Lecture 21 - Vertex congestion and comparison theorems

Monday, October 4

Today we apply Lemma 20.3 to prove mixing time theorems involving not just the edge edge expansion, but also the vertex expansion.

21.1 Edge and Vertex expansion effects on mixing times

First, let us give a few applications of Lemma 20.3 which give a good intuition into what governs mixing.

Corollary 21.1 (Edge expansion / flow). *If $\mathcal{M}$ and $\mathcal{M}'$ are finite, ergodic, lazy Markov chains with the same stationary distribution π, if $f : [0,1] \to \mathbb{R}^+$ is a concave function, and $\mathcal{M}'$ has smaller pointwise flow, that is*

$$\forall A \subset \Omega,\, \forall v \in A^c :\, \mathsf{Q}(A,v) \geq \mathsf{Q}'(A,v)$$

then

$$\forall A \subset \Omega :\, 1 - \mathcal{C}_f(A) \geq 1 - \mathcal{C}'_f(A)\,.$$

This says that if the edge expansion is smaller then the mixing time is worse.

In order to consider vertex expansion we need to define exactly what this term means. One reasonable definition is to say that the flow is well distributed if cutting it off at some threshold does not cut off too much, that is if the threshold is u then $\sum_{v \in A^c} \min\{u\pi(v),\, \mathsf{Q}(A,v)\}$ is about the same size as $\mathsf{Q}(A, A^c)$, and likewise with a sum over $v \in A$.

Corollary 21.2 (Vertex expansion). *If $\mathcal{M}$ and $\mathcal{M}'$ are finite, ergodic, lazy Markov chains with the same stationary distribution π, if $f : [0,1] \to \mathbb{R}^+$ is a concave function, and $\mathcal{M}'$ has more well distributed flow, that is*

$$\forall A \subset \Omega,\, \forall u \in [0, 1/2] : \sum_{v \in A^c} \min\{u\pi(v),\, \mathsf{Q}'(A,v)\} \geq \sum_{v \in A^c} \min\{u\pi(v),\, \mathsf{Q}(A,v)\}$$

then

$$\forall A \subset \Omega :\, 1 - \mathcal{C}'_f(A) \geq 1 - \mathcal{C}_f(A)$$

Likewise, this says that if the flow is more evenly distributed among vertices (vertex expansion is higher) then the mixing time is better.

Proofs of Corollary 21.1 and 21.2. Figure 1 gives a visual "proof." This is easily made rigorous and so the details are left to the interested reader. $\square$

The corollaries seem to only require require flow into A^c to be good. However, by considering the subset $A^c \subset \Omega$ then the flow into $(A^c)^c = A$ must also be good. Alternatively, it suffices to consider only sets with $\pi(A) \leq 1/2$ if flow into A and A^c are both considered.

21.2 Canonical paths

Recall the setting of canonical paths results is as follows. Consider two Markov chains $\tilde{\mathcal{M}}$ and $\mathcal{M}$, with the same state space (set of vertices V) but different transition probabilities (edges $\tilde{E}$ and E). For each (directed) edge $\tilde{e} = (x,y) \in \tilde{E}$, define a path γ_{xy} from x to y along (directed) edges E of $\mathcal{M}$ and transport $\tilde{pi}(x)\tilde{\mathsf{P}}(x,y)$ units from x to y along this path. The paths should be well distributed so that not too many pass through the same vertices or edges, or it will become a transportation bottleneck.

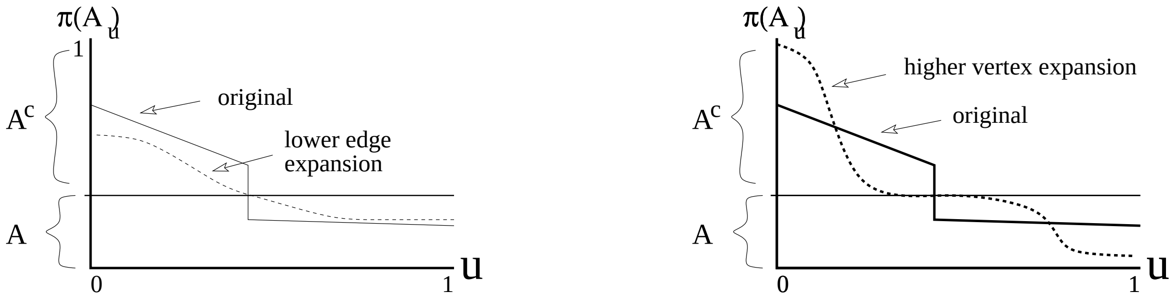

Figure 1: The $\pi(A_u)$ for $\mathcal{M}$ vs. lower edge expansion $\mathcal{M}'$, and vs. higher vertex expansion $\mathcal{M}'$.

