Session 6: Eigenvalues, eigenvectors & Markov chains

Optimization and Computational Linear Algebra for Data Science

While you wait: exercise: let S be a subspace of $\mathbb{R}^n$ and P_S the orthogonal projection onto S. Show that

$$\boxed{\operatorname{rank}(P_S) = \operatorname{Tr}(P_S)}$$

Léo Miolane

Contents

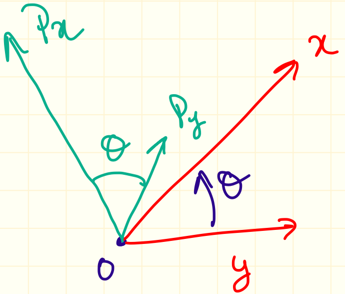
1. Orthogonal matrices
2. Eigenvalues & eigenvectors
3. Properties of eigenvalues and eigenvectors
4. Markov chains

Orthogonal matrices

Orthogonal matrices

If $P \in \mathbb{R}^{n \times n}$ is orthogonal

- P "preserves the norm" $\|Px\| = \|x\|$
- P "preserves the "angles" " $\langle Py, Px \rangle = \langle x, y \rangle$

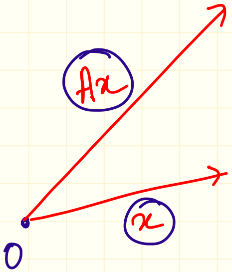


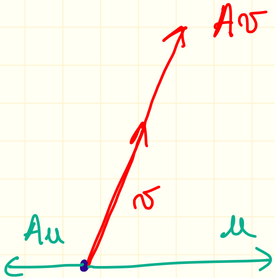
Thats why we should understand orthogonal matrices as "rotations" in $\mathbb{R}^n$

Eigenvalues & eigenvectors

Introduction

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





Here x is not an eigenvector of A

Here v and Av are colinear: we say that $\underset{u}{v}$ is an eigenvector of A

Definition

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

$$\boxed{Av = \lambda v.}$$

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

$\mathbb{R}^{n\times m}$

$\mathbb{R}^m \longrightarrow \mathbb{R}^n$

$\boxed{Id\, x = 1 \cdot x}$

Remark: If $v \in \mathrm{Ker}(A)$ and if $\boxed{v \neq 0}$ then v is an eigenvector of A associated to the eigenvalue 0:

$$Av = 0 = 0 \cdot v$$

Example: diagonal matrices

$$D = \mathrm{Diag}(\lambda_1, \lambda_2, \dots \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

$D e_1 = \lambda_1 e_1$ e_1 is an eigenvector of D

λ_1 is the associated eigenvalue

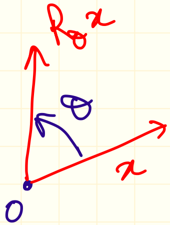
$\vdots$

$D e_n = \lambda_n e_n$ e_n ______

λ_n ______

Matrix with no eigenvalues/vectors

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



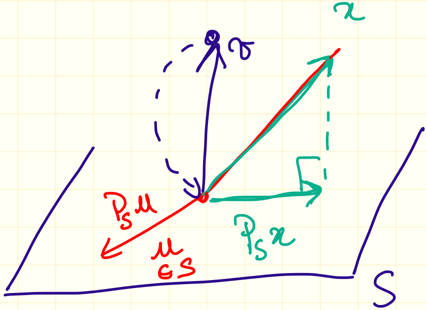
$x \neq 0$

For all $x \in \mathbb{R}^2$, for all $\lambda \in \mathbb{R}$

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

Therefore R_θ does not have any (real) eigenvalues

Example: orthogonal projection



Question: Give me eigen values | vectors of P_S!

- if $u \in S$, $u \neq 0$ then u is an eigenvector of P_S with eigenvalue 1

- if $v \in S^\perp$ then $v \in \mathrm{Ker}(P_S)$ so if $v \neq 0$ then v is an eigenvector of P_S with eigenvalue 0.

Eigenspaces

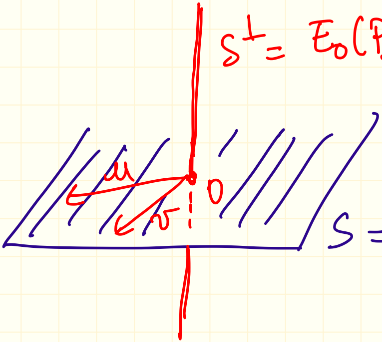
Definition

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

$(A - \lambda Id)x = 0$

$$E_\lambda(A) = \{x \in \mathbb{R}^n \mid Ax = \lambda x\} = \mathrm{Ker}(A - \lambda Id)$$

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



$S^\perp = E_0(P_S) \leftarrow$ dim 1

$S = E_1(P_S)$ dim 2

Properties

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.

$Ax = \lambda x$

Fact #1

For all $\alpha \in \mathbb{R}$, $\alpha\lambda$ is an eigenvalue of the matrix αA and x is an associated eigenvector.

