Lecture 8 - Dirichlet forms and comparison of spectral gap

Wednesday, September 1

In the past several lectures we have found that the mixing time is controlled by the second largest eigenvalue in magnitude, $\lambda_{max} = \max\{\lambda_2, |\lambda_n|\}$. A good method is now needed to upper bound λ_{max}.

This will be much easier to cope with if $\lambda_2 \geq |\lambda_n|$. Suppose that $\forall x \in \Omega :\ \mathsf{P}(x,x) \geq \gamma$ for some $\gamma \in [0,1)$. Then observe that

$$\mathsf{P} = (1-\gamma)\,\tilde{\mathsf{P}} + (1-(1-\gamma))\,I \quad \text{where} \quad \tilde{\mathsf{P}} = \frac{1}{1-\gamma}\,\mathsf{P} + \left(1 - \frac{1}{1-\gamma}\right) I$$

is the Markov chain sped up by a factor of $1-\gamma$. The matrix $\tilde{\mathsf{P}}$ is still a transition matrix for a Markov chain and so $\tilde{\lambda}_n \geq -1$. Moreover, the eigenvalues λ_i of P satisfy

$$\lambda_i = (1-\gamma)\,\tilde{\lambda}_i + (1-(1-\gamma)) \quad \text{with eigenvector} \quad u_i = \tilde{u}_i$$

where $\tilde{\lambda}_i$ are the eigenvalues of $\tilde{\mathsf{P}}$ and $\tilde{u}_i$ are the eigenvectors. Then $\lambda_n \geq (1-\gamma)(-1) + \gamma = -1 + 2\gamma$ and so $|\lambda_n| \leq \min\{\lambda_2, 1-2\gamma\}$.

Remark 8.1. *From the discussion above it follows that*

- *If P is lazy then $\lambda_{max} = \lambda_2$.*
- *The transition matrix $\hat{\mathsf{P}} = \frac{1}{2}(I + \mathsf{P})$ has $\hat{\lambda}_{max} = \hat{\lambda}_2 = \frac{1}{2}(1+\lambda_2)$.*
- *If P has laziness γ and $c \geq \lambda_2$ is some upper bound on λ_2 with $c \geq 1-2\gamma$ then $\lambda_{max} \leq c$.*

Example 8.2. Consider the walk on the binary cube 2^d given by $\mathsf{P}(x,y) = \frac{1}{d+1}$ if $y \sim x$ or $y = x$. Then $\gamma = \frac{1}{d+1}$ and $\lambda_2 = 1 - \frac{2}{d+1}$ (use the fact that $\lambda_2 = 1 - 1/d$ for the lazy chain, given last class, to show the value for the new walk P). Then $\lambda_{max} \leq \max\{1-2\gamma, \lambda_2\} = 1 - \frac{1}{d+1}$. In fact, every upper bound $c \geq \lambda_2$ will satisfy $c \geq 1 - \frac{2}{d+1} = 1 - 2\gamma$, so no matter how bad the estimate it will always happen that $\lambda_{max} \leq c$.

Remark 8.3 (Issues in implementation). *Suppose that P has laziness γ and we want to run the Markov chain for $\tau(\mathcal{E})$ steps. Then it is faster to first generate $t = Binomial(\tau(\epsilon), 1-\gamma)$ and then take t steps of the Markov chain $\hat{\mathsf{P}} = \frac{1}{1-\gamma}\,\mathsf{P} + \left(1 - \frac{1}{1-\gamma}\right) I$. The expected number of steps $\mathsf{E}t = (1-\gamma)\tau(\epsilon)$ is smaller by a factor $(1-\gamma)$.*

Observe that "slowing down" the Markov chain by a factor of two in taking $\hat{\mathsf{P}} = \frac{1}{2}(I+\mathsf{P})$ at worst halves the value of $1-\lambda_{max}$, that is $1 - \hat{\lambda}_{max} = 1 - \hat{\lambda}_2 = \frac{1}{2}(1-\lambda_2) \geq \frac{1}{2}(1-\lambda_{max})$ and therefore the mixing time bound $\tau(\mathcal{E}) \leq \frac{1}{1-\lambda_{max}} \log \frac{1}{2\epsilon\sqrt{\pi_*}}$ worsens by at most a factor of 2. However, $\mathsf{E}t = (1-\gamma)\tau(\epsilon)$ and so this factor of two is regained by the binomial argument.

In short making a Markov chain lazy does not effect the number of steps it takes to generate a good sample (at least via the λ_{max} bound). However it makes it possible to guarantee $\lambda_{max} = \lambda_2$, so we may as well assume that $\lambda_{max} = \lambda_2$ always.

The goal of the remainder of the lecture is to upper bound λ_2.

