CS294 Markov Chain Monte Carlo: Foundations & Applications **Fall 2009**

Lecture 7: September 29

Lecturer: Alistair Sinclair *Scribes: Alistair Sinclair*

In today's lecture we will continue the discussion on path coupling and see how it can be applied to bound the mixing time of a Markov chain defined on the space of all linear extensions of a partial order. Towards the end of the lecture we will introduce yet another type of coupling called monotone coupling.

7.1 Mixing time using path coupling

Suppose we have a pre-metric d on the state space Ω of an ergodic Markov chain. If we can define a coupling $(X, Y) \mapsto (X', Y')$ for pairs of adjacent (in the pre-metric) states (X, Y) such that

$$\mathbf{E}[d(X', Y')|X, Y] \leq (1-\alpha)d(X, Y) \tag{7.1}$$

for some $0 \leq \alpha \leq 1$, then this coupling can be extended to all pairs (X, Y) which also satisfy (7.1). Moreover, if $\alpha > 0$ and d is integer-valued then we have $\tau_{\mathrm{mix}} = \mathrm{O}(\alpha^{-1}\log D)$, where $D = \max_{x,y} d(x, y)$ is the maximum distance between any pair of states under the metric extension of d. The proof of this follows along the same lines as the final step in our analysis of path coupling for the colorings Markov chain in the previous lecture.

Now consider the case in which (7.1) holds with $\alpha = 0$. Then one cannot expect the above result to be true since the distance does not decay geometrically as before. Still one can prove by standard martingale arguments (**exercise!**) that, when d is integral,

$$\tau_{\mathrm{mix}} = \mathrm{O}(\beta^{-1}D^2),$$

where $\beta = \min_{X,Y\in\Omega} \mathbf{E}\Big[\big(d(X', Y') - d(X, Y)\big)^2\Big]$ is the variance of the change in d. (Note that a crude bound on the above would be $\beta = \min_{X,Y\in\Omega} \mathbf{P}\big(|d(X', Y') - d(X, Y)| \geq 1\big)$.)

Caution: In the definition of β, it is **not** sufficient to take the minimum over *adjacent* pairs (X, Y) (**exercise**).

7.2 Linear Extensions of a Partial Order

In this section we illustrate path coupling in the context of sampling a linear extension of a partial order uniformly at random.

Input: A partial order[1] $\preceq$ on $V = \{1, 2, \ldots, n\}$.

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

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

Note that, given a partial order $\preceq$ on $V = \{1, 2, \ldots, n\}$, one can easily construct a linear extension of $\preceq$ (**exercise**). For example, in figure 7.1, the right hand picture represents a linear extension of the partial order shown on the left.

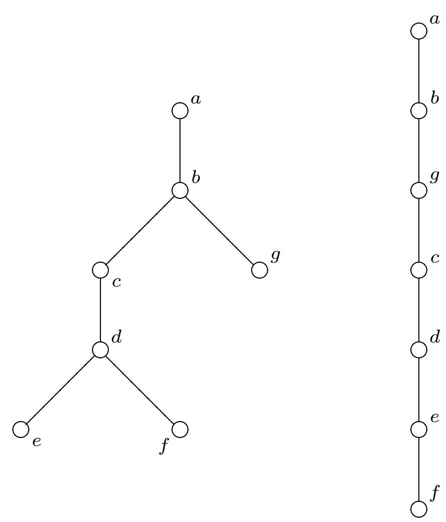

Figure 7.1: A partial order and one of its linear extensions

Goal: Sample a linear extension of $\preceq$ uniformly at random.

Let $\Omega = \Omega(\preceq)$ denote the set of all linear extensions of $\preceq$. Being able to sample from Ω has a variety of applications in combinatorics, near-optimal sorting and decision theory. Brightwell and Winkler [BW91] showed that counting linear extensions is $\#P$-complete. But if we have an efficient algorithm for sampling (almost) uniformly from Ω, then that can be used recursively, as described in lecture note 1, to get an efficient approximation algorithm for $|\Omega|$.

As usual, we propose to sample from Ω by constructing an ergodic Markov Chain on state space Ω, whose stationary distribution is uniform. Define a Markov Chain on Ω as follows.

Markov Chain:

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

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

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

Hence the above chain is ergodic and converges to the uniform distribution π on Ω. To apply path coupling, we need first to define a pre-metric on the state space Ω.

Pre-metric:
Two states X and Y in Ω are adjacent iff they differ at exactly two positions, say, i and j, $1 \le i < j \le n$. We denote this by $X = Y \circ (i,j)$. In this case, the distance $d(X,Y)$ between X and Y is defined to be $j - i$ (see figure 7.2).

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

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

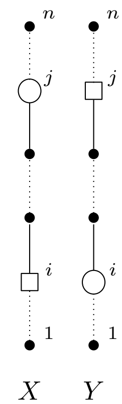

Figure 7.2: Two adjacent states

7.2.1 Coupling for adjacent pairs

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

Case 1: If $j \ne i+1$,

- With probability $\frac{1}{2}$ do nothing in both.
- Else choose the same p from $\{1,2,\ldots,n-1\}$ uniformly at random in both and exchange elements at position p and $p+1$ if "legal".

