Lecture 10: Mixing time and eigenvalues

John Sylvester Nicolás Rivera Luca Zanetti Thomas Sauerwald

Lent 2019

UNIVERSITY OF CAMBRIDGE

Plan

In the last lecture:

- a review of linear algebra
- introduced reversible Markov chains

Today:

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

Recap

Let $|\Omega| = n$, $\quad f, g\colon \Omega \to \mathbb{R}$, $\quad \pi\colon \Omega \to \mathbb{R}_+$

- $\langle f, g \rangle_\pi = \sum_{x \in \Omega} f(x) g(x) \pi(x)$
- $\|f\|_{p,\pi} = \left(\sum_{x \in \Omega} |f(x)|^p \pi(x) \right)^{1/p}$

Let M be a $|\Omega| \times |\Omega|$ matrix

- M is self-adjoint if $\langle Mf, g \rangle_\pi = \langle f, Mg \rangle_\pi$ for any f, g
- If M is self-adjoint then it has n real eigenvalues with corresponding orthonormal eigenvectors

Let P be the transition matrix of a Markov chain and assume P is self-adjoint. Then,

- The chain is reversible (i.e., it's a random walk on an undirected graph and $\pi(x)P(x,y) = \pi(y)P(y,x)$)
- It has eigenvalues $1 = \lambda_1 \geq \cdots \geq \lambda_n \geq -1$
- If $\lambda = \max_{i \neq 1} |\lambda_i| < 1$, the chain is irreducible and aperiodic (i.e., it converges to stationary)

Examples at the visualiser

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

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

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:

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

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

It holds that: $\tau(2\epsilon) \leq \tau_2(\epsilon) = O(\tau(2\epsilon) \log(1/\pi_*))$, where $\pi_* \triangleq \min_x \pi(x)$.

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

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

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

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

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

Variational characterisation of λ_2

Lemma

Let P be the transition matrix of a reversible Markov chain with stationary distribution π and eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Then,

$$1 - \lambda_2 = \min_{\mathbf{0} \neq f \perp \mathbf{1}} \frac{\sum_{x,y} (f(x) - f(y))^2 P(x,y) \pi(x)}{2 \|f\|_{2,\pi}^2}$$

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

It immediately follows from the lemma that $1 - \lambda_2 \geq 0 \implies \lambda_2 \leq 1$

If P represents a lazy random walk on an undirected graph $G = (V, E)$,

$$1 - \lambda_2 = \min_{\mathbf{0} \neq f \perp \pi} \frac{\sum_{\{x,y\} \in E} (f(x) - f(y))^2}{2 \sum_x d(x) f(x)^2}$$

where $d(x)$ is the degree of x.

Mixing time on regular graphs (1/2)

Lemma

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

Proof: Since G is regular, $\pi = \mathbf{1}/n$, and

$$1 - \lambda_2 = \min_{\mathbf{0} \neq f \perp \mathbf{1}} \frac{\sum_{\{x,y\} \in E} (f(x) - f(y))^2}{2d \sum_x f(x)^2}$$

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 $\sum_x f(x) = 0$. Hence, there exists $y \in V$ such that $\operatorname{sign}(f(y)) \neq \operatorname{sign}(f(x))$. 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,

$$\begin{aligned} (f(x) - f(y))^2 &= (f(u_0) - f(u_1) + f(u_1) - f(u_2) + \cdots + f(u_{\ell-1}) - f(u_\ell))^2 \\ &\leq \delta \sum_{i=0}^{\ell-1} (f(u_i) - f(u_{i+1}))^2 \qquad (1) \end{aligned}$$

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

□

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