6.896 Probability and Computation	February 23, 2011
Lecture 7	
Lecturer: Constantinos Daskalakis	*Scribe: Matthew Johnson, Brian Basham*

NOTE: The content of these notes has not been formally reviewed by the lecturer. It is recommended that they are read critically.

1 Recap: Path Coupling

Setup Given a Markov chain and pre-metric on Ω, our goal is to design a good coupling. Recall that a pre-metric on Ω is a weighted graph $(\Omega, \mathcal{E}_{\text{pre}})$ in which every edge is a shortest path, and we name the induced shortest path metric d.

In the previous lecture we proved the Path Coupling Theorem:

Theorem 1 (Path Coupling)**.** *Suppose we have a coupling for all $(x, y) \in \mathcal{E}_{pre}$ and suppose that there exists some $\alpha \in [0, 1]$ such that for all $(x, y) \in \mathcal{E}_{pre}$ we have*

$$\mathbb{E}\left[d(X', Y') | \, X = x, Y = y\right] \le (1 - \alpha) d(x, y). \tag{*}$$

Then we can construct a coupling for all pairs in Ω^2 such that the contraction (*) holds.

See the previous lecture for a proof.

2 Application of Path Coupling to Graph Colorings

Setup Recall we used Δ to denote the max degree of the graph on n nodes to be colored with colors $\{1, 2, \ldots, q\}$. We showed that the natural Markov chain (defined in the previous lecture) mixed in time $\tau_{\text{mix}} = O(n \log n)$ if we had $q \ge 4\Delta + 1$, and we were not able to improve the lower limit on q with our previous technique.

Theorem 2. *The natural Markov Chain has mixing time $\tau_{mix} = O(n \log n)$ if $q \ge 3\Delta + 1$.*

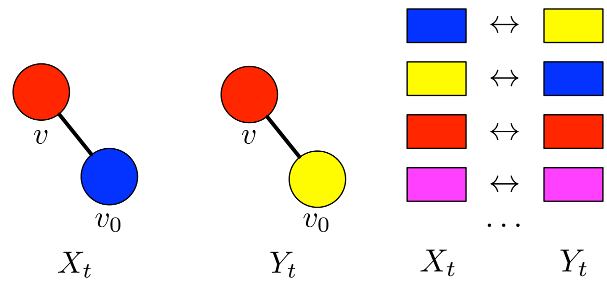
Proof:

Pre-metric Our first step is to define a suitable pre-metric. We say that a pair of legal colorings (x, y) is an edge in the pre-metric, i.e. $(x, y) \in \mathcal{E}_{\text{pre}}$, if they differ only at the color of a single vertex. Further, we set all the edge-weights in the pre-metric to be 1. Note that the diameter D of Ω with respect to the induced metric d satisfies $D \le 2n$.

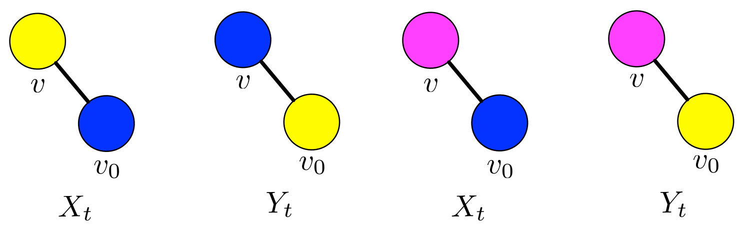
Next we define a coupling for adjacent states that satisfies a contraction property, as in (*).

Coupling in the Pre-Metric: Suppose X_t and Y_t disagree only at v_0, and use $N(v_0)$ to denote the neighborhood of node v_0, not including v_0. Our coupling is the following:

1. Pick the same vertex v in both chains, and:
   - if $v \notin N(v_0)$: pick the same color c in both chains and set v's color to c in both chains, if this change results in a valid coloring; note that either the change is allowed in both chains or in none.



(a) Initial configuration (b) Color matching



(c) A possible move that increases the distance from 1 to 3. (d) A possible move that keeps the distance at 1.

Figure 1: An example of the graph coloring coupling move when $v \in N(v_0)$.

