Chapter 4

Coupling and colourings

The outline of our programme is now clear: in order to count (approximately) it is enough to be able to sample (almost) uniformly; in order to sample we may simulate an appropriately defined MC. For this approach to be feasible, however, it is important that the MC in question is "rapidly mixing," i.e., that it converges to near-equilibrium in time polynomial (hopefully of small degree) in the size of the problem instance. Since the state space is usually of exponential size as a function of the problem size — think of the number of matchings in a graph as a function of the size of the graph — this is a distinctly non-trivial requirement. We shall presently formalise the rate of convergence to equilibrium in terms of the "mixing time" of the MC. The classical theory of MCs has little to say concerning non-asymptotic bounds on mixing time, and most of the techniques we use have been specially developed for the task in hand. However, there is one classical device, namely coupling, that can be applied in certain situations. As it is the most elementary approach to bounding mixing times, we study it first.

4.1 Colourings of a low-degree graph

Anil Kumar and Ramesh [3] present persuasive evidence that the coupling argument is not applicable to the MC on matchings that was defined at the end of the previous chapter. We therefore use a somewhat simpler example, namely colourings of a low-degree graph. Let $G = (V, E)$ be an undirected graph, and Q a set of q colours. A (proper) q-colouring of G is a an assignment $\sigma : V \to Q$ of colours to the vertices of G such that $\sigma(u) \neq \sigma(v)$ for all edges $\{u, v\} \in E$. In general, even deciding existence of a q-colouring in G is computationally intractable, so we need to impose

1. Select a vertex $v \in V$, u.a.r.
2. Select a colour $c \in Q \setminus X_0(\mathit{\Gamma}(v))$, u.a.r.
3. $X_1(v) \leftarrow c$ and $X_1(u) \leftarrow X_0(u)$ for all $u \neq v$.

Figure 4.1: Trial defining an MC on q-colourings.

some condition on G and q.

Denote by $\mathit{\Delta}$ the maximum degree[1] of any vertex in G. Brooks' theorem asserts that a q-colouring exists when $q \geq \mathit{\Delta}$, provided $\mathit{\Delta} \geq 3$ and G does not contain $K_{\mathit{\Delta}+1}$ as a connected component [7, 9].[2] The proof of Brooks' theorem is effective, and yields a polynomial-time algorithm for constructing a q-colouring. It is also best possible in the (slightly restricted) sense that there are pairs, for example $q = 3$, $\mathit{\Delta} = 4$, which just fail the condition of the theorem, and for which the problem of deciding q-colourability is NP-complete, even when restricted to graphs of maximum degree $\mathit{\Delta}$. So if we are aiming at an efficient sampling procedure for q-colourings we should certainly assume $q \geq \mathit{\Delta}$. Moreover, to approximate the number of q-colourings using the reduction of Exercise 3.5 we need to assume further that $q > \mathit{\Delta}$. Before we complete the work of this section, we shall need to strengthen this condition still further.

So let $G = (V, E)$ be a graph of maximum degree $\mathit{\Delta}$ and let $\mathit{\Omega}$ denote the set of all q-colourings of G, for some $q > \mathit{\Delta}$. Denote by $\mathit{\Gamma}(v) = \{u : \{u, v\} \in E(G)\}$ the set of vertices in G that are adjacent to v. Consider the (time-homogeneous) MC (X_t) on $\mathit{\Omega}$ whose transitions are defined by the experimental trial presented in Figure 4.1. Here we are considering a colouring as a function $V \to Q$, so $X_0(u)$ denotes the colour of vertex u in the initial state, and $X_0(\mathit{\Gamma}(v)) = \{X_0(u) : u \in \mathit{\Gamma}(v)\}$ denotes the set of all colours applied to neighbours of v. Note that the assumption $q > \mathit{\Delta}$ makes it easy to construct a valid initial state X_0.

Exercises 4.1. 1. Prove that the above MC is irreducible (and hence ergodic) under the (stronger) assumption $q \geq \mathit{\Delta} + 2$. Further prove, using Lemma 3.7, that its (unique) stationary distribution is uniform over $\mathit{\Omega}$.

2. [Alan Sokal.] Exhibit a sequence of connected graphs of increasing

[1]The *degree* of a vertex is the number of edges incident at that vertex.
[2]K_r denotes the complete graph on r vertices.

size, with $\Delta = 4$, such that the above MC fails to be irreducible when $q = 5$. (Hint: as a starting point, construct a "frozen" 5-colouring of the infinite square lattice, i.e., the graph with vertex set $\mathbb{Z}\times\mathbb{Z}$ and edge set $\{(i,j),(i',j') : |i-i'|+|j-j'| = 1\}$. The adjective "frozen" applied to a state is intended to indicate that the only transition available from the state is a loop (with probability 1) to the same state.)

3. Design an MC on q-colourings of an arbitrary graph G of maximum degree Δ that is ergodic, provided only that $q \geq \Delta+1$. The MC should be easily implementable, otherwise there is no challenge! (Hint: use transitions based on edge updates rather than vertex updates.)

