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

Lecture 18: November 5

Lecturer: Prof. Alistair Sinclair *Scribes: Sharmodeep Bhattacharyya and Bharath Ramsundar*

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.*

18.1 Continuation of Volume Estimation

In the past two lectures, we have been analyzing the construction of an f.p.r.a.s. to estimate the volume of a convex body $K \subseteq \mathbb{R}^n$ in the membership oracle model. We assumed a general upper bound on the mixing time of heat-bath ball walks, and used this assumption to analyze our construction. In this section, we prove the assumed bound. Recall that the Poincaré constant is defined by $\alpha = \inf_{\varphi: K \to \mathbb{R}} \frac{\int_K h_\varphi d\mu}{\mathrm{Var}_\mu \varphi}$, where the local variance is $h_\varphi = \frac{1}{2}\int_K (\varphi(x) - \varphi(y))^2 p(x, dy)$. This is a continuous analog of the discrete version introduced in an earlier lecture.

First we need to generalize our definition of mixing time to the continuous setting:

Definition 18.1 *For $B \subseteq \mathbb{R}^n$, let $\tau_B(\epsilon) := \min\{t \in \mathbb{N}^+ | \forall$ measurable $S \subseteq K,\ |p^t(B,S) - \mu(S)| \le \epsilon\}$ where μ is the stationary measure for the given Markov chain, and $p^t(B,S)$ denotes the probability that the chain is in S after t steps starting from the uniform distribution on B.*

Claim 18.2 *Let $B \subseteq K$ be a δ-ball contained in K. (This exists since K is assumed to contain the unit ball.) Then*

$$\tau_B(\epsilon) \le O\left(\frac{1}{\alpha}\left(n \ln\left(\frac{D}{2\delta}\right) + \ln(\epsilon^{-1})\right)\right),$$

where D is the diameter of K.

Proof: Let μ be the stationary measure on K for the given Markov chain. By mimicking the proof in the discrete case, we have that $\mathrm{Var}_\mu P\varphi \le \mathrm{Var}_\mu \varphi - \alpha \mathrm{Var}_\mu \varphi$, where the operator P is defined in analogous fashion to the discrete case, i.e., $P\varphi(x) = \int_K p(x, dy)\varphi(y))$. Thus, by induction, $\mathrm{Var}_\mu P^t \varphi \le \exp(-\alpha t) \mathrm{Var}_\mu \varphi$.

Now let S be any measurable subset of K, and φ the indicator function of S. Since $\mathrm{Var}_\mu \varphi \le 1$ we have

$\mathrm{Var}_\mu P^t\varphi \le \exp(-\alpha t)$. Also,

$$\begin{aligned}
\mathrm{Var}_\mu P^t\varphi &= \int_K (P^t\varphi(y) - \mu(S))^2 \mu(dy) \\
&= \int_B (P^t\varphi(y) - \mu(S))^2 \mu(dy) \\
&\ge \frac{2}{5}\frac{\mathrm{vol}\, B}{\mathrm{vol}\, K}\int_B \frac{1}{\mathrm{vol}\, B}(P^t\varphi(y) - \mu(S))^2 dy \\
&\ge \frac{2}{5}\frac{\mathrm{vol}\, B}{\mathrm{vol}\, K}\left(\int_B \frac{1}{\mathrm{vol}\, B}(P^t\varphi(y) - \mu(S)) dy\right)^2
\end{aligned}$$

In the second line we may restrict the integral to B because we start with density zero outside B. In the third line we have used our curvature assumption to write $\mu(dy) \ge \frac{2}{5}\frac{dy}{\mathrm{vol}\, K}$. The last inequality follows from the fact $E[X^2] \ge E[X]^2$.

