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

Lecture 11: October 13

Lecturer: 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.*

11.1 Review

Recall from the previous lecture the definition of the *Poincaré constant*

$$\alpha := \inf_{\varphi \text{ non-constant}} \frac{\mathcal{E}_\pi(\varphi,\varphi)}{\mathrm{Var}_\pi \varphi} \ ,$$

where the infimum is over all non-constant functions $\varphi : \Omega \to \mathbb{R}$. In the above expression, the numerator $\mathcal{E}_\pi(\varphi,\varphi)$ is the "local variance" and the denominator is the global variance of φ w.r.t. the stationary distribution π.

In the previous lecture we proved the following theorem relating the mixing time to α.

Theorem 11.1 *For any lazy ergodic P and any initial state $x \in \Omega$,*

$$\tau_x(\varepsilon) \le \frac{1}{\alpha}\left(2\ln\varepsilon^{-1} + \ln\pi(x)^{-1}\right) \ .$$

We then proved the following combinatorial bound on α in terms of multicommodity flows on the Markov chain (viewed as a network).

Theorem 11.2 *For any ergodic P and any flow f for P*

$$\alpha \ge \frac{1}{\rho(f)\ell(f)} \ ,$$

where $\rho(f)$ is the cost of f and $\ell(f)$ is the length of f.

The above two results lead to the following immediate corollary:

Corollary 11.3 *For any lazy, ergodic Markov chain P and any flow f, the mixing time from any initial state $x \in \Omega$ is bounded by*

$$\tau_x(\varepsilon) \le \rho(f)\ell(f)\left(2\ln\varepsilon^{-1} + \ln\pi(x)^{-1}\right) \ .$$

We briefly mention here, without proof, some (partial) converses to the above bounds.

- Converse for Theorem 11.1: $\tau_{mix} \ge \text{constant} \cdot \frac{1-\alpha}{\alpha}$.
- Converse for Theorem 11.2: There exists a flow f with $\alpha \le \text{constant} \cdot \frac{\log|\Omega|}{\rho(f)}$. (This is a consequence of the $\log n$ approximation to sparsest cut devised by Leighton-Rao using multicommodity flows; see [Si92].)

- Converse for Corollary 11.3: There exists a flow f with $\rho(f) \leq \text{constant} \cdot \tau_{\text{mix}}$. (This flow is exactly that generated by the probability mass flow in the Markov chain itself; see [Si92].)

In this lecture we will apply the above flow technology to obtain bounds on the mixing time of various Markov chains. We begin with a few very simple warm-up examples.

11.2 Example: Random Walk on the Hypercube

Consider the usual lazy random walk on the hypercube $\{0,1\}^n$. Let $N = 2^n$ be the number of vertices. Then $\pi(x) = \frac{1}{N}$. The edge capacity is $C(u,v) = \pi(u)P(u,v) = \frac{1}{N}\frac{1}{2n} = \frac{1}{2Nn}$ for all edges (u,v). The demand between vertices x, y is $D(x,y) = \pi(x)\pi(y) = \frac{1}{N^2}$.

Now we need to construct a flow f satisfying the demands above. Moreover, to obtain the best bound possible we want to do this maximizing the ratio $\frac{1}{\rho(f)\ell(f)}$. Intuitively, this is achieved by spreading the flow between x and y evenly among all shortest paths from x to y. So we define the flow f in this way.

To compute $\rho(f)$ we now need to get a handle on $f(e)$ for all e. This is not difficult if we use the symmetry of the hypercube to notice that $f(e) = f(e')$ for all edges e, e', so that:

$$f(e) = \frac{\sum_{e \in E} f(e)}{|E|} = \frac{\frac{1}{N^2}\sum_{x,y \in V}\{\text{length of shortest path between } x, y\}}{Nn}$$

Notice $|E| = Nn$ as, in the context of flow, we consider edges to be directed. Observe that the average distance between two vertices in the hypercube is $n/2$ so that:

$$\sum_{x,y \in V}\{\text{length of shortest path between } x, y\} = N^2\frac{n}{2}$$

