Lecture 19 - Consequences of New Cheeger, Average Congestion

Wednesday, September 29

19.1 Consequences of New Cheeger bound

Let us look at a few more small consequences of past results.

Recall from Lecture 10 that Jerrum and Sinclair showed how to bound mixing times of reversible Markov chains in terms of conductance, and that Mihail [1] later improved this and generalized to the non-reversible case. We have not used reversibility and so our result applies to non-reversible chains as well.

Corollary 19.1. *If $\mathcal{M}$ is a lazy (non-reversible) Markov chain then*

$$\tau(\epsilon) \leq L^2(2\epsilon) \leq \min\left\{\frac{1}{\Phi^2}, \frac{2}{\tilde{\Phi}^2}\right\} \log \frac{1}{2\epsilon\sqrt{\pi_*}}$$

and if $\mathcal{M}$ is reversible then moreover $\lambda \geq \max\{\Phi^2, \tilde{\Phi}^2/2\}$.

One application of the corollary is to the canonical paths method discussed in Lecture 10. Recall that $\Phi \geq 1/(2\rho)$; if one checks the proof it is clear that it in fact shows $\tilde{\Phi} \geq 1/\rho$. This gives a factor of 2 improvement Corollary 10.7, and a factor 4 improvement over the statement $\lambda \geq 1/8\rho^2$ found in most papers.

Corollary 19.2 (Canonical Paths). *If $\mathcal{M}$ is a lazy, reversible, ergodic Markov chain then*

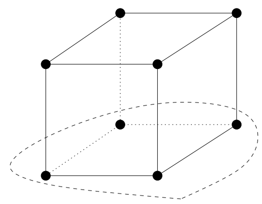
$$\lambda \geq 1 - \sqrt{1 - \rho^{-2}} \geq \frac{1}{2\rho^2}\,.$$

Further generalizations on canonical path theorems can be found in [3] and might be discussed later, time permitting.

19.2 Average Congestion

The main point of today's class is to study how the behavior of $\mathcal{C}(A)$ with set sizes effects the mixing time. So far our study of evolving sets has focused on a global congestion measure $\mathcal{C}_f$. However, if a Markov chain starts at a single point then it makes more sense to first consider the behavior of small sets, and then gradually build up to big sets.

A good example of this is the lazy random walk on the boolean cube $\{0, 1\}^d$.



At the set A given by a single vertex the Markov chain steps out half the time. However, when it is the lower half space then it steps out only $1/2d$ of the time. Therefore we might expect that the Markov chain quickly spreads out over the space (due to small sets being good), and only then do the slow asymptotics start to kick in.

The following theorem shows just that. This can lead to a major improvement because small sets often have significantly higher expansion than large sets (the cube example is worked out in the next class). Our method is almost identical to that of Morris and Peres [4], however I have extended it to the various distances that we now know how to study [2].

Corollary 19.3. *If $\mathcal{M}$ is a Markov chain with stationary distribution π then*

$$\begin{aligned}
\tau(\epsilon) &\leq \int_{\pi_*}^{1/2} \frac{dx}{(1-x)\,\psi_{TV}(x)} + \frac{\log(1/2\epsilon)}{\psi_{TV}(1/2)}, && \text{if } z\psi_{TV}(1-z) \text{ is convex} \\
\mathsf{D}(\epsilon) &\leq \int_{\pi_*}^{e^{-\epsilon}} \frac{dx}{x\log(1/x)\,\psi_D(x)}, && \text{if } z\psi_D(e^{-z}) \text{ is convex} \\
L^2(\epsilon) &\leq \int_{\pi_*}^{1/2} \frac{dx}{2x(1-x)\psi_{L^2}(x)} + \frac{\log(1/\epsilon)}{\psi_{L^2}(1/2)}, && \text{if } z\psi_{L^2}\left(\tfrac{1}{1+z^2}\right) \text{ is convex}
\end{aligned}$$

where $\psi_{TV}(x) \leq \min_{0<\pi(A)\leq x} 1 - \mathcal{C}_{TV}(A)$, $\psi_D(x) \leq \min_{0<\pi(A)\leq x} 1 - \mathcal{C}_D(A)$ *and* $\psi_{L^2}(x) \leq \min_{0<\pi(A)\leq x} 1 - \mathcal{C}_{L^2}(A)$.