We shall show that (X_t) is rapidly mixing, provided $q \geq 2\Delta + 1$, which we assume from now on. (The reader may be assured that this is the very last time we shall strengthen the lower bound on the number of colours!) This result will provide us with a simple and efficient sampling procedure for q-colourings in low-degree graphs.

Suppose (X_t) is any ergodic MC on countable state space Ω, with transition matrix P and initial state $X_0 = x \in \Omega$. For $t \in \mathbb{N}$, the distribution of X_t (the t step distribution) is naturally denoted $P^t(x,\cdot)$. Let π denote the the stationary distribution of the MC, i.e., the limit of $P^t(x,\cdot)$ as $t \to \infty$. Recall the definition of total variation distance from (3.2). We measure the rate of convergence to stationarity of (X_t) by its *mixing time* (from initial state x):

$$\tau_x(\varepsilon) := \min\{t : \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon\}. \tag{4.1}$$

Lemma 4.2. *The total variation distance $\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}}$ of the t-step distribution from stationarity is a non-increasing function of t.*

Exercise 4.3. Prove Lemma 4.2. (A proof is given at the end of the chapter.)

In the light of Lemma 4.2, the following definition of mixing time is equivalent to (4.1):

$$\tau_x(\varepsilon) := \min\{t : \|P^s(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon,\ \text{for all } s \geq t\}.$$

In other words, once the total variation distance becomes smaller than ε it stays smaller than ε.

Often we would like to make a statement about mixing time that is independent of the initial state, in which case we take a worst-case view and write

$$\tau(\varepsilon) = \max_{x\in\Omega} \tau_x(\varepsilon);$$

we shall refer to $\tau(\varepsilon)$ simply as the *mixing time*.

Remark 4.4. Sometimes the further simplification of setting ε to some constant, say $\varepsilon = \frac{1}{4}$, is made. The justification for this runs as follows. If τ is the first time t at which $\|P^t(x,\,\cdot\,) - \pi\|_{\mathrm{TV}} \le \frac{1}{4}$, then it can be shown [2, Chap. 2, Lemma 20] that $\|P^{k\tau}(x,\,\cdot\,) - \pi\|_{\mathrm{TV}} \le 2^{-k}$ for every $k \in \mathbb{N}$.

Our aim in the next section is to show that the mixing time $\tau(\varepsilon)$ of the MC on colourings is bounded by a polynomial in n and $\log \varepsilon^{-1}$.

Proposition 4.5. *Suppose G is a graph on n vertices of maximum degree Δ. Assuming $q \ge 2\Delta + 1$, the mixing time $\tau(\varepsilon)$ of the MC of Figure 4.1 is bounded above by*

$$\tau(\varepsilon) \le \frac{q - \Delta}{q - 2\Delta}\, n \ln\left(\frac{n}{\varepsilon}\right).$$

Taking the instance size n into account is a prominent feature of applications of MCs in computer science, especially as compared with classical Markov chain theory. Observe that Proposition 4.5, combined with Proposition 3.4, implies the existence of an FPRAS for q-colourings in graphs of low enough degree.

Corollary 4.6. *Suppose G is a connected graph of maximum degree Δ, and $q \ge 2\Delta + 1$. Then there is an FPRAS for counting q-colourings in G. Denote by n the number of vertices in G and by m the number of edges. Then the running time of this FPRAS as a function of n, m and the error tolerance ε (regarding Δ and q as fixed) is bounded by $cnm^2\varepsilon^{-2}\max\{\ln(m/\varepsilon), 1\}$ for some constant c.*

4.2 Bounding mixing time using coupling

Coupling as a proof technique was discovered by Doeblin in the 1930s. However, its more recent popularity as a tool for bounding mixing time owes much to Aldous. Actually, we shall be using only a restricted form of coupling, namely Markovian coupling.

We start with a ground (time homogeneous) MC (Z_t) with state space Ω and transition matrix P. A *(Markovian) coupling* for (Z_t) is an MC (X_t, Y_t) on $\Omega \times \Omega$, with transition probabilities defined by:

$$\begin{aligned} \Pr[X_1 = x' \mid X_0 = x, Y_0 = y] &= P(x, x'), \\ \Pr[Y_1 = y' \mid X_0 = x, Y_0 = y] &= P(y, y'). \end{aligned} \tag{4.2}$$

Equivalently, with $\widehat{P} : \Omega^2 \to \Omega^2$ denoting the transition matrix of the coupling,

$$\sum_{y' \in \Omega} \widehat{P}((x,y),(x',y')) = P(x,x'),$$
$$\sum_{x' \in \Omega} \widehat{P}((x,y),(x',y')) = P(y,y').$$

Thus, the sequence of r.v.'s (X_t) viewed in isolation forms an MC with transition matrix P, as does the sequence (Y_t).

The easy way to achieve (4.2) would be to assume independence of (X_t) and (Y_t), i.e., that

$$\widehat{P}((x,y),(x',y')) = P(x,x')P(y,y').$$

But this is not necessary, and for our application not desirable. Instead, we are after some correlation that will tend to bring (X_t) and (Y_t) together (whatever their initial states) so that $X_t = Y_t$ for some quite small t. Note that once $X_t = Y_t$, we can arrange quite easily for X_s to be equal to Y_s, for all $s \geq t$, while continuing to satisfy (4.2): just choose a transition from X_s and let Y_s copy it.