Proof $\quad (\alpha A)x = \alpha Ax = (\alpha\lambda)x$

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 #2

For all $\alpha \in \mathbb{R}$, $\lambda + \alpha$ is an eigenvalue of the matrix $A + \alpha \mathrm{Id}$ and x is an associated eigenvector.

Proof:

$$
\begin{aligned}
\underline{\underline{(A + \alpha \mathrm{Id})}}\, x &= Ax + \alpha \mathrm{Id}\, x \\
&= \lambda x + \alpha x \\
&= \underline{\underline{(\lambda + \alpha)}}\, x
\end{aligned}
$$

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.

Proof:

$$Ax = \lambda x$$

$$A^2 x = A(\lambda x) = \lambda A x = \lambda^2 x$$

$$A^3 x = A(\lambda^2 x) = \lambda^3 x$$

$$\vdots$$

$$A^k \underline{x} = \lambda^k \underline{x}$$

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.

Proof: exercise

Spectrum

Definition

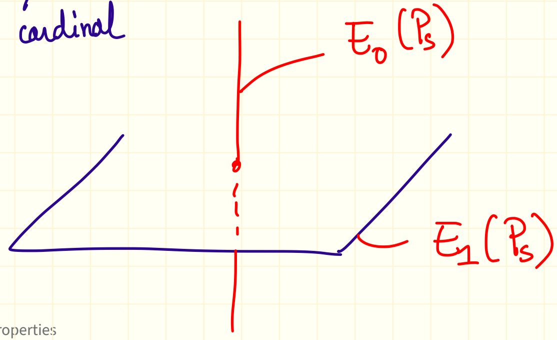
The set of all eigenvalues of A is called the *spectrum* of A and denoted by $\mathrm{Sp}(A)$.

$\mathrm{Sp}(\mathrm{Diag}(\lambda_1, \dots, \lambda_n)) = \{\lambda_1, \dots, \lambda_n\}$

Theorem

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

cardinal



$\mathrm{Sp}(P_s) = \{0, 1\}$

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.

Assuming that the proposition holds.

If $\lambda_1 — \lambda_k$ are distinct eigenVal. A associated with $v_1 — v_k$: $\underbrace{v_1 — v_k}_{\in \mathbb{R}^n}$ lin indep

This implies that $\boxed{k \leq n}$.

Proof of the proposition

Assume first $\boxed{\lambda_1 > \dots > \lambda_k}$ ≥ 0 assumption.

Let $\alpha_1 \dots \alpha_k \in \mathbb{R}$ such that $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_k v_k = 0$ $\times A^m$

$A \dots A$

$$\alpha_1 A^m v_1 + \dots + \alpha_k A^m v_k = 0$$

$$\alpha_1 \lambda_1^m v_1 + \dots + \alpha_k \lambda_k^m v_k = 0 \quad / \lambda_1^m$$

$$\underbrace{\alpha_1 v_1 + \alpha_2 \Big(\underbrace{\frac{\lambda_2}{\lambda_1}}_{\in [0,1)}\Big)^m v_2 + \dots + \alpha_k \Big(\underbrace{\frac{\lambda_k}{\lambda_1}}_{\in [0,1)}\Big)^m v_k}_{\xrightarrow[m \to +\infty]{} 0} = 0$$

$$\boxed{\alpha_1 v_1 = 0}$$

Proof of the proposition

$\alpha_1 v_1 = 0$ but $v_1 \neq 0$ therefore $\alpha_1 = 0$

Repeating this, we get $\alpha_1 = \alpha_2 = \dots = \alpha_k = 0$.

In the general case, let $\lambda_{min} = \min(\lambda_1 \dots \lambda_k)$

We apply what we have proved to the matrix

$(A - \lambda_{min} Id) \leftarrow$ eigenvalues $\lambda_1 - \lambda_{min} \dots \lambda_k - \lambda_{min}$ $\geqslant 0$

eigenvectors $v_1 \dots v_k$

I get that $v_1 \dots v_k$ are lin. independent.

Even better!

Theorem

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

$\downarrow$ Stronger!

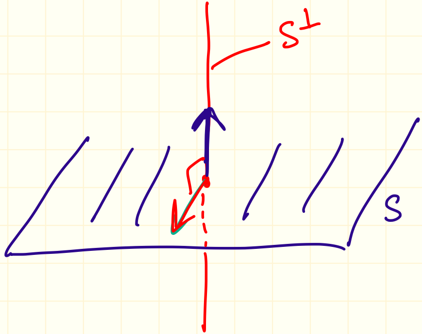
Theorem

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

$m_1 = \dim\left(E_{\lambda_1}(A)\right)$

$$\boxed{m_1 + \cdots + m_k \leq n.}$$

Example



- $S \subset \underline{E_1(P_S)}$
- $S^\perp \subset E_0(P_S)$

$\rightarrow$ $\dim E_1(P_S) \geq 2$
$\dim E_0(P_S) \geq 1$

By the Theorem: $\underbrace{\dim E_1(P_S)}_{\text{multip. of } 1} + \dim E_0(P_S) \leq 3$

