CS294-2 Markov Chain Monte Carlo: Foundations & Applications **Fall 2002**

Lecture 12: October 10

Lecturer: Alistair Sinclair *Scribes: Itzik Parnafes and Alice Zheng*

Reminder from the last lecture: We're working in a setup with an ergodic and reversible Markov Chain. On the induced graph we define the following quantities:

Capacity For every edge $e = (x, y)$ define $C(e) = C(x, y) = \pi(x) \cdot P(x, y)$. Reversibility of the MC ensures that $C(x, y) = C(y, x)$.

Demand For each pair of (not necessarily adjacent) states x, y we define the demand $D(x, y) = \pi(x) \cdot \pi(y)$.

Let $\mathcal{P}_{x,y}$ denote the set of all simple paths between nodes x and y, and $\mathcal{P} = \cup_{x,y}\mathcal{P}_{x,y}$. A *flow* on the graph is a function $f : \mathcal{P} \mapsto \mathcal{R}^+$ assigning a non-negative weight to all simple paths in the network such that

$$\sum_{p \in \mathcal{P}_{x,y}} f(p) = D(x, y),$$

i.e., the flow needs to satisfy the demand between any two states.

Each flow is associated with a cost and a length.

Cost of a flow is

$$\rho(f) = \max_e \frac{f(e)}{C(e)}$$

where we define the flow through the edge e as

$$f(e) = \sum_{p \ni e} f(p)$$

Length of a flow f is simply the longest path that carries non-zero flow:

$$l(f) = \max_{p: f(p) > 0} |p|$$

12.1 The main theorem

In this lecture, we will upper bound the mixing time of a Markov Chain by properties of any flow on the induced graph.

Theorem 12.1 *If a Markov Chain MC is ergodic, reversible and lazy, then for any flow f on the induced graph, we have*

$$\tau_x(\epsilon) \leq \left(\log((2\sqrt{\pi(x)})^{-1}) + \log(\epsilon^{-1})\right) \cdot \rho(f) \cdot l(f).$$

As an immediate corollary we get

$$\tau_{mix} \leq \text{const} \cdot \log(\pi_{min}^{-1}) \cdot \rho(f) \cdot l(f),$$

where $\pi_{min} = \min_x \pi(x)$.

The game plan: we will prove the theorem using two lemmas. First we will give a general upper bound on the variation distance in terms of the "spectral gap" of the Markov chain (which does not involve flows). Then we will bound the spectral gap using the flow.

Recall that, for an ergodic, reversible transition matrix P, the eigenvalues are real and satisfy

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_N > -1.$$

If in addition P is lazy, we have that $\lambda_N \geq 0$. Lemma 12.2 below will show that

$$\Delta_x(t) \leq \frac{1}{2\sqrt{\pi(x)}} \cdot \lambda_2^t$$

which implies that

$$\tau_x(\epsilon) \leq \left(\log((2\sqrt{\pi(x)})^{-1}) + \log(\epsilon^{-1})\right) \cdot \frac{1}{1-\lambda_2}$$

where the factor $1 - \lambda_2$ is usually called the *spectral gap.*

In Lemma 12.3 we will show that for any flow f in the network,

$$1 - \lambda_2 \geq \frac{1}{\rho(f) \cdot l(f)}.$$

12.2 The First Lemma

We will now prove the first part:

Lemma 12.2 *For an ergodic, reversible, lazy MC, the variation distance can be upper bounded by*

$$\Delta_x(t) \leq \frac{1}{2\sqrt{\pi(x)}} \cdot \lambda_2^t.$$

Proof: Throughout the proof we will assume that P is symmetric. If it is asymmetric (but still reversible) we can apply the same argument to the matrix DPD^{-1}, where $D = diag(\sqrt{\pi(x)})$, which is symmetric. (See Problem Set 3, Q4.) We also assume that the state space Ω is numbered $1, \ldots, N$.

