Lecture 5

Lecturer: Michael Choi *Scribe: Michael Choi*

1 Goal of this lecture

In this lecture, we will look at the notion of mixing time under the relative entropy and ℓ^2 distance. We shall see how the log-Sobolev constant come into play to give tighter bound on these mixing times than previous techniques that we learnt in earlier lectures. Classical references are Bobkov and Tetali (2006); Diaconis and Saloff-Coste (1996); Montenegro and Tetali (2006).

2 Relative entropy and ℓ^2 mixing time

In this lecture, we will stick to the continuous-time setting as the results are cleaner. Let P be an ergodic transition matrix on a finite state space $\mathcal{X}$ with stationary distribution π, and as we have seen in our previous lecture we take the generator to be $G = P - I$. The continuized chain is a Markov chain $(X_t)_{t\geq 0}$ with transition semigroup

$$H^t = e^{tG} = \sum_{n=0}^{\infty} \frac{t^n G^n}{n!}.$$

Let

$$h_t^x(y) = \frac{H^t(x,y)}{\pi(y)}$$

be the density of $H_t(x,y)$ with respect to π at time t. Recall in lecture 1 that we are working in the Hilbert space $\ell^2(\pi)$ with inner product $\langle f, g\rangle_\pi = \sum_{x,y} f(x)g(x)\pi(x)$ and its norm $||f||_\pi^2 := \langle f, f\rangle_\pi$. Define the variance and relative entropy to be respectively

$$\begin{aligned}
\mathrm{Var}_\pi(h_t^x) &:= \|h_t^x - 1\|_\pi^2 = \sum_{y\in\mathcal{X}} \pi(y)\left(h_t^x(y) - 1\right)^2, \\
\mathrm{Ent}_\pi(h_t^x) &:= \sum_{y\in\mathcal{X}} H^t(x,y)\log\frac{H^t(x,y)}{\pi(y)} = \sum_{y\in\mathcal{X}} h_t^x(y)\log h_t^x(y)\pi(y).
\end{aligned}$$

The ℓ^2 **and relative entropy mixing time** are defined as follows, for $\epsilon > 0$,

$$\begin{aligned}
t_2(\epsilon) &:= \inf\{t \geq 0;\ \max_x \mathrm{Var}_\pi(h_t^x) \leq \epsilon\}, \\
t_{Ent}(\epsilon) &:= \inf\{t \geq 0;\ \max_x \mathrm{Ent}_\pi(h_t^x) \leq \epsilon\}.
\end{aligned}$$

The **Dirichlet form** of G is given by

$$\mathcal{E}(f,g) = \langle f, (-G)g\rangle_\pi = \sum_{x,y} f(x)(g(x) - g(y))P(x,y)\pi(y).$$

The **spectral gap λ, log-Sobolev constant ρ and modified log-Sobolev constant** ρ_0 are defined to be respectively:

$$\lambda := \inf_{\mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\pi(f)},$$
$$\rho := \inf_{\mathrm{Ent}_\pi(f^2)\neq 0} \frac{\mathcal{E}(f,f)}{\mathrm{Ent}_\pi(f^2)},$$
$$\rho_0 := \inf_{f\geq 0;\ \mathrm{Ent}_\pi(f)\neq 0} \frac{\mathcal{E}(f,\log f)}{\mathrm{Ent}_\pi(f)}.$$

3 Mixing time bounds via λ and ρ_0

Our main result in this lecture is the following:

Theorem 1 *Let $\pi_{min} := \min_x \pi(x)$. Then for $\epsilon > 0$,*

$$t_2(\epsilon) \leq \frac{1}{\lambda}\left(\frac{1}{2}\log\left(\frac{1-\pi_{min}}{\pi_{min}}\right) + \log\frac{1}{\epsilon}\right),$$
$$t_{Ent}(\epsilon) \leq \frac{1}{\rho_0}\left(\log\log\frac{1}{\pi_{min}} + \log\frac{1}{\epsilon}\right).$$

We first state two lemmas that will help our proof.

Lemma 2 (Kolmogorov forward equation) *For any $x, y \in \mathcal{X}$ and $t \geq 0$,*

$$\frac{d}{dt}h_t^x(y) = G^* h_t^x(y),$$

where G^* is the adjoint operator of G in $\ell^2(\pi)$.

Lemma 3 (Variance flow and entropy flow)

$$\frac{d}{dt}\mathrm{Var}_\pi(h_t^x) = -2\mathcal{E}(h_t^x, h_t^x),$$
$$\frac{d}{dt}\mathrm{Ent}_\pi(h_t^x) = -\mathcal{E}(h_t^x, \log h_t^x).$$

Proof:

$$\begin{aligned}
\frac{d}{dt}\mathrm{Var}_\pi(h_t^x) &= \sum_{y\in\mathcal{X}} \pi(y)\frac{d}{dt}\left(h_t^x(y)-1\right)^2 \\
&= 2\sum_{y\in\mathcal{X}} \pi(y)\left(h_t^x(y)-1\right)G^* h_t^x(y) \\
&= 2\langle G h_t^x, h_t^x\rangle_\pi - \langle G\mathbf{1}, h_t^x\rangle_\pi \\
&= -2\mathcal{E}(h_t^x, h_t^x),
\end{aligned}$$

where we use Lemma 2 in the second equality, and the fourth equality follows from $G\mathbf{1} = 0$.

