MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J Lecture N — April 5, SP17

Lecturer: Yury Polyanskiy — Scribe notes by Jennifer Tang

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They are posted to serve class purposes.*

Eigenvalues

1

Recall that $P(x, y)$ is $\mathcal{X} \times \mathcal{X}$ matrix with properties:

1. P is irreducible $\Leftrightarrow$ the eigenvalue $\lambda = 1$ has multiplicity 1.
2. P is aperiodic $\Leftrightarrow$ the spectrum $sp(P)$ does not have $(\gamma_a > 0)$ roots of unity
3. P is reversible $\Rightarrow$ all eigenvalues are real

Definition 1. *Absolute spectral gap* $\gamma_a = 1 - \max_{\lambda \neq 1} |\lambda|$

Theorem 1. $d(t) = (1 - \gamma_a)^{t+o(t)}$

This bound is typically useless. We want more explicit bounds.

Proof. Define $||A||_{\infty\to\infty} = \max_{x\neq 0} \frac{||Ax||_\infty}{||x||_\infty}$ and let $\mathbb{E}_\pi$ = the matrix whose rows are π.

$$d(t) = \frac{1}{2}||P_t - \mathbb{E}_\pi||_{\infty\to\infty}$$
$$d_{TV}(P, Q) = \sup_{||f||_\infty \leq 1} \mathbb{E}_P f - \mathbb{E}_Q f$$

P and $\mathbb{E}_\pi$ commute, that is $P\mathbb{E} = \mathbb{E}P$, and

$$(P - \mathbb{E}_\pi)^t = P_t - \mathbb{E}_\pi$$

which we can show by induction using

$$(P - \mathbb{E}_\pi)^2 = P^2 - \mathbb{E}_\pi P - P\mathbb{E}_\pi + \mathbb{E}_\pi = P^2 - \mathbb{E}_\pi$$

Thus

$$d(t) = \frac{1}{2}||(P - \mathbb{E}_\pi)^2||_{\infty\to\infty}$$

We have for every matrix norm, $\rho(A) = \lim_{t\to\infty} ||A^t||^{1/t}$. Since

$$\rho(P - \mathbb{E}) = 1 - \gamma_a$$

then

$$(d(t))^{1/t} \to \rho(P - \mathbb{E}_\pi) = 1 - \gamma_a$$

□

The issue with Thm 1 is that the constant hides the behavior for small values of t.

Definition 2. *The inner product is defined as* $(f, g) = \sum_x \pi(x)f(x)g(x)$.

1.1 Reversible chains

Proposition 2. *1. P is reversible $\Leftrightarrow (Pf, g) = (f, Pg), \forall f, g$*

2. There exists orthonormal eigenbasis where

$$1 = \lambda_1 \leftrightarrow f_1 = 1\lambda_j \leftrightarrow f_j$$

and

- $(f_i, f_j) = \mathbb{I}\{i = j\}$
- $Pf_i = \lambda f_i$
- *f_i is real valued*
- $||f_i||_2 = 1$

3. $P_t g = \sum_j \lambda_j^t (f_j, g) f_j$

Proof. 1) Since P is reversible, $\sum \pi(x)P(x, y)f(y)g(x) = \sum \pi(y)P(y, s)f(y)g(x)$

2) Use linear algebra

3) $P_t g = \sum \lambda_j^t f_i(f_j, g)$ and let $\delta_{x_0}(x) = \mathbb{I}\{x = x_0\}$, then

$$(\delta_{x_0}, f_j) = \pi(x_0)f_j(x_0)$$
$$\sum_j f_j(x_0)f_j(x) = \frac{1}{\pi(x_0)}\mathbb{I}\{x = x_0\}$$

□

2 χ^2 distance

Definition 3. $\chi^2(P||Q) = \sum_x Q(x)(\frac{P(X)}{Q(X)} - 1)^2 = \sum_x \frac{P^2(x)}{Q(x)} - 1$

Proposition 3. $d_{TV}(P, Q) \leq \frac{1}{2}\sqrt{\chi^2(P||Q)}$

Proof. $\frac{1}{2}\sum_x Q(x)(\frac{P(x)}{Q(x)} - 1) \leq \frac{1}{2}\sqrt{\sum_x Q(x)(\frac{P(x)}{Q(x)} - 1)^2}$ □

Theorem 4.

