University of Chicago **Autumn 2003**

CS37101-1 Markov Chain Monte Carlo Methods

Lecture 4: October 21, 2003

Bounding the mixing time via coupling

Eric Vigoda

4.1 Introduction

In this lecture we'll use the coupling method to get a priori bounds on the mixing time of various Markov chains. Our first example generates a random spanning tree. We then introduce the path coupling framework which simplifies our work. Finally, we apply path coupling to analyze a Markov chain for generating a random k-colorings of a graph.

4.2 Random Spanning Trees

4.2.1 Introduction

Given an undirected graph $G = (V, E)$ we are interested in generating a spanning tree of G uniformly at random. A spanning tree is a subset $S \subset E$ such that $|S| = |V| - 1$ and S contains no cycles, thus S has a unique path between every pair of vertices. Let $\Omega(G)$ denote the set of spanning trees of G.

It is convenient to transform G to a directed graph $\overleftrightarrow{G}$, and look at arborescences of G. From G, we form $\overleftrightarrow{G}$ by replacing each edge (u, v) by a pair of anti-parallel edges $\overrightarrow{uv}$ and $\overrightarrow{vu}$. An arborescence rooted at a vertex r of a directed graph is a subset S of edges, such that $|S| = |V| - 1$ and every vertex, except r, has exactly one edge directed away from it, thus S forms a tree rooted at r and all edges pointing towards r. Let $\Omega_r(\overleftrightarrow{G})$ denote the arborescences rooted at r.

It suffices to consider arborescences rooted at any vertex.

Lemma 4.1 *For $G = (V, E)$, $r \in V$, there is a bijection between $\Omega(G)$ and $\Omega_r(\overleftrightarrow{G})$.*

Proof: For a spanning tree S, form its associated arborescence by simply pointing all edges towards r (observe there is a unique way of doing this). For the reverse direction, given an arborescence, simply remove the orientations and the resulting undirected edges form a spanning tree. ∎

4.2.2 Markov chains

For simplicity, let us assume G is a d-regular graph.

Here is a simple Markov chain $\mathcal{MC}$ on the set of all arborescences $\Omega(\overleftrightarrow{G}) = \bigcup_{v \in V} \Omega_v(\overleftrightarrow{G})$. From $X_t \in \Omega(\overleftrightarrow{G})$ with root u,

- Choose a random neighbor v of u.
- Let $\overrightarrow{vw}$ denote the unique edge oriented away from v in X_t. Set

  $$X' = X_t \setminus \overrightarrow{vw} \bigcup \overrightarrow{uv}.$$

  Thus, X' is an arborescence rooted at v.
- Let

  $$X_{t+1} = \begin{cases} X' & \text{with probability } 1/2 \\ X_t & \text{otherwise.} \end{cases}$$

Since the graph is regular, the chain is symmetric. The "self-loop" probability of 1/2 (i.e., staying at the same state) ensures the chain is aperiodic. Therefore, the chain is ergodic and the unique stationary distribution is uniformly distributed over $\Omega(\overleftrightarrow{G})$.

The above chain was originally analyzed by Broder and Aldous. Propp and Wilson [5] introduced a slight variant of the above chain which we'll use.

Given a root $r \in V$, the following chain $\mathcal{MC}_r$ has state space $\Omega_r(\overleftrightarrow{G})$. From $Y_t \in \Omega_r(\overleftrightarrow{G})$,

- Run the chain $\mathcal{MC}$ with initial state $X_0 = Y_t$. Stop at the first time $t' > 0$ where $X_{t'} \in \Omega_r(\overleftrightarrow{G})$. In words, stop as soon as the root returns to r.
- Set $Y_{t+1} = X_{t'}$.

Exercise 4.2 *For a regular graph G, prove that the stationary distribution of $\mathcal{MC}_r$ is uniformly distributed over $\Omega_r(\overleftrightarrow{G})$.*

4.2.3 Analysis

To analyze the mixing time we'll bound the coupling time of the chain. For every pair $(Y_t, Y_t') \in \Omega_r(\overleftrightarrow{G})^2$, we'll define a coupling $(Y_t, Y_t') \to (Y_{t+1}, Y_{t+1}')$. The following lemma shows how the time till the chains couple gives a bound on the mixing time.