$$\begin{aligned}
\frac{d}{dt}\mathrm{Ent}_\pi(h_t^x) &= \sum_{y\in\mathcal{X}} \pi(y)\frac{d}{dt}h_t^x(y)\log h_t^x(y) \\
&= \sum_{y\in\mathcal{X}} \pi(y)(\log h_t^x(y)+1)G^* h_t^x(y) \\
&= \langle G\log h_t^x, h_t^x\rangle_\pi + \langle G\mathbf{1}, h_t^x\rangle_\pi \\
&= -\mathcal{E}(h_t^x, \log h_t^x),
\end{aligned}$$

where we use Lemma 2 in the second equality, and the fourth equality follows from $G\mathbf{1} = 0$.

□

We can now state the proof of Theorem 1.

Proof of Theorem 1: By Lemma 3 and the definition of λ and ρ_0, we have

$$\begin{aligned}
\frac{d}{dt}\mathrm{Var}_\pi(h_t^x) &= -2\mathcal{E}(h_t^x, h_t^x) \le -2\lambda \mathrm{Var}_\pi(h_t^x), \\
\frac{d}{dt}\mathrm{Ent}_\pi(h_t^x) &= -\mathcal{E}(h_t^x, \log h_t^x) \le -\rho_0 \mathrm{Ent}_\pi(h_t^x).
\end{aligned}$$

Desired results follow by noting that

$$\mathrm{Var}_\pi(h_0^x) \le \frac{1-\pi_{min}}{\pi_{min}}, \quad \mathrm{Ent}_\pi(h_0^x) \le \log\frac{1}{\pi_{min}}.$$

□

4 Bounds on the log-Sobolev constant ρ

In Lecture 2, we have seen how we apply geometric bounds such as Poincaré's or Cheeger's inequality in bounding the SLEM. In the following, we bound ρ_0 in terms of ρ and λ:

Theorem 4

$$2\rho \le \rho_0 \le 2\lambda.$$

We first state a useful lemma:

Lemma 5 *If $f \ge 0$, then*

$$2\mathcal{E}(\sqrt{f}, \sqrt{f}) \le \mathcal{E}(f, \log f).$$

Proof: First, observe that, for $a, b, c > 0$,

$$a(\log a - \log b) = 2a\log\frac{\sqrt{a}}{\sqrt{b}} \ge 2a\left(1 - \frac{\sqrt{b}}{\sqrt{a}}\right) = 2\sqrt{a}(\sqrt{a}-\sqrt{b})$$

by the relation $\log c \ge 1 - c^{-1}$. Then

$$\begin{aligned}
\mathcal{E}(f, \log f) &= \sum_{x,y} f(x)(\log f(x) - \log f(y))\mathrm{P}(x,y)\pi(x) \\
&\ge 2\sum_{x,y} f^{1/2}(x)\left(f^{1/2}(x) - f^{1/2}(y)\right)\mathrm{P}(x,y)\pi(x) \\
&= 2\mathcal{E}(\sqrt{f}, \sqrt{f}).
\end{aligned}$$

$\square$

Proof of Theorem 4: The first inequality is immediate from Lemma 5. For the second inequality, we take $g \in \ell^2(\pi)$ to be an arbitrary function with $\mathbb{E}_\pi(g) = \langle 1, g\rangle_\pi = 0$. Let $f = 1 + \epsilon g$, where $\epsilon \ll 1$ such that $f \geq 0$. Using Taylor expansion, we have $\log(1+\epsilon g) = \epsilon g - \frac{1}{2}(\epsilon)^2 g^2 + o\left(\epsilon^2\right)$, and so

$$\mathrm{Ent}_\pi(f) = \sum_{y\in\mathcal{X}} \pi(y) f(y) \log f(y) = \frac{1}{2}\epsilon^2 \pi\left(g^2\right) + o\left(\epsilon^2\right),$$

$$\mathcal{E}(f, \log f) = -\epsilon \mathbb{E}_\pi((Gg)\log(1+\epsilon g)) = \epsilon^2 \mathcal{E}(g,g) + o\left(\epsilon^2\right).$$

As a result, we have

$$\mathcal{E}(f,\log f) = \epsilon^2\mathcal{E}(g,g) + o\left(\epsilon^2\right) \geq \rho_0 \mathrm{Ent}_\pi(f) = \frac{\rho_0}{2}\epsilon^2\pi\left(g^2\right) + o\left(\epsilon^2\right).$$

Dividing by ϵ^2 and take $\epsilon \to 0$ yields

$$\frac{\mathcal{E}(g,g)}{\pi(g^2)} \geq \frac{\rho_0}{2}.$$

Desired result follows since this holds for arbitrary g with $\mathbb{E}_\pi g = 0$. $\square$

References

S. G. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006.

P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006.