Session 6 bis: Markov Chains and PageRank

Optimization and Computational Linear Algebra for Data Science

Léo Miolane

Be accurate !

Let x, v be vectors, S a subspace of $\mathbb{R}^n$ and M an $n \times n$ matrix.

NO

- $\cancel{x = S}$ ~~or~~ $\cancel{x \subset S}$
- $\cancel{S \in \mathbb{R}^n}$
- $\cancel{\mathrm{Span}(x, v) = \{ax + bv\}}$
- $\cancel{\dim(M)}$ ~~or~~ $\cancel{\dim(x)}$
- $\cancel{\mathrm{Ker}(M) = 0}$
- $\cancel{x + M}$

YES

- $x \in S$
- $S \subset \mathbb{R}^n$
- $\{ax + bv \mid a, b \in \mathbb{R}\}$
- $\mathrm{rank}(M)$
- $\mathrm{Ker}\, M = \{0\}$

Contents

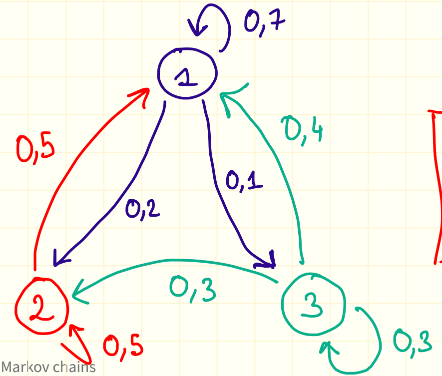
Markov chains

An example

- A cat has 3 states: (1) Sleeping (2) Eating (3) Playing
- We represent its evolution in time with a sequence:

$$\underbrace{X_0, X_1, X_2, \dots\, X_t \;\dots}_{\text{“Markov chain”}}$$

$X_t \in \{1,2,3\}$ state at time t



Transition matrix:

$$\boxed{P = \begin{pmatrix} 0{,}7 & 0{,}5 & 0{,}4 \\ 0{,}2 & 0{,}5 & 0{,}3 \\ 0{,}1 & 0 & 0{,}3 \end{pmatrix}}$$

$$\boxed{\mathbb{P}(X_{t+1} = i \mid X_t = j) = P_{i,j}}$$

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 $1 \leq i, j \leq n$. (non-negative entries)
2. $\sum_{i=1}^{n} P_{i,j} = 1$, for all $1 \leq j \leq n$. (each column sum to 1)

Given a stochastic matrix $P \in \mathbb{R}^{n \times n}$, one can define a Markov chain over n states, and vice-versa.

Probability vectors

Question: after t steps, what is the probability of being in a given state $j \in \{1, \dots n\}$?

- We have to compute:

$$\boxed{x^{(t)} = \begin{pmatrix} \mathbb{P}(X_t = 1) \\ \vdots \\ \mathbb{P}(X_t = n) \end{pmatrix} \in \mathbb{R}^n}$$

(definition)

- $$x^{(t)} \in \Delta_n = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} x_i \geq 0 \quad \text{for all } i \\ \sum_{i=1}^{n} x_i = 1 \end{array} \right\}$$

set of all "probability vectors"

The key equation

Proposition

For all $t \geq 0$

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

Proof:

$$x_i^{(t+1)} = \mathbb{P}(X_{t+1} = i)$$

$$= \sum_{j=1}^{n} \underbrace{\overbrace{\mathbb{P}(X_{t+1} = i \mid X_t = j)}^{P_{i,j}} \overbrace{\mathbb{P}(X_t = j)}^{x_j^{(t)}}}$$

$$= \sum_{j=1}^{n} P_{i,j} \, x_j^{(t)} = \left(P x^{(t)}\right)_i$$

Long-term behavior

- Numerical simulations suggest that

$$\boxed{x^{(t)} \xrightarrow[t \to +\infty]{} \mu \quad \text{for some } \mu \in \Delta_n.}$$

- Since

$$\underbrace{x^{(t+1)}}_{\downarrow\, t \to +\infty} = \underbrace{P\, x^{(t)}}_{\downarrow\, t \to +\infty}$$

$$\boxed{\mu = P\mu}$$

- μ has to be an eigenvector of P associated with the eigenvalue 1.

Perron-Frobenius Theorem

Invariant measure

"prob. vectors"

Definition

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

$$\boxed{\mu = P\mu,}$$

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

"invariant": if X_t is distibuted according to μ. ($x^{(t)} = \mu$) then

$$x^{(t+1)} = P x^{(t)} = P\mu = \mu$$

Therefore X_{t+1} is also distibuted according to μ.

