

## Lecture 9: Eigenvalues and mixing time

John Sylvester   Nicolás Rivera   [Luca Zanetti](#)   Thomas Sauerwald

Lent 2019



UNIVERSITY OF  
CAMBRIDGE

## Plan for the next four lectures

---

The topic of the next four lectures is “Probability meets Linear Algebra”

*Linear algebra is the study of linear maps on finite-dimensional vector spaces (matrices)*

### Why:

- Discrete probability distributions can be represented by vectors and Markov chains by matrices
- Data has become large: we need to be able to think in higher dimension.
- Linear algebra primitives can be exploited to design fast algorithms

### Our plan:

- Review (briefly) main concepts of linear algebra
- Borrow tools from linear algebra to analyse Markov chains
- Connect Markov chains to the problem of graph clustering



In previous lectures we have seen that:

- Any aperiodic and irreducible finite Markov Chain (e.g., lazy random walks) converges to a unique stationary distribution
- The mixing time captures how long it takes for convergence to happen
- Bounds on mixing using the coupling method

**Q?** Can we find a parameter, which is easy to compute, and from which we can derive good bounds on the mixing time?

**Yes!** It turns out the eigenvalues of the transition matrix govern convergence of the chain



## A brief linear algebra review



## Vectors and inner products

---

$\Omega$ : finite dimensional state space ( $|\Omega| = n$ )

We will represent a vector  $f \in \mathbb{R}^n$  as a function  $f: \Omega \rightarrow \mathbb{R}$

Let  $f, g: \Omega \rightarrow \mathbb{R}$  and  $\pi: \Omega \rightarrow \mathbb{R}_+$ . We define an inner product  $\langle \cdot, \cdot \rangle_\pi$  as

$$\langle f, g \rangle_\pi = \sum_{x \in \Omega} f(x)g(x)\pi(x)$$

For example,  $\langle \cdot, \cdot \rangle_1$  is the usual inner-product in  $\mathbb{R}^n$ .

If  $\pi$  is a probability measure,  $\langle f, g \rangle_\pi = E_\pi(fg)$ .

After fixing  $\pi$ , we denote with  $\ell_2(\Omega, \pi) = \{f: \Omega \rightarrow \mathbb{R}\}$  the set of all functions from  $\Omega$  to  $\mathbb{R}$  equipped with  $\langle \cdot, \cdot \rangle_\pi$ .

We say  $f, g \in \ell_2(\Omega, \pi)$  are orthogonal ( $f \perp g$ ) if  $\langle f, g \rangle_\pi = 0$



## Norms and distances

---

Let  $f \in \ell_2(\Omega, \pi)$  and  $p > 0$ . The  $\ell_p(\pi)$  norm of  $f$  is:

$$\|f\|_{p,\pi} = \left( \sum_{x \in \Omega} |f(x)|^p \pi(x) \right)^{\frac{1}{p}}.$$

In particular,  $\|f\|_{2,\pi} = \sqrt{\langle f, f \rangle_\pi}$ .

We say that  $f_1, \dots, f_k$  are **orthonormal** if, for any  $i \neq j$ ,  $\|f_i\|_{2,\pi} = 1$  and  $f_i \perp f_j$ .

Let  $f, g \in \ell_2(\Omega, \pi)$ . The  $\ell_p(\pi)$ -distance between  $f$  and  $g$  is

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

**Pythagorean theorem:**  $f \perp g \implies \|f + g\|_{2,\pi}^2 = \|f\|_{2,\pi}^2 + \|g\|_{2,\pi}^2$



## Matrices as linear operators

A  $|\Omega| \times |\Omega|$  matrix  $M$  is just a representation of a **linear** operator  $M: \ell_2(\Omega, \pi) \rightarrow \ell_2(\Omega, \pi)$ .

$$M(\alpha f + \beta g) = \alpha Mf + \beta Mg$$

For any  $f \in \ell_2(\Omega, \pi)$ ,  $x \in \Omega$ ,

$$Mf(x) = \sum_{y \in \Omega} M(x, y)f(y)$$

$$fM(x) = \sum_{y \in \Omega} M(y, x)f(y)$$

Definition

$M: \ell_2(\Omega, \pi) \rightarrow \ell_2(\Omega, \pi)$  is called **self-adjoint** if, for any  $f, g \in \ell_2(\Omega, \pi)$ ,

$$\langle Mf, g \rangle_\pi = \langle f, Mg \rangle_\pi.$$

Note: if  $\pi = \mathbf{1}$ , which implies  $\langle \cdot, \cdot \rangle_\pi = \langle \cdot, \cdot \rangle$ , we recover the usual definition of symmetric matrix.