The following simple lemma, which is the basis of the coupling method, was perhaps first made explicit by Aldous [1, Lemma 3.6]; see also Diaconis [18, Chap. 4, Lemma 5].

Lemma 4.7 (Coupling Lemma). *Let (X_t, Y_t) be any coupling, satisfying (4.2), based on a ground MC (Z_t) on Ω. Suppose $t : [0,1] \to \mathbb{N}$ is a function satisfying the condition: for all $x, y \in \Omega$, and all $\varepsilon > 0$*

$$\Pr[X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} \mid X_0 = x, Y_0 = y] \leq \varepsilon.$$

Then the mixing time $\tau(\varepsilon)$ of (Z_t) is bounded above by $t(\varepsilon)$.

Proof. Denote by P the transition matrix of (Z_t). Let $A \subseteq \Omega$ be arbitrary. Let $X_0 = x \in \Omega$ be fixed, and Y_0 be chosen according to the stationary distribution π of (Z_t). For any $\varepsilon \in (0,1)$ and corresponding $t = t(\varepsilon)$,

$$\begin{aligned}
P^t(x,A) &= \Pr[X_t \in A] \\
&\geq \Pr[X_t = Y_t \wedge Y_t \in A] \\
&= 1 - \Pr[X_t \neq Y_t \vee Y_t \notin A] \\
&\geq 1 - (\Pr[X_t \neq Y_t] + \Pr[Y_t \notin A]) \\
&\geq \Pr(Y_t \in A) - \varepsilon \\
&= \pi(A) - \varepsilon.
\end{aligned}$$

1. Select a vertex $v \in V$ u.a.r.
2. Select a pair of colours (c_x, c_y) from some joint distribution on $\big(Q \setminus X_0(\Gamma(v))\big) \times \big(Q \setminus Y_0(\Gamma(v))\big)$ that has the "correct" marginal distributions; specifically, the distribution of c_x (respectively c_y) should be uniform over $Q \setminus X_0(\Gamma(v))$ (respectively $Q \setminus Y_0(\Gamma(v))$. This joint distribution will be chosen so as to maximise $\Pr[c_x = c_y]$.
3. Set $X_1(v) \leftarrow c_x$ and $Y_1(v) \leftarrow c_y$.

Figure 4.2: A coupling for the MC on colourings

Hence, by the second part of definition (3.2), $\|P^t(x,\,\cdot\,) - \pi\|_{\mathrm{TV}} \leq \varepsilon$. □

Remark 4.8. Actually we established the stronger conclusion

$$\|P^t(x,\,\cdot\,) - P^t(y,\,\cdot\,)\|_{\mathrm{TV}} \leq \varepsilon, \quad \text{for all pairs } x, y \in \Omega.$$

This slightly different notion of l_1-convergence corresponds to a slightly different notion of mixing time. This new mixing time has certain advantages, notably submultiplicativity: see Aldous and Fill [2] for more detail.

Let's now see how these ideas may be applied to the q-colouring MC of Figure 4.1. We need to define a coupling on Ω^2 such that the projections onto the first and second coordinates are faithful copies of the original MC in the sense of (4.2). Moreover, we wish the coupling to *coalesce*, i.e., reach a state where $X_t = Y_t$, as soon as possible. Figure 4.2 presents what seems at first sight to be a reasonable proposal. Note that if you hide the random variable Y_1 then the companion random variable X_1 is distributed exactly as if we had used the trial presented in Figure 4.1. (By symmetry, a similar statement could be made about Y_1.) Thus the coupling condition (4.2) is satisfied.

We have argued that the coupling in Figure 4.2 is correct, but how efficient is it? Intuitively, provided we can arrange for $\Pr[c_x = c_y]$ in step 2 to be large, we ought to reach a state with $X_t = Y_t$ (i.e., coalescence) in not too many steps. The Coupling Lemma will then provide a good upper bound on mixing time. In order to understand what is involved in maximising $\Pr[c_x = c_y]$, the following exercise may be useful.

Exercise 4.9. Suppose that $Q = \{0, 1, \dots, 6\}$, $X_0(\Gamma(v)) = \{3, 6\}$ and $Y_0(\Gamma(v)) = \{4, 5, 6\}$. Thus the sets of legal colours for v in X_1 and Y_1

are $c_x \in \{0,1,2,4,5\}$ and $c_y \in \{0,1,2,3\}$, respectively. Construct a joint distribution for (c_x, c_y) such that c_x is uniform on $\{0,1,2,4,5\}$, c_y is uniform on $\{0,1,2,3\}$, and $\Pr[c_x = c_y] = \frac{3}{5}$. Show that your construction is optimal.

The best that can be done in general is as follows.

Lemma 4.10. *Let U be a finite set, A, B be subsets of U, and Z_a, Z_b be random variables, taking values in U. Then there is a joint distribution for Z_a and Z_b such that Z_a (respectively Z_b) is uniform and supported on A (respectively B) and, furthermore,*

$$\Pr[Z_a = Z_b] = \frac{|A \cap B|}{\max\{|A|, |B|\}}$$

Exercise 4.11. Prove Lemma 4.10 and show that the result is best possible. (Assuming your construction in Exercise 4.9 is reasonably systematic, it should be possible to adapt it to the general situation.)

