

Let

$$\lambda = \frac{\frac{1}{8}mn^{-1/3}}{\sqrt{2m}}.$$

Note that $\lambda = \omega(1)$ as a function of n. Also,

$$\mathbf{E}_{\mathcal{L}_{\mathrm{Gl},t}}(Z) - \lambda\sqrt{\mathrm{var}_{\mathcal{L}_{\mathrm{Gl},t}}(Z)} \geq q^{-1}m + \frac{5}{8}mn^{-1/3} - \lambda\sqrt{2m} = q^{-1}m + \tfrac{1}{2}mn^{-1/3}.$$

Thus by Chebyschev's inequality we have

$$\Pr_{\mathcal{L}_{\mathrm{Gl},t}}\left(Z \leq q^{-1}m + \tfrac{1}{2}mn^{-1/3}\right) \leq \Pr_{\mathcal{L}_{\mathrm{Gl},t}}\left(Z \leq \mathbf{E}_{\mathcal{L}_{\mathrm{Gl},t}}(Z) - \lambda\sqrt{\mathrm{var}_{\mathcal{L}_{\mathrm{Gl},t}}(Z)}\right) \leq \lambda^{-2} = o(1).$$

Combining this with Equation (25), we find that $d_{\mathrm{TV}}(\mathcal{L}_{\mathrm{Gl},t}, \pi) \geq 1 - o(1)$ so Glauber dynamics does not mix in t steps for any $t \leq (qnr)/(2e(q-1))$ Thus, we obtain the following theorem.

Theorem 24. *Let G be the n-vertex path. Let $q \geq 4$. Consider the Markov chain $\mathcal{M}_{\mathrm{Gl}}$ on the state space Ω. For any fixed $\varepsilon < 1$,*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}}, \varepsilon) \geq \frac{qn\left(\frac{1}{3}\log_{q-1} n - 2\right)}{2e(q-1)}.$$

Lemma 25. *For $i \neq j$ $\mathrm{cov}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i, Z_j) \leq m^{-1}$.*

Proof. We will show that Z_i and Z_j have low covariance in the t-step distribution by showing that Glauber dynamics (over t steps) is quite close to a "clamped distribution" in which some vertex between M_i and M_j is held fixed. This "disagreement percolation" argument is similar to the argument in Section 6.4.7. The only difference is that, in order to get a sufficiently small upper bound on the covariance, we have to look at a "clamped process" that is slightly different from $\mathcal{M}^{\mathrm{c}}_{\mathrm{Gl}}$. In particular, in $\mathcal{M}^{\mathrm{c}}_{\mathrm{Gl}}$, M_i is only r vertices away from the nearest "clamped vertex", R_i. Here we need to spread the clamped vertices out more symmetrically with respect to the vertices M_i. Let

$$\Gamma = \{M_i + k/2 \mid i \in \{0, \ldots, m-1\}\}.$$

Consider a process $\sigma(0), \sigma(1), \ldots$ which evolves according to $\mathcal{M}_{\mathrm{Gl}}$. Let $\rho(0), \rho(1), \ldots$ be a process which evolves according a clamped version of $\mathcal{M}_{\mathrm{Gl}}$ in which those moves involving vertices in Γ are rejected. We refer to this dynamics as $\mathcal{M}^{\mathrm{sc}}_{\mathrm{Gl}}$ (where "sc" is intended to indicate "symmetric clamped"). Consider the joint process $(\sigma(t), \rho(t))$ which starts with $\rho(0) = \sigma(0)$ and progresses according to the identity coupling (the same vertices and colours are chosen in the transition $\sigma(t-1) \to \sigma(t)$ and in the transition $\rho(t-1) \to \rho(t)$). Now the probability that $\sigma(t)_{M_i} \neq \rho(t)_{M_i}$ or $\sigma(t)_{M_j} \neq \rho(t)_{M_j}$ (or both) is at most

$$4\binom{t}{k/2}\left(\frac{1}{n}\right)^{k/2},$$

since an ordered sequence of $k/2$ vertices would need to be chosen either from the left towards M_i or from the right towards M_i or from the left or right towards M_j. The probability that a particular vertex is chosen at any step is $1/n$. Since $t \leq (qnr)/(2e(q-1))$ and $k \geq 8er$, this is at most

$$\left(\frac{8te}{kn}\right)^{k/2} \leq e^{-k/2} \leq n^{-24}.$$

Now

$$\begin{aligned}
\mathrm{cov}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i, Z_j) &= \mathbf{E}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i Z_j) - \mathbf{E}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i)\,\mathbf{E}_{\mathcal{L}_{\mathrm{Gl},t}}(Z_j) \\
&= \Pr_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i = 1 \wedge Z_j = 1) - \Pr_{\mathcal{L}_{\mathrm{Gl},t}}(Z_i = 1)\Pr_{\mathcal{L}_{\mathrm{Gl},t}}(Z_j = 1) \\
&\leq \Pr_{\mathcal{L}^{\mathrm{sc}}_{\mathrm{Gl},t}}(Z_i = 1 \wedge Z_j = 1) + n^{-24} \\
&\qquad - (\Pr_{\mathcal{L}^{\mathrm{sc}}_{\mathrm{Gl},t}}(Z_i = 1) - n^{-24})(\Pr_{\mathcal{L}^{\mathrm{sc}}_{\mathrm{Gl},t}}(Z_j = 1) - n^{-24}) \\
&\leq \mathrm{cov}_{\mathcal{L}^{\mathrm{sc}}_{\mathrm{Gl},t}}(Z_i, Z_j) + 4n^{-24} \\
&= 4n^{-24}.
\end{aligned}$$

