Session 6: Eigen values & Markov Chains

Optimization and Computational Linear Algebra for Data Science

Marylou Gabrié (based on material by Léo Miolane)

Contents

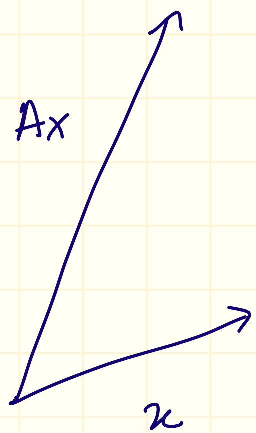
1. Eigenvalues & eigenvectors
    - 1.1 Definition
    - 1.2 Useful facts
    - 1.2 Eigenspaces
    - 1.3 Spectrum
2. Markov chains
    - 2.1 Stochastic matrices and key equation
    - 2.2 Perron-Frobenius Theorem
    - 2.3 Application: PageRank
3. The spectral theorem

1. Eigenvalues & eigenvectors

Introduction

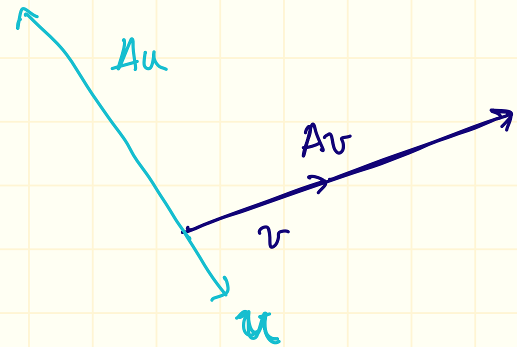
let $A \in \mathbb{R}^{n\times n}$

Generic case



(Ax, x)

linearly independant



(Av, v) linearly dependent

(Au, u) linearly dependent

u and v are "eigen vectors" of A.

1.1 Definition

$v \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

Definition

Let $A \in \mathbb{R}^{n \times n}$. A **non-zero** vector $v \in \mathbb{R}^n$ is said to be an *eigenvector* of A is there exists $\lambda \in \mathbb{R}$ such that

$$Av = \lambda v.$$

The scalar λ is called the eigenvalue (of A) associated to v.

Examples: I_d? matrix A with $\ker(A) \neq \{0\}$?

(•) $A = Id_n$, let $x \in \mathbb{R}^n$, $Id_n x = x$, x is an eigenvector of Id_n and its eigenvalue is 1

(•) A such that $\mathrm{Ker}(A) \neq \{0\}$

let $x \in \mathrm{Ker}(A)$, $Ax = 0 = 0x$ then x is an eigenvector of A and its eigenvalue is 0.

Example: diagonal matrices

let $D \in \mathbb{R}^{n \times n}$ with $D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$

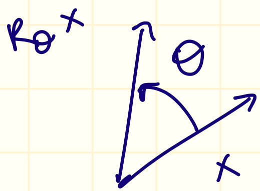
$$De_1 = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \lambda_1 e_1$$

e_1 is an eigenvector of D with associated eigenvalue λ_1.

e_i is an eigenvector of D with associated eigenvalue λ_i.

Matrix with no eigenvalues/vectors

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad \text{for } \theta \in (0, \pi)$$

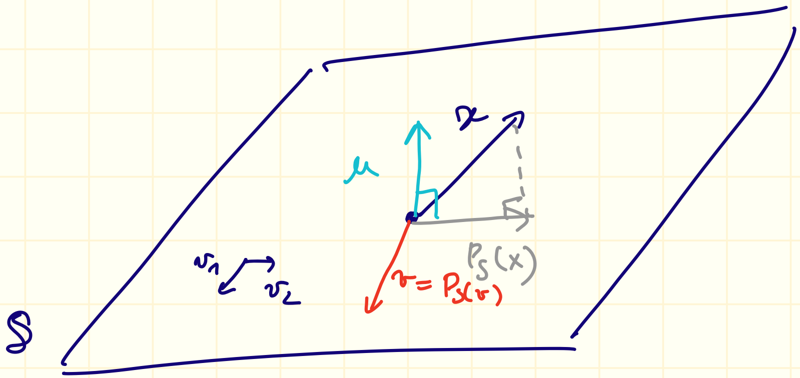


for all $\begin{cases} x \in \mathbb{R}^2 \\ x \neq 0 \end{cases}$, for any $\lambda \in \mathbb{R}$

$$R_\theta x \neq \lambda x.$$

R_θ does not have any real eigenvalues.