Remark 4.12. The term "coupling" does not have a precise agreed meaning, but its general sense is the following. A pair or perhaps a larger collection of r.v's is given. A coupling is a joint distribution of the several variables that has the correct marginals — i.e., each r.v., when observed independently of the others, has the correct probability distribution — but, taken together, the variables are seen to be correlated. Usually the correlation aims to "bring the r.v's closer together" in some sense. Lemma 4.10 is a special example of an optimal coupling of two r.v's that Lindvall calls the *γ-coupling* [42, §I.5]. The coupling of MCs, as captured in condition (4.2), is another example of the concept.

We are now well prepared for the main result of the chapter.

Proof of Proposition 4.5. We analyse the coupling of Figure 4.2 using the joint distribution for the colour-pair (c_x, c_y) that is implicit in Lemma 4.10. Thus, letting

$$\xi := |Q \setminus X_0(\Gamma(v))| \qquad (= \#\text{ legal colours for } v \text{ in } X_1),$$
$$\eta := |Q \setminus Y_0(\Gamma(v))| \qquad (= \#\text{ legal colours for } v \text{ in } Y_1),$$

and

$$\zeta := \big|\big(Q \setminus X_0(\Gamma(v))\big) \cap \big(Q \setminus Y_0(\Gamma(v))\big)\big| \quad (= \#\text{ common legal colours}),$$

the probability that the same colour is chosen for both X_1 and Y_1 in step 2 is just

$$\Pr[c_x = c_y] = \frac{\zeta}{\max\{\xi, \eta\}}. \tag{4.3}$$

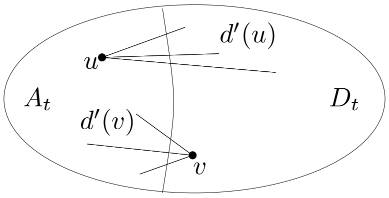

Figure 4.3: Two ways to count the edges spanning the cut (A_t, D_t).

Consider the situation that obtains after the coupling has been run for t steps. Let $A_t \subseteq V$ be the set of vertices on which the colourings X_t and Y_t agree, and $D_t = V \setminus A_t$ be the set on which they disagree. Let $d'(v)$ denote the number of edges incident at vertex v that have one endpoint in A_t and one in D_t. Clearly,

$$\sum_{v \in A_t} d'(v) = \sum_{u \in D_t} d'(u) = m',$$

where m' is the number of edges of G that span A_t and D_t. (The situation is visualised in Figure 4.3.) We want to prove that the disagreement set D_t tends to get smaller and smaller.

In one transition, the size of the disagreement set D_t changes by at most one. We therefore need to consider just three cases: increasing/decreasing by one or remaining constant. In fact, we just need to compute the probability of the first two, since the third can be got by complementation.

Consider first the probability that $|D_{t+1}| = |D_t| + 1$. For this event to occur, the vertex v selected in step 1 must lie in A_t, and the new colours c_x and c_y selected in step 2 must be different. Observing that the quantities ξ, η and ζ satisfy the linear inequalities

$$\begin{aligned} \xi - \zeta &\leq d'(v), \\ \eta - \zeta &\leq d'(v), \quad \text{and} \\ \xi, \eta &\geq q - \mathit{\Delta}, \end{aligned} \tag{4.4}$$

we deduce, from (4.3), that

$$\begin{aligned} \Pr[c_x = c_y] &\geq \frac{\max\{\xi,\eta\} - d'(v)}{\max\{\xi,\eta\}} \\ &\geq 1 - \frac{d'(v)}{q - \mathit{\Delta}}, \end{aligned}$$

conditional on v being selected in step (1). Thus

$$\begin{aligned} \Pr\big[|D_{t+1}| = |D_t| + 1\big] &= \frac{1}{n} \sum_{v \in A_t} \Pr\big[c_x \neq c_y \mid v \text{ selected}\big] \\ &\leq \frac{1}{n} \sum_{v \in A_t} \frac{d'(v)}{q - \Delta} = \frac{m'}{(q-\Delta)n}. \end{aligned} \tag{4.5}$$

Now consider the probability that $|D_{t+1}| = |D_t| - 1$. For this event to occur, the vertex v selected in step 1 must lie in D_t, and the new colours c_x and c_y selected in step 2 must be the same. The analogues of inequalities (4.4) in this case are

$$\begin{aligned} \xi - \zeta &\leq \Delta - d'(v), \\ \eta - \zeta &\leq \Delta - d'(v), \quad \text{and} \\ \xi, \eta &\geq q - \Delta. \end{aligned}$$

