Lecture 18 - Simple random walk on cycles, new Cheeger inequalities

Monday, September 27

Last class we showed how to use a modified conductance $\tilde{\phi}(A)$ to upper bound mixing time in total variation, informational divergence and in L^2 distances. Today we look at some consequences of this, and a few more examples. Once again, these are taken from [1].

First, let us interpret the results from the previous class. Recall that

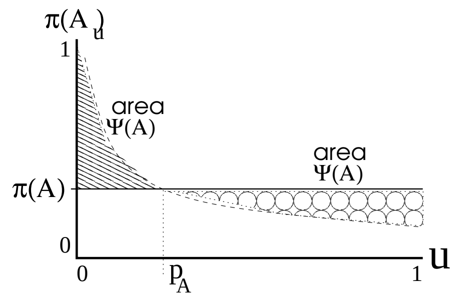
$$\tilde{\phi}(A) \geq 1 - \mathcal{C}_{\chi^2}(A) \geq \tilde{\phi}(A)^2/2$$

Why should $\tilde{\phi}(A)$ be more appropriate to consider than the conductance $\tilde{\Phi}(A)$?

One interpretation of conductance bounds on mixing are as follows. In the worst case a Markov chain starts at a point. Now, inductively, given that a Markov chain is reasonably well mixed within a set A we would like to measure how soon the Markov chain moves into a larger set than A. The conductance Φ is just the conditional probability of stepping from A to A^c, given initial point drawn from A according to π, so the inverse is the expected number of steps it takes to step out of A. Once this event occurs then the Markov chain is fairly well mixed in a slightly larger set, and the procedure can be repeated.

However, if the Markov chain is periodic then the set might not be any bigger. Consider the simple random walk on the complete bipartite graph $K_{n,n}$, a periodic Markov chain. Every subset has many edges so Φ is large, and yet the Markov chain will never converge. The problem arises because if A is one of the bipartition's then the set of points the Markov chain can reach in one step is just another bipartition of equal size, i.e. the Markov chain can bounce from A to A^c but this does not expand it into a larger set. Therefore it is better to consider how much flow from A reaches a strictly larger set, that is the worst flow into a set B where $\pi(B) = \pi(A^c)$.

This is exactly what $\Psi(A)$ does. Consider the behavior of $\pi(A_u)$, as below.



The two shaded regions have equal area, $\Psi(A) = \int_0^{\wp_A} (\pi(A_u) - \pi(A))\,du$, where $\wp_A = \sup\{u : \pi(A_u) \geq \pi(A)\} = \inf\{u : \pi(A_u) < \pi(A)\}$. By definition of the A_u, the set $A_{\wp_A}$ is the set of size $\pi(A)$ which is most connected to A, i.e. least connected to A^c. Likewise, $\Omega \setminus A_{\wp_A}$ is the set of size $\pi(A^c)$ which is least connected to A. In fact, it is not too hard to argue that if π is uniform then

$$\Psi(A) = \min_{\pi(B)=\pi(A^c)} \mathsf{Q}(A, B) \tag{1}$$

is the smallest ergodic flow into a set of size $\pi(A^c)$. If π is non-uniform then the same result holds but with the space treated as if it were continuous (so partial vertices can be taken):

$$\Psi(A) = \min_{\substack{B \subset V,\, v \in V,\\ \pi(B) \leq \pi(A^c),\, \pi(B \cup v) > \pi(A^c)}} \mathsf{Q}(A, B) + (\pi(A^c) - \pi(B))\,\frac{\mathsf{Q}(A, v)}{\pi(v)}\,,$$

This captures important properties quite well. For instance, on a reversible chain $\Psi(A) = 0$ if and only if A is one of the bipartitions of a periodic walk; the minimum in $\Psi(A)$ is then achieved by $B = A$, and $\Psi(A) = \mathsf{Q}(A, B) = 0$ rather than $\mathsf{Q}(A, A^c) > 0$ as with conductance.

For lazy chains the minimum in $\Psi(A) = \min_{\pi(B)=\pi(A^c)} \mathsf{Q}(A, B)$ is $B = A^c$, so $\tilde{\phi}(A) = \tilde{\Phi}(A)$ if the chain is lazy.

Here are a few example demonstrating how to use these various results.

Example 18.1. Consider the walk on the complete graph K_n with laziness $\gamma = 1/n$ ($\alpha = 0$ in Example 17.2). Then $\tilde{\phi}(A) = \inf_{\pi(B)=\pi(A^c)} \frac{\mathsf{Q}(A,B)}{\pi(A)\pi(A^c)} = 1$ and so $1 - \mathcal{C}_{L^2} = 1 - \mathcal{C}_{TV} = 1$. It follows from Theorem 17.3 that $1 - \mathcal{C}_{TV} = 1 - \mathcal{C}_D = 1 - \mathcal{C}_{\chi^2} = 1$, exactly as we found in Example 17.2.