Lemma 8.4. *For a reversible Markov chain*

$$1 - \lambda_2 = \inf_{\substack{f:\Omega\to\mathbb{R},\\ f\neq constant}} \frac{\mathcal{E}(f,f)}{Var_\pi(f)} = \inf_{f\neq constant} \frac{\frac{1}{2}\sum_{x,y\in\Omega}(f(x)-f(y))^2\,\pi(x)\mathsf{P}(x,y)}{\frac{1}{2}\sum_{x,y\in\Omega}(f(x)-f(y))^2\,\pi(x)\pi(y)}$$

Proof. This is not particularly hard to prove but due to time constraints I will skip it. See Lecture 12 of Sinclair's notes [2]. $\square$

The quantity $\mathcal{E}(f,f)$ is known as a *Dirichlet form.* More generally,

$$\mathcal{E}(f,g) = \frac{1}{2}\sum_{x,y\in\Omega} (f(x)-f(y))\,(g(x)-g(y))^2\,\pi(x)\mathsf{P}(x,y)\,.$$

The quantity $\lambda = 1-\lambda_2$ is known as the *spectral gap.*

Example 8.5. Consider the lazy walk on the complete graph K_n. Then $\pi(x) = 1/n$ for all $x \in \Omega$, and $\pi(x)\,\mathsf{P}(x,y) = \frac{1}{n}\,\frac{1}{2(n-1)}$ for all $y \neq x \in \Omega$. It follows that

$$1-\lambda_2 = \frac{\frac{1}{n}\frac{1}{2(n-1)}}{\frac{1}{n}\,\frac{1}{n}}\ \inf_{f\neq constant}\frac{\sum_{x,y\in\Omega}(f(x)-f(y))^2}{\sum_{x,y\in\Omega}(f(x)-f(y))^2} = \frac{n}{2(n-1)}$$

and so $\lambda_2 = 1-\frac{n}{2(n-1)} = \frac{n-2}{2(n-1)}$.

Also, for an arbitrary probability distribution π on K_n let $\mathsf{P}(x,y)=\pi(y)$ (the single-step-to-stationary walk). Then $\mathcal{E}(f,f) = Var_\pi(f)$ and therefore $1-\lambda_2 = 1$ and $\lambda_2 = 0$.

We now show how to bound $1-\lambda_2$ by comparing the Dirichlet forms of two Markov chains, one $\mathcal{M}'$ with known spectral gap and the other $\mathcal{M}$ with unknown gap.

Let $\mathcal{M}$ and $\mathcal{M}'$ be Markov chains on the same state space Ω. For every edge $e=(x,y)\in E'$ (the edge set of $\mathcal{M}'$, i.e. pairs (x,y) with $\mathsf{P}'(x,y)>0$) associate a path in E (the edge set of $\mathcal{M}$) from x to y. Preferably not too many paths should intersect at any given edge.

Theorem 8.6 (Comparison Theorem).

$$\mathcal{E}(f,f) \geq \frac{\mathcal{E}'(f,f)}{\overline{A}} \quad \textit{where} \quad \overline{A} = \max_{e=(a,b)\in E}\frac{1}{\mathsf{Q}(e)}\sum_{\gamma_{xy}\ni e}\mathsf{Q}'(x,y)\,|\gamma_{xy}|$$

and $\mathsf{Q}(e=(a,b)) = \pi(a)\mathsf{P}(a,b)$, *likewise* $\mathsf{Q}'(e=(x,y)) = \pi(x)\mathsf{P}(x,y)$.

Proof. The proof is surprisingly simple. It was first observed by Diaconis and Saloff-Coste [1].

$$\begin{aligned}
\mathcal{E}'(f,f) &= \frac{1}{2}\sum_{x,y\in\Omega}(f(x)-f(y))^2\,\pi'(x)\mathsf{P}'(x,y)\\
&= \frac{1}{2}\sum_{x,y\in\Omega}\left(\sum_{e=(a,b)\in\gamma_{xy}} f(a)-f(b)\right)^2 \pi'(x)\mathsf{P}'(x,y)\\
&\leq \frac{1}{2}\sum_{x,y\in\Omega}\left[\sum_{e=(a,b)\in\gamma_{xy}}(f(a)-f(b))^2\,|\gamma_{xy}|\right]\pi'(x)\mathsf{P}'(x,y)\\
&= \frac{1}{2}\sum_{e=(a,b)\in E}(f(a)-f(b))^2\,\pi(a)\mathsf{P}(a,b)\,\left[\frac{1}{\pi(a)\mathsf{P}(a,b)}\sum_{\gamma_{xy}\ni e}\pi(x)\mathsf{P}(x,y)\,|\gamma_{xy}|\right]\\
&\leq A\,\mathcal{E}(f,f)
\end{aligned}$$

The first inequality followed from Cauchy-Schwartz. $\square$

This theorem can be a bit better understood by letting $\ell = \max|\gamma_{xy}|$ denote the length of the longest path. Then

$$\overline{A} \le A\,\ell \quad \text{where} \quad A = \max_{e=(a,b)\in E} \frac{1}{\mathsf{Q}(e)} \sum_{\gamma_{xy}\ni e} \mathsf{Q}'(x,y)\,.$$

To interpret A, suppose that the Markov chain represents a transportation network. To every pair of vertices (x, y) send a total of $\mathsf{Q}'(x, y)$ units from x to y along the path in $\mathcal{M}$, where each edge $e = (a, b)$ has capacity $\mathsf{Q}(e)$. Then A measures the congestion at the worst place in the network. Now, the known Markov chain $\mathcal{M}'$ sent a total of $\mathsf{Q}'(x, y)$ units directly from x to y along the edge (x, y). If A is close to 1 then this shows that sending these units along $\mathcal{M}$ is not much harder than what happened in $\mathcal{M}'$, so the Dirichlet forms won't differ too much.

Observe that if $\pi = \pi'$ then the variances satisfy $Var_\pi(f) = Var_{\pi'}(f)$.

Corollary 8.7. *If $\mathcal{M}$ and $\mathcal{M}'$ have the same stationary distributions, that is $\pi' = \pi$, then*

$$1 - \lambda_2 \ge \frac{1 - \lambda_2'}{\overline{A}}\,.$$

Next class we will see how to apply this theorem to more examples.

References

[1] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.

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