1. $\chi^2(P_t(x, \cdot)||\pi) = \sum_{j \neq 1} \lambda_j^{2t} f_j^2(x)$
2. *If the markov chain is transitive* $\chi^2(P_t(x, \cdot)||\pi) = \sum_{j \neq 1} \lambda_j^{2t}$
3. $\chi^2(\nu P||\pi) \leq (1 - \gamma_a)^2 \chi^2(\nu||\pi)$
   $d_2(t) = \sup_x \sqrt{\chi^2(P_t(x, \cdot)||\pi)}$
4. $d_2(t) \leq (1 - \gamma_a)^t \sqrt{\frac{1}{\pi_{\min}} - 1}$

Proof. 1) Suppose ν is a distribution on $\mathcal{X}$, $\nu(x) = h(x)\pi(x)$ (ν has density h).

$$\begin{aligned}
\nu P_t(x) &= \sum_y \nu(y) P_t(y, x) \\
&= \sum_y h(y) P_t(x, y) \pi(x) \\
&= \pi(x)(P_t h)(x)
\end{aligned}$$

$P_t(x_0, \cdot)$ has density $P_t h_{x_0}(\cdot)$

$$\begin{aligned}
h_{x_0}(x) &= \frac{1}{\pi(X_0)} \mathbb{I}\{x = x_0\} = \sum_j f_j(x_0) f_j(x) \\
P_t h_{x_0} &= \sum_{j=1}^n f_j(x_0) \lambda_j^t f_j(x)
\end{aligned}$$

$$\begin{aligned}
\chi^2(P_t(x_0, \cdot)||\pi) &= ||P_t h_{x_0} - 1||_2^2 = (P_t h_{x_0}, P_t h_{x_0}) - 1 \\
&= \sum_{j \neq 1} \lambda_j^{2t} f_i^2(x_0)
\end{aligned}$$

Definition 4. *Permutation* $g : \mathcal{X} \to \mathcal{X}$ *is a* sym *of* P *if* $P(x, y) = P(g(x), g(y))$

P is transitive if $\forall x \neq y, \exists g-$ sym such that $g(x) = y$ ("All points are born equal").

2) $\mathbb{E}_{x_0 \sim \pi} \Rightarrow \mathbb{E} f_j(x_0)^2 = 1$ by orthonormal basis

3) Let h be density of ν

$$\frac{\chi^2(\nu P||\pi)}{\chi^2(\nu||\pi)} = \frac{\sum_{j\neq 1} \lambda_j^2 (h, f_j)^2}{\sum_{j \neq 1} (h, f_j)^2} \leq (1-\gamma_a)^2$$

Taking $\sup_\nu$ will give $(1-\gamma_a)^2$. Inequality is sharp when ν is close to π.

4) Notice that $\chi^2(\delta_{x_0}||\pi) = \frac{1}{\pi(x_0)} - 1$. Then iterate 3. □

Theorem 5. $(\ln \frac{1}{\varepsilon}) \cdot (\frac{1}{\gamma_a} - 1) \leq t_{mix}(\varepsilon) \leq \frac{1}{\gamma_a}(\ln \frac{1}{2\varepsilon} + \frac{1}{2} \ln \frac{1}{\pi_{\min}})$

Proof. Upper bound:

$$d_{TV}(t) \leq \frac{1}{2} d_2(t) \leq (1-\gamma_a)^t \sqrt{\frac{1}{\pi_{\min}}} = \varepsilon$$
$$\Rightarrow t = \frac{\ln \frac{1}{2\varepsilon} + \ln \frac{1}{\sqrt{\pi_{\min}}}}{\ln \frac{1}{1-\gamma_a}}$$

Lower bound: If $d(t) \leq \varepsilon$, then $\forall f$ such that $\mathbb{E}_\pi f = 0$

$$||P_t f||_\infty \leq \varepsilon ||f||_\infty$$

and

$$\mathbb{E}_P f - \mathbb{E}_Q f \leq d_{TV} ||f||_\infty$$

Take $f = f_j$ such that $|\lambda_j| = 1 - \gamma_a$,

$$P_t f_j = \lambda_j^t f_j \Rightarrow \lambda_j^t ||f_j||_\infty \leq \varepsilon ||f_j||_\infty$$

$$t \ln(1-\gamma_a) \leq \ln \varepsilon$$

$$t_{mix}(\varepsilon) \geq \frac{\ln \frac{1}{\varepsilon}}{\ln \frac{1}{1-\gamma_a}} \geq (\frac{1}{\gamma_a} - 1) \ln \frac{1}{\varepsilon}$$

□

Definition 5. $t_{rel} = \frac{1}{\gamma_a}$

3 Mu Fa Chen

Theorem 6 (Mu Fa Chen)**.** *Suppose that for some metric ρ,*

$$W_\rho, (P(x,\cdot),, P(y,\cdot) \leq e^{-\alpha}\rho(x,y)$$

then

$$\gamma_{\geq} 1 - e^{-\alpha}$$

Shows W_ρ contradiction implies absolute spectral gap.