Example 18.2. The walk on a cycle of even length has $\tilde{\phi} = 0$, with the worst set A given by choosing $n/2$ alternating points around the cycle, and $B = A$ in equation (1). Therefore $0 = \tilde{\phi} \geq 1 - \mathcal{C} \geq 0$ for all of the quantities dealt with in Theorem 17.3, which gives the correct $1 - \mathcal{C} = 0$ because the walk is periodic.

Now consider a walk on a cycle of odd length. In Lecture 11 (Seminar) it was found that for a walk with $\Psi(A)$ constant over set sizes that it is best to consider a sinusoidal distance. Then

$$\|\mathsf{P}^t(x,\cdot) - \pi\|_{TV} \leq \frac{1}{\pi(x)} \mathsf{E}\pi(S_t)(1-\pi(S_t)) \leq \frac{1}{\pi(x)} \frac{\pi(x)(1-\pi(x))}{\sin(3.14\,\pi(x))} \mathsf{E}\sin(3.14\,\pi(S_t))$$

$$\leq \frac{\sin(3.14\,\pi(x))}{\pi(x)} \frac{\pi(x)(1-\pi(x))}{\sin(3.14\,\pi(x))} \mathcal{C}_{sine}^t = (1-\pi(x))\,\mathcal{C}_{sine}^t$$

where 3.14 is used to represent the number π and

$$\mathcal{C}_{sine} = \max_{A \subset V} \int_0^1 \frac{\sin(3.14\,\pi(A_u))}{\sin(3.14\,\pi(A))}\,du\,.$$

Now, Jensen's inequality, as in Theorem 17.3, shows that $\mathcal{C}_{sine}(A) \leq \cos(2\pi\Psi(A))$ if $\wp_A = 1/2$. It is easily seen that $\pi(A_u) > \pi(A)$ if $u < 1/2$, while $\pi(A_u) < \pi(A)$ if $u > 1/2$, so $\wp_A = 1/2$ for all sets A on the odd cycle (consider the figure below and convince yourself that this is true on the figure). Then $\Psi(A)$ is minimized by taking alternating points around the cycle, as in the figure below

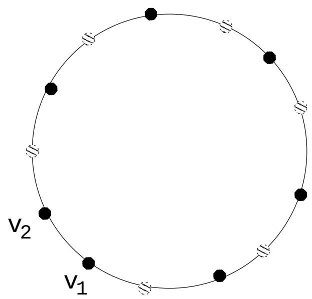


Figure 1: Let A be the white points, $B = A \cup v_1$, so $\Psi(A) = \mathsf{Q}(A, B) = 1/2n$.

Therefore $\mathcal{C}_{sine}(A) \leq \cos(\pi/n)$. The mixing time bound given above is then

$$\Delta_{TV}(t) \leq (1-\pi_*)\,\mathcal{C}_{sine}^t \leq (1-1/n)\,\cos^t(\pi/n)\,.$$

To show a lower bound, observe that $\lambda_{max} \leq \mathcal{C}_f \leq \cos(\pi/n)$ (see proof of Corollary 18.3 for this). Moreover, $\cos\left(\frac{\pi(n-1)}{n}\right)$ is an eigenvalue with eigenvector $f(j) = \cos\left(\frac{2\pi(n-1)j}{n}\right)$, so $\lambda_{max} \geq \left|\cos\left(\frac{\pi(n-1)}{n}\right)\right| = \cos(\pi/n)$. Then $\lambda_{max} = \cos(\pi/n)$, which combined with Theorem 7.4 implies a lower bound on distance as well

$$\frac{1}{2}\cos^t(\pi/n) \leq \Delta_{TV}(t) \leq (1-1/n)\,\cos^t(\pi/n)\,. \tag{2}$$

These are fairly close upper and lower bounds and are even equal at $n = 3$.

Recall from Lectures 6 and 7 that spectral decomposition can be used to show that

$$\frac{1}{2}\cos^t(\pi/n) \le \Delta_{TV}(t) \le \frac{1}{2}\Delta_{\chi^2}^{1/2}(t) \le e^{-\pi^2 t/2n^2} \quad if\ t \ge n^2/40\,.$$

The upper bound of (2) is at most $(1-1/n)e^{-\pi^2 t/2n^2}$, slightly better overall and with no conditions on t.