Lemma 4.3 *Given a coupling $(Y_t, Y_t') \to (Y_{t+1}, Y_{t+1}')$ for all $(Y_t, Y_t') \in \Omega_r(\overleftrightarrow{G})^2$, such that for all $X_0, Y_0 \in \Omega$, there exists a time T where*

$$\Pr(\ X_T \neq Y_T \ \mid \ X_0, Y_0\ ) \leq 1/4,$$

then

$$T_{\text{mix}}(1/4) \leq T.$$

This Lemma, with slightly different language, was proven in Lecture 2.

We will bound the first time T the chains couple via the cover time. The cover time of a graph G is the number of steps till a random walk on G visits every vertex at least once. Denote this as T_{cover}.

Lemma 4.4 (Propp and Wilson [5])

$$\mathrm{E}(\ T\ ) \leq 4\mathrm{E}(\ T_{\text{cover}}\ ).$$

Proof: Denote the Hamming distance (or the difference) between the pair of chains by

$$H(Y_t, Y_t') = |Y_t \setminus Y_t'|.$$

Once $H(Y_t, Y_t') = 0$ the chains have coupled. It will be straightforward to bound the time till the distance reaches zero.

The coupling is the obvious choice. A step $Y_t \to Y_{t+1}$ of $\mathcal{MC}_r$ involves a sequence of transitions $X_0 \to X_1 \dots X_{t'-1} \to X_{t'}$ of chain $\mathcal{MC}$. Call this sequence an excursion. Both chains Y_t, Y_t' use the identical excursion.

Consider the effect of an excursion on the resulting arborescence Y_{t+1}. Say vertex $v \neq r$ is visited during the excursion, if $X_s \in \Omega_v(\overleftrightarrow{G})$ for some time $s < t'$ (i.e., the root moved to v for some intermediate arborescence). For the last such time s, let w denote the next root. Thus, we know $\overrightarrow{vw} \in X_{s+1}$ and $\overrightarrow{vw} \in X_{t'}$ (since it's the last visit to v). This implies $\overrightarrow{vw} \in Y_{t+1}$. Similarly, $\overrightarrow{vw} \in Y_{t+1}'$. Thus, if a vertex $v \neq r$ is visited during an excursion, then the two chains agree on the edge coming out of v, i.e., $\overrightarrow{vw} \in Y_{t+1}$ and $\overrightarrow{vw} \in Y_{t+1}'$.

Therefore, once every vertex is visited on some excursion, the two chains are identical on every edge. The expected time till every vertex is visited is the cover time. Running at most a further cover time brings us back to the root r. There is an extra factor of two coming from the "self-loop" probability of 1/2 in the definition of $\mathcal{MC}$. This proves the lemma. ■

We can then bound the mixing time

$$T_{\text{mix}}(1/4) \leq 16\mathrm{E}\left(T_{\text{cover}} \right),$$

by Markov's inequality.

4.2.4 Cover time

We conclude with a few final remarks about the cover time. For any graph, Feige [2, 3] proved

$$n \log n \leq \mathrm{E}\left(T_{\text{cover}} \right) \leq O(|V||E|).$$

For some specific class of graphs, such as a random graph, more precise results are known.

4.3 Path Coupling

In the above example the distance is non-increasing. Typically things are not that easy, and even defining a good coupling can be complicated. The following lemma of Bubley and Dyer [1], known as the path coupling lemma, significantly simplifies the definition and analysis of a coupling.

Before stating and proving the lemma we need some basics about couplings.

4.3.1 Composing Couplings

Consider a coupling α of μ (on Ω_1) with ν (on Ω_2), and a coupling β of ν with ω (on Ω_3). Define their composition as, for all $x \in \Omega_1, z \in \Omega_3$,

$$\gamma(x,z) = (\alpha \circ \beta)(x,z) := \sum_{y \in \Omega_2} \frac{\alpha(x,y)\beta(y,z)}{\nu(y)}.$$

Alternatively, we can construct γ algorithmically in the following manner. First, sample (X,Y) from α. Given $(X,Y) = (x,y)$, we then sample (Y,Z) from β conditional on $Y = y$. We output (X,Z).