Example: orthogonal projection



$$V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$$

$$P_S(x) = VV^T$$

Examples of eigenvector:

$$P_S(x) = \lambda x$$

- let $x \in S$, $P_S(x) = x$ then x is an eigenvector of P_S with associated eigenvalue 1
- let $u \perp S$, $P_S(u) = 0$ then u is an e.v with e.v 0

$$= \langle u, v_1 \rangle v_1 + \cdots + \langle u, v_n \rangle v_n$$

with $(v_1 \ldots v_m)$

Example: orthogonal projection

1.2 Some useful facts

Let $A \in \mathbb{R}^{n \times n}$. Suppose that A has an eigenvalue $\lambda \in \mathbb{R}$ and let $x \in \mathbb{R}^n$ be an eigenvector associated to λ.

Fact #3

For all $k \in \mathbb{N}$, λ^k is an eigenvalue of the matrix A^k and x is an associated eigenvector.

1.2 Some useful facts

Let $A \in \mathbb{R}^{n \times n}$. Suppose that A has an eigenvalue $\lambda \in \mathbb{R}$ and let $x \in \mathbb{R}^n$ be an eigenvector associated to λ.

Fact #4

If A is invertible then $1/\lambda$ is an eigenvalue of the matrix inverse A^{-1} and x is an associated eigenvector.

1.3 Eigenspaces

Definition

If $\lambda \in \mathbb{R}$ is an eigenvalue of $A \in \mathbb{R}^{n \times n}$, the set

$$E_\lambda(A) = \{x \in \mathbb{R}^n \,|\, Ax = \lambda x\}$$

is called the eigenspace of A associated to λ. The dimension of $E_\lambda(A)$ is called the multiplicity of the eigenvalue λ.

Examples: Eigenvalue 1 for I_d? Eigenvalue 0 for $\ker(A)$?

1.4 Spectrum

Definition

The set of all eigenvalues of A is called the *spectrum* of A and denoted by $\mathrm{Sp}(A)$.

Theorem

A $n \times n$ matrix A admits at most n different eigenvalues: $\#\mathrm{Sp}(A) \leq n$.

Proof that $\#\mathrm{Sp}(A) \leq n$

Proposition

Let $v_1, \ldots, v_k$ be eigenvectors of A corresponding (respectively) to the eigenvalues $\lambda_1, \ldots, \lambda_k$.
If the λ_i are all distinct ($\lambda_i \neq \lambda_j$ for all $i \neq j$) then the vectors $v_1, \ldots, v_k$ are linearly independent.

Proof of the proposition

Proof of the proposition

Even better!

Theorem

A $n \times n$ matrix A admits at most n different eigenvalues: $\#\mathrm{Sp}(A) \leq n$.

Theorem

Let $A \in \mathbb{R}^{n \times n}$. If $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of A of multiplicities $m_1, \dots, m_k$ respectively, then

$$m_1 + \cdots + m_k \leq n.$$

Example

2. Markov chains

An example

Consider a "cat" with 3 "states": 1. Eating 2. Sleeping 3. Playing

2.1 Stochastic matrices

Definition

A matrix $P \in \mathbb{R}^{n \times n}$ is said to be *stochastic* if:

1. $P_{i,j} \geq 0$ for all $1 \leq i, j \leq n$.
2. $\sum_{i=1}^{n} P_{i,j} = 1$, for all $1 \leq j \leq n$.