Since P is symmetric, we can diagonalize it using an orthonormal basis composed of its eigenvectors $e_1, \ldots, e_N$:

$$\begin{aligned} P &= E\Lambda E^T \\ \Lambda &= \text{diag}(\lambda_i) \\ E^T &= \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}, \end{aligned}$$

where e_i is the eigenvector corresponding to λ_i. The e_i are normalized to have length 1, and are orthogonal to each other, so that $EE^T = E^T E = I$. Note also that $e_1 = \sqrt{\pi}$, and π is uniform since P is symmetric.

We now expand P^t in terms of the eigenvalues: $P^t = E\Lambda^t E^T$, and thus

$$P^t(x,y) = \sum_z E(x,z)\lambda_z^t E^T(z,y) = \sum_z E(x,z)\lambda_z^t E(y,z).$$

Separating the case $z = 1$, and using the facts that $\lambda_1 = 1$ and $E(x,1) = E(y,1) = \sqrt{\pi(x)} = \sqrt{\pi(y)}$, we get

$$P^t(x,y) = \pi(y) + \sum_{z\geq 2} E(x,z)E(y,z)\lambda_z^t.$$

Specializing to the case $x = y$ we get

$$P^t(x,x) = \pi(x) + \sum_{z\geq 2} E(x,z)^2\lambda_z^t.$$

The second term can be bounded by λ_2^t times the norm of the vector e_x, which is 1, so we finally get

$$P^t(x,x) \leq \pi(x) + \lambda_2^t. \tag{12.1}$$

To relate the values of $P^t(x,y)$ to the diagonal values $P^t(x,x)$ we observe that

$$\sum_y P^t(x,y)^2 = \sum_y P^t(x,y)\cdot P^t(y,x) = P^{2t}(x,x), \tag{12.2}$$

so we can write

$$\begin{aligned}
2\cdot\Delta_x(t) &= \sum_y |P^t(x,y) - \pi(y)| \\
&= \sum_y \frac{|P^t(x,y)-\pi(y)|}{\sqrt{\pi(y)}}\cdot\sqrt{\pi(y)} \\
\text{by Cauchy-Schwarz} \quad &\leq \sqrt{\sum_y \frac{(P^t(x,y)-\pi(y))^2}{\pi(y)}\cdot\sum_y \pi(y)} \\
&= \sqrt{\sum_y \frac{P^t(x,y)^2}{\pi(y)} - 2\sum_y P^t(x,y) + \sum_y \pi(y)^2} \\
\text{by Eqn (12.2)} \quad &= \sqrt{\frac{P^{2t}(x,x)}{\pi(x)} - 2 + 1} \\
\text{by Eqn (12.1)} \quad &\leq \sqrt{\frac{\lambda_2^{2t}}{\pi(x)}} = \frac{\lambda_2^t}{\sqrt{\pi(x)}}.
\end{aligned}$$

[Here we've made use of the Cauchy-Schwartz inequality, which states that $\sum_i a_i b_i = (\sum_i a_i^2 \sum_i b_i^2)^{1/2}$.] ■

12.3 The Second Lemma

We now want to bound the spectral gap in terms of a flow on the network. The proof of Lemma 12.3, due to [DS91,Si92], will use the variational characterization of eigenvalues, which will be briefly reviewed.

Lemma 12.3 *Let P be the transition matrix of an ergodic, reversible, and lazy Markov chain, and f any flow on the induced graph. Then*

$$1-\lambda_2 \geq \frac{1}{\rho(f)\cdot l(f)}.$$

Proof: Again assume for simplicity that P is symmetric (see Problem Set 3, Q4 for a generalization to reversible P). Let us define the Laplacian $L = I - P$ with eigenvaues $\mu_i = 1 - \lambda_i$,

$$0 = \mu_1 < \mu_2 \leq \mu_3 \ldots \leq \mu_N \leq 1$$

With the usual inner product of real vectors $<\phi,\psi>=\sum_x \phi(x)\psi(x)$, we can write the smallest eigenvalue μ_1 as

$$\mu_1 = \inf_{\phi\neq 0}\frac{<\phi, L\phi>}{<\phi,\phi>}$$

where ϕ is any non-zero real N-dimensional vector.