Definition

$f \in \ell_2(\Omega, \pi)$  is a left (right) **eigenvector** with **eigenvalue**  $\lambda \in \mathbb{C}$  for  $M: \ell_2(\Omega, \pi) \rightarrow \ell_2(\Omega, \pi)$  if

$$f M = \lambda f \quad (Mf = \lambda f)$$

Examples:

- Let  $P$  be the transition matrix of a finite Markov chain on  $\Omega$  with stationary distribution  $\pi$ :

$$P\mathbf{1} = \mathbf{1} \quad \pi P = \pi$$

- Suppose  $f$  is a left (right) eigenvector with eigenvalue  $\lambda$  for  $M$ . Then, for any  $k \geq 0$ ,

$$f M^k = \lambda^k f \quad (M^k f = \lambda^k f)$$





## The spectral theorem

Recall:  $M: \ell_2(\Omega, \pi) \rightarrow \ell_2(\Omega, \pi)$  is self-adjoint if, for any  $f, g \in \ell_2(\Omega, \pi)$ ,  
 $\langle Mf, g \rangle_\pi = \langle f, Mg \rangle_\pi$ .

The spectral theorem

Let  $\mathbf{M}: \ell_2(\Omega, \pi) \rightarrow \ell_2(\Omega, \pi)$  be self-adjoint and  $|\Omega| = n$ . Then,  $M$  has  $n$  real eigenvalues  $\lambda_1 \geq \dots \geq \lambda_n$  and a corresponding set of  $n$  orthonormal right eigenvectors  $\{f_1, \dots, f_n\}$ .

Consequences:

- We can express any  $g \in \ell_2(\Omega, \pi)$  as  $g = \sum_{i=1}^n \langle g, f_i \rangle_\pi f_i$ .
- $M$  can be decomposed as:  $\frac{M(x, \cdot)}{\pi} = \sum_{i=1}^n \lambda_i f_i(x) f_i$

Proof: 
$$\frac{M(x, \cdot)}{\pi} = \sum_{i=1}^n \left\langle \frac{M(x, \cdot)}{\pi}, f_i \right\rangle_\pi f_i$$
$$\left\langle \frac{M(x, \cdot)}{\pi}, f_i \right\rangle_\pi = \sum_y \frac{M(x, y) f_i(y) \pi(y)}{\pi(y)} = Mf_i(x) = \lambda_i f_i(x)$$

□



## Back to Markov chains



## Reversible Markov chains

Definition

A Markov chain on  $\Omega$  with transition matrix  $P$  and stationary distribution  $\pi$  is called **reversible** if, for any  $x, y \in \Omega$ ,

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

- Any Markov chain can be described as random walk on a weighted directed graph.
- Reversible Markov chains are equivalent to random walks on weighted undirected graphs.

Lemma

A Markov chain with transition matrix  $P$  is **reversible** if and only if  $P$  is **self-adjoint** w.r.t. its stationary distribution  $\pi$ .

Proof:

$$\begin{aligned}(\implies) \quad \langle Pf, g \rangle_\pi &= \sum_{x, y \in \Omega} P(x, y) f(y) g(x) \pi(x) \stackrel{\text{REV.}}{=} \sum_{x, y \in \Omega} P(y, x) f(y) g(x) \pi(y) \\ &= \langle f, Pg \rangle_\pi\end{aligned}$$

$$(\impliedby) \quad \pi(y)P(y, x) = \langle P1_x, 1_y \rangle_\pi \stackrel{\text{S.A.}}{=} \langle 1_x, P1_y \rangle_\pi = \pi(x)P(x, y). \quad \square$$



## Basic facts about eigenvalues

---

Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  be the eigenvalues of the transition matrix  $P$  of a reversible Markov chain.

- $\lambda_1 = 1$
- $\lambda_2 < 1$  if and only if the chain is irreducible
- $\lambda_n \geq -1$
- $\lambda_n > -1$  if and only if the chain is aperiodic

This implies the fundamental theorem of finite Markov chains (i.e., convergence to stationarity) holds whenever

$$\lambda \triangleq \max_{i \neq 1} |\lambda_i| < 1.$$

You will be asked to prove these facts in the exercises.