- if $v \in N(v_0)$: match up colors as follows

$$\begin{array}{rcl} X_t & & Y_t \\ c_{X_t}(v_0) & \leftrightarrow & c_{Y_t}(v_0) \\ c_{Y_t}(v_0) & \leftrightarrow & c_{X_t}(v_0) \\ c & \leftrightarrow & c \quad \forall c \neq c_{X_t}(v_0), c_{Y_t}(v_0), \end{array}$$

  where $c_Z(\cdot)$ denotes the color of a node in chain Z; pick a random pair of colors for the two chains according to this pairing of colors and try to assign the chosen colors to vertex v in the respective chains, if allowed. See Figure 1 for an example.

Good moves: Choose vertex $v = v_0$ and a color $c \notin N(v_0)$. In such a case, the distance between the two chains decreases by 1. There are $(q - \Delta)$ good moves.

Bad moves: Choose a vertex $v \in N(v_0)$ and choose color $c_{Y_t}(v)$ for the X_t chain and $c_{X_t}(v)$ for the Y_t chain. [1] In this case, observe that the distance of the two chains increases by 2. [2] There are Δ bad moves.

[1] Notice that the symmetric case where color $c_{X_t}(v)$ is chosen for the X_t chain and color $c_{Y_t}(v)$ is chosen for the Y_t chain does not result in any change in the states of the two chains.

[2] Note that the distance change of 2, rather than 1, is exactly the weakness of this argument that yields $q \geq 3\Delta + 1$ colors, rather than $q \geq 2\Delta + 1$ colors.

Accounting: We therefore have, for all $(X_t, Y_t) \in \mathcal{E}_{\text{pre}}$,

$$\begin{aligned}\mathbb{E}\left[d(X_{t+1}, Y_{t+1})|\, X_t, Y_t\right] &\le \left[1 + \left(\frac{q-\Delta}{qn}\right)(-1) + \left(\frac{\Delta}{qn}\right)(+2)\right] d(X_t, Y_t) \\ &= \left(1 - \frac{q - 3\Delta}{qn}\right) d(X_t, Y_t).\end{aligned}$$

We can apply the Path Coupling Theorem to extend the coupling to arbitrary pairs of states, thus yielding

$$\mathbb{E}\left[d(X_t, Y_t)|X_0, Y_0\right] \le \left(1 - \frac{q-3\Delta}{qn}\right)^t d(X_0, Y_0).$$

Hence we have:

$$\begin{aligned}\Delta(t) &\le \max_{x_0, y_0} \mathbb{P}\left[X_t \ne Y_t|\, X_0 = x_0, Y_0 = y_0\right] \\ &= \max_{x_0, y_0} \mathbb{P}\left[d(X_t, Y_t) > 0|\, X_0 = x_0, Y_0 = y_0\right] \\ &= \max_{x_0, y_0} \mathbb{P}\left[d(X_t, Y_t) \ge 1|\, X_0 = x_0, Y_0 = y_0\right] \\ &\le \max_{x_0, y_0} \mathbb{E}\left[d(X_t, Y_t)|\, X_0 = x_0, Y_0 = y_0\right] \\ &\le \left(1 - \frac{q-3\Delta}{qn}\right)^t 2n,\end{aligned}$$

where we have used the observation that the diameter of Ω under $d(\cdot,\cdot)$ is $D \le 2n$. Therefore we have $q \ge 3\Delta + 1 \implies \tau_{\text{mix}} = O(n \log n)$. □

With a little effort we can strengthen the previous theorem to $q \ge 2\Delta + 1$, which is a weaker condition on the number of colors.

Theorem 3. *The natural MC for graph coloring has $\tau_{mix} = O(n \log n)$ whenever $q \ge 2\Delta + 1$.*

Proof: We consider the set $\widetilde{\Omega} \supseteq \Omega$ of all (not necessarily proper) colorings, and the Markov chain that picks a vertex v and color c uniformly at random and changes the color of vertex v to c if c does not appear in $N(v)$.

Observe that, if we start this Markov chain at a point in Ω (i.e. a proper coloring), then we never leave Ω and are hence running a copy of our previous natural Markov chain.

Pre-metric on $\widetilde{\Omega}$: x, y are adjacent if they differ at exactly one vertex in which case $d(x, y) = 1$. The induced metric is then the Hamming metric.

Coupling in the Pre-Metric: We use the same coupling as in the previous proof.

Good moves: Decrease distance to zero. There are at least $q - \Delta$ good moves, as before.

Bad moves: These are the same as before, except they increase the distance to 2 (rather than to 3 in our earlier proof). There are at most Δ bad moves, as before.

