Lecture 9 - Canonical Paths

Friday, September 3

In the previous class we showed how to compare Dirichlet forms. The most important corollary of this was shown by Diaconis and Stroock [1] and Sinclair [2].

Corollary 9.1 (Canonical Paths). *Given a reversible Markov chain $\mathcal{M}$, to every pair of states $x \neq y \in \Omega$ associate a path from x to y along edges ("canonical paths"). Then*

$$1-\lambda_2 \geq 1/\overline{\rho} \quad \textit{where} \quad \overline{\rho} = \max_{(a,b)\in E} \frac{1}{\pi(a)\mathsf{P}(a,b)} \sum_{\gamma_{xy}\ni(a,b)} \pi(x)\pi(y)\,|\gamma_{xy}|\,.$$

Proof. Consider the Markov chain $\mathcal{M}'$ with stationary distribution $\pi' = \pi$ and transitions $\mathsf{P}'(x,y) = \pi(y)$. It was shown last class that $1-\lambda_2' = 1$. To every edge (x,y) in $\mathcal{M}'$ associate the canonical path given in the problem. Then the comparison theorem applies and $\overline{A} = \overline{\rho}$. □

This is correct for the walk on the uniform two-point space, as $\overline{\rho} = 1/2(1-\gamma)$ which implies $1-\lambda_2 \geq 2(1-\gamma)$, the correct value.

As with the comparison case, the easiest way to find $\overline{\rho}$ is typically to use the bound

$$\overline{\rho} \leq \rho\,\ell \quad \text{where} \quad \rho = \max_{(a,b)\in E} \frac{1}{\pi(a)\mathsf{P}(a,b)} \sum_{\gamma_{xy}\ni(a,b)} \pi(x)\pi(y)$$

where ℓ is again the length of the longest path. In most applications the stationary distribution is uniform $\pi = 1/|\Omega|$ and the transition probabilities are constant with $\mathsf{P}(a,b) = 0$ *or* $\wp$. Then $\rho = \frac{1}{|\Omega|\,\wp} \max_{(a,b)\in E} |\{(x,y) \in \Omega\times\Omega : \gamma_{xy} \ni (a,b)\}|$ and it suffices to find which edge has the most paths through it.

Example 9.2 (Odd Cycle). Consider the simple random walk on a cycle C_n of odd length, as discussed before.

Given a pair of states $x,\, y \in [1 \dots n]$ a natural choice of path is to take the shortest route around the cycle. The longest path is of length $\frac{n-1}{2}$, the stationary distribution is uniform at $\pi = 1/n$, and the transitions are all 0 or 1/2, so it suffices to know the number of paths through a single edge. To count the number of paths through an edge e suppose the edge is $e = (a-1, a)$ and observe that the paths with an endpoint at $a+i$ must have begun somewhere from $a+i-\frac{n-1}{2}$ to $a-1$, i.e. there are $a-1-a-i+\frac{n-1}{2}+1 = \frac{n-1}{2}-i$ such paths, for a total of

$$\#\ \textit{paths} = \sum_{i=0}^{(n-1)/2} \frac{n-1}{2} - i = \frac{(n-1)(n+1)}{8}\,.$$

Then

$$\overline{\rho} \leq \frac{n-1}{2}\,\frac{1}{n\,(1/2)}\,\frac{(n-1)(n+1)}{8} = \frac{(n-1)^2(n+1)}{8n} < \frac{n^2}{8}$$

and so

$$1-\lambda_2 \geq \frac{1}{\overline{\rho}} \geq \frac{8}{n^2}\,.$$

Recall from Lecture 6 that $1-\lambda_2 = 1-\cos(2\pi/n) \approx \frac{2\pi^2}{n^2}$. The canonical path bound was of the correct order, and off by only a factor of $\frac{\pi^2}{4} \approx 2.5$.

It is not hard to compute $\overline{\rho}$ directly for this problem. It is $\overline{\rho} = \frac{1}{n\,(1/2)} \sum_{i=1}^{(n-1)/2} i^2 = \frac{(n-1)(n+1)}{12} < \frac{n^2}{12}$. The bound $1-\lambda_2 \geq \frac{12}{n^2}$ is off by only a factor of $\frac{\pi^2}{6} \approx 1.6$.

Example 9.3 (Boolean cube). Consider again the lazy walk on the boolean cube 2^d. This can be considered as a walk on $\{0,1\}^d$ with transitions given by choosing a coordinate uniformly at random and flipping it with probability 1/2.