□

7 Optimal mixing for Glauber and scan when $q > 4$

Let G be the n-vertex path. For $q > 4$, Lemma 1 of [20] shows that Glauber dynamics mixes in $O(n \log n)$ steps. For scan, we use the coupling from Section 6.3. Consider a pair $(\sigma, \tau) \in S$ which disagrees at a single vertex i. Obtain σ' and τ' by scanning left-to-right, starting at vertex $\max\{1, i-1\}$. Lemma 16 shows that $\mathbf{E}[\mathrm{Ham}(\sigma', \tau')] \leq 3/4$. This implies the following theorem (by path coupling).

Theorem 26. *Let G be the n-vertex path. Let $q > 4$. Consider the Markov chain $\mathcal{M}_{\rightarrow}$ on the state space Ω^+.* $\mathrm{Mix}(\mathcal{M}_{\rightarrow}, \varepsilon) \leq 4\log(n\varepsilon^{-1})$.

8 H-Colouring: $O(n^5)$ updates or scans suffice

Let H be a fixed graph, possibly with self-loops, and let Ω be the set of H-colourings of the graph G. These are the homomorphisms from G to H — see [6, 14, 18] for details. We can extend the dynamics $\mathcal{M}_{\mathrm{Gl}}$ and $\mathcal{M}_{\rightarrow}$ to the domain of H-colouring by modifying the procedure Metropolis(v) from Section 2. In particular, a proposed colour c (which is a vertex of H) is accepted if and only if every neighbour w of v is coloured with some neighbour c_w of c. The original dynamics corresponds to the situation in which H is a q-clique with no self-loops.

Suppose that H is connected. Let G be the n-vertex path. If H has an odd cycle then Glauber dynamics and systematic scan are both ergodic on Ω, the set of H-colourings of G. In this case we say that any two colourings, $\sigma \in \Omega$ and $\tau \in \Omega$ are *compatible.* If H does not have an odd cycle then it is bipartite. Neither dynamics is ergodic on Ω. However the H-colourings can be partitioned in a natural way into two subsets, such that Glauber and scan are both ergodic on either subset. In particular, the H-colourings are partitioned as follows. Two H-colourings $\sigma \in \Omega$ and $\tau \in \Omega$ are compatible if σ_1 and τ_1 are chosen from the same side of the bipartition of H. Our aim is to show rapid mixing on the set(s) of compatible colourings.

Let $h = |V(H)|$. Define t as follows.

$$t = \begin{cases} 4h-1, & \text{if } H \text{ is not bipartite and } n \text{ is even;} \\ 2h-1, & \text{if } H \text{ is bipartite and } n \text{ is even;} \\ 4h, & \text{if } H \text{ is not bipartite and } n \text{ is odd;} \\ 2h, & \text{if } H \text{ is bipartite and } n \text{ is odd.} \end{cases}$$

Note that $n + t$ is always odd.

Lemma 27. *In any two compatible H-colourings σ and τ, there is a t-edge path in H from σ_n to τ_1.*

Proof. We look at each of the four cases.

H is not bipartite and n is even: Let c be some point on an odd-length cycle. Go from σ_n to c in at most $h-1$ edges. Also go from c to τ_1 in at most $h-1$ edges. If the constructed path has an even number of edges, go around the cycle using at most h more edges. Now go back and forth on the last edge to make the total length equal to t.

H is bipartite and n is even: Note that σ_n and τ_1 are on opposite sides of the bipartition. Go from σ_n to τ_1 in at most $h-1$ edges and go back and forth on the last edge.

H is not bipartite and n is even: Let c be some point on an odd-length cycle. Go from σ_n to c in at most $h-1$ edges. Also go from c to τ_1 in at most $h-1$ edges. If the constructed path has an odd number of edges, go around the cycle using at most h more edges. Now go back and forth on the last edge to make the total length equal to t.

H is bipartite and n is odd: Note that σ_n and τ_1 are on the same side of the bipartition. Go from σ_n to τ_1 in at most $h-1$ edges and go back and forth on the last edge.

□

8.1 Constructing canonical paths

Let Ω' be the state space. It is either the set of all proper colourings (if H is not bipartite) or it is one of the two maximum sets of compatible colourings (if H is bipartite). We will use the canonical paths method, which can be viewed as a special case of comparison in which we compare $\mathcal{M}_{\mathrm{Gl}}$ to the uniform random walk on Ω'. Thus, for each $\sigma \in \Omega'$ and $\tau \in \Omega'$ we will construct a canonical path $\gamma_{\sigma,\tau}$ from σ to τ.