Proceeding as in the previous case,

$$\begin{aligned} \Pr[c_x = c_y] &\geq \frac{\max\{\xi,\eta\} - \Delta + d'(v)}{\max\{\xi,\eta\}} \\ &= 1 - \frac{\Delta - d'(v)}{\max\{\xi,\eta\}} \\ &\geq \frac{q - 2\Delta + d'(v)}{q - \Delta}, \end{aligned}$$

conditional on v being selected in step (1). Hence

$$\begin{aligned} \Pr\big[\,|D_{t+1}| = |D_t| - 1\big] &\geq \frac{1}{n} \sum_{v \in D_t} \frac{q - 2\Delta + d'(v)}{q - \Delta} \\ &\geq \frac{q - 2\Delta}{(q-\Delta)n}\,|D_t| + \frac{m'}{(q-\Delta)n} \end{aligned} \tag{4.6}$$

Define

$$a = \frac{q - 2\Delta}{(q-\Delta)n} \quad \text{and} \quad b = b(m') = \frac{m'}{(q-\Delta)n},$$

so that $\Pr\big[\,|D_{t+1}| = |D_t| + 1\big] \leq b$ and $\Pr\big[\,|D_{t+1}| = |D_t| - 1\big] \geq a\,|D_t| + b$. Provided $a > 0$, i.e., $q > 2\Delta$, the size of the set D_t tends to decrease with t, and hence, intuitively at least, the event $D_t = \emptyset$ should occur with high probability for some $t \leq T$ with T not too large. Since $D_t = \emptyset$ is precisely the event that coalescence has occurred, it only remains to confirm this intuition, and quantify the rate at which D_t converges to the empty set. From equations (4.5) and (4.6),

$$\begin{aligned} \mathrm{E}\left[\,|D_{t+1}|\;\middle|\; D_t\right] &\le b(|D_t|+1) + (a|D_t|+b)(|D_t|-1) \\ &\qquad + (1-a|D_t|-2b)|D_t| \\ &= (1-a)|D_t|. \end{aligned}$$

Thus $\mathrm{E}\,|D_t| \le (1-a)^t|D_0| \le (1-a)^t n$, and, because $|D_t|$ is a non-negative integer random variable, $\Pr[\,|D_t| \ne 0] \le n(1-a)^t \le ne^{-at}$. Note that $\Pr[D_t \ne \emptyset] \le \varepsilon$, provided $t \ge a^{-1}\ln(n\varepsilon^{-1})$, establishing the result. □

Remark 4.13. With a little care, the argument can be pushed to $q = 2\Delta$, though the bound on mixing time worsens by a factor of about n^2. (The r.v. D_t behaves in the boundary case rather like an unbiased random walk, and therefore its expected time to reach the origin $D_t = 0$ is longer; refer, e.g., to Dyer and Greenhill [24], in particular their Theorem 2.1.)

The (direct) coupling technique described here has been used in a number of other applications, such as approximately counting independent sets in a low-degree graph (Luby and Vigoda [46])[3] and estimating the volume of a convex body (Bubley, Dyer and Jerrum [15]).[4] In practice, the versatility of the approach is limited by our ability to design couplings that work well in situations of algorithmic interest. The next section reports on a new technique that promises to extend the effective range of the coupling argument by providing us with a powerful design tool.

4.3 Path coupling

The coupling technique described and illustrated in the previous section is conceptually very simple and appealing. However, in applying the method to concrete situations we face a technical difficulty, which began to surface even in §4.2: how do we encourage (X_t) and (Y_t) to coalesce, while satisfying the demanding constraints (4.2)? Path coupling is an engineering solution to this problem, invented by Bubley and Dyer [11, 12]. Their idea is to define the coupling only on pairs of "adjacent" states, for which the task of satisfying (4.2) is relatively easy, and then to extend the coupling to arbitrary

[3]Though the subsequent journal article [47] uses the more sophisticated path coupling method, which will be described presently.

[4]The latter application draws inspiration from Lindvall and Rodgers's [43] idea of coupling diffusions by reflection.

1. Select $p \in [n-1]$ according to the distribution f, and $r \in \{0,1\}$ u.a.r.
2. If $r = 1$ and $X_0 \circ (p, p+1) \in \Omega$, then $X_1 := X_0 \circ (p, p+1)$; otherwise, $X_1 := X_0$.

Figure 4.4: Trial defining an MC on linear extensions of a partial order $\prec$.

pairs of states by composition of adjacent couplings along a path. The approach is not entirely distinct from classical coupling, and the Coupling Lemma still plays a vital role.

We illustrate path coupling in the context of a MC on linear extensions of a partial order. We are given a partially ordered set $(V, \prec)$, where $V = [n] = \{0, 1, \ldots, n-1\}$. Denote by $\operatorname{Sym} V$ the symmetric group on V. We are interested in sampling, u.a.r., a member of the set

$$\Omega = \bigl\{ g \in \operatorname{Sym} V : g(i) \prec g(j) \Rightarrow i \leq j, \text{ for all } i, j \in V \bigr\}$$

of linear extensions of $\prec$. In forming a mental picture of the the set Ω, the following characterisation may be helpful: $g \in \Omega$ iff the linear order

$$g(0) \sqsubset g(1) \sqsubset \cdots \sqsubset g(n-1) \tag{4.7}$$

extends, or is consistent with, the partial order $\prec$.