To see why this is true, expand ϕ as a linear combination of the orthonormal eigenvectors e_i, $\phi=\sum_i \alpha_i e_i$, and get

$$\inf_{\phi\neq 0}\frac{<\phi, L\phi>}{<\phi,\phi>} = \inf_{\phi\neq 0}\frac{\sum\alpha_i^2\mu_i}{\sum\alpha_i^2}.$$

This expression is clearly minimized when ϕ is parallel to the principal eigenvector e_1, i.e., $\alpha_1 > 0, \alpha_2 = \alpha_3 = \ldots = \alpha_N = 0$.

To get μ_2, we have to minimize the same quotient but over all ϕ that are orthogonal to e_1. Equivalently, we can take the minimum over all ϕ that have a component orthogonal to e_1, and subtract off the projection of ϕ on e_1:

$$\mu_2 = \inf_{\phi\perp e_1}\frac{<\phi, L\phi>}{<\phi,\phi>} = \inf_{\phi\not\parallel e_1}\frac{<\phi, L\phi>}{<\phi,\phi>-<\phi,e_1>^2}. \tag{12.3}$$

[In the asymmetric reversible case, we need to go through the same calculation, with a modified inner product weighted by π:

$$<\phi,\psi>_\pi=\sum_x \phi(x)\psi(x)\pi(x).$$

We then get

$$\mu_2 = \inf_{\phi \text{ not constant}}\frac{<\phi, L\phi>_\pi}{<\phi,\phi>_\pi-<\phi,\mathbf{1}>_\pi^2}$$

For the details, see Problem Set 3.]

We now rewrite the numerator and denominator in Eqn 12.3 in a more symmetrical form:

$$\begin{aligned}
<\phi, L\phi> &= \sum_x \phi(x)(L\phi)(x) = \sum_{x,y}\phi(x)L(x,y)\phi(y)\\
&= \sum_{x,y}\phi(x)[I-P](x,y)\phi(y)\\
&= \sum_x \phi(x)^2 - \sum_{x,y}\phi(x)P(x,y)\phi(y)\\
&= \sum_{x,y}\phi(x)^2P(x,y) - \sum_{x,y}\phi(x)P(x,y)\phi(y)\\
&= \frac{1}{2}\sum_{x,y}(\phi(x)-\phi(y))^2P(x,y).
\end{aligned}$$

Similarly, the denominator can be written as

$$\begin{aligned} < \phi, \phi > - < \phi, e_1 >^2 &= \sum_x \phi(x)^2 - \left(\sum_x \phi(x) \cdot \sqrt{\pi(x)} \right)^2 \\ &= \sum_{x,y} \phi(x)^2 \pi(y) - \sum_{x,y} \phi(x)\phi(y)\pi(x) \\ &= \frac{1}{2} \sum_{x,y} (\phi(x) - \phi(y))^2 \, \pi(x). \end{aligned}$$

Thus we have

$$1 - \lambda_2 \equiv \mu_2 = \inf_{\phi \text{ not constant}} \frac{\sum_{x,y} (\phi(x) - \phi(y))^2 P(x,y) \pi(x)}{\sum_{x,y} (\phi(x) - \phi(y))^2 \, \pi(x)\pi(y)},$$

where we have multiplied all terms on top and bottom by the uniform vector $\pi(x)$. Exactly the same expression holds for reversible but non-symmetric P (see Problem Set 3).