Thus we have:

$$f(e) = \frac{\frac{1}{N^2}\frac{n}{2}N^2}{Nn} = \frac{1}{2N}\ .$$

Hence, $\rho(f) = \max_{e \in E}\frac{f(e)}{C(e)} = \frac{1}{2N}2Nn = n$. Moreover, as all the flow goes along shortest paths we have also that $\ell(f) = n$, so from Corollary 11.3 we get (for some constant c)

$$\begin{aligned} \tau_x(\varepsilon) &\leq \rho(f)\ell(f)\left(c + \ln \pi(x)^{-1}\right) \\ &= n \cdot n \cdot O(n) = O(n^3). \end{aligned}$$

This is not tight, as we saw in an earlier lecture that $\tau_{\text{mix}} \sim \frac{n}{2}\ln n$ (and indeed obtained a similar bound, up to a constant factor, by coupling). However, our flow argument does yield a polynomial upper bound on the mixing time despite the fact that the size of the cube is exponential in n, so this is a non-trivial result. The slackness in our bound is typical of this method, which is heavier-duty than coupling.

Let's look in a bit more detail at where we lose here. In the case of the hypercube, it is actually possible to compute exactly the Poincaré constant α (or, equivalently since this is a reversible chain) the spectral gap $1 - \lambda_2$) to be $\sim \frac{1}{n}$. This shows that both Theorems 11.1 and 11.2 are not tight, as:

- Using the above exact value of α, Theorem 11.1 gives $\tau_{mix} = O(n^2)$, which is off by a factor of almost n from the true value. This error results from the factor $\ln \pi(x)^{-1}$, which arises essentially because we approximate the mixing time using only the second eigenvalue, treating the other eigenvalues pessimistically.
- Theorem 11.2 bounds α by $\frac{1}{\rho(f)\ell(f)} = 1 - \frac{1}{n^2}$, which is also loose as the true value is $\alpha \sim \frac{1}{n}$. This demonstrates the potential slackness in our bound relating α to flows.

11.3 Example: Random Walk on a Line

We consider the Markov chain for lazy random walk on the line $\{1, 2, \ldots, N\}$, with self-loop probability 1/2 at every state (except at the endpoints where it is 3/4). (Thus the chain moves left or right from each position with probability 1/4 each.) The stationary distribution is uniform: $\pi(x) = 1/N$. The edge capacities and demands are

$$C(e) \;=\; \pi(x)P(x,y) \;=\; \frac{1}{4N} \;\text{ for all non-self loop edges },$$

$$D(x,y) \;=\; \pi(x)\pi(y) \;=\; \frac{1}{N^2}\;\forall x,y \;.$$

There is only one simple path between each pair of vertices, so there is a unique flow f here. The amount of flow on any edge $(i, i+1)$ is given by

$$f\big((i,i+1)\big) \;=\; i(N-i)\frac{1}{N^2} \;\leq\; \frac{1}{4} \;.$$

(The maximum is achieved on the middle edge.) The cost of the flow f is

$$\rho(f) \;=\; \max_e \frac{f(e)}{C(e)} \;\leq\; \frac{1/4}{1/(4N)} \;=\; N \;,$$

and the length is

$$\ell(f) \;=\; N \;.$$

Thus from Theorem 11.2 the Poincar'e constant is bounded by

$$\alpha \;\geq\; \frac{1}{\rho(f)\ell(f)} \;\geq\; \frac{1}{N^2} \;,$$

which happens to be asymptotically tight for this Markov chain. From Corollary 11.3 the bound on the mixing time is

$$\tau_{mix} = O(N^2 \log N).$$

This is off from the true answer only by the $O(\log N)$ factor (which again arises from Theorem 11.1).

11.4 Example: Random Walk on $K_{2,N}$

Consider lazy random walk on the complete bipartite graph $K_{2,N}$, with a self-loop of 1/2 at every vertex. Label the two vertices on the small side of the graph s, t respectively.

The stationary distribution is

$$\pi(s) \;=\; \pi(t) \;=\; \frac{1}{4}\;, \qquad \pi(x) \;=\; \frac{1}{2N}\;\forall x \neq s, t.$$

The capacity $C(e) = 1/(8N)$ for all edges and the demands are given by:

$$\begin{aligned}
D(s,t) \;&=\; D(t,s) \;=\; 1/16, \\
D(s,x) \;&=\; D(t,x) \;=\; D(x,s) \;=\; D(x,t) \;=\; 1/(8N), \\
D(x,y) \;&=\; 1/(4N^2).
\end{aligned}$$

Consider two strategies to route flow between any pair of states:

(1) Send flow along a single shortest path for each pair of states. Then exactly four edges will carry 1/16 units of flow for the $s \to t$ and $t \to s$ paths. Construct the flow f such that the edges carrying the maximum flow only carry flows from/to s and t. Then,

$$\max_e f(e) \;\leq\; \frac{1}{8N} + \frac{1}{16} \;,$$

which gives,

$$\rho(f) \;\leq\; \frac{1/(8N) + 1/16}{1/(8N)} \;=\; \textit{constant} \times N \;.$$

Also $\ell(f) = 2$. Hence (assuming we start at either s or t), we get a bound on the mixing time of $O(N)$. This is a severe over-estimate as the true mixing time (starting from any state) is $\Theta(1)$ (why?)

(2) Distribute $s \to t$ and $t \to s$ flows along all shortest paths evenly. Then,

$$\max_e f(e) \;\leq\; \frac{1}{8N} + \frac{1}{16N} + \frac{1}{4N^2}\frac{N-1}{2} \;\leq\; \textit{constant} \times \frac{1}{N} \;,$$

which gives,

$$\rho(f) \;\leq\; \frac{\textit{constant} \times 1/N}{1/(8N)} \;=\; \textit{constant} \;.$$

Hence, the mixing time starting from s or t is $O(1)$, which is correct.

This simple example shows that it is sometimes necessary to spread the flow over many paths in order to get a good value for $\rho(f)$.

11.5 Generic Flow Calculation

Our examples so far have been on symmetric graphs where computing the cost of a flow was easy. But in general, it is difficult to determine how much flow is carried on an edge. To solve this, we will develop technology to count paths and calculate flows in a generic setup.

Let $|\Omega| = N$ be the size of the state space of an ergodic, lazy Markov chain, where N is exponential in n, the natural measure of problem size. Assume the stationary distribution is uniform ($\pi(x) = 1/N$). Also assume $P(u,v) \;\geq\; 1/\textit{poly}(n)$ for all non-zero transition probabilities. (This implies that the degree of the underlying graph is not huge, and holds in most of our examples.) Then the capacity of any edge is

$$C(u,v) \;=\; \pi(u)P(u,v) \;\approx\; \frac{1}{N\textit{poly}(n)} \;.$$

Our goal is to construct a flow f such that

$$\frac{f(e)}{C(e)} \;\leq\; \textit{poly}(n), \qquad \ell(f) \;\leq\; \textit{poly}(n) \;,$$

in order to provide a polynomial upper bound on mixing time. Hence we must have

$$f(e) \;\leq\; \frac{\textit{poly}(n)}{N} \;. \tag{11.1}$$

On the other hand, the number of edges in the graph is less than $N \times \textit{poly}(n)$ and the total flow along all paths is $\sum_{x,y} \frac{1}{N^2} \approx 1$. Hence *some* edge must carry at least $1/(N \times \textit{poly}(n))$ flow, i.e.,

$$f(e) \;\geq\; \frac{1}{N \times \textit{poly}(n)} \;.$$

Comparing this with Eq. (11.1) we see that any good flow has to be optimal up to a polynomial factor.

Finally, suppose the flow $x \to y$ goes along a *single* path $\gamma_{x,y}$. (Again, in many of our examples this will be the case.) Let *paths*(e) denote the set of paths through edge e under flow f. Then

$$f(e) \ = |\ paths(e)\ | \ \ \times \frac{1}{N^2} \ .$$

Since we wanted $f(e) \leq poly(n)/N$, this implies that the flow must satisfy

$$|\ paths(e)\ | \leq \ N \times poly(n) \ .$$

The goal, then, is to set up a flow such that the number of paths along any edge is $N \times poly(n)$. This will give us a polynomial bound on the mixing time using Theorem **??**. However, we typically don't know N, the size of the state space; in fact, this is often what we are trying to compute! So we can't hope to explicitly count the paths through an edge and compare this value with N. The way we get around this is to construct an *injective map*,

$$\eta_e : \ paths(e) \hookrightarrow \Omega \ .$$

Such a map (if we can construct it) will show implicitly that $|paths(e)| \leq |\Omega|$, which is what we want, without any counting. We continue this discussion in the next lecture.

References

[Si92] A. Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinatorics, Probability and Computing* **1**, 1992, pp. 351–370.