Proof. Note that $\forall f$, $||Pf||_{Lip} \leq ||f||_{Lip}\rho(x,y)$. We want to prove

$$Pf(x) - Pf(y) \leq e^{-\alpha}||f||_{Lip}\rho(x,y)$$

Let $X \sim P(x,\cdot)$ and $Y \sim P(y,\cdot)$.

$$\begin{aligned}
\mathbb{E}f(X_1) - \mathbb{E}f(Y_1) &= \mathbb{E}[f(X_1) - f(Y_1)] \\
&\leq \mathbb{E}[||f||_{Lip}\rho(X_1, Y_1)] \\
&= ||f||_{Lip}\mathbb{E}(\rho(X_1, Y_1)) \\
&\leq e^{-\alpha}\rho(x,y)||f||_{Lip}
\end{aligned}$$

Let f in above be eigenbasis $f_j^* \leftrightarrow |\lambda_j^*| = 1 - \gamma_a$. □

4 Examples

1. Lazy random walk (LRW) on n-cycle C_n. (Let n be even.)

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 0 & \dots & 0 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \dots & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{4} & 0 & 0 & \dots & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Eigenfunctions are $\cos\frac{2\pi x(j-1)}{n}$ and $\sin\frac{2\pi x(j-1)}{n}$ for $j = 1, \dots, \frac{n}{2}+1$. The eigenvalues are $\frac{1}{2}(1 + \cos\frac{2\pi(j-1)}{n})$.

$$\gamma_a = 1 - \lambda_2 \approx \frac{c}{n^2}$$

2. Lazy random walk on path P_{k-1}.

   Project Markov chain of C_n where $n = 2k - 2$.

$$\begin{aligned} sp(\text{proj}P) &\subset sp(P) \\ \gamma_a(\text{proj}P) &\geq \gamma_a(P) \\ \gamma_a(P_{k-1}) &\geq \frac{c}{k^2} \end{aligned}$$

3. Product Chains. Let P be a Markov Chain on $\mathcal{X}$, then we define a new Markov Chain on $\mathcal{X}^n$

$$P^{(n)} = \frac{1}{n}(P \otimes I \otimes \cdots \otimes I + I \otimes P \otimes \cdots \otimes I + \dots)$$

   Theorem 7. $sp(P^{(n)}) = \{\frac{1}{n}\sum_{i=1}^{n} \mu_i : \mu_i \in sp(P)\}$

$$\gamma_a(P^{(n)}) = \frac{1}{n}\gamma_a(P)$$

   Proof. Eigenfunctions of $P^{(n)}$ are $f_{i1}(x_1), \dots, f_{in}(x_n)$. □

   LRW on H_n (hypercube) is a n-product of LRW on P_1 (path) where

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

   and $sp(P) = \{0, 1\}$.

$$sp(\text{LRW } H_n) = \begin{cases} \frac{0}{n} & \text{multiplicity } 1 \\ \frac{1}{n} & \text{multiplicity } n \\ \vdots & \vdots \\ \frac{k}{n} & \text{multiplicity } \binom{n}{k} \\ \vdots & \vdots \end{cases}$$

Consequence: $\gamma_a(\text{LRW } H_n) = \frac{1}{n}$.

$$\begin{aligned}
d(t) &\le \frac{1}{2} d_2(t) \\
&= \frac{1}{2}\sqrt{\sum_{j\neq 1} \lambda_j^{2t}} \\
&= \frac{1}{2}\sqrt{\sum_{k=0}^{n-1} \binom{n}{k}\left(\frac{k}{n}\right)^{2t}} \\
&= \frac{1}{2}\sqrt{\sum_{m=1}^{n-1} \binom{n}{m}\left(1-\frac{m}{n}\right)^{2t}} \\
\frac{1}{2} &\le \sqrt{\sum_{m=1}^{\infty} e^{-m(\frac{2t}{n} - \ln n)}} \\
&\le \varepsilon
\end{aligned}$$

if $t = \frac{1}{2} n \log n + c_\varepsilon n$

	LRW Cycle	LRW Hypercube
Upper bound	$t_{mix}(\varepsilon) \le c_\varepsilon n^2$	$t_{mix}(\varepsilon) \le n \log n + c_\varepsilon n$
Lower bound	$t_{mix}(\varepsilon) \ge c_\varepsilon n^2$	$t_{mix}(\varepsilon) \ge \frac{1}{2} n \log n + c_\varepsilon n$
	$t_{rel} = cn^2$	$t_{rel} = n$

All Markovian couplings only achieve $n \log n$.

For the hypercube, $d(t)$ has a cutoff point. The rule of thumb (not entirely true) is if t_{rel} is not tight, there is a cutoff.