Exercise 4.5 *Prove γ is a coupling of μ with ω.*

4.3.2 Statement of Lemma

Consider a Markov chain on Ω. The path coupling lemma says it suffices to define and analyze a coupling for pairs in a subset of $\Omega \times \Omega$ assuming the subset "connects" the state space.

Lemma 4.6 (Bubley and Dyer [1]) *Let $S \subseteq \Omega \times \Omega$ such that (Ω, S) is connected. For $(X, Y) \in \Omega \times \Omega$, define $\mathrm{dist}(X, Y)$ as the shortest path length between X and Y in the graph (Ω, S). (Thus, if $(X, Y) \in S$, then $\mathrm{dist}(X, Y) = 1$.)*

- *If there exists $\beta < 1$, and for all $(X_t, Y_t) \in S$ a coupling $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$ such that*
$$\mathrm{E}\left(\mathrm{dist}(X_{t+1}, Y_{t+1} \mid X_t, Y_t \right) \le \beta \mathrm{dist}(X_t, Y_t), \tag{4.1}$$
then
$$T_{\mathrm{mix}}(\epsilon) \le \frac{\log(d_{\max}/\epsilon)}{1-\beta}, \tag{4.2}$$
where $d_{\max}$ is the diameter of the graph (Ω, S).

- *If (4.1) holds with $\beta = 1$ and there exists $\alpha > 0$ such that for this coupling for all $(X_t, Y_t) \in \Omega^2$,*
$$\Pr\left(\mathrm{dist}(X_{t+1}, Y_{t+1}) \neq \mathrm{dist}(X_t, Y_t) \mid X_t, Y_t \right) \ge \alpha$$
then
$$T_{\mathrm{mix}}(\epsilon) \le \lceil \frac{e{d_{\max}}^2}{\alpha} \rceil \lceil \log(1/\epsilon) \rceil,$$

Proof: We'll only prove the $\beta < 1$ case.

We first construct a coupling for all $(X_t, Y_t) \in \Omega \times \Omega$ which satisfies (4.1). We do this by simply "composing" couplings along a shortest path between X_t and Y_t. Let $Z_t^0 = X_t, Z_t^1, \ldots, Z_t^j = Y_t, j = \mathrm{dist}(X_t, Y_t)$ denote an arbitrary such shortest path.

From the hypothesis of the lemma we know

$$\mathrm{E}\left(\mathrm{dist}(Z_{t+1}^i, Z_{t+1}^{i+1} \mid Z_t^i, Z_t^{i+1} \right) \le \beta \mathrm{dist}(Z_t^i, Z_t^{i+1}).$$

We then have

$$\begin{aligned}
\mathrm{E}\left(\mathrm{dist}(X_{t+1}, Y_{t+1} \mid X_t, Y_t \right) &\le \sum_{1 \le i < j} \mathrm{E}\left(\mathrm{dist}(Z_{t+1}^i, Z_{t+1}^{i+1} \mid Z_t^i, Z_t^{i+1} \right) \\
&\le \beta \sum_i \mathrm{dist}(Z_t^i, Z_t^{i+1}) \\
&= \beta \mathrm{dist}(X_t, Y_t)
\end{aligned}$$

Inductively we have

$$\begin{aligned}\Pr(\ X_t \neq Y_t \ \mid\ X_0, Y_0\ ) &\le \mathrm{E}(\ \mathrm{dist}(X_t, Y_t)\ \mid\ X_0, Y_0\ ) \\ &\le \beta^t d_{\max} \\ &\le \epsilon \qquad \text{for } t \ge \log(d_{\max}/\epsilon)/(1-\beta),\end{aligned}$$

where the last inequality uses that $\log(1/\beta) > 1 - \beta$.

■

4.4 Random Graph Colorings

For $G = (V, E)$, let $\Omega = \{1, 2, \ldots, k\}^V$. Let $N(v)$ denote the neighbors of vertex v. The following Markov chain is known as the Glauber dynamics. From $X_t \in \Omega$,

- Choose z uniformly at random from V, and c uniformly at random from $\{1, \ldots, k\}$.
- For all $w \neq z$, set $X_{t+1}(z) = X_t(z)$.
- If no neighbors of z have color c (i.e., $c \notin X_t(N(z))$), then set $X_{t+1}(z) = c$, otherwise set $X_{t+1}(z) = X_t(z)$.

