Lecture 10: Mixing time and eigenvalues

Nicolás Rivera John Sylvester Luca Zanetti Thomas Sauerwald

Lent 2020

UNIVERSITY OF CAMBRIDGE

Plan

In the last lecture:

- a review of linear algebra
- reversible Markov chains

Today:

- relate mixing time to eigenvalues of reversible chains
- show how to obtain bounds on eigenvalues for some family of graphs

Convergence to stationarity

Mixing time (revisited)

Recall the definition of mixing time: $\tau(\epsilon) = \min \left\{ t\colon\ \max_x \left\| P_x^t - \pi \right\|_{TV} \leq \epsilon \right\}$, where

$$\left\| P_x^t - \pi \right\|_{TV} = \frac{1}{2} \sum_y \left| P^t(x,y) - \pi(y) \right| = \frac{1}{2} \sum_y \left| \frac{P^t(x,y)}{\pi(y)} - 1 \right| \pi(y)$$

$$= \frac{1}{2} \left\| \frac{P^t(x,\cdot)}{\pi} - 1 \right\|_{1,\pi}.$$

This is also called the ℓ_1-mixing time.

When dealing with spectral properties of P, it is actually easier to consider a stronger notion of mixing: the ℓ_2-mixing time:

$$\tau_2(\epsilon) = \min \left\{ t\colon\ \max_x \left\| \frac{P^t(x,\cdot)}{\pi} - 1 \right\|_{2,\pi} \leq \epsilon \right\}$$

where $\left\| \frac{P_x^t}{\pi} - 1 \right\|_{2,\pi} = \sqrt{\sum_y \left(\frac{P^t(x,y)}{\pi(y)} - 1 \right)^2 \pi(y)} = \sqrt{\mathrm{Var}_\pi \left(\frac{P^t(x,\cdot)}{\pi} \right)}$.

It holds that: $\tau_2(2\epsilon) \geq \tau(\epsilon)$.

Addendum: comparison between mixing times

Lemma

Let $\tau_1(\epsilon) = \min\{t\colon \max_x \|P^t(x,\cdot)-\pi\|_{TV} \le \epsilon\}$ (ℓ_1 mixing time) and $\tau_2(\epsilon) = \min\left\{t\colon \max_x \left\|\frac{P^t(x,\cdot)}{\pi} - 1\right\|_{2,\pi} \le \epsilon\right\}$ (ℓ_2 mixing time). Then,

$$\tau_2(2\epsilon) \ge \tau(\epsilon).$$

Proof: We just need to show that, for any $\epsilon > 0$,

$$\|P^t(x,\cdot)-\pi\|_{TV} > \epsilon \implies \left\|\frac{P^t(x,\cdot)}{\pi} - 1\right\|_{2,\pi} > 2\epsilon.$$

Assume (LHS). Notice that,

$$\epsilon < \left\|P^t(x,\cdot)-\pi\right\|_{TV} = \frac{1}{2}\sum_y \left|P^t(x,y)-\pi(y)\right| = \frac{1}{2}\sum_y \left|\frac{P^t(x,y)}{\pi(y)} - 1\right|\pi(y)$$

$$\le \frac{1}{2}\sqrt{\sum_y \left(\frac{P^t(x,y)}{\pi(y)} - 1\right)^2 \pi(y)} = \frac{1}{2}\left\|\frac{P^t(x,\cdot)}{\pi} - 1\right\|_{2,\pi}$$

(The inequality follows from $\mathbf{E}[X^2] - (\mathbf{E}[X])^2 \ge 0$ for any discrete r.v. X.)
Therefore,

$$\left\|\frac{P^t(x,\cdot)}{\pi} - 1\right\|_{2,\pi} > 2\epsilon.$$

□

Mixing time and eigenvalues

Let P be a transition matrix of a reversible Markov chain with stationary distribution π and eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$.
Suppose $\lambda = \max_{i \neq 1} |\lambda_i| < 1$.

