Lecture 7 - Exponential decay and lower bounding mixing time

Monday, August 30

Today we show a general method for lower bounding mixing time. This will be used to show that the upper bound on the cycle walk from the previous lecture is essentially sharp.

First we need a definition and lemma.

Definition 7.1. Let $\Delta_x(t) = \|\mathsf{P}^t(x,\cdot) - \pi\|_{TV}$ and $\Delta(t) = \max_{x\in\Omega} \Delta_x(t)$.

Lemma 7.2. *If $t = \sum_{i=1}^k t_i$ then*

$$\Delta(t) \leq 2^{k-1} \prod_{i=1}^k \Delta(t_i)\,.$$

Proof. Consider the case when $t = t_1 + t_2$. Then

$$\begin{aligned}
2\Delta_x(t_1+t_2) &= \sum_{y\in\Omega} \left|\mathsf{P}^{t_1+t_2}(x,y) - \pi(y)\right| \\
&= \sum_{y\in\Omega} \left|\sum_{z\in\Omega} \mathsf{P}^{t_1}(x,z)\,\mathsf{P}^{t_2}(z,y) - \pi(y)\right| \\
&= \sum_{y\in\Omega} \left|\sum_{z\in\Omega} (\mathsf{P}^{t_1}(x,z) - \pi(z))(\mathsf{P}^{t_2}(z,y) - \pi(y))\right| \\
&\leq \sum_{y\in\Omega}\sum_{z\in\Omega} \left|(\mathsf{P}^{t_1}(x,z) - \pi(z))(\mathsf{P}^{t_2}(z,y) - \pi(y))\right| \\
&\leq \sum_{z\in\Omega} \left|\mathsf{P}^{t_1}(x,z) - \pi(z)\right| \sum_{y\in\Omega} \left|\mathsf{P}^{t_2}(z,y) - \pi(y)\right| \\
&\leq 2\Delta_x(t_1)\, 2\Delta(t_2)
\end{aligned}$$

The final equality can be checked by multiplying out terms and adding over $z \in \Omega$ (reversibility is not needed). The first inequality was the triangle inequality $|\sum x_i| \leq \sum |x_i|$.

By induction $2\Delta_x(\sum_{i=1}^k t_i) \leq 2\Delta_x(t_1) \prod_{i=2}^k 2\Delta(t_i)$. The lemma follows immediately. □

In particular, we have

Corollary 7.3. *For every $\delta < 1/2$ the mixing time satisfies $\tau(\epsilon) \leq \tau(\delta) \lceil \log_{\frac{1}{2\delta}} (1/2\epsilon) \rceil$ and in particular $\tau(\epsilon) \leq \tau(1/2e) \lceil \log(1/2\epsilon) \rceil$.*

This shows that once the distance drops below 1/2 then it will subsequently drop exponentially fast.

Remark: The typical method for proving this is a coupling argument. I have always seen the bound stated as $\tau(\epsilon) \leq \tau(1/2e) \lceil \log(1/\epsilon) \rceil$. However, a careful look at the coupling proof shows that in fact a factor 2 improvement is possible and the coupling proof also shows $\tau(\epsilon) \leq \tau(1/2e) \lceil \log(1/2\epsilon) \rceil$, as well as the more general δ form given by Corollary 7.3.

We are now in a situation to prove the main result.

Theorem 7.4. *A reversible ergodic Markov chain satisfies $\Delta(t) \geq \frac{1}{2} \lambda_{max}^t$, or equivalently*

$$\tau(\epsilon) \geq \frac{\log(1/2\epsilon)}{\log(1/\lambda_{max})} \geq \frac{\lambda_{max}}{1-\lambda_{max}} \log(1/2\epsilon)\,.$$