To bound the mixing time, we want to find t such that $|p^t(B, S) - \mu(S)| \le \epsilon$. Note that $p^t(B, S) - \mu(S) = \int_B \frac{1}{\mathrm{vol}\, B}(P^t\varphi(y) - \mu(S))dy$. Thus, it follows that if $\mathrm{Var}_\mu P^t\varphi \le \epsilon^2 \frac{2}{5}\frac{\mathrm{vol}\, B}{\mathrm{vol}\, K}$, then our bound is proven. But $\mathrm{Var}_\mu P^t\varphi \le \exp(-\alpha t)$, so we need only set $t = \frac{1}{\alpha}\ln(\frac{5}{2\epsilon^2}\frac{\mathrm{vol}\, K}{\mathrm{vol}\, B})$. Since the volume of K is bounded above by the volume of the n-dimensional ball of diameter D, we have that $\frac{\mathrm{vol}\, K}{\mathrm{vol}\, B} \le (\frac{D}{2\delta})^n$. Thus, $t \le O(\frac{1}{\alpha}(n\ln(\frac{D}{2\delta}) + \ln(\frac{1}{\epsilon})))$, as required. ■

18.2 Conductance and Sparsest Cut

Once again, we return to the framework of looking at a discrete Markov chain as a multicommodity flow network. In earlier lectures, we saw that any flow gave us an upper bound on mixing time. We will now look at *cuts*, the dual of flows, to develop a method for lower-bounding the mixing time of a Markov chain. Recall that in the single source/single sink model, the size of a maximum flow equals the size of a minimum cut. This statement isn't necessarily true when there are multiple sources and sinks, but we can still get bounds.

Definition 18.3 *For any subset $S \subseteq \Omega$,*

$$\Phi(S) := \frac{\sum_{\substack{x\in S \\ y\in S}} C(x,y)}{\pi(S)} = \frac{C(S, \overline{S})}{\pi(S)}$$

We may think of this quantity as the normalized flow passing out of subset S, or equivalently as the probability of escaping from S in one step of the Markov chain in the stationary distribution, conditional on starting in S.

Definition 18.4 *Define the* conductance Φ *as*

$$\Phi \equiv \min_{S: 0 < \pi(S) \le 1/2} \Phi(S)$$

The conductance corresponds to the "sparsest cut" in the graph. Combinatorially, a small cut is a bottleneck in the graph. Probabilistically, it is a subset which the Markov chain is unlikely to leave in one move. Consequently, if Φ is very small, mixing time should be very large and vice versa. We note that restricting attention to sets with $\pi(S) \leq 1/2$ is natural, for if S is large then we should not be too concerned if it takes a long time to leave it. However, we can also define the following more symmetrical quantities:

Definition 18.5 *For any subset $S \subseteq \Omega$,*

$$\Phi'(S) = \frac{C(S,\overline{S})}{\pi(S)\pi(\overline{S})} = \frac{C(S,\overline{S})}{D(S,\overline{S})}$$

$$\Phi' = \min_{S:0<\pi(S)\leq 1/2} \Phi'(S)$$

Here $D(S,\overline{S}) = \sum_{x\in S,\, y\in\overline{S}} \pi(x)\pi(y)$ is the total demand between vertices in S and in $\overline{S}$.

For $\pi(S) \leq 1/2$, it is clear that $\Phi(S) \leq \Phi'(S) \leq 2\Phi(S)$. Hence Φ and Φ' differ by at most a factor of 2. It is also obvious that for any flow f, $\rho(f) \geq \frac{1}{\Phi'}$ (the quantity on the r.h.s is the maximum ratio of demand to capacity for sets, so in any flow there must be at least one edge that has this ratio of demand to capacity). Less obviously, we have that $\rho(f) \leq O(\frac{\log|\Omega|}{\Phi'})$. This result follows from the approximation to sparsest cut devised by Leighton and Rao [LR99]; see [Si92] for the application to Markov chains. This bound is tight, as the following exercise asks you to verify.

Exercise 18.6 *Verify that in a constant degree expander, we have $\rho(f) = \Theta(\log|\Omega|)$, but Φ' is constant.*

We now state two theorems relating conductance and mixing time. The first is an upper bound on mixing time:

Theorem 18.7 *For any lazy, reversible Markov chain,*

$$\tau_X(\epsilon) \leq \text{const} \times \left[\frac{1}{\Phi^2}\left(\log \pi(x)^{-1} + \log \epsilon^{-1}\right)\right] \tag{18.1}$$

