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

Lecture 6: September 22

Lecturer: Prof. Alistair Sinclair *Scribes: Alistair Sinclair*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

In today's lecture we discuss *path coupling*, a variation on the coupling technique that makes it much easier to apply in practice. We will use path coupling to show that the mixing time for the graph coloring Markov chain discussed last time remains $O(n \log n)$ under a much weaker condition on the number of colors.

6.1 Path coupling

"Path coupling," an idea introduced by Bubley and Dyer [BD97], is a powerful engineering tool that makes it much easier to design couplings in complex examples.

Definition 6.1 *A* pre-metric *on* Ω *is a connected undirected graph with positive edge weights with the following property: every edge is a shortest path. We call two elements* $x, y \in \Omega$ adjacent *if* (x, y) *is an edge in the pre-metric.*

Notice that a pre-metric extends to a metric in the obvious way (just take shortest path distances in the graph of the pre-metric).

Path coupling says that, when defining a coupling for a Markov chain on Ω, it is enough to specify the coupling *only for pairs of states that are adjacent in the pre-metric.* This will usually be much easier than specifying the coupling for arbitrary pairs of states.

This fact is expressed in the following theorem:

Theorem 6.2 *Suppose there exists a coupling* $(X, Y) \to (X', Y')$ *defined only on pairs* (X, Y) *that are adjacent in the pre-metric such that*

$$\mathrm{E}[d(X', Y')|X, Y] \leq (1 - \alpha) d(X, Y) \text{ for some } \alpha \in [0, 1], \tag{6.1}$$

where d *is the metric extending the pre-metric. Then this coupling can be extended to a coupling which satisfies (6.1) on* all *pairs* (X, Y).

Note that (6.1) says that the distance between X and Y (as measured in the metric d) decreases in expectation by a factor $(1 - \alpha)$. Just as in our analysis of the graph colorings Markov chain in the previous lecture, assuming d takes non-negative integer values this immediately leads to a bound on the mixing time of $\tau_{\text{mix}} = O(\frac{1}{\alpha} \log D)$, where D is the maximum distance between any two states. (In that application, we had $\alpha = \frac{1}{qn}$ and $D = n$.)

Proof: Let (X, Y) be arbitrary, not necessarily adjacent. Consider any shortest path in the pre-metric

$$X = Z_0 \to Z_1 \to Z_2 \to \dots Z_{k-1} \to Z_k = Y.$$

We construct a coupling of one move of (X, Y) by composing couplings for each adjacent pair (Z_i, Z_{i+1}) along this path, as follows:

- Map (Z_0, Z_1) to (Z'_0, Z'_1) according to the coupling.
- For each $i \geq 1$ in sequence, map (Z_i, Z_{i+1}) to (Z'_i, Z'_{i+1}) according to the coupling, but conditional on Z'_i already being chosen.

This process constructs a coupling $(X, Y) \to (X', Y') = (Z'_0, Z'_k)$. [**Exercise:** convince yourself that this is indeed a valid coupling!]

Now the expected change in distance under the coupling can be analyzed as follows:

$$\begin{aligned} \mathrm{E}[d(X', Y')] &\leq \mathrm{E}\left[\sum_{i=0}^{k-1} d(Z'_i, Z'_{i+1}) \,\middle|\, X, Y\right] \\ &\leq (1-\alpha)\sum_{i=0}^{k-1} d(Z_i, Z_{i+1}) \\ &\leq (1-\alpha) d(X, Y), \end{aligned}$$

which establishes (6.1) as required. ∎

6.2 Application to graph coloring

We now apply path coupling to the graph coloring Markov chain from the previous lecture, and see how it leads to a simpler, more elegant and tighter analysis.

As before, let $G = (V, E)$ be a graph of maximum degree Δ, and assume that we have q colors at our disposal. A (proper) coloring is an assignment of a color to each node such that no two vertices linked by and edge have the same color.

We proved last time that for $q \geq 4\Delta + 1$, the mixing time is $O(n \log n)$. We will now refine this result using path coupling:

Theorem 6.3 *Provided $q \geq 2\Delta + 1$, $\tau_{\mathrm{mix}} = O(n \log n)$.*

Proof: We will use the following pre-metric: two colorings X, Y are adjacent iff they differ at exactly one vertex. In this case, we set $d(X, Y) = 1$. We will extend this pre-metric to the *hamming metric*: $d(X, Y)$ is the number of vertices at which X, Y differ.