Theorem 21.3 (Comparison with Conductance Profile (see [3])). *Suppose that $\tilde{\mathcal{M}}$ and $\mathcal{M}$ are lazy Markov chains with the same set of vertices V, and to every edge $\tilde{e} = (x,y) \in \tilde{E}$ associate a path $\gamma_{xy} \subset E$. Let*

$$\rho_e = \max_{e\in E} \frac{1}{\mathsf{Q}(e)} \sum_{\gamma_{xy}\ni e} \tilde{\pi}(x)\,\tilde{\mathsf{P}}(x,y) \qquad \textit{and} \qquad \rho_v = \max_{v\in V} \frac{1}{\pi(v)} \sum_{\gamma_{xy}\ni v} \tilde{\pi}(x)\,\tilde{\mathsf{P}}(x,y)\,.$$

If $A \subset V$ a proper subset and $\pi = \tilde{\pi}$ then

$$\begin{aligned}
1 - \mathcal{C}_{L^2}(A) &\geq 2\,\frac{\rho_v}{\rho_e}\left(1 - \sqrt{1 - \left(\frac{\tilde{\Phi}_{\tilde{\mathcal{M}}}(A)}{2\rho_v}\right)^2}\,\right) \geq \frac{\tilde{\Phi}_{\tilde{\mathcal{M}}}(A)^2}{4\,\rho_v\rho_e} \\
1 - \mathcal{C}_D(A) &\geq \frac{\Phi_{\tilde{\mathcal{M}}}(A)^2}{\rho_e\,\rho_v\,\log(1/\pi(A))} \\
1 - \mathcal{C}_{TV}(A) &\geq \frac{2\,\tilde{\Phi}_{\tilde{\mathcal{M}}}(A)^2\,\pi(A)\pi(A^c)}{\rho_v\,\rho_e}
\end{aligned}$$

If $\pi \neq \tilde{\pi}$ then replace $\tilde{\Phi}_{\tilde{\mathcal{M}}}(A)$ by $\frac{\mathsf{Q}_{\tilde{\mathcal{M}}}(A,A^c)}{\pi(A)\pi(A^c)}$ and likewise $\Phi_{\tilde{\mathcal{M}}}(A)$ by $\frac{\mathsf{Q}_{\tilde{\mathcal{M}}}(A,A^c)}{\pi(A)}$.

Proof. To do this we need only use Lemma 20.3 to construct the worst case from the paths, and then do some simplification. By Corollary 21.1 being too pessimistic about the ergodic flow $\mathsf{Q}(A,A^c)$ will help lower bound $1 - \mathcal{C}_f(A)$, so we may pessimistically assume that the canonical paths include all of the ergodic flow, and so $\mathsf{Q}(A,A^c) = \mathsf{Q}_{\tilde{\mathcal{M}}}(A,A^c)/\rho_e$. Likewise, ρ_v says that this flow is not too concentrated at any points, in particular the worst case of this flow at a single point is $\forall v \in A:\ \frac{\mathsf{Q}(A^c,v)}{\pi(v)} \leq \frac{\tilde{\mathsf{Q}}(A,A^c)/\rho_e}{\tilde{\mathsf{Q}}(A,A^c)/\rho_v} = \frac{\rho_v}{\rho_e}$, and likewise for $v \in A^c$. It follows that if $M := \rho_v/\rho_e$ then $\forall u \in [M, 1-M]:\ \pi(A_u) = \pi(A)$. Subject to this constraint, in Lemma 20.3 the integral $\int_0^t \pi(A_u)\,du$ is minimized by

$$\pi(A_u) = \begin{cases} \pi(A) + M^{-1}\,\mathsf{Q}(A,A^c) & \text{if } u < M \\ \pi(A) & \text{if } u \in [M, 1-M] \\ \pi(A) - M^{-1}\,\mathsf{Q}(A,A^c) & \text{if } u > 1-M \end{cases}$$

Again, it is perhaps easier to understand this with Figure 2.

To finish the proof apply the worst case constructed above to each of the $\mathcal{C}_f(A)$ quantities, and simplify. The simplification step is basically the same as in Theorem 17.3, hence the similar bounds. □

With Theorem 21.3, if the conductance profile of one Markov chain is known then it can be used to study the mixing time of a second chain via comparison. Alternatively, comparison to K_n gives a canonical path theorems.

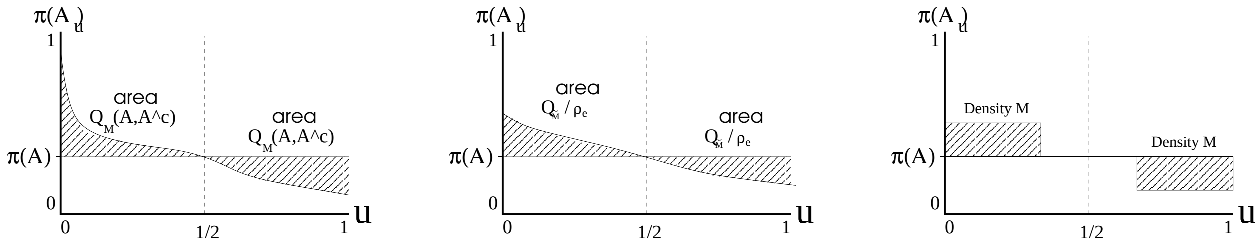

Figure 2: Correct flow for $\mathcal{M}$, flow bounded by $\mathsf{Q}_{\tilde{\mathcal{M}}}(A, A^c)/\rho_e$, then adjust for vertex congestion.

