Lecture 10 - Conductance and the spectral gap

Wednesday, September 8

Many of the early results in studying mixing times were derived by geometric methods. These include the original version for approximating the permanent, estimating volume of a convex body, counting total orderings of a partially ordered set, etc.

Recall that a Markov chain can be represented by a directed weighted graph, with vertices given by elements of the state space and edge weights $w(x,y) = \mathsf{P}(x,y)$. It is generally more appropriate to use a different edge weighting, with $q(x,y) = \pi(x)\mathsf{P}(x,y)$. A reversible Markov chain is then represented by an undirected graph, as $q(x,y) = \pi(x)\mathsf{P}(x,y) = \pi(y)\mathsf{P}(y,x) = q(y,x)$.

The Markov chain can be considered as a random walk on this weighted graph, and the edge weights represent the *ergodic flow* $q(x,y)$ which travels along the edge at each step if the chain is in the stationary distribution. Intuition tells us that if there are a lot of edges and no bottlenecks then the flow should spread out quickly and the Markov chain mix rapidly.

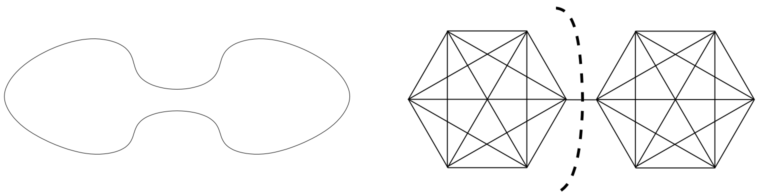

Figure 1: The bottleneck in the center is the most important thing for mixing here.

Definition 10.1. If $A \subset \Omega$ then let the *ergodic flow* from a set A to a set B be given by

$$\mathsf{Q}(A,B) = \sum_{x\in A,\, y\in A^c} q(x,y) = \sum_{x\in A,\, y\in A^c} \pi(x)\mathsf{P}(x,y)$$

Then $\mathsf{Q}(A,A^c)$ measures the weight of the edges from a set A to its complement.

Also, the *conductance* is given by

$$\Phi = \max_{A\subset\Omega} \frac{\mathsf{Q}(A,A^c)}{\min\{\pi(A),\pi(A^c)\}}$$

This is also known as the *cutset-expansion*, *Cheeger constant* or *edge-isoperimetric number.*

Cheeger's inequality is a widely used tool for bounding the smallest non-trivial eigenvalue for the Laplacian of a range of structures, including manifolds, graphs, and Markov chains. Jerrum and Sinclair [1] and Lawler and Sokal [2] independently proved a Markov chain version of this.

Theorem 10.2 (Cheeger's Inequality). *If $\mathcal{M}$ is a reversible Markov chain then*

$$2\Phi \ge 1-\lambda_2 \ge 1-\sqrt{1-\Phi^2} \ge \Phi^2/2\,.$$

Proof. Recall the variational form of the spectral gap $\lambda = 1-\lambda_2$. Let $f = 1_A$ be the indicator function for a set A. Then

$$\begin{aligned}
\mathcal{E}(1_A,1_A) &= \frac{1}{2}\sum_{x,y\in\Omega} (1_A(x)-1_A(y))^2\,\pi(x)\mathsf{P}(x,y) = \frac{1}{2}\left(\mathsf{Q}(A,A^c)+\mathsf{Q}(A^c,A)\right) = \mathsf{Q}(A,A^c) \\
Var_\pi(1_A) &= \frac{1}{2}\sum_{x,y\in\Omega} (1_A(x)-1_A(y))^2\,\pi(x)\pi(y) = \pi(A)\pi(A^c)
\end{aligned}$$

Then

$$1 - \lambda_2 = \inf_{f \neq constant} \frac{\mathcal{E}(f, f)}{Var_\pi(f)} \leq \inf_{A \subset \Omega} \frac{\mathcal{E}(1_A, 1_A)}{Var_\pi(1_A)} \leq 2\,\Phi$$

A variation on the lower bound will be proven in the next class (the seminar). □

Corollary 10.3. *If $\mathcal{M}$ is a lazy reversible Markov chain then*

$$\frac{1 - 2\Phi}{2\Phi} \log(1/2\epsilon) \leq \tau(\epsilon) \leq \left\lceil \frac{1}{\Phi^2} \left(\frac{1}{2} \log \frac{1 - \pi_*}{\pi_*} + \log(1/2\epsilon) \right) \right\rceil$$

Proof. Recall from Theorem 7.5 that

$$\frac{\lambda_{max}}{1 - \lambda_{max}} \log(1/2\epsilon) \leq \tau(\epsilon) \leq \left\lceil \frac{1}{1 - \lambda_{max}} \left(\frac{1}{2} \log(1/\pi_*) + \log(1/2\epsilon) \right) \right\rceil$$

On a lazy Markov chain $\lambda_{max} = \lambda_2$, which when combined with Cheeger's inequality implies $2\Phi \geq 1 - \lambda_{max} \geq \Phi^2/2$. This gives the theorem, but with the upper bound off by a factor of two.