First, let $\sigma_n c_1 \cdots c_{t-1} \tau_1$ be some t-edge path from σ_n to τ_1 and let $z_1 z_2 \cdots z_{2n+t-1}$ denote $\sigma_1 \cdots \sigma_n c_1 \cdots c_{t-1} \tau_1 \cdots \tau_n$. Let Z_i denote the H-colouring $z_i z_{i+1} \cdots z_{i+n-1}$ so $Z_1 = \sigma$ and $Z_{n+t} = \tau$. The path $\gamma_{\sigma,\tau}$ passes through $Z_1, Z_3, Z_5, \ldots, Z_{n+t}$. Moving from Z_i to Z_{i+2} can be implemented by n Glauber transitions (applied to vertices 1 to n in order). Let

$$A = \max_{\alpha,\beta} \frac{1}{\pi(\alpha) P_{\mathrm{Gl}}(\alpha,\beta)} \sum_{\sigma,\tau} \pi(\sigma)\pi(\tau)\,|\gamma_{\sigma,\tau}|, \tag{29}$$

where the max is over all Glauber-dynamics transitions (α,β) and the sum is over all pairs (σ,τ) such that (α,β) is on the canonical path $\gamma_{\sigma,\tau}$. By Theorem 2.1 of [8] we have $\lambda(\mathcal{M}_{\mathrm{Gl}}) \geq 1/A$. We now derive an upper bound on A.

The three stationary probabilities in (29) are all $1/|\Omega'|$. Furthermore, every canonical path $\gamma_{\sigma,\tau}$ satisfies $|\gamma_{\sigma,\tau}| \leq \frac{n+t}{2} n$. Finally, $P_{\mathrm{Gl}}(\alpha,\beta) = \frac{1}{nh}$. Plugging this into Equation (29) we get

$$A \leq \frac{n+t}{2}\, n \,\frac{nh}{|\Omega'|} \max_{\alpha,\beta} \sum_{\sigma,\tau} 1.$$

We will show that the number of pairs (σ,τ) using transition (α,β) is in $O(n\,|\Omega'|)$, from which we can conclude

$$A = O(n^4). \tag{30}$$

The method we use is standard: each canonical path through (α,β) will be assigned a unique "encoding" chosen from a set of $O(n\,|\Omega'|)$ encodings.

So now fix (α,β) and consider the set of all canonical paths that use transition (α,β). We show how to encode a typical such path, from σ to τ say. Let $\tau_n c'_1 \cdots c'_{t-1} \sigma_1$ be some t-edge path in H from τ_n to σ_1 and let $\hat{z}_1 \hat{z}_2 \cdots \hat{z}_{2n+t-1}$ denote the path $\tau_1 \cdots \tau_n c'_1 \cdots c'_{t-1} \sigma_1 \cdots \sigma_n$. Let $\hat{Z}_i$ denote the H-colouring $\hat{z}_i \hat{z}_{i+1} \cdots \hat{z}_{i+n-1}$.

The encoding of the canonical path from σ to τ consists of the following information.

- i, indicating that the current transition is on the path from Z_i to Z_{i+2}, and the colours z_i and z_{i+1};
- $\hat{Z}_i$, and
- the colours $\sigma_{i-t+1}, \ldots, \sigma_{i-1}$ and $\tau_{i-t+1}, \ldots, \tau_{i-1}$.

Given the transition (α,β) and the values of i, z_i and z_{i+1}, we can deduce Z_i. From Z_i and $\hat{Z}_i$ and the extra colours we can deduce σ and τ. Thus, the number of pairs (σ,τ) using the given transition is at most the number of encodings, which is in $O(n\,|\Omega'|)$ as required, so we have now established (30).

Note that $\mathcal{M}_{\mathrm{Gl}}$ is reversible. Let $1 = \beta_0 > \beta_1 \geq \cdots \geq \beta_{|\Omega'|-1} > -1$ be the eigenvalues of its transition matrix P_{Gl}. Since $1 - \beta_1 = \lambda(\mathcal{M}_{\mathrm{Gl}})$, we have $1/(1-\beta_1) \leq A$. To bound the mixing time of $\mathcal{M}_{\mathrm{Gl}}$, we also need an upper bound on $\frac{1}{1+\beta_{|\Omega'|-1}}$. This is an easy application of Proposition 2 of [9] since for every $\sigma \in \Omega'$ we have $P_{\mathrm{Gl}}(\sigma,\sigma) \geq 1/h$.