If the convexity condition is not met then

$$\begin{aligned}
\tau(\epsilon) &\leq \frac{\log\frac{1-\pi_*}{\epsilon}}{1-\mathcal{C}_{TV}} \\
\mathsf{D}(\epsilon) &\leq \int_{\sqrt{\pi_*}}^{e^{-\epsilon/2}} \frac{2\,dx}{x\log(1/x)\,\psi_D(x)} \\
L^2(\epsilon) &\leq \int_{\pi_*}^{1/2} \frac{dx}{2x(1-x)\psi_{L^2}(x)} + \frac{\log(2/\epsilon)}{\psi_{L^2}(1/2)}
\end{aligned}$$

Proof. Our argument is taken from the proof Morris and Peres [4] for studying L^2 mixing. We have made only slight changes to fit our slightly different mixing time bounds.

Recall that

$$\|\mathsf{P}^t(x,\cdot) - \pi\|_{2,\pi} \leq \frac{1}{\pi(x)}\, \mathsf{E}_{\{x\}} \sqrt{\pi(S_t)(1-\pi(S_t))} \tag{1}$$

We now try to see how quickly this ratio drops.

Consider the random walk on sets given by

$$\hat{\mathsf{K}}(S,A) = \frac{\pi(A)}{\pi(S)}\, \mathsf{K}(S,A)$$

where $\mathsf{K}(S,A)$ is the transition matrix (kernel) for the evolving sets walk. Then $\hat{\mathsf{K}}(S,A)$ is a transition matrix because $\sum_{A\subset\Omega} \hat{\mathsf{K}}(S,A) = 1$. Also, by induction

$$\hat{\mathsf{K}}^t(S,A) = \frac{\pi(A)}{\pi(S)}\, \mathsf{K}^t(S,A)\,.$$

Therefore,

$$\hat{\mathsf{E}}_S f(S_t) = \mathsf{E}_S\left[\frac{\pi(S_t)}{\pi(S)}\, f(S_t)\right]$$

where $\hat{\mathsf{E}}$ is the expectation for the transition matrix $\hat{\mathsf{K}}^t(S,\cdot)$. In the setting of equation (1) this means that

$$\frac{1}{\pi(x)}\,\mathsf{E}_{\{x\}}\sqrt{\pi(S_t)(1-\pi(S_t))} = \hat{\mathsf{E}}_{\{x\}}\sqrt{\frac{1-\pi(S_t)}{\pi(S_t)}}$$

Therefore, to bound the mixing time we need only find how small $\hat{\mathsf{E}}_{\{x\}}Z_t$ is, for $Z_t = \sqrt{\frac{1-\pi(S_t)}{\pi(S_t)}}$.

First, notice that we know the rate at which Z_t drops, as

$$\hat{\mathsf{E}}_{\{x\}}\left(\frac{Z_{t+1}}{Z_t}\,\middle|\,S_n\right) = \mathsf{E}_{\{x\}}\left(\frac{\sqrt{\pi(S_{t+1})(1-\pi(S_{t+1}))}}{\sqrt{\pi(S_t)(1-\pi(S_t))}}\,\middle|\,S_t\right) = \mathcal{C}_{L^2}(S_t) \le 1-\psi_{L^2}(\pi(S_t))\,.$$

This can be rewritten in terms of Z_t because $\pi(S_t) = 1/(1+Z_t^2)$, and so

$$\hat{\mathsf{E}}_{\{x\}}\left(\frac{Z_{t+1}}{Z_t}\,\middle|\,S_n\right) \le 1-\psi_{L^2}(1/(1+Z_t^2))$$