To improve the upper bound we reverse the argument used in Lecture 8 where it was argued that laziness implies $\lambda_{max} = \lambda_2$. The matrix $\mathsf{P}' = 2\mathsf{P} - I$ is the transition matrix for a Markov chain twice as fast as $\mathcal{M}$. This still has stationary distribution π but the new spectral gap is $1 - \lambda_2' = 2(1 - \lambda_2)$ and new conductance is $\Phi' = 2\Phi$ (the original eigenvectors $\overrightarrow{u_i}$ work with eigenvalues $\lambda_i' = -1 + 2\lambda_i$). Cheeger's inequality implies that $1 - \lambda_2' \geq \Phi'^2/2 = 2\Phi^2$. Therefore, $1 - \lambda_2 = \frac{1 - \lambda_2'}{2} \geq \Phi^2$. □

Remark 10.4. *Mihail [4] generalized this argument and showed that the same upper bound holds even if the Markov chain is non-reversible. We will eventually show that a related bound also holds for non-lazy chains.*

Remark 10.5. *Conductance entirely characterizes* **rapid mixing** *(mixing in time polynomial in the problem size). The theorem above shows that the mixing time is polynomial in the problem size if and only if the conductance is inverse polynomial. In fact, almost all proofs of* **slow mixing** *work by showing that the conductance is exponentially small in the problem size.*

Originally canonical path theorems were used to bound the conductance, only later was it realized that a direct argument exists to bound λ_2. The original argument works like this:

Example 10.6. Consider again the lazy walk on the boolean cube 2^d. Recall that 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. Recall that canonical paths were constructed in Lecture 9, and it was found that $\rho = d$.

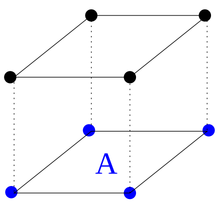

Figure 2: The lower half plane is the worst cutset for the walk on the boolean cube.

Now, let us bound conductance. If a path connects vertices $x, y \in \Omega$ then route a flow of $\pi(x)\pi(y)$ along the path, for a sum total of 1. Recall the definition

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

Then ρ is the highest amount of flow sent through an edge (a, b) relative to its capacity $q(a, b) = \pi(a)\mathsf{P}(a, b)$.

Consider a set $A \subset \Omega$. Restrict attention to paths from $x \in A$ to $y \in A^c$, so the total flow is $\pi(A)\pi(A^c)$. Certainly ρ is still an upper bound on the flow sent through any edge $(a, b) \in \Omega \times \Omega$ relative to its capacity $q(a, b)$. However, every path from A to A^c must pass through some edge $(a, b) \in A \times A^c$, so the total flow routed through edges from A to A^c is at least $\pi(A)\pi(A^c)$. It follows that

$$\mathsf{Q}(A, A^c) \geq \frac{\pi(A)\pi(A^c)}{\rho}$$

Therefore

$$\Phi \geq \inf_{A \subset \Omega} \frac{\pi(A)\pi(A^c)}{\rho \min\{\pi(A), \pi(A^c)\}} \geq \frac{1}{2\rho}$$

and consequently for the lazy walk on the cube

$$1 - \lambda_2 \geq \Phi^2 \geq \frac{1}{4\rho^2} \geq \frac{1}{4d^2}$$

This is essentially the same (poor) bound that we showed before by use of the canonical paths theorem.

We can summarize these results as follows.

Corollary 10.7. *If $\mathcal{M}$ is a lazy reversible Markov chain then*

$$1 - \lambda_{max} = 1 - \lambda_2 \geq \Phi^2 \geq 1/4\rho^2$$

Remark 10.8. *A generalization of the canonical path method is the method of* **multicommodity flows**. *This is basically the same but the flow $\pi(x)\pi(y)$ may be divided between multiple paths, which can sometimes be helpful. It was shown by Leighton and Rao [3] that there is always some choice of multicommodity flows such that*

$$\frac{1}{4\rho^2} \leq 1 - \lambda_2 \leq c \, \frac{\log(1/\pi_*)}{\rho^2}$$

for some universal constant c. The paths we gave on the boolean cube match this, as $\pi_* = 2^{-d}$ and our bound was within factor of $\log(1/\pi_*) = d \log 2$ from the correct bound.

References

[1] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of the permanent resolved. *Proc. 20nd Annual ACM Symposium on Theory of Computing*, pages 235–243, 1988.

[2] G. Lawler and A. Sokal. Bounds on the l^2 spectrum for markov chains and markov processes: a generalization of cheeger's inequality. *Trans. Amer. Math. Soc.*, 309:557–580, 1988.

[3] T. Leighton and S. Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. *Proc. 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 422–431, 1988.

[4] M. Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. *30th Annual Symposium on Foundations of Computer Science*, pages 526–531, 1989.