Lecture 14 - Logarithmic Sobolev inequalities, tensorization

Friday, September 17

Nash inequalities, discussed briefly in the previous lecture, can be used to derive fairly good bounds on mixing times. However, they are often difficult to compute and may give weak bound if the time s over which they are used is too long. We now proceed to break down the quantity $N(s)$ into more manageable pieces.

Our arguments follow the approach of Aldous-Fill book [1]. The theory in this section originates outside the rapid mixing community, but in the finite Markov chains setting it was largely developed in work of Diaconis and Saloff-Coste [3].

Now, recall that

$$\begin{aligned}\Delta(t) &\leq \frac{1}{2} \max_{\|f_0\|_1=1} \sqrt{Var_\pi(f_t)} \leq \frac{1}{2} \max_{\|f_0\|_1=1} \|H_s\, f_0\|_2\, \|H_{t-s} - E\|_{2\to 2} \\ &\leq \frac{1}{2} \|H_s\|_{1\to 2}\, e^{-(t-s)\lambda}\end{aligned}$$

This can be rewritten in a (possibly) better form via the following lemma.

Lemma 14.1. *Given any operator* A *let* $\mathsf{A}^*_{ij} = (\pi_j\, a_{ji}/\pi_i)$ *denote its adjoint with respect to* π*. Then for any* $1 \leq q_1,\, q_2 \leq \infty$

$$\|\mathsf{A}\|_{q_1\to q_2} = \|\mathsf{A}\|_{q_2^*\to q_1^*}$$

where $\frac{1}{q} + \frac{1}{q^*} = 1$*. In particular, if* A *is a reversible transition matrix then*

$$\|\mathsf{A}\|_{2\to q} = \|\mathsf{A}\|_{q^*\to 2}$$

This is a standard duality result. For a proof see Lemma 13 in Chapter 8 of Aldous-Fill book [1].

Then $N(s) = \|H_s\|_{1\to 2} = \|H_s\|_{2\to\infty}$. Rather than work with $N(s)$, a slightly better form for our purposes will be to observe that if $\frac{1}{q} + \frac{1}{q_*} = 1$ then

$$\|H_s\, f_0\|_2 \leq \|f_0\|_{q_*}\, \|H_s\|_{q_*\to 2} = \|f_0\|_{\frac{q}{q-1}}\, \|H_s\|_{2\to q}$$

and so

$$\|\mathsf{p}^{(t)} - \pi\|_{TV} \leq \frac{1}{2} \sqrt{Var_\pi(f_t)} \leq \frac{1}{2} \|f_0\|_{\frac{q}{q-1}}\, \|H_s\|_{2\to q}\, e^{-(t-s)\lambda}$$

As $s \to \infty$ then the distance converges to zero, or equivalently H_s converges to E. Let

$$s_q := inf\{s \geq 0 :\, \|H_s\|_{2\to q} = 1\}$$

Certainly $s_2 = 0$, but somewhat surprisingly for every finite q the value s_q is in fact finite.

Definition 14.2. The *logarithmic Sobolev constant* ρ is given by

$$\rho = \inf_{Ent_\pi(f^2)\neq 0} \frac{2\mathcal{E}(f,f)}{Ent_\pi(f^2)}$$

Theorem 14.3. *For any finite, irreducible, reversible Markov chain*

$$\rho = \inf_{2<q<\infty} \frac{\log(q-1)}{2s_q}\,,$$

i.e. $\forall q \geq 2 :\, s_q \leq (2\rho)^{-1}\, \log(q-1)$.

The proof is not too bad but it is too long for the limited time we have. See Section 8, Theorem 24 of Aldous-Fill book for a proof of this.

