Barcelona Graduate School in Mathematics, Fall 2013

Random Structures and the Probabilistic Method

Session 19: Conductance, flows and spectral gap

Lecturer: Elitza Maneva

The speed of convergence of a Markov chain is closely related to how well connected the state-space is. This leads us to the study of the "flow of probability" and its dual - the presence or absence of "bottlenecks" in the state-space.

First we'll talk about the bottlenecks and how we can use them to give lower bounds to the mixing time. Then we will look at how we can define flows on the state-space and use them to give upper bounds on the mixing time.

1. Conductance

The *capacity* of an edge is $C(x, y) := \pi(x)P(x, y)$, and that between sets is $C(S) := \sum_{x\in S, y\notin S} C(x, y)$.

The conductance of a set S is defined to be:

$$\Phi(S) := \frac{C(S)}{\pi(S)},$$

and the conductance of the chain is $\Phi = \min_{S:\pi(S)\le\frac{1}{2}} \Phi(S)$.

There is a related symmetric quantity $\Phi'(S) = \frac{C(S)}{\pi(S)\pi(\bar{S})}$, known as the *sparsity* of the cut $(S, \bar{S})$, and it is usually used in relation to flows. For S with $\pi(S) \le 1/2$, $\Phi(S) \le \Phi'(S) \le 2\Phi(S)$.

Theorem 1. $t_{\text{mix}} \ge \frac{1}{4\Phi}$.

The key observation is that $\Phi(S)$ can be thought of as the change in the distribution in one step of the chain, if initially it is concentrated on S. Let $\mu_0(x) = \frac{\pi(x)}{\pi(s)}$, for $x \in S$, and 0 otherwise. Let $\mu_1 = \mu_0 P$. Then $||\mu_1 - \mu_0||_{TV} = \Phi(S)$.

As an example of an application, on the binary tree this theorem gives us a lower bound of $(n - 2)/4$, which is tight up to a constant factor. On the other hand, on the cycle it gives us $n/8$, which is far from n^2.

There is also an upper bound for the mixing time in terms of conductance but it's harder to use because it means we have to bound $\Phi(S)$ for all S, and not just find one as in the lower bound. It is

$$t_{\text{mix}} \le c \times \frac{1}{\Phi^2} \log(\pi_{\min}^{-1}),$$

where $\pi_{\min} = \min_{x\in\Omega} \pi(x)$.

2. Flows of probability

We define the *demand* between two vertices to be $D(x, y) := \pi(x)\pi(y)$. We can imagine that different commodities have to be routed from x to y along the paths from x to y, and this is done for all pairs of vertices (x, y) at the same time. A *flow* f is an assignment of an amount to each path $f : P \to \mathbb{R}^+ \cup \{0\}$, where $P = \cup_{xy} P_{xy}$ and P_{xy} denotes the set of all simple paths from x to y, so that

$$\sum_{p\in P_{xy}} f(p) = D(x, y).$$

The total flow on an edge e is denoted by $f(e) = \sum_{p \ni e} f(p)$. The *cost* of a flow is defined to be $\rho(f) := \max_e f(e)/C(e)$. The *length* of a flow is defined to be $\ell(f) := \max_{p: f(p)>0} |p|$.

Theorem 2. *For any lazy (self-loop probability at least $1/2$), irreducible Markov chain P and any flow f,*

$$\tau_x(\epsilon) \leq \rho(f)\ell(f) \times (\log \pi(x)^{-1} + 2 \log \epsilon^{-1}).$$

In particular for the mixing time the bound is $t_{\text{mix}} \leq \rho(f)\ell(f) \times (\log(\pi_{\min}^{-1}) + 4 \log 2)$.

We will consider the rate of decay of the *discrepancy* from π with respect to an arbitrary event $A \subseteq Q$, which we define as $\varphi_t(x) := P_x^t(A) - \pi(A)$, where P_x^0 is the indicator function for x being in A. Notice that $\varphi_{t+1}(x) = \sum_y P(x, y)\varphi_t(y)$.

We use the following squared norm with respect to π which is also known as the *global variance*:

$$||\varphi||_\pi^2 := \sum_x \pi(x)\varphi(x)^2.$$

It can be interpreted as variance if the expectation is defined as $\mathbf{E}_\pi \varphi := \sum_x \pi(x)\varphi(x)$ and noticing that in our case $\mathbf{E}_\pi \varphi_i = 0$ for all i.

The adjective *global* is justified by the following identity which holds whenever $\mathbf{E}_\pi \varphi = 0$:

$$||\varphi||_\pi^2 = \frac{1}{2} \sum_{x,y} \pi(x)\pi(y) \times (\varphi(x) - \varphi(y))^2.$$

In contrast, the *local variance* is obtained by considering only the edges of the chain. It is also known as the Dirichlet form and is denoted by

$$\mathcal{E}_\pi(\varphi) := \frac{1}{2} \sum_{x,y} \pi(x)P(x, y) \times (\varphi(x) - \varphi(y))^2.$$

We will prove the following two lemmas which together give us decay of the norm of φ_i of the form $||\varphi_{i+1}||_\pi^2 \leq (1 - \alpha)||\varphi_i||_\pi^2$ and imply the theorem.

Lemma 3. *If P is a lazy chain, i.e. $P = \frac{1}{2}(I + \hat{P})$ for a stochastic matrix $\hat{P}$, then*

$$||\varphi_i||_\pi^2 - ||\varphi_{i+1}||_\pi^2 \ \geq \ \mathcal{E}_\pi(\varphi_i).$$

Lemma 4. *If $\mathbf{E}_\pi \varphi = 0$ then*

$$||\varphi||_\pi^2 \ \leq \ \rho(f)\ell(f)\mathcal{E}_\pi(\varphi).$$

The first lemma is proved by splitting one step of the chain into two half-steps. In the first half-step on every edge we record an average over its endpoints, and in the second half-step the averages on the edges incident on a vertex are added.

For the second lemma we split the global variance expression according to the flow f and for each path $u_0, u_1, \ldots u_k$ we bound $(\varphi(u_0) - \varphi(u_k))^2$ by $k \times \sum_{i=0}^{k-1} (\varphi(u_i) - \varphi(u_{i+1}))^2$ (using Cauchy-Schwarz) thus converting the global sum into local (edge) contributions.

3. Exercise

(1) Let P be a reversible, irreducible and aperiodic Markov chain. Show that its eigenvalues satisfy $1 = \lambda_1 > \lambda_2 > \cdots > \lambda_n > -1$. The *spectral gap* of P is defined to be $1 - \lambda_2$. Prove that

$$1 - \lambda_2 = \inf_{\varphi \text{ non-constant}} \frac{\mathcal{E}_\pi(\varphi)}{||\varphi||_\pi^2}.$$

4. References

(1) Scribes of "Markov chain Monte Carlo: Foundations and Applications" taught by Alistair Sinclair, UC Berkeley, Fall 2009, http://www.cs.berkeley.edu/ sinclair/cs294/f09.html (Lectures 10 and 18.2)

(2) Milena Mihail, "Conductance and convergence of Markov chains - a combinatorial treatment of expanders", FOCS, 1989.

(3) Alistair Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow", Combinatorics, Probability and Computing 1, 1992, pp. 351-370.

(4) Persi Diaconis and Daniel Stroock, "Geometric bounds for eigenvalues of Markov chains", Annals of Applied Probability 1, 1991, pp. 36-61.