Note, however, that in order to have the hamming distance indeed be the extension of the pre-metric described above, we need to let our state space include non-valid colorings as well. This is because the shortest path from one coloring to another might involve temporarily assigning illegal colors to vertices. We keep the transitions as before, i.e., we do not allow the MC to make a transition to an invalid coloring. Thus the state space is not irreducible, but rather it consists of a single irreducible component (namely, all proper colorings) plus some transient components consisting of invalid colorings. It is easy to see that this Markov chain converges to the uniform distribution on the proper colorings, as before; and, moreover, that a bound on the mixing time derived using coupling extends to a bound on the mixing time of the original Markov chain (without the invalid colorings). [**Exercise:** Verify this by going back to the basic coupling bound in Lecture 5.]

Now let X, Y be any two (not necessarily proper) colorings that are adjacent; this means that they differ at only one vertex, say v_0. We define our coupling for (X, Y) as follows:

- Pick the same vertex v in both chains.
- If v is not in the neighborhood of the unique disagreeing vertex v_0, then pick the same color c in both chains.
- If $v \in N(v_0)$, match up the choice of colors as follows:

$$\begin{aligned} c_X &\longleftrightarrow c_Y \\ c_Y &\longleftrightarrow c_X \\ c &\longleftrightarrow c, \end{aligned}$$

  where c_X, c_Y are the colors of v_0 in X, Y respectively, and $c \notin \{c_X, c_Y\}$.
- Having chosen a vertex v and a color c in both chains, recolor v with c in each chain if possible.

This is clearly a valid coupling because each of X, Y, viewed in isolation, makes a move according to the Markov chain. Note that we don't need to explicitly define the coupling on other pairs (X, Y) as this is taken care of by path coupling.

To analyze the above coupling, we count the numbers of "good" moves and "bad" moves. A "good" move corresponds to choosing the disagreeing vertex v_0 and some color not present among its neighbors; hence there are at least $q - \Delta$ good moves. A "bad" move corresponds to picking a neighbor $v \in N(v_0)$ together with one choice of color (namely, the combination c_Y in X and c_X in Y, for then v_0 is recolored to different colors in the two chains and becomes a new disagreeing vertex; note that the complementary choice c_X in X and c_Y in Y is not a bad move because neither chain will recolor v in this case). Hence the number of bad moves is at most Δ (i.e., the number of neighbors of v_0). All other moves are neutral (do not change the distance).

Since each move occurs with the same probability $\frac{1}{qn}$, and since $d(X_t, Y_t) = 1$ for all adjacent pairs X_t, Y_t, we can conclude that

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

Thus, assuming $q \geq 2\Delta + 1$, (6.1) holds with $\alpha \geq \frac{q-2\Delta}{qn}$ and $D = n$, so we get $\tau_{\text{mix}} = O(n \log n)$ as claimed. ■

Theorem 6.3 was first proved by Jerrum [Je95] (see also Salas and Sokal [SS97]) using a direct coupling rather than path coupling. The resulting analysis, though still elementary, is quite a bit messier than that above.

6.3 Going beyond the 2Δ bound

The ultimate goal of research on MCMC for graph coloring is to resolve the following:

Conjecture 6.4 *τ_{mix} is $O(n \log n)$ for all $q \geq \Delta + 2$.*

Note that this is precisely the range of q for which the Markov chain is guaranteed to be connected. The Conjecture has been recently proved for the case of Δ-regular trees [MSW06].

There is a further interesting related conjectured connection between colorings on trees and on general graphs. The range $q \geq \Delta + 1$ is precisely the region in which the Gibbs measure on the infinite Δ-regular tree is *unique*, i.e., if we set any boundary condition (fixed coloring) on the leaves of the tree, the asymptotic influence of this boundary condition on the distribution of the color at the root of the tree goes to zero with distance if and only if $q \geq \Delta+1$. It is believed that the same holds for general graphs (i.e., among all graphs of maximum degree Δ, the decay of influence with distance is slowest for the tree). Resolving either this conjecture, or the related one above for Markov chain mixing, would be of great interest in combinatorics, computer science and statistical physics.

We now briefly examine some recent work that approaches Conjecture 6.4.

6.3.1 Ideas for improving $2\Delta \to 1.76\Delta$

Recall our path coupling analysis in the proof of Theorem 6.3 above. Our bound of $q-\Delta$ on the number of good moves there was pessimistic, because in a typical coloring we would expect significantly fewer than Δ colors (the maximum possible number) to be represented in the neighborhood $N(v_0)$.