Proof: **[Sketch]** The proof follows from the following bound on eigenvalue gap (sometimes known as "Cheeger's inequality" because of its more classical continuous analog):

$$1 - \lambda_2 \geq \text{const}.\Phi^2 \tag{18.2}$$

So, by Fact 10.9 in lecture 10, $\alpha = 1 - \lambda_2 \geq \text{const}.\Phi^2$ and then by Thm 10.5 in lecture 10, we get that

$$\tau_x(\epsilon) \leq \text{const}\left[\frac{1}{\Phi^2}\left(\log \pi(x)^{-1} + \log \epsilon^{-1}\right)\right].$$

■

We will not prove (18.2) (which is quite non-trivial) here; for a proof, see [SJ89]. We also will not dwell on the above theorem as it is usually less useful than the analogous upper bound based on flows (Corollary 11.3

from Lecture 11). The reason for this is that, while *any* flow gives us an upper bound on the mixing time, in order to use Theorem 18.7 we need to quantify over *all* cuts, which is inherently harder. Thus, with the exception of a few geometric examples where Φ can be bounded directly using an *isoperimetric inequality* (i.e., a bound on surface area vs volume), Theorem 18.7 is not usually the best path to upper bounds.

On the other hand, by the same token, conductance *is* very useful for proving *lower* bounds on mixing times, because any set S provides such a bound as the following easy result confirms:

Theorem 18.8 *For any Markov chain, and any $S \subseteq \Omega$ with $\pi(S) \le \frac{1}{2}$,*

$$\tau_X(1/4) \ge \frac{1}{4\Phi(S)}. \tag{18.3}$$

Proof: Consider initial distribution

$$p^0(x) = \begin{cases} \frac{\pi(x)}{\pi(S)}, & \text{if } x \in S \\ 0, & \text{if } x \notin S \end{cases}$$

Then the distribution p^1 after 1 step satisfies

$$\begin{aligned} ||p^1 - p^0||_{TV} &= \frac{1}{2}\sum_y \left|\sum_x p^0(x)P(x,y) - p^0(y)\right| \\ &= \sum_{y\in\bar{S}} \left|\sum_x p^0(x)P(x,y) - p^0(y)\right| \\ &= \sum_{x\in S}\sum_{y\in\bar{S}} p^0(x)P(x,y) \qquad (\because p^0(y) = 0, \text{ for } y \in \bar{S}) \\ &= \Phi(S). \end{aligned}$$

The second equality above follows from the fact that, $||p^1 - p^0||_{TV} = \sum_{y\in A} \left|p^1(y) - p^0(y)\right|$, where $A = \{y \in \Omega : |p^1(y) - p^0(y)| > 0\}$, together with the observation that in this case $A = \bar{S}$. **Exercise:** check this!

We leave the following simple observation as an exercise:

Exercise 18.9 $||p^{t+1} - p^t||_{TV} \le ||p^t - p^{t-1}||_{TV}$

Iterating the result of this Exercise gives $||p^t - p^0|| \le t\Phi(S)$. Thus, using the triangle inequality and the fact that $||p^0 - \pi|| = \pi(\overline{S}) \ge \frac{1}{2}$, we get

$$\begin{aligned} ||p^t - \pi|| &\ge ||p^o - \pi|| - ||p^t - p^0|| \\ &\ge \frac{1}{2} - t\Phi(S) \\ &\ge \frac{1}{4} \qquad (\text{assuming } t \le \frac{1}{4\Phi(S)}). \end{aligned}$$

So, $\tau_x\left(\frac{1}{4}\right) \ge \frac{1}{4\Phi(S)}$, as claimed. ∎

Corollary 18.10 *If we can find $S \subseteq \Omega$ with $\pi(S) \leq \frac{1}{2}$ and $\frac{\pi(\partial S)}{\pi(S)} \leq \delta$, where ∂S denotes the interior boundary of S (i.e., points in S that are connected to a point in $\overline{S}$), then $\tau_{mix} \geq \frac{1}{4\delta}$.*

The proof of the Corollary follows immediately from the easy observation that $\frac{\pi(\partial S)}{\pi(S)} \geq \frac{C(S,\bar{S})}{\pi(S)}$.

18.3 Independent Sets

Input: Graph, $G = (V, E)$, Δ =maximum degree of the graph.
Goal: Sample an independent set (of any size) of vertices in G u.a.r. (or, equivalently, give a f.p.r.a.s. for counting independent sets).