Canonical paths can be defined as follows. If $\vec{x} = (x_1, x_2, \ldots, x_d)$ and $\vec{y} = (y_1, y_2, \ldots, y_d)$ then match the coordinates one at a time, that is follow the path

$$(x_1, x_2, \ldots, x_d) \to (y_1, x_2, \ldots, x_d) \to (y_1, y_2, \ldots, x_d) \to \cdots \to (y_1, y_2, \ldots, y_d)$$

Once again the transition probabilities are a constant $\wp = 1/2d$, the stationary distribution $\pi = 1/|\Omega| = 2^{-d}$, and the paths have max length d. To count paths, suppose that coordinate i is being changed. Then the paths through this edge may have started at anything in coordinates 1 to $i-1$ (for 2^{i-1} choices), and may be going to anything in coordinates $i+1$ to d (for 2^{d-i} choices). In total 2^{d-1} paths may pass through any particular edge.

Therefore

$$\overline{\rho} \leq \ell\,\rho \leq d\,\frac{1}{2^d\,(1/2d)}\,2^{d-1} = d^2$$

and so

$$1 - \lambda_2 \geq 1/d^2\,.$$

The actual value is $1 - \lambda_2 = 1/d$, so our bound is not too great.

Example 9.4 (Metropolis). As our final example consider the Metropolis method discussed earlier to generate from a distribution π. Recall that a different Markov chain $\mathcal{M}'$ makes transitions, and then these are accepted with some probability, so that $\mathsf{P}(x,y) = \mathsf{P}'(x,y)\,\min\{1, \frac{\pi(y)}{\pi(x)}\}$.

This is a natural problem for comparison, to find $1 - \lambda_2$ in terms of the spectral gap $1 - \lambda_2'$ of the base chain $\mathcal{M}'$. The easiest choice of paths is that if (a,b) is an edge in E' then simply take the path $\gamma_{ab} = (a,b)$ in E to be exactly the same edge. Then

$$\begin{aligned}\overline{A} = &\max_{e=(a,b)\in E} \frac{\pi'(a)\mathsf{P}'(a,b)}{\pi(a)\mathsf{P}(a,b)} &= \max_{e=(a,b)\in E} \frac{\pi'(a)\mathsf{P}'(a,b)}{\pi(a)\mathsf{P}'(a,b)\min\{1,\pi(b)/\pi(a)\}} \\ = &\max_{e=(a,b)\in E} \frac{\pi'(a)}{\min\{\pi(a),\pi(b)\}} &= \left(\min_{a\in\Omega} \frac{\pi(a)}{1/|\Omega|}\right)^{-1}\end{aligned}$$

just measures how much smaller the Metropolis distribution may be compared to the uniform distribution of $\mathcal{M}'$. It follows that $\mathcal{E}(f,f) \geq \mathcal{E}'(f,f)/\overline{A}$.

To bound the spectral gap, observe that

$$\begin{aligned}Var_\pi(f) &= \frac{1}{2}\sum_{x,y\in\Omega} (f(x)-f(y))^2\,\pi(x)\pi(y) \\ &\leq \left(\max_{a\in\Omega}\frac{\pi(a)}{1/|\Omega|}\right)^2 \frac{1}{2}\sum_{x,y\in\Omega}(f(x)-f(y))^2\,\frac{1}{|\Omega|}\,\frac{1}{|\Omega|} \\ &= \left(\max_{a\in\Omega}\frac{\pi(a)}{1/|\Omega|}\right)^2 Var_{\pi'}(f)\end{aligned}$$

It follows that

$$1 - \lambda_2 = \sup_{f\neq constant} \frac{\mathcal{E}(f,f)}{Var_\pi(f)} \geq \frac{\min_{a\in\Omega}\frac{\pi(a)}{1/|\Omega|}}{\left(\max_{a\in\Omega}\frac{\pi(a)}{1/|\Omega|}\right)^2}\,(1-\lambda_2')$$

If Metroplis changes the distribution by at most a factor of k then the Markov chain slows by at most a factor of k^3. Why does increasing $\pi(a)$ from uniform causes a much bigger penalty than decreasing $\pi(a)$?

Canonical paths is one of the most widely used methods for studying the mixing time of Markov chains. Numerous applications can be found in the literature. Week 7 of Eric Vigoda's notes covers Approximating the Permanent (one of the most important results in "rapid mixing"), while Sinclair's notes 13 to 17 cover various applications such as Monomer-Dimer systems.

References

[1] P. Diaconis and D. Stroock. Geometric bounds for eigenalues of markov chains. *The Annals of Applied Probability*, 1:36–61, 1991.

[2] A. Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370, 1992.