In particular, for every $\sigma \in \Omega'$, we define the (odd-length) canonical path from σ to itself to be single transition $P_{\mathrm{Gl}}(\sigma,\sigma)$. Proposition 2 of [9] then gives

$$\frac{1}{1+\beta_{|\Omega'|-1}} \leq \frac{1}{2}\max_{\sigma} \frac{1}{P_{\mathrm{Gl}}(\sigma,\sigma)} \leq \frac{h}{2}.$$

Combining this with (30), Proposition 1(i) of [24] gives

$$\mathrm{Mix}_{\sigma}(\mathcal{M}_{\mathrm{Gl}},\varepsilon) \leq \frac{1}{1-\beta_{\max}}\left(\ln \pi(\sigma)^{-1} + \ln \varepsilon^{-1}\right) = O(n^4)\left(\ln \pi(\sigma)^{-1} + \ln \varepsilon^{-1}\right).$$

Thus, we have the following theorem.

Theorem 28. *Let H be a fixed connected graph. Let G be the n-vertex path. Let Ω' be the state space of $\mathcal{M}_{\mathrm{Gl}}$. It is either the set of all proper H-colourings of G (if H is not bipartite) or it is one of the two maximum sets of compatible colourings (if H is bipartite). Consider the Markov chain $\mathcal{M}_{\mathrm{Gl}}$ on the state space Ω'.*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}},\varepsilon) = O(n^5 \ln \varepsilon^{-1}).$$

In Section 10.1, we will show how to use our lower bound $\lambda(\mathcal{M}_{\mathrm{Gl}}) \geq 1/A$ to get a corresponding lower bound on $\lambda(\mathcal{M}_{\rightarrow})$. This will imply that the mixing time of systematic scan is also in $O(n^5 \ln \varepsilon^{-1})$, though for technical reasons (since scan is not reversible), we state the result in continuous time. See Theorem 32 in Section 10.1 for details.

8.2 Corollary

In this section, we have shown that both Glauber and scan have polynomial mixing time for H-colourings on the line. As a corollary, we get polynomial mixing for ordinary colourings on narrow grids (with a constant number of rows). Thus, the following is a corollary of Theorem 28.

Corollary 29. *Let q and d be constants and let G be the $n \times d$ grid. Consider the Glauber dynamics $\mathcal{M}_{\mathrm{Gl}}$ on Ω, the set of q-colourings of G.*

$$\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}},\varepsilon) = O(n^5 \ln \varepsilon^{-1}).$$

A similar corollary of Theorem 32 holds for scan.

8.3 Special case

Suppose that H is an odd cycle of length k. We noted at the beginning of Section 8 that Glauber and scan are ergodic on Ω and Section 8.1 shows that the mixing time is in $O(n^5)$. In fact, the analysis for 3-colouring translates directly to the case of a k-cycle so we get the following analog of Theorems 1 and 3.

Theorem 30. *Let H be an odd cycle. Let G be the n-vertex path. Let Ω' be the set of H-colourings of G. Consider the Markov chain $\mathcal{M}_{\mathrm{Gl}}$ on the state space Ω'.* $\mathrm{Mix}(\mathcal{M}_{\mathrm{Gl}}, 7/8) = \Theta(n^3 \log n)$.

The generalisation of the proofs of Theorems 1 and 3 is straightforward. In Section 4.1, each configuration $X \in \Upsilon$ corresponds to k colourings. In Section 4.2, the height h_i of every vertex i satisfies $h_i = \sigma_i \pmod k$. The quantity B in Section 4.3 is increased by a factor of k. A similar result holds for scan.

9 Directed H-colouring

It is natural to ask whether the H-colouring results could be generalized, for example to directed H-colouring. The answer is no. To illustrate this, we give an example of a directed H that is not ergodic on the n-vertex path G, and another example of a directed H for which Glauber is ergodic, but mixes slowly.

For the first example, let H have vertex set $\{x, y, z\}$ and edge set $\{(x, y), (y, z), (z, x)\}$. Now the three possible colourings of G are $xyzxyz\cdots$, $yzxyzx\cdots$ and $zxyzxy\cdots$. These are not connected by either Glauber or scan moves.

For the second example, let the vertices of H be $\{x, b_1, \dots, b_k, c_1, \dots, c_k\}$. Let the edges of H consist of an edge from x to every vertex (including itself), a directed clique on $B = \{b_1, \dots, b_k\}$, and a directed clique on $C = \{c_1, \dots, c_k\}$. Let X be the singleton set $\{x\}$. The H-colourings of G correspond to the length-n strings satisfying the regular expression $X^*B^* \cup X^*C^*$. Let A be the set of H-colourings satisfying the regular expression X^*B^+. (A colouring in A starts out with a possibly empty sequence of colour-x vertices, then contains a non-empty sequence of vertices with colours from B.) Let M be the set of all colourings with at most one colour from $B \cup C$. Since B and C are the same size, $\pi(A) \le 1/2$. Furthermore, Furthermore, for $\sigma \in A \backslash M$ and $\tau \in \overline{A} \backslash M$, $P_{\mathrm{Gl}}(\sigma, \tau) = 0$ and $P_{\rightarrow}(\sigma, \tau) = 0$. Claim 2.3 of [11] shows that the mixing time of both of these chains is at least $\pi(A)/8\pi(M)$. Now

$$\pi(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{2|A| + 1} \ge \frac{|A|}{3|A|} = \frac{1}{3}.$$

Also

$$\pi(M) = \frac{|M|}{|\Omega|} = \frac{2k + 1}{|\Omega|} \le \frac{2k + 1}{k^n},$$

which finishes the proof.