Facts 18.11

1. *In general there can be no polynomial time sampling algorithm (or equivalently, no f.p.r.a.s.) for the problem unless $NP = RP$.*
2. *The same holds even for graphs with maximum degree $\Delta \geq 25$.*
3. *A simple MCMC algorithm (with local moves) solves the problem in polynomial time for graphs with $\Delta \leq 5$.*
4. *No MCMC algorithm (from a very wide class) can work in polynomial time for $\Delta = 6$.*

We will prove Fact 1 (which is very easy) below. Fact 2, due to Dyer, Frieze and Jerrum [DFJ02], is a more refined version of Fact 1, which replaces the simple NP-hardness of independent set by a hardness of approximation derived from the PCP theorem; we will not prove this. Fact 3 is a recent, and rather substantial result of Weitz [We06]; we may prove this later if there is time. Finally, we will prove the negative Fact 4 (also due to [DFJ02]) in the next lecture, using the lower bound on mixing times given in Theorem 18.8 above.

Before proceeding, note that the above Facts naturally suggest the following intriguing conjecture:

Conjecture 18.12 *There can be no polynomial time algorithm for the problem even for graphs with maximum degree $\Delta \geq 6$.*

This would extend the negative result of Fact 2 all the way down to $\Delta = 6$, and provide a sharp dichotomy since by Fact 3 the problem is tractable for $\Delta \leq 5$. In the next lecture we will prove Fact 4, which can be viewed as strong evidence for the Conjecture (since MCMC is the only algorithmic tool we know here, and the proof of Facts 3 and 4 also suggest a qualitative change in the problem as Δ moves from 5 to 6).

We end today's lecture with the proof of the easy Fact 1.

Proof: (of (1)) Let $G = (V, E)$ be arbitrary, $|V| = n$. Let $I \subseteq V$ be an independent set of size k in G. Construct graph G' by replacing each vertex of G by a group of r vertices with no edges between them, and

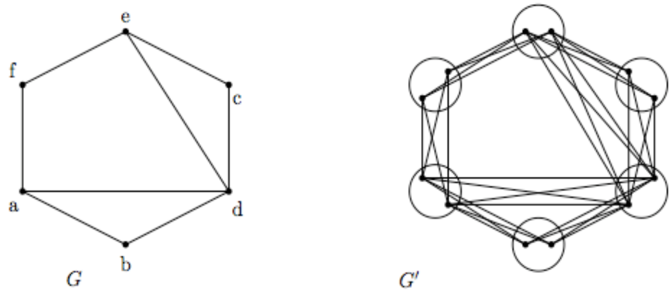

Figure 18.1: G and G' for $r = 2$.

connecting all pairs of vertices in two such groups iff there is an edge between the two corresponding vertices of G. The construction of G' is explained in Fig. 1 for $r = 2$.

Thus each independent set I of size k in G gives rise to $(2^r - 1)^k$ independent sets in G' corresponding to I (since we can choose any non-empty subset of each of the k associated groups). So, in G', the number of independent sets of size $< k$ is $\leq 2^n(2^r - 1)^{k-1}$.

Now choose r such that $(2^r - 1)^k \gg 2^n(2^r - 1)^{k-1}$. (Taking $r = cn$ for a suitable constant c suffices; note that G' can be constructed in polynomial time from G.) In other words, when translated to G', even a single independent set of size k in G completely swamps all the smaller independent sets.

Thus we see that, if we could approximately count (even very crudely) the number of independent sets in G', then we could determine the maximum size of an independent set in G. Since this latter problem is NP-hard, this would imply that $NP = P$ (or, if the approximate counting algorithm is randomized, that $NP = RP$). ■

References

[DFJ02] M. Dyer, A. Frieze and M. Jerrum. On counting independent sets in sparse graphs. *SIAM Journal on Computing* **31** (2002), pp. 1527–1541.

[LR99] F.T. Leighton and S.B. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM* **46** (1999), pp. 787–832.

[Si92] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing* **1**, 1991, pp. 351–370.

[SJ89] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information & Computation* **82** (1989), pp. 93–133.

[We06] D. WEITZ, Counting Independence Sets up to the tree threshold. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006, Session 4A, pp. 140–149.