Recall that $P(x,y)\pi(x) = C(x,y)$ is the capacity of the edge between x and y, and $\pi(x)\pi(y) = D(x,y)$ is the demand. Thus

$$\mu_2 = \inf_{\phi \text{ not constant}} \frac{\sum_{x,y} (\phi(x) - \phi(y))^2 C(x,y)}{\sum_{x,y} (\phi(x) - \phi(y))^2 \, D(x,y)}$$

We can view the numerator as the "local variation" of ϕ and the denominator as the "global varation" of ϕ. A bound on the ratio of local and global variations is often called a *Poincaré inequality*. A Poincaré inequality in terms of flow will be our final ingredient.

Our task now is to show that the above quotient is bounded below by $\frac{1}{\rho(f)l(f)}$. First we rewrite the denominator using the definition of a flow f:

$$\sum_{x,y} (\phi(x) - \phi(y))^2 \, D(x,y) = \sum_{x,y} (\phi(x) - \phi(y))^2 \sum_{p \in \mathcal{P}_{x,y}} f(p).$$

Any path p between x and y can be broken into a sequence of edges, and we will write $\phi(x) - \phi(y)$ as a telescoping sum along the path:

$$\begin{aligned} \sum_{x,y} (\phi(x) - \phi(y))^2 \sum_{p \in \mathcal{P}_{x,y}} f(p) &= \sum_{x,y} \sum_{p \in \mathcal{P}_{x,y}} f(p) \cdot \left[\sum_{e \in p} \phi(e^+) - \phi(e^-) \right]^2 \\ \text{[by Cauchy-Schwartz]} \quad &\leq \sum_{x,y} \sum_{p \in \mathcal{P}_{x,y}} f(p) \cdot \sum_{e \in p} \left[\phi(e^+) - \phi(e^-) \right]^2 \cdot \sum_{e \in p} 1^2 \\ &= \sum_{x,y} \sum_{p \in \mathcal{P}_{x,y}} f(p) \cdot |p| \cdot \sum_{e \in p} \left[\phi(e^+) - \phi(e^-) \right]^2 \\ &= \sum_{e} \sum_{p \ni e} f(p) \cdot |p| \cdot \left[\phi(e^+) - \phi(e^-) \right]^2 \\ \text{[by definition of } \rho(f)] \quad &\leq l(f) \cdot \rho(f) \cdot \sum_{e} C(e) \left[\phi(e^+) - \phi(e^-) \right]^2 \\ &= l(f) \cdot \rho(f) \cdot \sum_{x,y} (\phi(x) - \phi(y))^2 \, C(x,y). \end{aligned}$$

The summation here is exactly the numerator in our quotient, so we're left with

$$1 - \lambda_2 = \mu_2 \geq \frac{1}{l(f)\rho(f)}.$$

■

12.4 Concluding Remarks

This result means that we can try to find a good (i.e., low cost) flow for a Markov Chain and use it to bound the mixing time of that chain. In the next few lectures we will consider some chains where it is hard to determine the mixing time using simpler arguments such as coupling. We will bound the mixing times for these chains by finding suitable flows.

We conclude this lecture by mentioning some related bounds. One can show the following alternative bound to Lemma 12.3:

Lemma 12.4 $1 - \lambda_2 \geq \frac{1}{8(\rho(f))^2}$

This result will be proved in a later lecture. It is clean in that it does not involve the quantity $l(f)$. On the other hand, it is usually weaker in applications since typically $\rho(f)$ is larger than $l(f)$.

One can also prove the following converse bounds (see [Si92]):

Theorem 12.5 *There exists a flow f with $\rho(f) \leq const \times \tau_{mix}$. Also, there exists a flow f with $1 - \lambda_2 \leq const \times \frac{ln\ N}{\rho(f)}$.*

Since we know from the previous lecture that Lemma 12.2 is tight up to a factor $O(\log \pi_{min}^{-1})$, we see that there is always a flow that is optimal up to this factor and a square. (Of course, this flow may be very hard to find.)

References

[DS91] P. Diaconis and D. Stroock, "Geometric bounds for eigenvalues of Markov chains," *Annals of Applied Probability* **1**, 1991, pp. 36–61.

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