Probability vectors

2.1 The key equation

Proposition

For all $t \geq 0$

$$x^{(t+1)} = Px^{(t)} \quad \text{and consequently,} \quad x^{(t)} = P^t x^{(0)}.$$

Long-term behavior

Invariant measure

Definition

A vector $\mu \in \Delta_n$ is called an invariant measure for the transition matrix P if

$$\mu = P\mu,$$

i.e. if μ is an eigenvector of P associated with the eigenvalue 1.

2.2 Perron-Frobenius Theorem

Theorem

Let P be a stochastic matrix such that there exists $k \geq 1$ such that all the entries of P^k are strictly positive. Then the following holds:

1. 1 is an eigenvalue of P and there exists an eigenvector $\mu \in \Delta_n$ associated to 1.
2. The eigenvalue 1 has multiplicity 1: $\mathrm{Ker}(P - \mathrm{Id}) = \mathrm{Span}(\mu)$.
3. For all $x \in \Delta_n$, $P^t x \xrightarrow[t\to\infty]{} \mu$.

Consequence

Corollary

Let P be a stochastic matrix such that there exists $k \geq 1$ such that all the entries of P^k are strictly positive.
Then there exists a unique invariant measure μ and for all initial condition $x^{(0)} \in \Delta_n$,

$$x^{(t)} = P^t x^{(0)} \xrightarrow[t \to \infty]{} \mu.$$

2.3 PageRank: Ordering the Web

Naive attempt

First idea: rank the webpages according to their number of *incomming links*. (The more incomming links, the more the webpage is important).

The random surfer

PageRank Algorithm

This defines a Markov chain of transition matrix:

$$P_{i,j} = \begin{cases} 1/\deg(j) & \text{if there is a link } j \to i \\ 0 & \text{otherwise,} \end{cases}$$

- After a long time, the surfer is more likely to be on an *important webpage*.
- If μ is the invariant measure of P (provided P verifies the hypotheses of Perron-Frobenius), we take

$$\mu_i = \text{« importance of webpage } i \text{ ».}$$

PageRank Algorithm

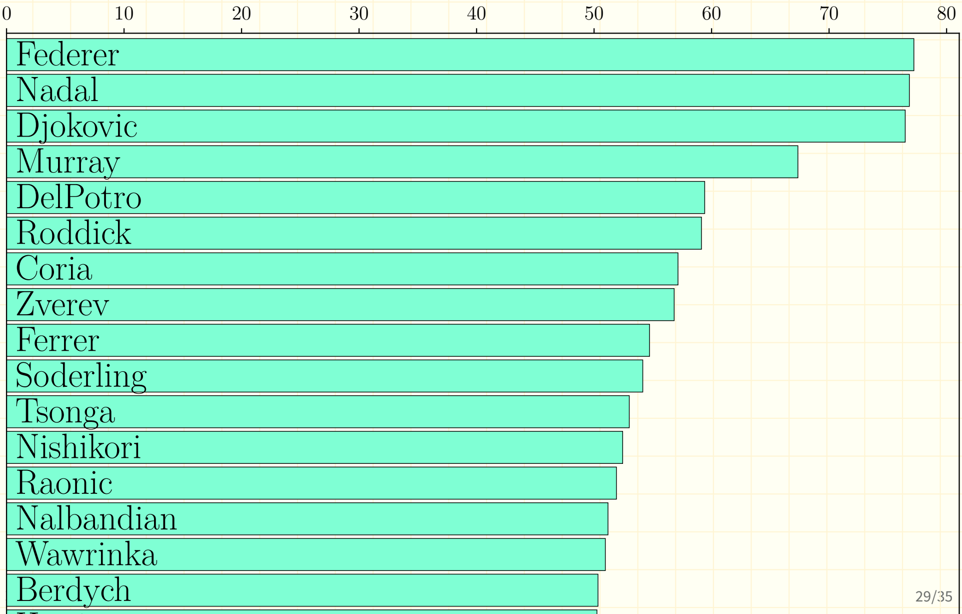
Application: ranking Tennis players

Goal: rank the following players:

Federer, Nadal, Djokovic, Murray, Del Potro, Roddick, Coria, Zverev, Ferrer, Soderling, Tsonga, Nishikori, Raonic, Nalbandian, Wawrinka, Berdych, Hewitt, Tsitsipas, Monfils, Gonzalez, Thiem, Ljubicic, Davydenko, Cilic, Pouille, Safin, Isner, Dimitrov, Medvedev, Ferrero, Goffin, Bautista Agut, Sock, Gasquet, Simon, Blake, Monaco, Coric, Stepanek, Khachanov, Almagro, Robredo, Verdasco, Anderson, Youzhny, Baghdatis, Dolgopolov, Kohlschreiber, Fognini, Melzer, Paire, Querrey, Tomic, Basilashvili.

Data: Head-to Head records (number of times that player x has defeated player y)

Ranking by % of victories



The random spectator

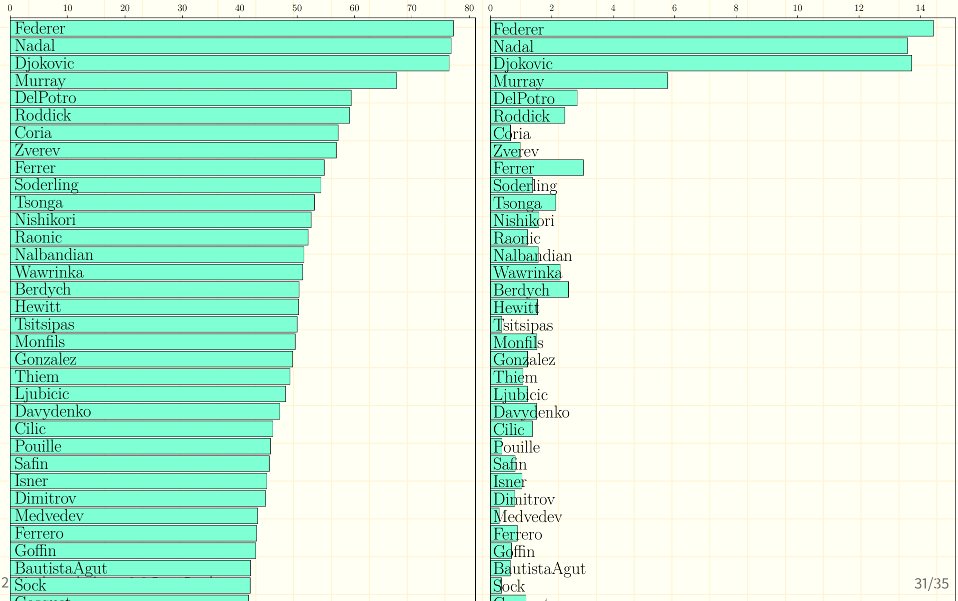
Imagine the following « random spectator »:

- At time t, the spectator believes that player j is the best: $X_t = j$.
- Then, he picks a game of player j uniformly at random:
  - if player j wins, then the spectator still believes that j is the best: $X_{t+1} = j$.
  - otherwise, the spectator changes his mind and now believes that player i who defeated j is the best: $X_{t+1} = i$.

This defines a transition matrix P. We rank the players according to the stationary distribution μ of

$$M = \alpha P + \frac{1-\alpha}{N} J$$

Naive ranking vs PageRank



3. The spectral theorem

The spectral theorem

Theorem

Let $A \in \mathbb{R}^{n \times n}$ be a **symmetric** matrix. Then there is a orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of A.

That means that if A is symmetric, then there exists an orthonormal basis $(v_1, \dots, v_n)$ of $\mathbb{R}^n$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that

$$Av_i = \lambda_i v_i \qquad \text{for all} \quad i \in \{1, \dots, n\}.$$

The spectral orthonormal basis

Matrix formulation

Theorem (Matrix formulation)

Let $A \in \mathbb{R}^{n \times n}$ be a **symmetric** matrix. Then there exists an orthogonal matrix P and a diagonal matrix D of sizes $n \times n$ such that

$$A = PDP^{\mathsf{T}}.$$

Questions?

Questions?