$\lambda_i^t \to 0$ as $t \to \infty$

Recall the spectral decomposition

$$\frac{P^t(x, \cdot)}{\pi} = \sum_{i=1}^{n} \lambda_i^t f_i(x) f_i = \mathbf{1} + \sum_{i=2}^{n} \lambda_i^t f_i(x) f_i.$$

$f_i = \mathbf{1}$ and $\lambda_1 = 1$

Theorem

For any $\epsilon \in (0, 1)$,

$$\left(\frac{1}{1-\lambda} - 1\right) \log\left(\frac{1}{\epsilon}\right) \leq \tau_2(\epsilon) \leq \log\left(\frac{1}{\epsilon\sqrt{\pi_*}}\right) \frac{1}{1-\lambda},$$

where $\pi_* \triangleq \min_x \pi(x)$.

Mixing time and eigenvalues (2)

Theorem

Let P be the transition matrix of a reversible Markov chain with stationary distribution π and $\lambda = \max_{i \neq 1} |\lambda_i| < 1$. Then, for any $\epsilon \in (0,1)$,

$$\tau_2(\epsilon) \leq \log\left(\frac{1}{\epsilon\sqrt{\pi_*}}\right)\frac{1}{1-\lambda},$$

Proof: From the spectral decomposition:

$$\frac{P^t(x,\cdot)}{\pi} = \sum_{i=1}^n \lambda_i^t f_i(x) f_i = 1 + \sum_{i=2}^n \lambda_i^t f_i(x) f_i.$$

$$\left\| \frac{P_x^t}{\pi} - 1 \right\|_{2,\pi}^2 = \left\| \sum_{i=2}^n \lambda_i^t f_i(x) f_i \right\|_{2,\pi}^2 \leq \lambda^{2t} \left\| \sum_{i=2}^n f_i(x) f_i \right\|_{2,\pi}^2$$

Now notice that $\frac{1_x}{\pi} = \sum_{i=1}^n \langle \frac{1_x}{\pi}, f_i \rangle_\pi f_i = \sum_{i=1}^n f_i(x) f_i$. Hence,

$$\left\| \frac{P_x^t}{\pi} - 1 \right\|_{2,\pi}^2 \leq \lambda^{2t} \left\| \frac{1_x}{\pi} \right\|_{2,\pi}^2 = \lambda^{2t} \cdot \frac{1}{\pi(x)}$$

Finally, take t such that $\frac{\lambda^{2t}}{\pi(x)} \leq \epsilon^2$.

□

How to obtain bounds on the spectral gap

Lazy random walks

From now on we will focus on lazy random walks:

- a particle moves on an undirected graph $G = (V, E)$
- at each time-step, it can either stay with probability $1/2$ or move to an adjacent vertex picked uniformly at random.

Let P be the transition matrix for the lazy walk, and P' for the simple walk on the same graph G. Then,

$$P = \frac{1}{2}(\mathbb{I} + P')$$

Therefore $\lambda_n \geq 0$ and $\lambda = \lambda_2$.

Moreover, $\pi(x) = \frac{d(x)}{2|E|}$ and $\pi_* = \Omega(n^{-2})$. Therefore,

$$\tau(\epsilon) = O\left(\frac{\log(n/\epsilon)}{1 - \lambda_2}\right).$$

Courant-Fischer min-max formula

Courant-Fischer formula

Let M be an n by n self-adjoint matrix (with respect to $\langle\cdot,\cdot\rangle_\pi$) with eigenvalues $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ and corresponding orthonormal eigenvectors $f_1, \ldots, f_n$. Then,

$$\mu_1 = \min_{f \in \mathbb{R}^n \setminus \{\underline{0}\}} \frac{\langle f, Mf\rangle_\pi}{\|f\|_{2,\pi}^2} \qquad \mu_2 = \min_{\substack{f \in \mathbb{R}^n \setminus \{\underline{0}\} \\ f \perp f_1}} \frac{\langle f, Mf\rangle_\pi}{\|f\|_{2,\pi}^2}$$

The eigenvectors corresponding to λ_1 and λ_2 minimise such expressions.