10 Comparisons of scan and Glauber for general graphs

From the results obtained so far, it seems as if one sweep of systematic scan is equivalent to a linear number of Glauber updates. In the majority of cases examined (Sections 4–7) we have obtained tight asymptotic bounds, and we know the equivalence is exact. Where we don't have tight bounds (Section 8) at least our results are consistent with this supposed equivalence. It is natural to wonder whether a result can be framed that relates scan and Glauber in a more general setting, where the graph G is arbitrary.

In this section we use the comparison method of Diaconis and Saloff-Coste [8] to compare the optimal Poincaré constant $\lambda(\mathcal{M}_{\mathrm{Gl}})$ of Glauber dynamics to the optimal Poincaré constant $\lambda(\mathcal{M}_{\rightarrow})$ of scan. Ideally, we might hope for $\lambda(\mathcal{M}_{\rightarrow}) = \Theta(n\lambda(\mathcal{M}_{\mathrm{Gl}}))$. In fact, the best bounds we can prove lose a factor n in either direction so we have a lower bound for $\lambda(\mathcal{M}_{\rightarrow})$ of $\Omega(\lambda(\mathcal{M}_{\mathrm{Gl}}))$, and an upper bound of $O(n^2\lambda(\mathcal{M}_{\mathrm{Gl}}))$. Moreover, for the lower bound we need to assume G has bounded degree.

10.1 Comparing scan to Glauber

Theorem 31. *Suppose G has maximum degree Δ. Let $\mathcal{M}_{\mathrm{Gl}}$ and $\mathcal{M}_{\rightarrow}$ be the Glauber dynamics or systematic scan applied to H-colourings of G for a fixed but arbitrary H. Then $\lambda(\mathcal{M}_{\mathrm{Gl}}) \le 4q^{\Delta+1}\lambda(\mathcal{M}_{\rightarrow})$.*

Proof. Suppose $\sigma \to \sigma'$ is a possible Glauber transition, i.e., $P_{\mathrm{Gl}}(\sigma,\sigma') > 0$. Let i be the unique vertex satisfying $\sigma_i \neq \sigma'_i$. Say that $\tau \in \Omega$ is *between* σ and σ' if $\sigma \to \tau$ is a possible scan transition, and additionally: (i) $\tau_i = \sigma'_i$ and (ii) $\tau_j = \sigma_j$ for all $j \sim i$, where $\sim$ denotes adjacency in G. Denote by $B(\sigma,\sigma')$ the set of states between σ and σ'. Consider a scan transition from state σ, or reverse scan transition from state σ', and denote by $\mathcal{E}_i$ the event that, for all $k \in \{i\} \cup \{j : j \sim i\}$, the colour proposed by Metropolis(k) is $\sigma'(k)$. The following observations are easy to verify.

- Conditioned on $\mathcal{E}_i$, a scan transition from state σ is certain to result in a state $\tau \in B(\sigma,\sigma')$.
- For all $\tau \in B(\sigma,\sigma')$: conditioned on $\mathcal{E}_i$, the probability that reverse scan makes a transition from τ to σ' is equal to the probability that scan makes a transition from σ to τ.
- $\Pr(\mathcal{E}_i) = q^{-(\Delta+1)}$.

It follows from these three observations that

$$\sum_{\tau \in B(\sigma,\sigma')} \min\left\{P_{\rightarrow}(\sigma,\tau), P_{\leftarrow}(\tau,\sigma')\right\} \ge q^{-(\Delta+1)}.$$