As usual, we propose to sample from Ω by constructing an ergodic MC on state space Ω, whose stationary distribution is uniform. The transitions from one linear extension to another are obtained by pre-composing the current linear extension with a random transition $(p, p+1)$. Instead of selecting $p \in [n-1]$ uniformly, we select p from a probability distribution f on $[n-1]$ that gives greater weight to values near the centre of the range. It is possible that this refinement actually reduces the mixing time; in any case, it leads to a simplification of the proof. Formally, the transition probabilities of the MC are defined by the experimental trial presented in Figure 4.4. Note that composition "$\circ$" is to be read right to left, so that (assuming $r = 1$): $X_1(p) = X_0(p+1)$, $X_1(p+1) = X_0(p)$ and $X_1(i) = X_0(i)$, for all $i \notin \{p, p+1\}$.

Provided the probability distribution f is supported on the whole interval $[n-1]$, this MC is irreducible and aperiodic. It is easy to verify, for example using Lemma 3.7, that the stationary distribution of the MC is uniform. As in §3.3, the explicit loop probability of $\frac{1}{2}$ is introduced mainly for convenience in the proof. However, some such mechanism for destroying

1. Select $p \in [n-1]$ according to the distribution f, and $r_x \in \{0,1\}$ u.a.r. If $j - i = 1$ and $p = i$, set $r_y := 1 - r_x$; otherwise, set $r_y := r_x$.

2. If $r_x = 1$ and $X_0 \circ (p, p+1) \in \Omega$ then set $X_1 := X_0 \circ (p, p+1)$; otherwise, set $X_1 := X_0$.

3. If $r_y = 1$ and $Y_0 \circ (p, p+1) \in \Omega$ then set $Y_1 := Y_0 \circ (p, p+1)$; otherwise, set $Y_1 := Y_0$.

Figure 4.5: A possible coupling for the MC on linear extensions.

periodicity is necessary in any case if we wish to treat the empty partial order consistently.

Our analysis of the mixing time of the MC using path coupling will closely follow that of Bubley and Dyer [13]. To apply path coupling, we need first to decide on an adjacency structure for the state space Ω. In this instance we decree that two states g and g' (linear extensions of $\prec$) are adjacent iff $g' = g \circ (i, j)$ for some transposition (i, j) with $0 \leq i < j \leq n-1$; in this case, the distance $d(g, g')$ from g to g' is defined to be $j - i$. Note that the notions of adjacency and distance are symmetric with respect to interchanging g and g', so we can regard this imposed adjacency structure as a weighted, undirected graph on vertex set Ω; let us refer to this structure as the adjacency graph. It is easily verified that the shortest path in the adjacency graph between two adjacent states is the direct one using a single edge. Thus d may be extended to a metric on Ω by defining $d(g, h)$, for arbitrary states g and h, to be the length of a shortest path from g to h in the adjacency graph.

Next we define the coupling. We need to do this just for adjacent states, as the extension of the coupling via shortest paths to arbitrary pairs of states will be automatic. Suppose that $(X_0, Y_0) \in \Omega^2$ is a pair of states related by $Y_0 = X_0 \circ (i, j)$ for some transposition (i, j) with $0 \leq i < j \leq n - 1$. then the transition to (X_1, Y_1) in the coupling is defined by the experimental trial presented in Figure 4.5. We need to show:

Lemma 4.14. *For adjacent states* X_0 *and* Y_0*,*

$$\mathrm{E}\left[d(X_1, Y_1) \mid X_0, Y_0\right] \leq \varrho\, d(X_0, Y_0), \tag{4.8}$$

where $\varrho < 1$ *is a constant depending on* f*. For a suitable choice for* f*, one has* $\varrho = 1 - \alpha$*, where* $\alpha = 6/(n^3 - n)$*.*

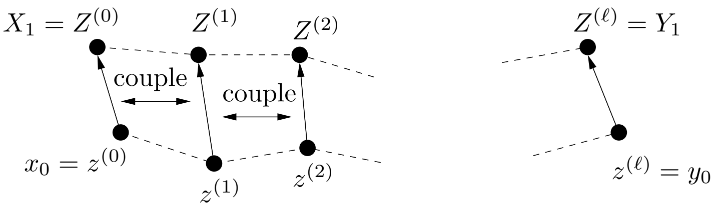

Figure 4.6: Extending a coupling along a shortest path

Informally, Lemma 4.14 says that distance between pairs of states in the coupled process tends to decrease: exactly the situation we encountered earlier in the context of the MC on q-colourings. Before proceeding with the proof of Lemma 4.14, let us pause to consider why it is sufficient to establish (4.8) just for adjacent states.

