Lecture 9: Linear algebra review and Markov chains

Nicolás Rivera John Sylvester Luca Zanetti Thomas Sauerwald

Lent 2020

UNIVERSITY OF CAMBRIDGE

Plan for the next four lectures

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

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

Today's lecture

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

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

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

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

Let $f, g\colon \Omega \to \mathbb{R}$ and $\pi\colon \Omega \to \mathbb{R}_+$. 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, $\langle \cdot, \cdot \rangle_{\mathbf{1}}$ is the usual inner-product in $\mathbb{R}^n$.
If π is a probability measure, $\langle f, g \rangle_\pi = \mathrm{E}_\pi(fg)$.

After fixing π, we denote with $\ell_2(\Omega, \pi) = \{f\colon \Omega \to \mathbb{R}\}$ the set of all functions from Ω 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 $\ell_2(\pi)$ norm of f is:

$$\|f\|_{2,\pi} = \sqrt{\sum_{x \in \Omega} |f(x)|^2 \pi(x)} = \sqrt{\langle f, f \rangle_\pi}.$$

We say that $f_1, \ldots, f_k$ 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 $\ell_2(\pi)$-distance between f and g is

$$\|f - g\|_{2,\pi} = \sqrt{\sum_{x \in \Omega} |f(x) - g(x)|^2 \pi(x)}.$$

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

Cauchy-Schwarz inequality: $\langle f, g \rangle_\pi \leq \|f\|_{2,\pi} \cdot \|g\|_{2,\pi}$

Matrices as linear operators

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

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

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

$M\colon \ell_2(\Omega, \pi) \to \ell_2(\Omega, \pi)$ 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.

Eigenvalues and eigenvectors

Definition

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

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

Examples:

- Let P be the transition matrix of a finite Markov chain on Ω with stationary distribution π:
$$P\mathbf{1} = \mathbf{1} \qquad \pi P = \pi$$
- Suppose f is a left (right) eigenvector with eigenvalue λ for M. Then, for any $k \geq 0$,
$$f M^k = \lambda^k f \qquad \left(M^k f = \lambda^k f\right)$$

The spectral theorem

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

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

$$\left\langle \frac{M(x,\cdot)}{\pi}, f_i \right\rangle_\pi = \sum_y \frac{M(x,y) f_i(y) \pi(y)}{\pi(y)} = Mf_i(x) = \lambda_i f_i(x) \qquad \square$$

Back to Markov chains

Reversible Markov chains

Definition

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

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

- 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 π.

Proof:

$$(\Longrightarrow) \quad \langle Pf, g\rangle_\pi = \sum_{x,y\in\Omega} P(x,y)f(y)g(x)\pi(x) \stackrel{\text{REV.}}{=} \sum_{x,y\in\Omega} P(y,x)f(y)g(x)\pi(y)$$

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

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

Basic facts about eigenvalues

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 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

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

We will prove some of these facts in the next few lectures.