We now have

Theorem 14.4. *For a finite, irreducible, reversible Markov chain then for any state with* $\pi(i) \leq e^{-1}$ *then*

$$\begin{aligned} \|\delta_x \, \mathsf{P}^t - \pi\|_{TV} &\leq \frac{1}{2}\, e^{1-c} \quad \textit{if } t \geq \frac{1}{2\rho} \log\log(1/\pi(x)) + \frac{c}{\lambda} \\ \tau(\epsilon) &\leq \frac{1}{2\rho} \log\log(1/\pi_*) + \frac{1}{\lambda}\,(1 + \log(1/2\epsilon)) \\ \left\| 1 - \frac{\mathsf{p}^{(t)}}{\pi} \right\|_{2,\pi} &\geq e^{-1} \quad \textit{if}\, t \leq 1/\rho \end{aligned}$$

Proof. Recall that

$$\|\mathsf{p}^{(t)} - \pi\|_{TV} \leq \frac{1}{2}\, \|f_0\|_{\frac{q}{q-1}}\, \|H_s\|_{2\to q}\, e^{-(t-s)\lambda}$$

The worst case for f_0 is a point mass, i.e. $f_0 = \delta_x$, at which point $\|f_0\|_{q/(q-1)} \leq \pi(x)^{-1/q}$. Also, the result above shows that $\|H_s\|_{2\to q} = 1$ for all $s \geq (2\rho)^{-1} \log(q-1)$. Let $q(s) = 1 + e^{2\rho\, s}$, then

$$\|\delta_x \, \mathsf{P}^t - \pi\|_{TV} \leq \frac{1}{2}\, \pi(x)^{-1/q(s)}\, e^{-(t-s)\lambda}$$

Let $s = \frac{1}{2\rho} \log\log(1/\pi(x))$ and so $q(s) = 1 + \log(1/\pi(x)) = \log(e/\pi(x))$ and thus

$$\|\delta_x \, \mathsf{P}^t - \pi\|_{TV} \leq \frac{1}{2}\, e^{1-(t-s)\lambda}$$

when $t \geq s$.

The lower bound on L^2 mixing time is too complicated for this course. The proof can be found in the paper of Diaconis and Saloff-Coste [3]. □

The lower bound on L^2 distance is perhaps clearest if written as follows:

$$\frac{1}{\rho} \leq \chi^2(1/e) \leq \frac{4 + \log\log(1/\pi_*)}{2\rho} \qquad \textit{where } \chi^2(\epsilon) = \max_{x\in\Omega} \min\{t : \ \|1 - \delta_x\, H_t/\pi\|_{2,\pi} \leq \epsilon\}$$

So the time for L^2 distance to reach e^{-1} is fairly tightly bounded both above and below by ρ. In contrast, the upper and lower bounds on $\tau(1/e)$ in terms of spectral gap differ by a factor of $\log(1/\pi_*)$. When the state space has exponential size then this is a substantial difference.

Finally, all our analytic methods for bounding mixing time are done. It turns out that the various quantities studied in the last few lectures, ρ, ρ_0 and λ have a strict ordering of sizes.

Following [2] we have:

Theorem 14.5.

$$\frac{2(1 - 2\pi_*)}{\log\left(\frac{1-\pi_*}{\pi_*}\right)}\lambda \leq \rho \leq \rho_0 \leq \lambda$$

Proof. We will not consider the first inequality. This can be found in [3].

For the second inequality consider any non-negative function f. Then $\mathcal{E}(\sqrt{f}, \sqrt{f}) \leq \frac{1}{4}\,\mathcal{E}(f, \log f)$. To show this it suffices to verify that $(x-y)^2 \leq \frac{1}{2}\,(x^2-y^2)(\log x - \log y)$ for every $x, y \geq 0$. This is easily verified. It follows that

$$\frac{2\mathcal{E}(\sqrt{f}, \sqrt{f})}{Ent_\pi(f)} \leq \frac{\mathcal{E}(f, \log f)}{2\,Ent_\pi(f)}$$

and taking the infinum over all such f shows that $\rho \leq \rho_0$.

The inequality $\rho_0 \leq \lambda$ will be given as a homework problem. □

The mixing time bounds in terms of ρ and ρ_0 are essentially the same. Since $\rho_0 \geq \rho$ it may be more useful under certain circumstances, but in the asymptotics if $\rho_0 < \lambda$ then the log-Sobolev bound will be better.