Then for any $f : \Omega \to \mathbb{R}$,

$$\begin{aligned}
\mathcal{E}_{\mathcal{M}_{\mathrm{Gl}}}(f,f) &= \frac{1}{2} \sum_{\sigma,\sigma' \in \Omega} \pi(\sigma) P_{\mathrm{Gl}}(\sigma,\sigma')(f(\sigma) - f(\sigma'))^2 \\
&\le \frac{q^{\Delta+1}}{2} \sum_{\sigma,\sigma' \in \Omega} \pi(\sigma) P_{\mathrm{Gl}}(\sigma,\sigma') \sum_{\tau \in B(\sigma,\sigma')} \min\left\{P_{\rightarrow}(\sigma,\tau), P_{\leftarrow}(\tau,\sigma')\right\} (f(\sigma) - f(\sigma'))^2 \\
&\le q^{\Delta+1} \sum_{\sigma,\sigma' \in \Omega} \pi(\sigma) P_{\mathrm{Gl}}(\sigma,\sigma') \\
&\qquad \sum_{\tau \in B(\sigma,\sigma')} \left[P_{\rightarrow}(\sigma,\tau)(f(\sigma) - f(\tau))^2 + P_{\leftarrow}(\tau,\sigma')(f(\tau) - f(\sigma'))^2\right] && (31) \\
&= 2q^{\Delta+1} \sum_{\sigma,\sigma' \in \Omega} \pi(\sigma) P_{\mathrm{Gl}}(\sigma,\sigma') \sum_{\tau \in B(\sigma,\sigma')} P_{\rightarrow}(\sigma,\tau)(f(\sigma) - f(\tau))^2 && (32) \\
&= 2q^{\Delta+1} \sum_{\sigma,\tau \in \Omega} \pi(\sigma) P_{\rightarrow}(\sigma,\tau)(f(\sigma) - f(\tau))^2 \sum_{\sigma' : \tau \in B(\sigma,\sigma')} P_{\mathrm{Gl}}(\sigma,\sigma') \\
&\le 2q^{\Delta+1} \sum_{\sigma,\tau \in \Omega} \pi(\sigma) P_{\rightarrow}(\sigma,\tau)(f(\sigma) - f(\tau))^2 && (33) \\
&= 4q^{\Delta+1}\, \mathcal{E}_{\mathcal{M}_{\rightarrow}}(f,f).
\end{aligned}$$

Inequality (31) applies the fact that $\frac{1}{2}(a-b)^2 \le (a-\xi)^2 + (\xi-b)^2$ for all ξ; and equality (32) uses the fact that Glauber is time reversible, i.e., that $\pi(\sigma)P_{\mathrm{Gl}}(\sigma,\sigma') = \pi(\sigma')P_{\mathrm{Gl}}(\sigma',\sigma)$, for all $\sigma, \sigma' \in \Omega$ and the fact that $B(\sigma, \sigma') = B(\sigma', \sigma)$. Inequality (33) seems crude at first sight, but it is not obvious how to do better: the knowledge of τ does little to constrain σ'. $\square$

The inverse of $\lambda(\mathcal{M})$ is closely related to the mixing time of $\mathcal{M}$. Much is known about the precise relationship between these quantities, see for example the inequalities in [1, 8, 9, 11, 15, 21, 24]. Some known results only apply when $\mathcal{M}$ is reversible, or when the eigenvalues of its transition matrix P are positive. Our survey paper [13] gives inequalities between Poincaré constants and mixing times in both the general case and the reversible case. We will not repeat the details or trace the development of the ideas here, but we mention a few simple facts that are useful for us. Slightly stronger bounds can be obtained with more effort. Let $\widetilde{\mathcal{M}}_{\rightarrow}$ be the continuization of $\mathcal{M}_{\rightarrow}$ as defined in [2] Chapter 2, page 5. Essentially, this is just $\mathcal{M}_{\rightarrow}$ except that the holding time between discrete transitions is exponential with mean 1. It is a classical result (see, e.g., Jerrum [21, p.55, p.63]) that the mixing time of $\widetilde{\mathcal{M}}_{\rightarrow}$ is bounded as follows:

$$\mathrm{Mix}_x(\widetilde{\mathcal{M}}_{\rightarrow}, \varepsilon) \leq \frac{1}{\lambda(\mathcal{M}_{\rightarrow})}(2\ln(1/\varepsilon) + \ln(1/\pi(x))). \tag{34}$$

Combining Equation (34), Theorem 31, and the upper bound $1/\lambda(\mathcal{M}_{\mathrm{Gl}}) = O(n^4)$ from Section 8.1, we get the following.

Theorem 32. *Let H be a fixed connected graph. Let G be the n-vertex path. Let Ω' be the state space of $\mathcal{M}_{\mathrm{Gl}}$. It is either the set of all proper H-colourings of G (if H is not bipartite) or it is one of the two maximum sets of compatible colourings (if H is bipartite). Consider the Markov chain $\widetilde{\mathcal{M}}_{\rightarrow}$ on the state space Ω'.*

$$\mathrm{Mix}(\widetilde{\mathcal{M}}_{\rightarrow}, \varepsilon) = O(n^5 \ln \varepsilon^{-1}).$$

Let $\mathcal{M}_{\mathrm{Gl}}^{\mathrm{ZZ}}$ be the "lazy" version of Glauber dynamics from Page 53 of [21]. In each step, the lazy Markov chain stays where it is with probability 1/2, and otherwise makes the transition specified in the definition of $\mathcal{M}_{\mathrm{Gl}}$. We introduce the lazy chain to keep the eigenvalues positive. See [13] for inequalities which avoid this device. The following inequality from [13] is similar to Proposition 1(ii) of Sinclair [24].

$$\frac{1}{\lambda(\mathcal{M}_{\mathrm{Gl}})} = \frac{1}{\lambda(\mathcal{M}_{\mathrm{Gl}}^{\mathrm{ZZ}})} \leq \max_x \mathrm{Mix}_x\Big(\mathcal{M}_{\mathrm{Gl}}^{\mathrm{ZZ}}, \frac{1}{2e}\Big).$$

Combining this with Theorem 31 and with Equation (34) we find that, for bounded-degree graphs G, the mixing time of $\widetilde{\mathcal{M}}_{\rightarrow}$ is at most $O(n)$ times the mixing time of $\mathcal{M}_{\mathrm{Gl}}^{\mathrm{ZZ}}$. Perhaps this result can be improved by a factor of n^2.