Proof. Our proof is based on an argument of Sinclair [2]. If $\lambda_{max} = \lambda_2$ then let x be such that $(\overrightarrow{u_2})_x \neq 0$, while if $\lambda_{max} = |\lambda_n|$ then let x be such that $(\overrightarrow{u_n})_x \neq 0$. Clearly $\Delta_x(2t) = \frac{1}{2}\sum_{y\in\Omega}|\mathsf{P}^{2t}(x,y) - \pi(y)| \geq \frac{1}{2}(\mathsf{P}^{2t}(x,x) - \pi(x))$. From Theorem 5.2 it is known that $\mathsf{P}^{2t}(x,x) - \pi(x) = \sum_{m=2}^{n}\lambda_m^{2t}\,(\overrightarrow{u_m})_x^2$ and it follows that

$$\Delta_x(2t) \geq c\,\lambda_{max}^{2t}$$

for some $c > 0$ (in particular, $c \geq \frac{1}{2}\,(\overrightarrow{u_2})_x^2$ or $\frac{1}{2}\,(\overrightarrow{u_n})_x^2$).

Now, let $t = A\,\tau(\epsilon)$. Then by Lemma 7.2 it follows that $\frac{1}{2}\,(2\epsilon)^{2A} \geq \Delta_x(2t) \geq c\,\lambda_{max}^{2A\tau(\epsilon)}$ Taking $A \to \infty$ it follows that $2\epsilon \geq \lambda_{max}^{\tau(\epsilon)}$ and so $\tau(\epsilon) \geq \frac{\log(1/2\epsilon)}{\log(1/\lambda_{max})}$. Finish by applying the approximation $\log(1/x) = -\log x = -\log(1-(1-x)) \leq \frac{1-x}{x}$ when $x \in (0,1]$.

Now, let $\epsilon = \Delta(t)$, observe that $\tau(\epsilon) \leq t$ and rearrange the lower bound on $\tau(\epsilon)$ to show that $\Delta(t) = \epsilon \geq \frac{1}{2}\,\lambda_{max}^{\tau(\epsilon)} \geq \frac{1}{2}\,\lambda_{max}^{t}$. □

This theorem will be used in the seminar next Friday in which I go in the reverse direction and use upper bounds on the mixing time to prove upper bounds on λ_{max}.

Combining this with the upper bound of Theorem 5.2 gives

Theorem 7.5. *A reversible ergodic Markov chain satisfies*

$$\left\lceil \frac{\log(1/2\epsilon)}{\log(1/\lambda_{max})} \right\rceil \leq \tau(\epsilon) \leq \left\lceil \frac{\frac{1}{2}\,\log\frac{1-\pi_*}{\pi_*} + \log(1/2\epsilon)}{\log(1/\lambda_{max})} \right\rceil$$

where $\pi_* = \min_{x\in\Omega}\pi(x)$.

Observe that in the limit as $\epsilon \to 0^+$ that the top and bottom bounds converge to the same order, so $\tau(\epsilon) \xrightarrow{\epsilon\to 0^+} (1+o(1))\,\frac{\log(1/\epsilon)}{\log(1/\lambda_{max})}$. Therefore λ_{max} determines the asymptotic rate of convergence.

For the walk on the uniform two-point space the upper and lower bounds match at all ϵ and this implies that $\tau(\epsilon) = \left\lceil \frac{\log(1/2\epsilon)}{\log(1/\lambda_{max})} \right\rceil$. In fact, you can check that $\lambda_{max} = |2\gamma - 1|$ and so this gives exactly the same bound we found earlier.

A more conventional form is

$$\frac{\lambda_{max}}{1-\lambda_{max}}\,\log(1/2\epsilon) \leq \tau(\epsilon) \leq \frac{1}{1-\lambda_{max}}\left(\frac{1}{2}\,\log(1/\pi_*) + \log(1/2\epsilon)\right)$$

where it is understood that the upper bound should be rounded up.

We can now show the lower bound for the walk on the cycle.

Example 7.6. Consider the simple random walk on the cycle C_ℓ of odd length ℓ. A quick look at the eigenvalues given in the previous lecture shows that $\lambda_{max} = \cos(\pi/\ell) \geq 1 - \frac{\pi^2}{2\ell^2}$. The lower bound on $\tau(\epsilon)$ given above implies that $\tau(\epsilon) \geq \frac{2\ell^2}{\pi^2}\left(1 - \frac{\pi^2}{2\ell^2}\right)\log(1/2\epsilon)$. Combined with what was shown in the previous lecture it follows that for this walk

$$\left(\frac{2\ell^2}{\pi^2} - 1\right)\log(1/2\epsilon) \leq \tau(\epsilon) \leq \frac{2\ell^2}{\pi^2}\,\log(1/\epsilon) \quad \text{if } \epsilon \leq e^{-\pi^2/80} \approx 0.88$$

The difference between the upper and lower bounds is negligible.