This argument is nice in that it only required a little examination of edge expansion properties. Once one is used to working with evolving sets then the argument is also not too difficult.

The previous two examples show that the new evolving set bounds can be useful for studying mixing times. We can also do a role reversal of sorts, and use our mixing time bounds to prove a Cheeger-like inequality.

Corollary 18.3 (New Cheeger Inequality). *If $\mathcal{M}$ is a reversible Markov chain and then*

$$\tilde{\Phi} \ge \lambda \ge 2\left(1-\sqrt{1-\tilde{\Phi}^2/4}\right) \ge \frac{\tilde{\Phi}^2}{4}\,.$$

or more generally if the laziness is γ (i.e. $\forall x \in \Omega:\ \mathsf{P}(x,x) \ge \gamma \in [0,1]$) then

$$\tilde{\Phi} \ge \lambda \ge 2(1-\gamma)\left(1-\sqrt{1-\left(\frac{\tilde{\Phi}}{2(1-\gamma)}\right)^2}\right) \ge \frac{\tilde{\Phi}^2}{4(1-\gamma)}\,,$$

where $\tilde{\Phi} = \min_{A\subset\Omega}\tilde{\Phi}(A) = \min_{A\subset\Omega}\frac{\mathsf{Q}(A,A^c)}{\pi(A)\pi(A^c)}$.

Proof. The upper bound was proven in Lecture 10 when we showed the upper bound on Cheeger's inequality. Simply try test functions $f(x) = 1_A(x)$ in the variational form of the spectral gap λ.

We know from Lecture 7 that lazy Markov chains satisfy $\|\mathsf{P}^t(x,\cdot)-\pi\|_{TV} \ge \frac{1}{2}\lambda_{max}^t = \frac{1}{2}(1-\lambda)^t$. But then

$$\frac{1}{2}(1-\lambda)^t \le \|\mathsf{P}^t(x,\cdot)-\pi\|_{TV} \le \frac{1}{2}\|\mathsf{P}^t(x,\cdot)-\pi\|_{L^2(\pi)}^{1/2} \le \frac{1}{2}\sqrt{\frac{1-\pi(x)}{\pi(x)}}\,\mathcal{C}_{L^2}^t\,.$$

Taking $t\to\infty$ implies that $1-\lambda = \lambda_{max} \le \mathcal{C}_{L^2} \le \sqrt{1-\tilde{\Phi}^2}$ (recall that $\tilde{\phi} = \tilde{\Phi}$ for a lazy chain).

For the general (not necessarily lazy) Markov chain $\mathcal{M}$ reverse the procedure used in Corollary 10.3. Let $\mathcal{M}'$ be the lazy Markov chain given by $\mathsf{P}_{\mathcal{M}'} = \frac{1}{2}(\mathsf{I}+\mathsf{P})$, so $\lambda_{\tilde{\mathcal{M}}} \ge 1-\sqrt{1-\tilde{\Phi}_{\tilde{\mathcal{M}}}^2}$. Observe that if $\mathsf{P}' = c\,\mathsf{P}+(1-c)\mathsf{I}$ then the eigenvalues of P' satisfy the condition that $\lambda_i' = c\,\lambda_i + (1-c)$, and solving for the eigenvalues of P implies that $\lambda_i = 1 + c^{-1}(-1+\lambda_i')$. In particular, $\lambda_{\tilde{\mathcal{M}}} = \frac{1+\lambda}{2}$. Also, $\tilde{\Phi}_{\tilde{\mathcal{M}}} = \frac{1}{2}\tilde{\Phi}$, so solving for λ in terms of $\lambda_{\tilde{\mathcal{M}}}$ and then $\tilde{\Phi}$ gives the corollary.

For the general case consider instead $\mathsf{P}_{\mathcal{M}'} = \frac{1}{2(1-\gamma)}\mathsf{P} + \left(1-\frac{1}{2(1-\gamma)}\right)\mathsf{I}$. ☐

The upper and lower bounds match at $2 \ge \lambda \ge 2$ for the periodic walk on the uniform two point space $\{0,1\}$, with $\mathsf{P}(0,1) = \mathsf{P}(1,0) = 1$ and $\tilde{\Phi} = 2$, so our new Cheeger bound is sharp.

In the seminar I discussed a variation on the argument used in the above proof, and used it to derive the correct values of the top and bottom eigenvalues for many Markov chains, including a walk on the complete graph K_n, on the weighted two-point space, and on both odd and even length cycles. See the paper [2] for these examples and more general methods.

References

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

[2] R. Montenegro. Evolving sets and cheeger profile bounds on the top and bottom eigenvalues. *Preprint*, 2004.