Overall our coupling on pairs of adjacent states in the pre-metric satisfies

$$\mathbb{E}\left[d(X_{t+1}, Y_{t+1})|\, X_t = x, Y_t = y, d(x, y) = 1\right] \le \left(1 - \frac{q - 2\Delta}{qn}\right) d(X_t, Y_t).$$

The Path Coupling theorem implies then that we can extend this coupling to all pairs $(x, y) \in \widetilde{\Omega} \times \widetilde{\Omega}$ so that

$$\mathbb{E}\left[d(X_{t+1}, Y_{t+1})|\, X_t, Y_t\right] \le \left(1 - \frac{q - 2\Delta}{qn}\right) d(X_t, Y_t).$$

Observe that for all pairs of states $(x,y) \in \Omega \times \Omega$ the coupling produced by the Path Coupling theorem for our Markov chain constitutes a valid coupling also for our original natural Markov chain. (Indeed, for states on Ω our Markov Chain and the natural Markov chain have the same transition matrix.) From this point on, we apply the coupling produced by the path coupling theorem to our original natural Markov Chain, forgetting the Markov Chain on the larger state space $\widetilde{\Omega}$.

Let $X_0, Y_0 \in \Omega$. Our coupling satisfies:

$$\begin{aligned}
\mathbb{E}\left[d(X_t,Y_t)|X_0,Y_0\right] &\le \left(1-\frac{1-2\Delta}{qn}\right)^t d(X_0,Y_0)\\
&\le \left(1-\frac{q-2\Delta}{qn}\right)^t n.
\end{aligned}$$

$$\begin{aligned}
\implies \mathbb{P}\left[X_t \ne Y_t|X_0,Y_0\right] &\le \mathbb{P}\left[d(X_t,Y_t)\ge 1|X_0,Y_0\right]\\
&\le \mathbb{E}\left[d(X_t,Y_t)|X_0,Y_0\right]\\
&\le n\left(1-\frac{q-2\Delta}{qn}\right)^t.
\end{aligned}$$

Hence $\tau_{\text{mix}} = O(n\log n)$ when $q \ge 2\Delta+1$. □

Some Historical Remarks

- For Δ-regular trees, it has been shown that $\tau_{\text{mix}} = O(n\log n)$, whenever $q \ge \Delta+2$. [Martinelli, Sinclair, Weitz '06].
- It has also been shown that $\tau_{\text{mix}} = O(n\log n)$ for all graphs, whenever $q \ge \frac{11\Delta}{6}$, but for a slightly different Markov chain. [Vigoda '99]
- Hayes and Vigoda [HV '03] showed $\tau_{\text{mix}} = O_{1/\epsilon}(n\log n)$, whenever $q \ge (1+\epsilon)\Delta$ and $\Delta = \Omega(\log n)$.
- Dyer and Frieze [DF '03] showed that $\tau_{\text{mix}} = O(n\log n)$ for $q \ge \alpha\Delta$, $\alpha \approx 1.76$, as long as $\Delta = \Omega(\log n)$ and the girth of the graph is $g = \Omega(\log\log n)$.
- Dyer, Frieze, Hayes and Vigoda [DFHV 05] showed that $\tau_{\text{mix}} = O(n\log n)$, as long as
  - $q \ge \alpha\Delta, g \ge 5, \Delta = \Omega(1)$, where $\alpha \approx 1.76$; or
  - $q \ge \beta\Delta, g \ge 7, \Delta = \Omega(1)$, where $\beta = 1.49$.

3 Path Coupling CheatSheet

If we have the contraction (*) for some $\alpha > 0$ and the distances are integral, then we have $\tau_{\text{mix}} = O\left(\frac{1}{\alpha}\log D\right)$. It can also be shown that if we only have the contraction (*) for $\alpha = 0$, then we have $\tau_{\text{mix}} = O\left(\frac{D^2}{\beta}\right)$ where

$$\beta := \min_{X_t,Y_t\in\Omega} \mathbb{E}\left[\left(d(X_{t+1},Y_{t+1}) - d(X_t,Y_t)\right)^2 \middle| X_t, Y_t\right].$$

Exercise (1pt): Show the above mixing time result for $\alpha = 0$. *Hint: use the Optional Stopping Theorem.*

4 Sampling Linear Extensions

Input: A partial order $\preceq$ over V. (Think of a directed acyclic graph encoding this partial order.)

A *linear extension* of $\preceq$ is a total order $\le$ such that $i \preceq j \implies i \le j$. A topological sorting algorithm can produce a linear extension of a poset in polynomial time.