Lemma 4.15. *Suppose a coupling (X_t, Y_t) has been defined, as above, on adjacent pairs of states, and suppose that the coupling satisfies the contraction condition (4.8) on adjacent pairs. Then the coupling can be extended to all pairs of states in such a way that (4.8) holds unconditionally.*

Proof. Suppose $X_0 = x_0 \in \Omega$ and $Y_0 = y_0 \in \Omega$ are now arbitrary. Denote by $P(\cdot\,,\cdot)$ the transition probabilities of the MC on linear extensions. Let $x_0 = z^{(0)}, z^{(1)}, \ldots, z^{(\ell)} = y_0$ be a shortest path from x_0 to y_0 in the adjacency graph. (Assume a deterministic choice rule for resolving ties.) First select $Z^{(0)} \in \Omega$ according to the probability distribution $P(z^{(0)}, \cdot\,)$. Now select $Z^{(1)}$ according to the probability distribution induced by the transition $(z^{(0)}, z^{(1)}) \mapsto (Z^{(0)}, Z^{(1)})$ in the coupled process, conditioned on the choice of $Z^{(0)}$; then select $Z^{(2)}$ according to the probability distribution induced by the transition $(z^{(1)}, z^{(2)}) \mapsto (Z^{(1)}, Z^{(2)})$, conditioned on the choice of $Z^{(1)}$; and so on, ending with $Z^{(\ell)}$. (The procedure is visualised in Figure 4.6.)

Let $X_1 = Z^{(0)}$ and $Y_1 = Z^{(\ell)}$. It is routine to verify, by induction on path length ℓ, that Y_1 has been selected according to the (correct) distribution

$P(y_0, \cdot)$. Moreover, by linearity of expectation and (4.8),

$$\begin{aligned} \mathrm{E}\left[d(X_1, Y_1) \mid X_0 = x_0, Y_0 = y_0\right] &\le \sum_{i=0}^{\ell-1} \mathrm{E}\, d(Z^{(i)}, Z^{(i+1)}) \\ &\le \varrho \sum_{i=0}^{\ell-1} d(z^{(i)}, z^{(i+1)}) \\ &= \varrho\, d(x_0, y_0). \end{aligned}$$

□

So we see that it is enough to establish the contraction property (4.8) for adjacent pairs of states.

Proof of Lemma 4.14. If $p \notin \{i-1, i, j-1, j\}$ then the tests made in steps (2) and (3) either both succeed or both fail. Thus $Y_1 = X_1 \circ (i,j)$ and $d(X_1, Y_1) = j - i = d(X_0, Y_0)$. Summarising:

$$d(X_1, Y_1) = d(X_0, Y_0), \quad \text{if } p \notin \{i-1, i, j-1, j\}. \tag{4.9}$$

Next suppose $p = i - 1$ or $p = j$. These cases are symmetrical, so we consider only the former. With probability at least $\frac{1}{2}$, the tests made in steps (2) and (3) both fail, since $\Pr[r_x = r_y = 0] = \frac{1}{2}$. If this happens, clearly, $d(X_1, Y_1) = j - i = d(X_0, Y_0)$. Otherwise, with probability at most $\frac{1}{2}$, one or other test succeeds. If they both succeed, then

$$\begin{aligned} Y_1 &= Y_0 \circ (i-1, i) \\ &= X_0 \circ (i,j) \circ (i-1, i) \\ &= X_1 \circ (i-1, i) \circ (i,j) \circ (i-1, i) \\ &= X_1 \circ (i-1, j), \end{aligned}$$

and $d(X_1, Y_1) = j - i + 1 = d(X_0, Y_0) + 1$; if only one (say the one in step 2) succeeds, then $Y_1 = Y_0 = X_0 \circ (i,j) = X_1 \circ (i-1, i) \circ (i,j)$, and $d(X_1, Y_1) \le j - i + 1 = d(X_0, Y_0) + 1$. Summarising:

$$\mathrm{E}\left[d(X_1, Y_1) \mid X_0, Y_0, p = i-1 \vee p = j\right] \le d(X_0, Y_0) + \frac{1}{2}. \tag{4.10}$$

Finally suppose $p = i$ or $p = j - 1$. Again, by symmetry, we need only consider the former. There are two subcases, depending on the value of $j - i$. The easier subcase is $j - i = 1$. If $r_x = 1$ then $r_y = 0$ and

$$X_1 = X_0 \circ (i, i+1) = Y_0 \circ (i, i+1) \circ (i, i+1) = Y_0 = Y_1,$$

with a similar conclusion when $r_x = 0$. Thus $d(X_1, Y_1) = 0 = d(X_0, Y_0) - 1$. The slightly harder subcase is the complementary $j - i \geq 2$. The crucial observation is that $X_0 \circ (i, i+1), Y_0 \circ (i, i+1) \in \Omega$ and hence the tests in steps (2) and (3) either both succeed or both fail, depending only on the value of $r_x = r_y$. To see this, observe that

$$X_0(i) \not\prec X_0(i+1) = Y_0(i+1) \not\prec Y_0(j) = X_0(i),$$

from which we may read off the fact that $X_0(i)$ and $X_0(i+1)$ are incomparable in $\prec$. The same argument applies equally to $Y_0(i)$ and $Y_0(i+1)$. If $r_x = 0$ there is no change in state; otherwise, if $r_x = 1$,

$$\begin{aligned} X_1 &= X_0 \circ (i, i+1) \\ &= Y_0 \circ (i, j) \circ (i, i+1) \\ &= Y_1 \circ (i, i+1) \circ (i, j) \circ (i, i+1) \\ &= Y_1 \circ (i+1, j), \end{aligned}$$

and $d(X_1, Y_1) = j - i - 1 = d(X_0, Y_0) - 1$. Summarising both the $j - i = 1$ and $j - i \geq 2$ subcases:

$$\mathrm{E}\left[d(X_1, Y_1) \;\middle|\; X_0, Y_0, p = i \vee p = j - 1\right] \leq e(X_0, Y_0), \tag{4.11}$$

where

$$e(X_0, Y_0) = \begin{cases} 0, & \text{if } d(X_0, Y_0) = 1; \\ d(X_0, Y_0) - \frac{1}{2}, & \text{otherwise.} \end{cases}$$