10.2 Comparing Glauber to scan

Theorem 33. *Suppose G is arbitrary. Let $\mathcal{M}_{\mathrm{Gl}}$ and $\mathcal{M}_{\rightarrow}$ be the Glauber dynamics and systematic scan applied to H-colourings of G for a fixed but arbitrary H. Then $\lambda(\mathcal{M}_{\rightarrow}) \leq n^2 q \lambda(\mathcal{M}_{\mathrm{Gl}})$.*

Proof. Let $\sigma, \sigma' \in \Omega$ be a pair of states for which $P_{\rightarrow}(\sigma, \sigma') > 0$. There is a natural canonical path $\gamma_{\sigma,\sigma'} = (\sigma = \tau^0 \rightarrow \tau^1 \rightarrow \cdots \rightarrow \tau^n = \sigma')$ from σ to σ' using Glauber transitions, in which

τ^{i-1} differs from τ^i (if at all) only at vertex i. According to [8, Thm 2.1], the quantity we need to bound is

$$\begin{aligned} A &= \frac{1}{\pi(\tau)P_{\mathrm{Gl}}(\tau,\tau')} \sum_{\sigma,\sigma':(\tau,\tau')\in\gamma_{\sigma,\sigma'}} \pi(\sigma)P_{\rightarrow}(\sigma,\sigma')\,|\gamma_{\sigma,\sigma'}| \\ &= n^2 q \sum_{\sigma,\sigma':(\tau,\tau')\in\gamma_{\sigma,\sigma'}} P_{\rightarrow}(\sigma,\sigma'), \end{aligned}$$

where we have used the facts that π is uniform, $|\gamma_{\sigma,\sigma'}| = n$, and $P_{\mathrm{Gl}}(\tau,\tau') = 1/nq$. (Diaconis and Saloff-Coste state their theorem for time-reversible MCs, but their proof does not use time reversibility.) We shall demonstrate that $A \le n^2 q$, from which it follows that $\lambda(\mathcal{M}_{\rightarrow}) \le n^2 q \lambda(\mathcal{M}_{\mathrm{Gl}})$.

Regard τ and τ' as fixed, and suppose τ and τ' differ at vertex i. Denote by $\mathcal{E}_1^\sigma$ the event that the sequence

$$\mathrm{Metropolis}(1), \mathrm{Metropolis}(2), \ldots, \mathrm{Metropolis}(i-1)$$

takes σ to τ, and by $\mathcal{E}_2^{\sigma'}$ the event that

$$\mathrm{Metropolis}(i+1), \mathrm{Metropolis}(i+2), \ldots, \mathrm{Metropolis}(n)$$

takes τ' to σ'. Then $P_{\rightarrow}(\sigma,\sigma') \le \Pr(\mathcal{E}_1^\sigma \wedge \mathcal{E}_2^{\sigma'}) = \Pr(\mathcal{E}_1^\sigma)\Pr(\mathcal{E}_2^{\sigma'})$, and

$$\sum_{\sigma,\sigma':(\tau,\tau')\in\gamma_{\sigma,\sigma'}} P_{\rightarrow}(\sigma,\sigma') \le \sum_{\sigma} \Pr(\mathcal{E}_1^\sigma) \sum_{\sigma'} \Pr(\mathcal{E}_2^{\sigma'}).$$

The second sum above is clearly bounded by 1, since the events $\{\mathcal{E}_2^{\sigma'} : \sigma' \in \Omega\}$ are disjoint. In fact the first sum is also bounded by 1, since $\Pr(\mathcal{E}_1^\sigma)$ is equal to the probability that the sequence

$$\mathrm{Metropolis}(i-1), \mathrm{Metropolis}(i-2), \ldots, \mathrm{Metropolis}(1)$$

takes τ to σ. So the terms of the first sum may also be viewed as probabilities of disjoint events. Thus $A \le n^2 q$, as claimed. □

Combining Theorem 33 with inequalities of Diaconis and Stroock [9] and Sinclair [24] we get

$$\mathrm{Mix}_x(\mathcal{M}_{\mathrm{Gl}}^{\mathrm{ZZ}}, \varepsilon) \le n^2 q \frac{1}{\lambda(\mathcal{M}_{\rightarrow})} \ln \frac{1}{\varepsilon\pi(x)}. \tag{35}$$

This can be combined with the upper bound

$$\frac{1}{\lambda(\mathcal{M}_{\rightarrow})} \le \frac{2\big(\max_x \mathrm{Mix}_x(\mathcal{M}_{\rightarrow}, \frac{1}{e})\big)^2}{(\frac{1}{2} - \frac{1}{e})^2}. \tag{36}$$

The square of the mixing time in Equation (36) is necessary in the general non-reversible case (see [13]) though of course better inequalities might apply to the particular chain $\mathcal{M}_{\rightarrow}$. Combining Equations (35) and (36), we get a weak inequality which shows that the mixing time of (lazy) Glauber dynamics is at most $O(n^3)$ times the square of the mixing time of systematic scan.