Corollary 21.4 (Canonical Paths). *Suppose $\mathcal{M}$ is a finite, irreducible, lazy Markov chain with paths γ_{xy} between every pair of vertices $x,\, y \in V$ with $x \neq y$. Define*

$$\rho_e = \max_{e\in E} \frac{1}{\mathsf{Q}(e)} \sum_{\gamma_{xy}\ni e} \pi(x)\,\pi(y) \qquad \textit{and} \qquad \rho_v = \max_{v\in V} \frac{1}{\pi(v)} \sum_{\gamma_{xy}\ni v} \pi(x)\,\pi(y)\,.$$

Then

$$1 - \mathcal{C}_{L^2} \geq 2\,\frac{\rho_v}{\rho_e}\,\left(1 - \sqrt{1 - 1/4\rho_v^2}\right) \geq \frac{1}{4\,\rho_v\,\rho_e}\,,$$

$$\tau(\epsilon) \leq 4\,\rho_v\,\rho_e\,\log\frac{1}{2\epsilon\sqrt{\pi_*}}\,.$$

If $\mathcal{M}$ is reversible then it suffices to consider $\rho_v/2$, as unordered pairs of vertices suffice in the sum.

Here is an application showing how several of the results we have proven so far can be combined to get an improved bound on mixing time. See the papers referenced for details of the facts inequalities that are used.

Example 21.5. Feder and Mihail [1] studied a random walk for sampling spanning trees of an $(n+1)$-vertex, m-edge connected graph (more generally, "balanced matroids"). Given a spanning tree choose an edge to drop, a new edge from the graph to add, and make an exchange if and only if this makes a new spanning tree. Make this lazy by doing nothing half the time (a similar construction works for counting bases of a finite vector space as well).

Feder and Mihail showed a result equivalent to $\rho_e \leq n\,m$ and $\rho_v \leq 2n$. By Corollary 21.4 it follows that for simple balanced matroids then $\tau(\epsilon) \leq 8\,m\,n^2\,\log\frac{1}{2\sqrt{\pi_*}\epsilon} \leq 8\,m\,n^2\,\left(\frac{n}{2}\,\log m + \log(1/2\epsilon)\right)$, exactly the same upper bound obtained by Feder and Mihail [1] by a canonical path theorem.

This can be improved on further. In [4] Jung Bae Son and I showed that $\Phi(A) \geq \frac{\log_2(1/\pi(A))}{2\,m\,n}$, which by Average Conductance (Corollary 20.1) implies that implies that $\tau(\epsilon) \leq 4(\log 2)m^2\,n^2 + 8m^2\,n^2\,\log(1/2\epsilon)$. This is typically a worse bound, however, combining the two lower bounds on $1 - \mathcal{C}_{L^2}(A)$ implies that

$$1 - \mathcal{C}_{L^2}(A) \geq \max\left\{\frac{1}{8m\,n^2},\, \frac{\log_2^2(1/\pi(A))}{8m^2n^2}\right\} = \begin{cases} \log_2^2(1/\pi(A))/8m^2n^2 & \textit{if}\ \pi(A) \leq 2^{-\sqrt{m}} \\ 1/8m\,n^2 & \textit{if}\ \pi(A) > 2^{-\sqrt{m}} \end{cases}$$

and this time by Corollary 19.3 the mixing time is $\tau(\epsilon) \leq 8(\log 2)m^{3/2}\,n^2 + 8m\,n^2\,\log(1/2\epsilon)$ when $2^{-\sqrt{m}} \geq m^{-n}$. This is an improvement over the canonical paths bound as $m \leq \binom{n}{2}$.

We note that the correct bound is still much smaller at $\tau(\epsilon) = O\left(m\,n\,\log\frac{n}{\epsilon}\right)$ [2].

References

[1] T. Feder and M. Mihail. Balanced matroids. *Proc. 24th Annual ACM Symposium on Theory of Computing*, pages 26–38, 1992.

[2] M. Jerrum and J-B. Son. Spectral gap and log-sobolev constant for balanced matroids. *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 721–729, 2002.

[3] R. Montenegro. A geometric approach to canonical path and comparison theorems. *Preprint*, 2004.

[4] R. Montenegro and J-B. Son. Edge isoperimetry and rapid mixing on matroids and geometric markov chains. *Proc. 33rd Annual ACM Symposium on Theory of Computing*, pages 704–711, 2001.