Note that, in the case $j - i = 1$, inequality (4.11) covers just one value of p, namely $p = i = j - 1$, instead of two; however, this effect is exactly counterbalanced by an expected reduction in distance of 1 instead of just $\frac{1}{2}$. Combining (4.9)–(4.11) we obtain

$$\mathrm{E}\left[d(X_1, Y_1) \;\middle|\; X_0, Y_0\right] \leq d(X_0, Y_0) - \frac{-f(i-1) + f(i) + f(j-1) - f(j)}{2}.$$

Specialising the probability distribution $f(\cdot)$ to be $f(i) := \alpha(i+1)(n-i-1)$ — where $\alpha := 6/(n^3 - n)$ is the appropriate normalising constant — we have, by direct calculation, $-f(i-1) + f(i) + f(j-1) - f(j) = 2\alpha(j - i)$. Since $d(X_0, Y_0) = j - i$, we obtain (4.8) with $\varrho = 1 - \alpha$. □

From Lemmas 4.14 and 4.15 it is now a short step to:

Proposition 4.16. *The mixing time of the MC on linear extensions (refer to Figure 4.4) is bounded by*

$$\tau(\varepsilon) \leq (n^3 - n)(2 \ln n + \ln \varepsilon^{-1})/6.$$

Proof. By iteration, $\mathrm{E}\left[d(X_t, Y_t) \;\middle|\; X_0, Y_0\right] \le \varrho^t d(X_0, Y_0)$. For any pair of linear extensions g and h, there is a path in the adjacency graph using only *adjacent* transpositions (i.e., length one edges) that swaps each incomparable pair at most once. Thus $d(X_0, Y_0) \le \binom{n}{2} \le n^2$, and

$$\Pr[X_t \ne Y_t] \le \mathrm{E}\, d(X_t, Y_t) \le (1-\alpha)^t n^2.$$

The latter quantity is less than ε, provided $t \ge (n^3 - n)(2\ln n + \ln \varepsilon^{-1})/6$. The result follows directly from Lemma 4.7. □

David Wilson has recently derived a similar $O(n^3 \log n)$ bound on mixing time when f is uniform, i.e, when the transposition $(p, p+1)$ is selected u.a.r.

Exercises 4.17.

1. Use Proposition 4.16 to construct an FPRAS for linear extensions of a partial order.

2. Reprove Proposition 4.5 using path coupling. Note the significant simplification over the direct coupling proof.

New applications of path coupling are regularly being discovered. Bubley, Dyer and Greenhill [14] have presented an FPRAS for q-colourings of a low degree graph that extends the range of applicability of the one described earlier. They were able, for example, to approximate in polynomial time the number of 5-colourings of a graph of maximum degree 3, thus "beating the 2Δ bound" that appeared to exist following the result described in §4.1. Vigoda [67], in a path-coupling tour de force, was able to beat the 2Δ bound uniformly over all sufficiently large Δ; specifically, he proved rapid mixing whenever $q > \frac{11}{6}\Delta$. It is fair to say that neither of these improvements would have been possible without the aid of path coupling.

Dyer and Greenhill have also considered independent sets in a low degree graph [25], and obtained a result similar to, but apparently incomparable with, that of Luby and Vigoda [47]. Bubley and Dyer (again) applied path coupling to establish rapid mixing of a natural Markov chain on sink-free orientations of an arbitrary graph [10]. McShine [50] presents a particularly elegant application of path coupling to sampling tournaments. One further example must suffice: Cooper and Frieze [17] have applied path coupling to analyse the "Swendsen-Wang process," which is commonly used to sample configurations of the "random cluster" or ferromagnetic Potts model in statistical physics.

Finally, for those who skipped Exercise 4.3, here is the missing proof.

Proof of Lemma 4.2. The claim is established by the following sequence of (in)equalities:

$$
\begin{aligned}
2\,\|P^{t+1}(x,\,\cdot\,)-\pi\|_{\mathrm{TV}} &= \sum_{y\in\Omega}\big|P^{t+1}(x,y)-\pi(y)\big| \\
&= \sum_{y\in\Omega}\left|\sum_{z\in\Omega}P^t(x,z)\,P(z,y)-\sum_{z\in\Omega}\pi(z)\,P(z,y)\right| \\
&\le \sum_{y\in\Omega}\sum_{z\in\Omega}\big|P^t(x,z)-\pi(z)\big|\,P(z,y) \qquad (4.12) \\
&= \sum_{z\in\Omega}\big|P^t(x,z)-\pi(z)\big|\,\sum_{y\in\Omega}P(z,y) \\
&= 2\,\|P^t(x,\,\cdot\,)-\pi\|_{\mathrm{TV}},
\end{aligned}
$$

where (4.12) is the triangle inequality. □