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

- With probability $\frac{1}{2(n-1)}$ do nothing in X and pick $p = i$ in Y and exchange the elements at i and j in Y. (Note that the elements at i and j are incomparable in $\preceq$ and hence this exchange is always legal.)
- With probability $\frac{1}{2(n-1)}$ pick $p = i$ in X and make a similar exchange and do nothing in Y .
- With probability $\frac{n-2}{2(n-1)}$ do nothing in both.
- Else choose same p from $\{1,2,\ldots,n-1\} \setminus \{i\}$ uniformly at random in both and exchange elements at position p and $p+1$ if "legal".

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

7.2.2 Analysis

To analyze the Markov Chain we need to consider three cases. Note that in any coupled move there is at most one position p that is chosen by either or both of X,Y; we give a case analysis according to the value of p.

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

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

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

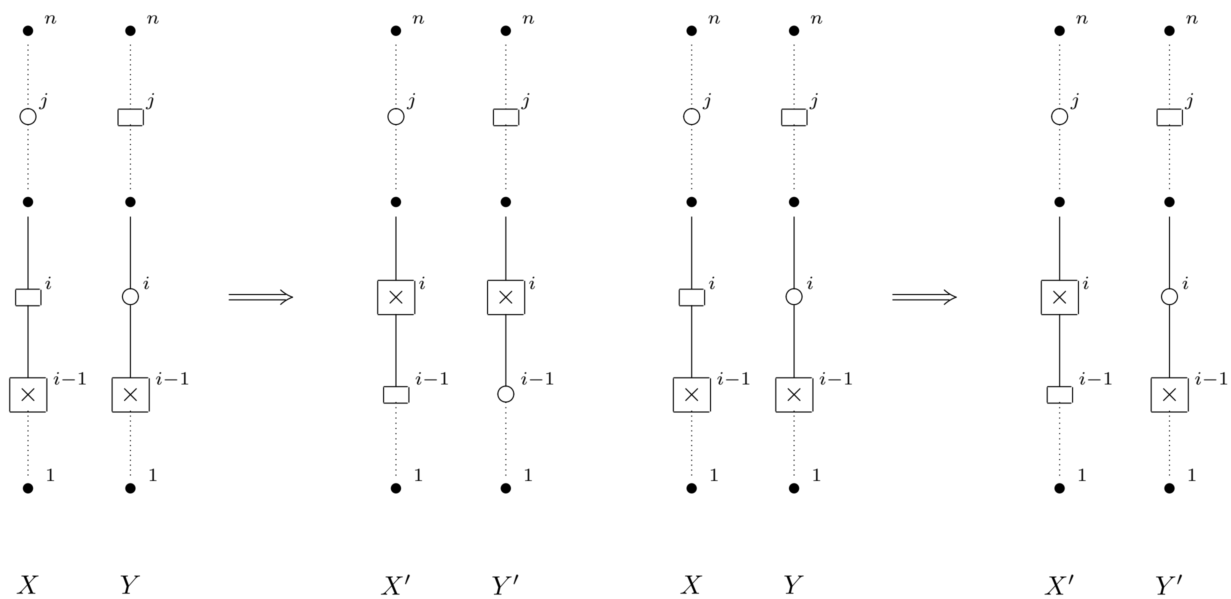

Figure 7.3: Graphical illustration of case II. *Left:* both exchanges are legal. *Right:* only exchange in X is legal.

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

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

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

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

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

Applying the path coupling lemma with $\alpha = 0$, we have $\tau_{\text{mix}} = \mathrm{O}(\beta^{-1} D^2)$ where $D = \max_{X,Y \in \Omega} d(X, Y) \leq \binom{n}{2}$ and $\beta = \min_{X,Y \in \Omega} \mathbf{P}(|d(X', Y') - d(X, Y)| \geq 1) \geq \frac{c}{n}$ for some constant $c > 0$ (**exercise**). Therefore,

$$\tau_{\text{mix}} = \mathrm{O}(n^5).$$

Bubley and Dyer [BD99] showed that the convergence can be accelerated by picking position p according to some other distribution instead of choosing it uniformly at random. We will discuss this now.

Modified Markov Chain

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

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

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

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

Plugging in, we get the expected change in distance

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

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

7.2.3 Remarks

1. Wilson [W04] improved the upper bound on the mixing time for the original Markov chain to $\mathrm{O}(n^3 \log n)$ and showed that this is tight. Interestingly, this is the correct order for the mixing time even in the case when $\preceq$ is empty (so that Ω includes all permutations of the elements); in this case the Markov chain is equivalent to shuffling cards by random adjacent transpositions.
2. It remains open whether the mixing time of the "improved" Markov chain can be strictly better than the original.