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. $(P\mu = \mu)$
2. The eigenvalue 1 has multiplicity 1: $\mathrm{Ker}(P - \mathrm{Id}) = \mathrm{Span}(\mu)$.
3. $\boxed{\text{For all } x \in \Delta_n, P^t x \xrightarrow[t\to\infty]{} \mu.}$

Consequence

Corollary

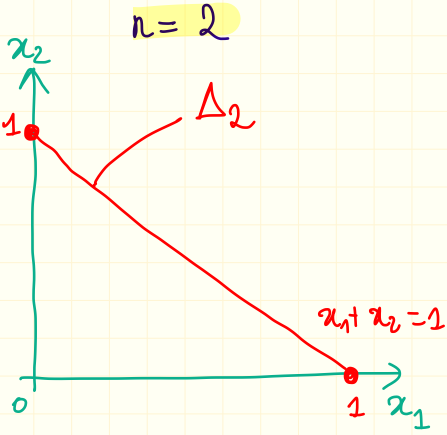
Let P be a stochastic matrix such that there exists $k > 1$ such that all the entries of P^k are strictly positive.
Then there exists a unique invariant measure μ and for all initial condition $x^{(0)} \in \Delta_n$,

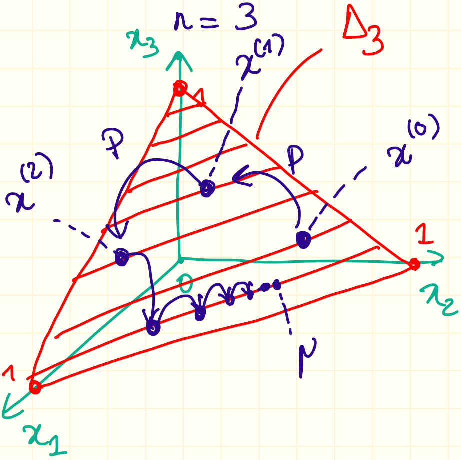
$$\boxed{x^{(t)} = P^t x^{(0)} \xrightarrow[t\to\infty]{} \mu.}$$

Point 1 Pt. 2 Pt 3.

Proof: Geometrical observations

$$\Delta_n = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} x_i \geq 0 \quad \text{for all } i \\ x_1 + \dots + x_n = 1 \end{array} \right\}$$





Proof: contraction

We will prove the theorem in the case where $P_{i,j} > 0$ for all i, j.

Lemma

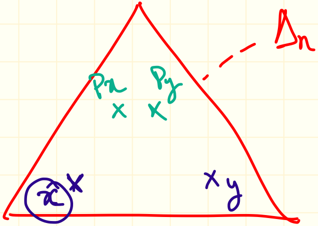
The mapping

$$\boxed{\begin{array}{cccc} \varphi: & \Delta_n & \to & \Delta_n \\ & x & \mapsto & Px \end{array}}$$

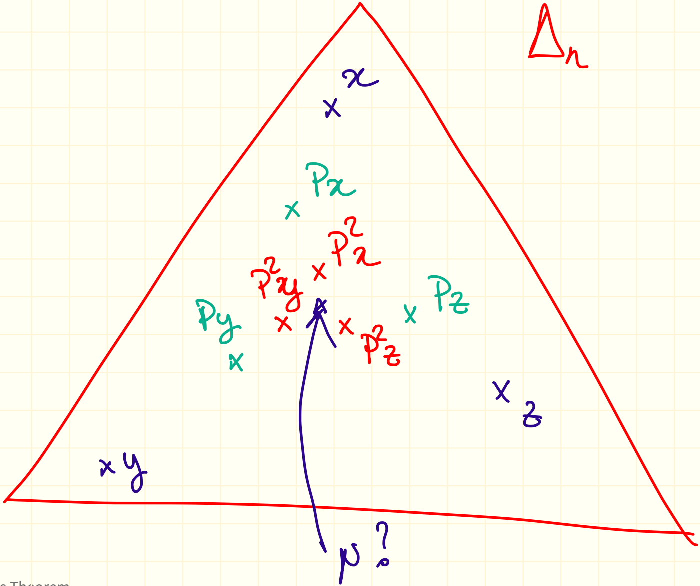
is a contraction mapping for the ℓ_1-norm: there exists $c \in (0,1)$ such that for all $x, y \in \Delta_n$:

$c \in (0,1)$

$$\boxed{\|Px - Py\|_1 \leq c\|x - y\|_1.}$$



Geometric picture



Proof of Perron-Frobenius

- Let $\nu \in \Delta_n$ be a minizer of $\underline{x \mapsto \|Px - x\|_1}$ on Δ_n. (we admit that ~~the~~ a minimizer exits)

- Then, $\boxed{P\nu = \nu}$. Indeed, if $\boxed{\|P\nu - \nu\|_1 > 0}$,

"contraction"

$$\|P(P\nu) - (P\nu)\|_1 \leq c\,\|P\nu - \nu\|_1 < \|P\nu - \nu\|_1$$

this contradicts the optimality of ν.

- Let $x \in \Delta_n$. $\|P^t x - \nu\|_1 = \|P^t x - P^t \nu\|_1$

$0 < c < 1$ $\underline{\qquad\qquad} \leq c^t \|x - \nu\|_1 \xrightarrow[t \to +\infty]{} 0$

Proof of Perron-Frobenius

- Let $x \in \mathbb{R}^n$ such that $\boxed{Px = x}$

$$x = P^t x$$
$$= P^t (x_1 e_1 + \dots + x_n e_1)$$
$$= x_1 \underbrace{P^t e_1}_{\xrightarrow{t \to +\infty} \nu} + \dots + x_n \underbrace{P^t e_n}_{\xrightarrow{t \to +\infty} \nu}$$

because $e_1 \dots e_n \in \Delta_n$

$$\xrightarrow[t \to +\infty]{} (x_1 + \dots + x_n)\,\nu \in \mathrm{Span}(\nu)$$

$$\boxed{x = (x_1 + \dots + x_n)\,\nu}$$

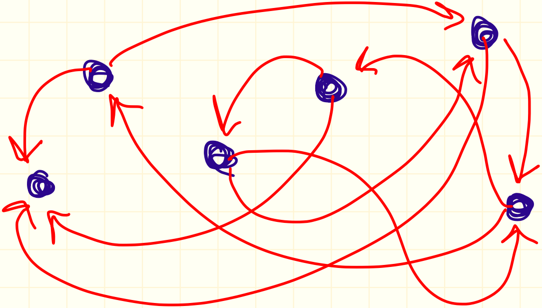
$$\boxed{x \in \mathrm{Span}(\nu)}$$

PageRank

Ordering the Web

World Wide Web.

Webpage j



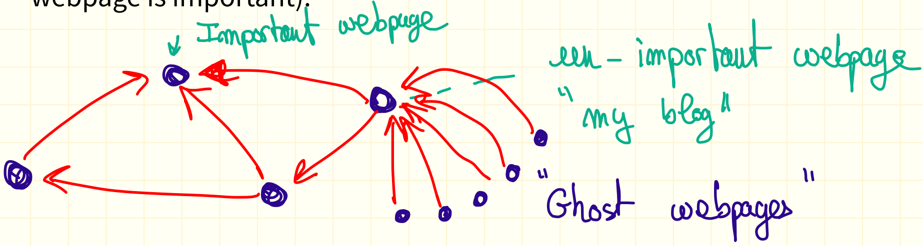
if there is a link from j to i

Webpage i

Goal: rank the webpages from the most "important" one to the less important one.

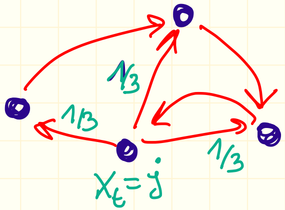
Naive attempt

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



That does not work: one can create thousand of "ghost webpages" pointing to a single webpage.

The random surfer



- Imagine a "drunk surfer"
- if at time t, the surfer is on a webpage $X_t = j$

he clicks on an outgoing link selected at random.

$$\boxed{\mathbb{P}\left(X_{t+1} = i \;\middle|\; X_t = j\right) = \begin{cases} 0 & \text{if there is no link } j \to i \\ \dfrac{1}{\deg(j)} & \text{otherwise} \end{cases}}$$

where $\deg(j)$ is the number of outgoing links from j

PageRank Algorithm

This defines a Markov chain of transition matrix:

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

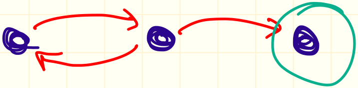
- After a long time, the surfer is more likely to be on an *important webpage*.
- If μ is the invariant measure of P (provided P verifies the hypotheses of Perron-Frobenius), we take

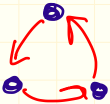
$$\mu_i = \text{« importance of webpage } i \text{ ».}$$

$\simeq$ prob. of being at i after a long time

PageRank Algorithm

Issue





Google considered the transition matrix:

$$G = \alpha P + \frac{1-\alpha}{N}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

$\alpha \simeq 0{,}85$

total number of pages

This avoids being trapped!