Let $A(X, v)$ denote the number of *available* colors at v in coloring X, i.e., the number of colors not represented in $N(v)$. Since the number of bad moves is (at most) Δ, the previous analysis will still go through provided we have $A(X, v) > \Delta$.

To get a handle on $A(X, v)$, suppose each vertex is colored independently and u.a.r. in X. Then by linearity of expectation we have

$$\mathrm{E}[A(X,v)] = q\Big(1-\frac{1}{q}\Big)^{\Delta} \approx qe^{-\frac{\Delta}{q}}$$

is the expected value. Thus we will have $\mathrm{E}[A(X, v)] > \Delta$ provided $qe^{-\frac{\Delta}{q}} > \Delta$, which is true whenever $q > \alpha\Delta$ where α is the solution to $x = e^{\frac{1}{x}}$. In particular, $\alpha \approx 1.76$. So we might hope that we get $O(n \log n)$ mixing time for $q > 1.76\Delta$.

This is the crux of the proof presented in [DF03]. However, quite a bit of work remains to be done: we need to justify why it is OK to work with a random coloring; in a random coloring, the neighbors of v are not colored independently; and we cannot work just with the *expected* value $\mathrm{E}[A(X, v)]$.

To sketch how to turn the above intuition into a real proof, we follow the development of [HV05].

Definition 6.5 *Say that a coloring X of G satisfies the* local uniformity *property if, for all v, $A(X, v)$ is at least $q(e^{-\Delta/q} - \delta)$.*

Here δ is an arbitrary small constant. Thus local uniformity says that the number of available colors at all vertices is not much less than the expected value we calculated above.

The following Fact captures formally the intuition from our informal calculation above:

Fact 6.6 *Let G be triangle-free and have maximum degree Δ. Assume that the number of colors satisfies $q \geq \max\{\Delta+2/\delta, C\log n\}$, where C is a constant that depends on δ. Then a random q-coloring of G satisfies the local uniformity property (w.r.t. δ) with probability at least $1 - O(n^{-4})$.*

The Fact is really only interesting in the case where $\Delta \geq C \log n$. In that case the condition on q is only $q \geq \Delta + \text{const}$, which will certainly hold in our application.

We will not prove this Fact here; the proof follows from a straightforward application of large deviation bounds on the colors of the neighbors of a given vertex v, conditional on an arbitrary fixed coloring of the rest of the graph. (Note that, since G is triangle-free, these colors are conditionally independent. And since the number of colors is $\geq C \log n$, large deviation bounds hold.) For the details, see [HV05].

To use this fact, we assume $\Delta \geq C \log n$ and we return to our original coupling idea (without path coupling). (As mentioned earlier, [Je95] shows how to obtain $q \geq 2\Delta + 1$ using a direct coupling.) Now we note that, rather than coupling two arbitrary initial states (X_0, Y_0), it is enough to couple an arbitrary state X_0 with a *stationary* Y_0, i.e., in the coupling we may assume that Y_0 is a uniformly random coloring. [**Exercise:** Check this by going back to our original justification of coupling at the beginning of Lecture 5.] So at all times t, Y_t is a uniformly random coloring, and thus by Fact 6.6 it is locally uniform with high probability. In fact, if we let $T = cn \log n$ (for some constant c), then applying a union bound over times t we see that Y_t satisfies local uniformity for $t = 0 \ldots T$ with probability $\geq 1 - O(n^{-2})$.

Now, exploiting local uniformity to bound the number of good moves as sketched earlier, we get an expected contraction $1 - O(\frac{1}{n})$ in distance at each step, from which we can conclude that if $q \geq 1.76\Delta$ then $\Pr[X_T \neq Y_T] \leq O(n^{-2})$. (Here we are choosing the constant c large enough that the probability of not having coupled is this small. Note that some work still needs to be done here; in particular, we need to check that it is enough that just one of the two colorings, Y_t, satisfies local uniformity.)