7.3 Monotone Coupling

Let P be the transition matrix of an ergodic Markov chain on state space Ω, and $\mathcal{F}$ be a probability distribution on functions $f : \Omega \to \Omega$ such that $\mathbf{P}[f(x) = y] = P(x, y)$ for all x, y. We say that the random function f (or, more correctly, the distribution $\mathcal{F}$) is *consistent* with the Markov chain. Note that $\mathcal{F}$ defines a *coupling* for the Markov chain, namely $(X, Y) \mapsto (f(X), f(Y))$. We call such a coupling a *complete coupling*, because it defines a move not only for each pair (X, Y) but for all states simultaneously. Note also that $\mathcal{F}$ actually specifies the Markov chain, via the relation

$$\mathbf{P}_{\mathcal{F}}(f(x) = y) = P(x, y) \quad \forall x, y \in \Omega.$$

It is easy to check that the couplings we defined for random walk on the hypercube and for graph coloring are actually complete couplings. [**Exercise:** Which other couplings in the course so far have been complete couplings?]

Definition 7.1 *Suppose the states in Ω have a partial order $\preceq$. A complete coupling is said to be* monotone *(w.r.t. $\preceq$) if $x \preceq y \implies f(x) \preceq f(y)$ with $\mathcal{F}$-probability* 1.

Let us now describe why monotone (complete) couplings are useful. Assume that $(\Omega, \preceq)$ has unique minimal and maximal elements, denoted by $\perp$ and $\top$ respectively. Then the following claim ensures that it is enough to consider these extremal states for analyzing the coupling time.

Claim: The coupling time $T_{X,Y}$ for any any pair (X, Y) is stochastically dominated by $T_{\perp,\top}$.

Proof: Let $F_t = f_t \circ f_{t-1} \circ \cdots \circ f_1$, where the f_i are independent samples from $\mathcal{F}$. After time t, (X, Y) moves to $(F_t(X), F_t(Y))$ and by monotonicity, $F_t(\perp) \preceq F_t(X), F_t(Y) \preceq F_t(\top)$. So, $F_t(\perp) = F_t(\top) \Rightarrow F_t(X) = F_t(Y)$.

This means that, to analyze the coupling time for the chain, it is enough to bound the coupling time for the two states $(\perp, \top)$. Moreover, even in cases where we cannot get a good analytical bound on this coupling time, we can perform numerical experiments to get at least a good statistical estimate of the coupling time just by observing how long it takes for $(\perp, \top)$ to meet.

We now give an important example of a monotone coupling.

7.3.1 Example: Ising model

Consider the "heat-bath" Markov chain for the Ising model. Let V be the set of vertices and let $\sigma \in \Omega = \{-, +\}^V$ be an Ising configuration. Recall that the Gibbs distribution for the Ising model is given by

$$\begin{aligned}\pi(\sigma) &\propto \exp\{\beta(a(\sigma) - d(\sigma))\} \\ &\propto \exp\{2\beta a(\sigma)\} = \lambda^{a(\sigma)},\end{aligned}$$

where $a(\sigma)$ = number of pairs of neighboring vertices whose spins agree in σ , $d(\sigma)$ = number of pairs of neighboring vertices whose spins disagree in σ, $\lambda = \exp(2\beta) \geq 1$ and β is inverse temperature. The heat-bath Markov chain makes transitions from any state $\sigma \in \Omega$ as follows:

- Pick a vertex v uniformly at random.
- Replace the spin σ_v at v by a random spin chosen from the distribution π conditioned on the spins σ_u of the vertices $u \in \mathrm{Nbd}(v)$. Specifically, set σ_v to '+' with probability $p_v^+(\sigma) = \lambda^{n_v^+}/(\lambda^{n_v^+} + \lambda^{n_v^-})$, where n_v^+ and n_v^- are the numbers of neighbors of v with spin '+' and '-' respectively; set σ_v to '$-$' with probability $p_v^-(\sigma) = 1 - p_v^+(\sigma)$.

Let us now define a partial order $\leq$ on Ω as follows:

For $\sigma, \tau \in \Omega$, $\sigma \preceq \tau$ iff $\sigma_v \leq \tau_v$ for all vertices $v \in V$. (Convention: $- \leq +$). Obviously, $\perp$ and $\top$ will be the configurations with all '$-$'s and all '$+$'s respectively.

We describe a complete coupling as follows, where σ is the current state:

- Pick vertex v uniformly at random.
- Pick $r \in [0, 1]$ uniformly at random.
- If $r \le p_v^+(\sigma)$ then set the spin σ_v to '+', else set it to '−'.

Exercise: Check that this coupling preserves the partial order, and hence is monotone.

References

[BW91] G. Brightwell and P. Winkler, "Counting linear extensions." *Order*, 8(3):225242, 1991.

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

[W04] D.B. Wilson, "Mixing times of lozenge tiling and card shuffling Markov chains," *The Annals of Applied Probability*, 14(1):274–325, 2004.