By Lemma 19.4 below it follows that if

$$n \ge \int_\delta^{Z_0} \frac{2\,dz}{z\,\psi_{L^2}(1/(1+(z/2)^2)}$$

then $\|\mathsf{P}^t(x,\cdot)-\pi\|_{L^2} \le \delta$. Make a change of variables, with $x = 1/(1+(z/2)^2)$ and $Z_0 = \sqrt{(1-\pi_*)/\pi_*}$, so $x(Z_0) \ge \pi_*$, while $\frac{dz}{z} = -\frac{dx}{2x(1-x)}$. Substituting this all in implies that

$$n \ge \int_{\pi_*}^{1/(1+(\delta/2)^2)} \frac{dx}{2x(1-x)\,\psi_{L^2}(x)} = \int_{\pi_*}^{1/2} \frac{dx}{2x(1-x)\,\psi_{L^2}(x)} + \frac{\log(2/\delta)}{\psi_{L^2}(1/2)}$$

The bounds for the convex case, as well as the other distances all follow similarly. □

The following lemma is still needed. The lemma and proof are taken verbatim from Morris and Peres [4].

Lemma 19.4. *Let f, $f_0 : [0,\infty) \to [0,1]$ be increasing functions. Suppose that $Z_n \ge 0$ for $n = 0,1,\dots$ are random variables with $Z_0 = L_0$. Denote $L_n = \mathsf{E}(Z_n)$.*

1. *If $L_n - L_{n+1} \ge L_n\,f(L_n)$ for all n, then for every $n \ge \int_\delta^{L_0}\frac{dz}{z\,f(z)}$, we have $L_n \le \delta$.*
2. *If $\mathsf{E}(Z_{n+1}|Z_n) \le Z_n(1-f(Z_n))$ for all n and the function $u \to u\,f(u)$ is convex on $(0,\infty)$ then the conclusion of (i) holds.*
3. *If $\mathsf{E}(Z_{n+1}|Z_n) \le Z_n(1-f_0(Z_n))$ for all n and $f(z) = f_0(z/2)/2$, then the conclusion of (i) holds.*

Proof. **(i)** It suffices to show that for every n we have

$$n \le \int_{L_n}^{L_0}\frac{dz}{z\,f(z)}\,.$$

We verify this by induction. Clearly this holds for $n=0$. Now, fix $n \ge 0$ and suppose that this holds. Then

$$L_{n+1} \le L_n\,[1-f(L_n)] \le L_n\,e^{-f(L_n)}$$

whence

$$\int_{L_{n+1}}^{L_n} \frac{dz}{z\,f(z)} \geq \frac{1}{f(L_n)} \int_{L_{n+1}}^{L_n} \frac{dz}{z} = \frac{1}{f(L_n)} \log \frac{L_n}{L_{n+1}} \geq 1\,.$$

Thus

$$\int_{L_{n+1}}^{L_0} \frac{dz}{z\,f(z)} \geq \int_{L_n}^{L_0} \frac{dz}{z\,f(z)} + 1 \geq n+1\,.$$

(ii) This is immediate from Jensen's inequality.

(iii) Fix $n \geq 0$. We have

$$\mathsf{E}(Z_n - Z_{n+1}) \geq \mathsf{E}[2Z_n f(2Z_n)] \geq \mathsf{E}[2Z_n \mathbf{1}_A] f(L_n)$$

where A is the even $\{2Z_n \geq L_n\}$. Clearly $\mathsf{E}[2Z_n \mathbf{1}_{A^c}] \leq L_n$, whence $\mathsf{E}[2Z_n \mathbf{1}_A] \geq L_n$. This, in conjunction with the previous equation yields the hypothesis of (i). □

References

[1] 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.

[2] R. Montenegro. Evolving sets and bounds on various mixing quantities. *Preprint*, 2004.

[3] R. Montenegro. Mixed path and conductance profile bounds for various measures of mixing time. *Preprint*, 2004.

[4] B. Morris and Y. Peres. Evolving sets and mixing. *Proc. 35th Annual ACM Symposium on Theory of Computing*, pages 279–286, 2003.