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: $X_{t+1} = i$.

The random spectator

Imagine the following « random spectator »:

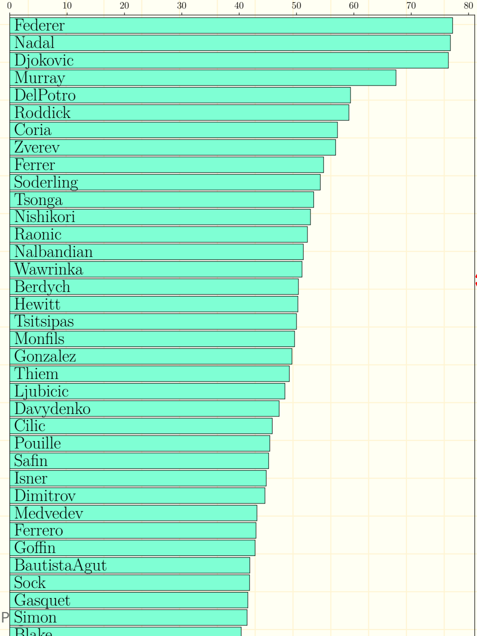
- At time t, the spectator believes that player j is the best: $X_t = j$.
- Then, he picks a game of player j uniformly at random:
  - if player j wins, then the spectator still believes that j is the best: $X_{t+1} = j$.
  - otherwise, the spectator changes his mind and now believes that player i who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

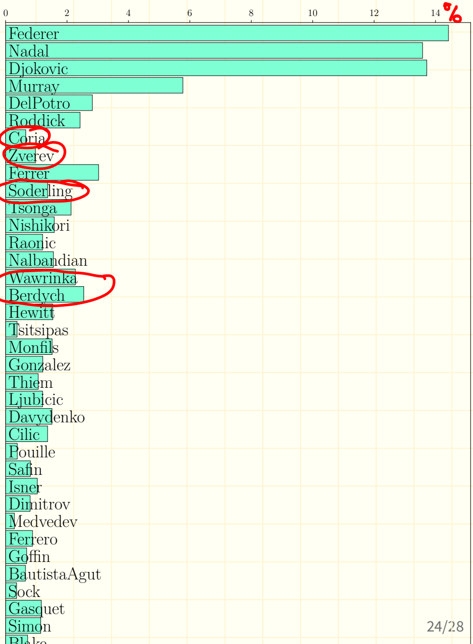
This defines a transition matrix P. We rank the players according to the stationary distribution μ of

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

$$\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

Naive ranking vs PageRank





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.

There exists $v_1 \dots v_n \in \mathbb{R}^n$ such that

- $(v_1 - v_n)$ is an orthonormal basis of $\mathbb{R}^n$
- $Av_i = \lambda_i v_i$ for all i

If we consider $x \in \mathbb{R}^n$

$$x = \langle v_1, x\rangle v_1 + \dots + \langle v_n, x\rangle v_n$$

The spectral theorem

$$Ax = \langle v_1, x\rangle A v_1 + \dots + \langle v_n, x\rangle A v_n$$

$$= \boxed{\underbrace{\langle v_1, x\rangle \lambda_1 v_1 + \dots + \langle v_n, x\rangle \lambda_n v_n}}$$

$$\boxed{P = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}}$$ (this is an orthogonal matrix)

$$P^T x = \begin{pmatrix} \langle v_1, x\rangle \\ \vdots \\ \langle v_n, x\rangle \end{pmatrix}$$

Let $D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$

$$DP^T x = \begin{pmatrix} \lambda_1 \langle v_1, x\rangle \\ \vdots \\ \lambda_n \langle v_n, x\rangle \end{pmatrix}$$

$$\boxed{Ax = PDP^T x}$$

$$A = P \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} P^T$$

Matrix formulation

Theorem

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

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

- the columns of P are eigenvectors of A
- the entries on the diagonal of D are associated eigenvalues.

Questions?

Questions?