Goal: We want to sample linear extensions uniformly at random.

Applications: Numerous, e.g. in combinatorics, sorting, rankings, decision theory, etc. Brightwell and Winkler [BW '91] showed that counting linear extensions of a poset is #P-complete. Nevertheless, if we have a good sampler we can use it to approximately count, as we are going to see in future lectures.

We are going to analyze the behavior of the following Markov Chain.

Markov chain

- With probability $\frac{1}{2}$, do nothing.
- Otherwise, pick a location $p \in \{1, \ldots, n-1\}$ uniformly at random and exchange the elements at locations p and $p+1$ if the exchange results in a valid linear extension.

The chain is symmetric and hence the uniform distribution is stationary. It is also aperiodic, due to the lazy step. As an exercise, show that it is also irreducible.

Exercise (1pt): Show that the Markov chain is irreducible; in particular, show that every linear extension can be reached from another in $\binom{n}{2}$ steps.

Theorem 4. *The mixing time of the Markov chain is $\tau_{mix} = O(n^5)$.*

Proof:

Pre-metric $(x, y) \in \mathcal{E}_{\text{pre}}$ if they differ at exactly two positions, $i < j$. In this case the weight of the edge is defined to be $d(x, y) = j - i$. Note that there may not be a Markov Chain step between neighbors in the pre-metric, and therefore the graph of the Markov chain and the graph of the pre-metric are different.

Exercise (1pt): Show that this is a valid pre-mertic.

Coupling in the Pre-Metric Say x and y differ at $i < j$. We consider two cases:

- $j \neq i + 1$:
  - With probability $\frac{1}{2}$, do nothing.
  - Otherwise, pick the same p.
- $j = i + 1$:
  - With probability $\frac{1}{2(n-1)}$ do nothing in X, pick i in Y and exchange elements at positions i and $i+1$ in Y (note that the exchange is legal).
  - With probability $\frac{1}{2(n-1)}$ do nothing in Y, pick i in X and exchange elements at positions i and $i+1$ in X (the exchange is legal).
  - With probability $\frac{n-2}{2(n-1)}$ do nothing in both.
  - Otherwise pick $p \in \{1, \ldots, n-1\} \setminus \{i\}$ uniformly at random and try to exchange elements at positions p and $p+1$ in both X and Y.

To analyze the coupling, we identify three cases in which a change is made:

Case I. $p \notin \{i-1, i, j-1, j\}$: in this case $d(X', Y') = d(X, Y)$ (i.e. the distance is not affected by the move).

Case II. $p = i - 1$ or $p = j$, which occurrs with probability $\frac{1}{n-1}$: in this case $d(X', Y') \leq d(X, Y) + 1$.

Case III. $p = i$ or $p = j - 1$, which occurrs with probability $\frac{1}{n-1}$: in this case $d(X', Y') = d(X, Y) - 1$.

Overall we have for an edge in the pre-metric

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

which by virtue of our discussion above implies

$$\tau_{\text{mix}} = O\left(\frac{1}{\beta}D^2\right) = O(n^5),$$

where we used that $D \leq \binom{n}{2}$ and the following exercise.

Exercise (1pt): Show $\beta \geq \max_{X,Y\in\Omega} \mathbb{P}\left[|d(X', Y') - d(X, Y)| \geq 1\right] \geq \frac{c}{n}$ for some constant $c > 0$. □

Improving τ_{mix} We can construct a different Markov chain:

- With probability $\frac{1}{2}$ do nothing.
- Otherwise pick $p \in \{1, 2, \ldots, n-1\}$ with probability $Q(p) = \frac{p(n-p)}{Z}$ where $Z = \frac{n^3-n}{6}$ and exchange elements at locations p and $p+1$ if legal.

We can use the same pre-metric and path coupling (with appropriately modified probabilities) to obtain

$$\begin{aligned}\mathbb{E}\left[d(X', Y') - d(X, Y)|X, Y\right] &\leq \frac{1}{2Z}\left[Q(i-1) + Q(j) - Q(i) - Q(j-1)\right] \\ &\leq -\frac{6}{n^3}d(X, Y).\end{aligned}$$

which gives $\tau_{\text{mix}} = O\left(\frac{1}{\alpha}\log D\right) = O(n^3 \log n)$.

Historical Remark: Wilson [W '04] tightened the analysis of the original Markov chain showing that the mixing time is in fact $\tau_{\text{mix}} = O(n^3 \log n)$, and this is tight.