Hence, we obtain that

$$E[d(X_T, Y_T)|X_0] \leq n(O(n^{-2}) + O(n^{-2})) = O(n^{-1}),$$

where the first $O(n^{-2})$ term bounds the probability that local uniformity fails, and the second bounds the probability that $\Pr[X_T \neq Y_T]$, and the factor n comes from the fact that we can always bound $d(X_T, Y_T)$ by its maximum value n. Thus, using Markov's inequality we get

$$\Pr[d(X_T, Y_T) > 0|X_0] = \Pr[d(X_T, Y_T) \geq 1|X_0] \leq E[d(X_T, Y_T)|X_0] \leq O(n^{-1}).$$

This establishes the result with $q \geq 1.76\Delta$ for triangle-free graphs of degree $\Delta \geq C \log n$.

See [HV05] for the missing details of the above argument.

6.3.2 Further Reading

Dyer and Frieze [DF03] were the first to obtain $\tau_{\text{mix}} = O(n \log n)$ for $q \geq \alpha\Delta$, with $\alpha \approx 1.76$, but using a more complicated argument than the one sketched above. (Rather than assuming Y_0 is stationary, they instead argued that, after sufficiently many steps T, Y_T satisfies a similar local uniformity condition.) In [M04], Molloy improves the $q \geq \alpha\Delta$ bound from [DF03,HV05] to $q \geq \beta\Delta$, where $\beta \approx 1.49$, for graphs meeting the above requirements (i.e., triangle-free and maximum degree $\Delta = \Omega(\log n)$). In [HV03], Hayes and Vigoda use a very delicate *non-Markovian* coupling to demonstrate $O(n \log n)$ mixing time for $q \geq (1 + \epsilon)\Delta$, but the $\Delta = \Omega(\log n)$ requirement remains, and moreover the associated constant in the O-expression depends on ϵ. [DFHV04] demonstrates that the $q \geq \alpha\Delta$ result of [DF03] holds for constant-degree graphs, provided $\Delta = \Omega(1)$ is sufficiently large. It is still open whether the stronger results also apply for constant-degree graphs. Random graphs with constant (average) degree are discussed in [DFFV06], where it is shown that, with high probability over the choice of the graph, $O(n \log n)$ mixing time holds for q down to about $o(\log \log n)$, which is much smaller than the maximum degree (which is of order $\frac{\log n}{\log \log n}$). Finally, while a graduate student at Berkeley, Eric Vigoda [Vi99] was the first to show that a (slightly more complicated) local Markov chain on colorings has mixing time $O(n \log n)$ for q below 2Δ; specifically, his result holds for $q \geq \frac{11}{6}\Delta$. This remains the best known result for a local Markov chain without any additional restrictions on the graph.

References

[BD97] R. Bubley and M.E. Dyer, "Path coupling: a technique for proving rapid mixing in Markov chains," *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science (FOCS)*, 1997, pp. 223–231.

[DFFV06] M. Dyer, A. Flaxman, A. Frieze and E. Vigoda, "Randomly coloring sparse random graphs with fewer colors than the maximum degree," *Random Structures and Algorithms* **29** (2006), pp. 450–465.

[DF03] M. Dyer and A. Frieze, "Randomly coloring graphs with lower bounds on girth and maximum degree," *Random Structures and Algorithms* 23, 2003, pp. 167–179.

[DFHV04] M. Dyer, A. Frieze, T.P. Hayes, E. Vigoda, "Randomly Coloring Constant Degree Graphs," *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, 2004 pp. 582–589

[HV03] T.P. Hayes and E. Vigoda, "A Non-Markovian Coupling for Randomly Sampling Colorings," *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003, pp. 618–627

[HV05] T.P. Hayes and E. Vigoda, "Coupling with the stationary distribution and improved sampling for colorings and independent sets," *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2005, pp. 971–979

[Je95] M.R. Jerrum, "A very simple algorithm for estimating the number of k-colourings of a low-degree graph," *Random Structures and Algorithms* **7** (1995), pp. 157–165.

[MSW06] F. Martinelli, A. Sinclair and D. Weitz, "Fast mixing for independent sets, colorings and other models on trees," *Random Structures & Algorithms*, 2006, to appear imminently.

[M04] M. Molloy, "The Glauber Dynamics on Colorings of a Graph with High Girth and Maximum Degree," *SIAM Journal on Computing* 33, 2004, pp. 721–737

[SS97] J. Salas and A.D. Sokal, "Absence of Phase Transition for Antiferromagnetic Potts Models via the Dobrushin Uniqueness Theorem," *Journal of Statistical Physics* **98** (2000), p. 551.

[Vi99] E. Vigoda, "Improved bounds for sampling colorings," *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science*, 1999, pp. 51–59.