Alternatively, in terms of variation distance Theorem 7.4 implies $\Delta(t) \geq \frac{1}{2}\cos^t(\pi/\ell)$. But if $x \leq 1/2$ then $\cos x \geq e^{-x^2/2 - x^4/11}$ and so

$$\frac{1}{2}\exp\left(-\frac{\pi^2\,t}{2\ell^2} - \frac{\pi^4\,t}{11\ell^4}\right) \leq \Delta(t) \leq \exp\left(-\frac{\pi^2\,t}{2\ell^2}\right)$$

if $\ell \geq 7$ and $t \geq \ell^2/40$. The upper and lower bounds on variation distance differ by only a factor of two.

Armed with Theorem 7.5 we may now show upper and lower bounds on mixing time for a variety of Markov chains.

Example 7.7. Consider the lazy random walk with loops on the binary cube 2^d. This can be interpreted as a walk on d-tuples in $\mathbb{Z}_2^d = \{0,1\}^d$, such $\forall x, y \in 2^d$ the transition probabilities are $\mathsf{P}(x,x) = 1/2$, $\mathsf{P}(x,y) = 1/2d$ if x and y differ in exactly one coordinate, and $\mathsf{P}(x,y) = 0$ otherwise. It is known that $\lambda_2 = 1 - 1/d$ for this Markov chain. By Theorem 7.5 it follows that

$$(d-1)\,\log(1/2\epsilon) \leq \tau(\epsilon) \leq \frac{\log 2}{2}\,d^2 + d\,\log(1/2\epsilon)\,.$$

The correct mixing time is $\tau(\epsilon) \approx \frac{1}{2}\,d\,\log d + d\,\log(1/2\epsilon)$ [1]. Our asymptotic rate was of course correct, but the "burn-in" (time to reach $\epsilon = 1/2e$) is far too slow.

Example 7.8. Consider the walk on the complete graph K_n with laziness $1/2$, that is $\forall x \neq y \in K_n$: $\mathsf{P}(x,x) = 1/2,\ \mathsf{P}(x,y) = 1/2(n-1)$. The spectrum is $\lambda_1 = 1,\, \lambda_2,\, \ldots,\, \lambda_n = \frac{n-2}{2(n-1)}$. Then

$$\frac{1}{2}\left(\frac{n-2}{2(n-1)}\right)^t = \frac{1}{2}\,\lambda_{max}^t \leq \Delta(t) \leq \frac{1}{2}\,\Delta_{2,\pi}(t) \leq \frac{1}{2}\sqrt{\frac{1-\pi_*}{\pi_*}}\,\lambda_{max}^t = \frac{1}{2}\,\sqrt{n-1}\,\left(\frac{n-2}{2(n-1)}\right)^t$$

where $\Delta_{2,\pi}(t) = \max_{x\in V}\left\|1 - \frac{\mathsf{P}^t(x,\cdot)}{\pi}\right\|_{2,\pi}$. The correct bounds are $\Delta(t) = \left(\frac{n-2}{2(n-1)}\right)^t(1-1/n)$ and $\Delta_{2,\pi}(t) = \left(\frac{n-2}{2(n-1)}\right)^t\sqrt{n-1}$. The lower bound was off by only a factor of two and the upper bound is correct.

However, $\Delta(t) \ll \Delta_{2,\pi}(t)$. This demonstrates that sometimes methods for bounding L_2 distance will give poor bounds on the total-variation distance. Generally this is not an issue, but we do still need methods for bounding total-variation distance more directly.

In the next class we will discuss how to upper bound λ_{max}. This makes it possible to bound mixing times for a much larger range of problems.

References

[1] P. Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *Lecture Notes - Monograph Series*. Institute of Mathematical Statistics, 1988.

[2] A. Sinclair. Notes for cs294-2: Markov chain monte carlo. *UC Berkeley*, 2002.