Let λ_2 be the second largest eigenvalue of the transition matrix of a lazy random walk on a d-regular graph $G = (V, E)$. Then,

$$1 - \lambda_2 = \min_{\substack{f \in \mathbb{R}^n \setminus \{\underline{0}\} \\ f \perp 1}} \frac{\langle f, (I-P)f\rangle}{\|f\|_2^2} = \min_{\substack{f \in \mathbb{R}^n \setminus \{\underline{0}\} \\ f \perp 1 \\ \|f\|_2 = 1}} \langle f, (I-P)f\rangle.$$

Variational characterisation of λ_2

Lemma

Let P be the transition matrix of a lazy random walk on a d-regular graph $G = (V, E)$. Then,

$$1 - \lambda_2 = \min_{\substack{f \perp 1 \\ \|f\|_2 = 1}} 1/(2d) \cdot \sum_{\{x,y\} \in E} (f(x) - f(y))^2$$

Moreover, $f^\star$ minimising the expression above is an eigenvector of P corresponding to λ_2.

Proof: Let A be the adjacency matrix of G.

$$\langle f, (I - P)f \rangle = f^T \left(\mathbb{I} - \left(\frac{1}{2}\mathbb{I} + \frac{1}{2d}A \right) \right) f = \frac{1}{2} \left(f^T f - \frac{1}{d} f^T A f \right)$$

$$= \frac{1}{2} \left(\sum_{u \in V} f(u)^2 - \frac{2}{d} \sum_{\{u,v\} \in E} f(u) f(v) \right)$$

$$= \frac{1}{2d} \sum_{\{u,v\} \in E} (f(u)^2 + f(v)^2 - 2f(u)f(v)) = \sum_{\{u,v\} \in E} \frac{(f(u) - f(v))^2}{2d}.$$

The Lemma follows from the Courant-Fischer formula. □

Bonus material (not seen in class)

Mixing time on regular graphs (1/2)

Lemma

Let $G = (V, E)$ be a d-regular graph of n vertices, with diameter δ. Then, a lazy random walk in G has mixing time $\tau(\epsilon) = O(d\delta n \log(n/\epsilon))$.

Proof: By the previous lemma,

$$1 - \lambda_2 = \min_{\substack{f \perp 1 \\ \|f\|_2 = 1}} 1/(2d) \cdot \sum_{\{x,y\} \in E} (f(x) - f(y))^2$$

Assume $\sum_x f(x)^2 = 1$. Then, there exists $x \in V$ such that $|f(x)| \geq 1/\sqrt{n}$.

$f \perp 1$ implies $\sum_u f_u = 0$. Hence, there exists $y \in V$ such that $\text{sign}(f(y)) \neq \text{sign}(f(x))$. Therefore, $(f(x) - f(y))^2 \geq 1/n$.

Since G is connected, there exists a path $x = u_0, u_1, \ldots, u_\ell = y$ such that $\{u_i, u_{i+1}\} \in E$ and $\ell \leq \delta$. Then,

$$(f(x) - f(y))^2 = (f_{u_0} - f_{u_1} + f_{u_1} - f_{u_2} + \cdots + f_{u_{\ell-1}} + f_{u_\ell})^2 \leq \ell\delta \sum_{i=0}^{\ell-1} (f_{u_i} - f_{u_{i-1}})^2 \quad (1)$$

and

$$1 - \lambda_2 \geq 1/(2d) \cdot \sum_{i=0}^{\ell-1} (f_{u_i} - f_{u_{i-1}})^2 \overset{(1)}{\geq} 1/(2d\delta) \cdot (f(x) - f(y))^2 \geq 1/(2d\delta n)$$

□

Mixing time on regular graphs (2/2)

Claim

Let $G = (V, E)$ be a regular graph of n vertices with degree d, and diameter δ. Then, $d \cdot \delta = O(n)$

Theorem

Let $G = (V, E)$ be a regular graph of n vertices. Then, a lazy random walk in G has mixing time $\tau(\epsilon) = O(n^2 \log(n/\epsilon))$.

Is this result tight?

- Almost. The best possible bound for general regular graphs is $\tau(1/10) = O(n^2)$.
- The cycle, in fact, has $\Theta(n^2)$ mixing time.
- For general graphs, mixing can take up to $O(n^3)$ steps.