Log-sobolev and modified log-Sobolev are notoriously difficult to bound. However, there are arguments that work in a few cases.

Definition 14.6. Given Markov chains $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_n$ the product chain $\mathcal{M} = \prod \mathcal{M}_i$ is the Markov chain on the Cartesian product $\prod \Omega_i$ with transition matrix

$$\mathsf{P}((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \begin{cases} \frac{1}{n}\,\mathsf{P}_i(x_i, y_i) & \text{if } x_i \neq y_i \text{ and } \forall j \neq i : x_j = y_j \\ \frac{1}{n}\,\sum_{i=1}^n \mathsf{P}_i(x_i, x_i) & \text{if } x_i = y_i \text{ for all } i \\ 0 & \text{otherwise} \end{cases}$$

In short, the product chain chooses a coordinate i uniformly at random and then makes a step on this coordinate according to the chain $\mathcal{M}_i$.

Theorem 14.7 (Tensorization). *For a product $\mathcal{M} = \prod \mathcal{M}_i$ of finite, irreducible, reversible Markov chain then*

$$\lambda = \frac{1}{n}\,\min \lambda_i,\ \ \rho = \frac{1}{n}\,\min \rho_i,\ \ N(s) = \prod N_i(s)$$

Proof. Consider $N(s)^2$. The bound on $N(s)^2$ follows quickly from the form $N(s)^2 = \max_i \frac{\mathsf{P}^{2s}(i,i)}{\pi(i)}$.

For the bounds on λ and ρ see Chapter 8, Lemma 35 and Theorem 36 in Aldous-Fill. A nicer proof can be found in the Ph.D. Dissertation of Stoyanov [4] (a former GATech student). □

This shows that if it is possible to determine ρ, λ or $N(s)$ then it is also possible to determine the values for the products.

Example 14.8. Consider the lazy simple walk on the boolean cube $\{0,1\}^d$ a product chain for the two-point space with a single step to uniform walk, i.e. $\mathsf{P}(0,0) = \mathsf{P}(0,1) = \mathsf{P}(1,0) = \mathsf{P}(1,1) = 1/2$. This is simple enough to determine everything exactly.

$$\lambda = \inf_{Var f \neq 0} \frac{\mathcal{E}(f, f)}{Var_\pi(f)}$$

Observe that given a constant c then

$$\frac{\mathcal{E}(c+f,\, c+f)}{Var_\pi(c+f)} = \frac{\mathcal{E}(f, f)}{Var_\pi(f)} = \frac{\mathcal{E}(c\,f,\, c\,f)}{Var_\pi(c\,f)}$$

Given f on $\{0,1\}$ then consider $f - f(0)$, so it can be assumed that $f(0) = 0$. Likewise, given that $f(0) = 0$ then consider $f/f(1)$, so it can further be assumed that $f(1) = 1$. It follows that the only function f of interest is $f(0) = 0$, $f(1) = 1$. Then

$$\lambda = \frac{(f(1) - f(0))^2 \, \pi(0)\mathsf{P}(0,1)}{(f(1) - f(0))^2 \, \pi(0)\pi(1)} = \frac{1/4}{1/4} = 1$$

Therefore

$$\lambda(\{0,1\}^d) = \frac{1}{d}$$

You will determine ρ and ρ_0 on the homework.

References

[1] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs (book to appear). *URL for draft http://www.stat.Berkeley.edu/users/aldous.*

[2] S. Bobkov and P. Tetali. Modified log-sobolev inequalities in discrete settings. *preprint available at http://www.math.gatech.edu/˜tetali/RESEARCH/pubs.html*, 2003.

[3] P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[4] T. Stoyano. *Isoperimetric and Related Constants for Graphs and Markov Chains.* Ph.d. thesis, Department of Mathematics, Georgia Institute of Technology, 2001.