A proper k-coloring is an assignment $\sigma \in \Omega$ such that all adjacent vertices receive different colors.

Exercise 4.7 *Let Δ denote the maximum degree of G. Prove that when $k > \Delta + 1$ the stationary distribution of the above Glauber dynamics is uniformly distributed over proper k-colorings.*

We'll prove fast convergence via path coupling. We begin with following theorem.

Theorem 4.8 *Let Δ denote the maximum degree of G. If $k > 3\Delta$, then*

$$\tau(\epsilon) \le nk \ln(n/\epsilon).$$

Proof: Let S denote the pairs of colorings which differ at exactly one vertex. Consider a pair of such colorings $(X_t, Y_t) \in S$ where $X_t(v) \neq Y_t(v)$, say $X_t(v) = c_X$ and $Y_t(v) = c_Y$.

The coupling is simply the identity coupling. Both chains attempt to update the same vertex z to the same color c. Observe that only updates with $z \in N(v)$ and $c \in \{c_X, c_Y\}$ might succeed or fail in exactly one chain. All other updates succeed or fail in both chains. Thus, these are the only updates which might increase the distance.

The only updates which might decrease the distance are successful recolorings of v. Since there are at most Δ colors in the neighborhood of v, there are at least $k-\Delta$ available colors for v.

Combining these contributions, we have, for $(X_t, Y_t) \in S$,

$$\mathrm{E}\left(\ \mathrm{dist}(X_{t+1}, Y_{t+1})\ \mid\ X_t, Y_t\ \right) \leq 1 + \frac{1}{kn}\left(2\Delta - (k-\Delta)\right) \leq 1 - 1/kn,$$

for $k > 3\Delta$.

Plugging this bound into the path coupling lemma proves the theorem. ∎

Modifying the above coupling we can achieve the following improvement.

Theorem 4.9 (Jerrum [4]) *If $k > 2\Delta$, then*

$$\tau(\epsilon) \leq nk \ln(n/\epsilon).$$

Proof: The set S remains the same, we simply modify the coupling slightly. If X_t attempts to recolor a neighbor w of v to color c_X, then Y_t attempts to recolor w to c_Y. Similarly, if X_t attempts to recolor $w \in N(v)$ to c_X, then Y_t attempts to recolor w to c_X. In all other cases, X_t and Y_t attempt to modify the same vertex to the same color.

Observe that $w \in N(v)$ can not receive color c_X in X_t (since this is v's color), and w can not receive c_Y in Y_t. Thus, such a coupled update has no effect on the pair of colorings. Therefore, there are at most Δ coupled recolorings which might increase the distance – attempting to recolor $w \in N(v)$ to c_Y in X_t and to c_X in Y_t.

We now have

$$\mathrm{E}\left(\ \mathrm{dist}(X_{t+1}, Y_{t+1})\ \mid\ X_t, Y_t\ \right) \leq 1 - \frac{1}{kn}\left(\Delta - (k-\Delta)\right) \leq 1 - 1/kn,$$

for $k > 2\Delta$. ∎

Exercise 4.10 *Why in the above proof did we need the Glauber dynamics to be defined on the set $\{1, \ldots, k\}^V$, instead of just on the set of proper k-colorings? We need it in order to claim that the distance increases by at most one (instead of two) after we attempt to recolor $w \in N(v)$ to c_Y in X_t and to c_X in Y_t.*

References

[1] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *38th Annual Symposium on Foundations of Computer Science*, pages 223–231, Miami Beach, FL, October 1997. IEEE.

[2] U. Feige. A tight lower bound on the cover time for random walks on graphs. *Random Structures Algorithms*, 6(4):433–438, 1995.

[3] U. Feige. A tight upper bound on the cover time for random walks on graphs. *Random Structures Algorithms*, 6(1):51–54, 1995.

[4] M. Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures Algorithms*, 7(2):157–165, 1995.

[5] J. G. Propp and D. B. Wilson. How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *J. Algorithms*, 27(2):170–217, 1998.