Note that the proofs of Theorems 31 and 33 are actually about Dirichlet forms rather than about Poincaré constants, so the same inequalities apply to the log-Sobolev constant.

Acknowledgement

The authors thank Persi Diaconis for comments on a draft of this article.

References

[1] D. Aldous, Some inequalities for reversible Markov chains, Journal of the London Math Society (Second Series) **25** (1982), 564–576.

[2] D. Aldous and J. Fill, Reversible Markov chains and random walks on graphs. `http://oz.berkeley.edu/users/aldous/RWG/book.html`

[3] Y. Amit, Convergence properties of the Gibbs sampler for perturbations of Gaussians, *Annals of Statistics* **24(1)** (1996), 122–140.

[4] I. Benjamini, N. Berger, C. Hoffman and E. Mossel, Mixing times of the biased card shuffling and the asymmetric exclusion process, *Annals of Applied Probability* to appear.

[5] R. Bubley and M. Dyer, Path coupling: A technique for proving rapid mixing in Markov chains, *Proceedings of the 38th IEEE Annual Symposium on Foundations of Computer Science* (IEEE, Los Alamitos, 1997), 223–231.

[6] C. Cooper, M. Dyer and A. Frieze, On Markov chains for randomly H-colouring a graph, *Journal of Algorithms,* **39(1)** (2001), 117–134.

[7] P. Diaconis and A. Ram, Analysis of systematic scan metropolis algorithm using Iwahori-Hecke Algebra techniques, *Michigan Mathematics Journal* **48** (2000), 157–190.

[8] P. Diaconis and L. Saloff-Coste, Comparison theorems for reversible Markov chains, *Annals of Applied Probability* **3** (1993), 696–730.

[9] P. Diaconis and D. Stroock, Geometric bounds for eigenvalues of Markov chains, *Annals of Applied Probability* **1(1)** (1991), 36-61.

[10] R. Durrett, *Probability: Theory and Examples* Brooks/Cole Publishing Company, 1991.

[11] M. Dyer, A. Frieze and M. Jerrum, On counting independent sets in sparse graphs, *SIAM Journal on Computing* **33(5)** (2002), 1527–1541.

[12] M. Dyer, L.A. Goldberg, C. Greenhill, M. Jerrum and M. Mitzenmacher, An extension of path coupling and its application to the Glauber dynamics for graph colourings, *SIAM Journal on Computing* **30(6)** (2001), 1962–1975.

[13] M. Dyer, L.A. Goldberg, M. Jerrum, R. Martin, Notes on comparison.

[14] M. Dyer, M. Jerrum and E. Vigoda, Rapidly mixing Markov chains for dismantleable constraint graphs. In J. Nesetril and P. Winkler, editors, *Proceedings of a DIMACS/DIMATIA Workshop on Graphs, Morphisms and Statistical Physics*, 2004.

[15] J.A. Fill, Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process, Annals of Applied Probability **1(1)** (1991), 62–87.

[16] G.S. Fishman, Coordinate selection rules for Gibbs sampling, *Annals of Applied Probability* **6** (1996), 444–465.

[17] R.J. Glauber, Time-dependent statistics of the Ising model, *Journal of Mathematical Physics* **4** (1963), 294–307.

[18] L.A. Goldberg, S. Kelk and M. Paterson, The complexity of choosing an H-colouring (nearly) uniformly at random, To appear in *SIAM Journal on Computing.*

[19] L.A. Goldberg, R. Martin and M. Paterson, Random sampling of 3-colourings in $\mathbb{Z}^2$, To appear in *Random Structures and Algorithms.*

[20] M. Jerrum, A very simple algorithm for estimating the number of k-colorings of a low-degree graph, *Random Structures and Algorithms* **7** (1995), 157–165.

[21] M. Jerrum, Counting, sampling and integrating: algorithms and complexity, Birkhäuser, 2003.

[22] C. Kenyon and D. Randall, personal communication.

[23] M. Luby, D. Randall and A.J. Sinclair, Markov chain algorithms for planar lattice structures. *SIAM Journal on Computing* **31** (2001), 167–192.

[24] A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow, *Combinatorics, Probability and Computing* **1** (1992), 351–370.

[25] A. Sinclair and M. Jerrum, Approximate counting, uniform generation and rapidly mixing Markov chains, Information and Computation **82** (1989), 93–133.

[26] J. van den Berg, A uniqueness condition for Gibbs measures, with application to the 2-dimensional Ising antiferromagnet. *Communications in Mathematical Physics* **152** (1993), 161–166.

[27] D. B. Wilson, Mixing times of lozenge tiling and card shuffling Markov chains, *Annals of Applied Probability* **14(1)** (2004), 274–325.