$\rightarrow$ $\boxed{\begin{array}{l} \dim E_1(P_S) = 2 \\ \dim E_0(P_S) = 1 \end{array}}$ $\longrightarrow$ $\boxed{\begin{array}{l} E_1(P_S) = S \\ E_0(P_S) = S^\perp \end{array}}$

Markov chains

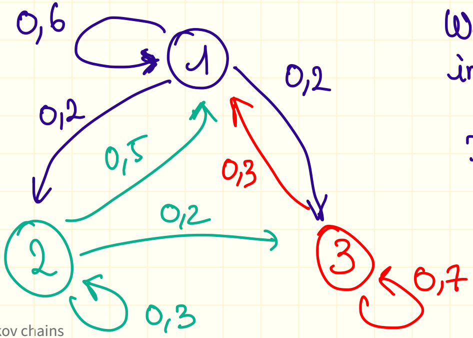
An example

Consider a "cat" that has only 3 "states":

① Sleeping ② Eating ③ Playing

We represent the evolution in time by the sequence

$X_0, X_1 \dots X_t \dots$ $\in \{1,2,3\}$ state of cat at time t



We can put these prob. in a "transition matrix"

$$P = \begin{pmatrix} 0,6 & 0,5 & 0,3 \\ 0,2 & 0,3 & 0 \\ 0,2 & 0,2 & 0,7 \end{pmatrix}$$

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

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

Probability vectors

<u>Question</u> : what is the prob. that the cat is

| sleeping / eating / playing at a time t

$$x^{(t)} = \begin{pmatrix} \mathbb{P}(X_t = 1) \\ \vdots \\ \mathbb{P}(X_t = 3) \end{pmatrix} \in \mathbb{R}^3$$ in my <u>example</u>

The key equation

Proposition

For all $t \geq 0$

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

Long-term behavior

We observe that $x^{(t)} \xrightarrow{t \to +\infty} \mu \in \mathbb{R}^3$

We know that $x^{(t+1)} = P x^{(t)}$

$x^{(t+1)} \xrightarrow{t \to +\infty} \mu$

$P x^{(t)} \xrightarrow{t \to +\infty} P\mu$

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

- μ has to verify $\mu = P\mu$
- μ is an <u>eigenvector</u> of P associated to the eigenvalue 1.

Next week

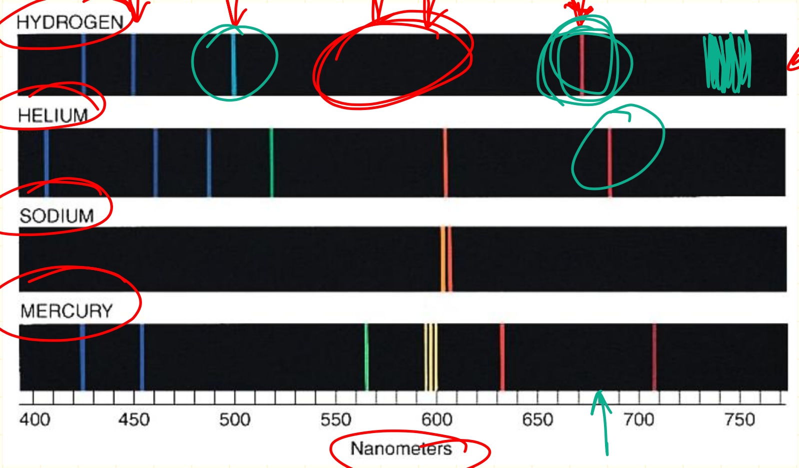
1. Does $x^{(t)}$ always converge?

2. Is there only one limiting distribution μ?

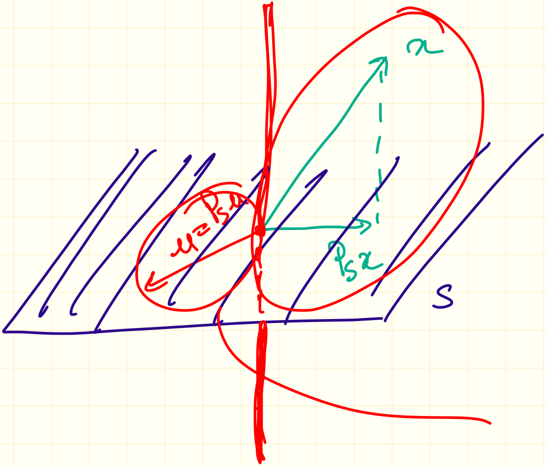
3. How do we make $$$ with that?

$\longrightarrow$ all answered next week.

Eigenvalues in physics



Questions?



P_S

? ?

$$A x = \lambda x$$

P_S

λ

$$P_S u = 1 \cdot u$$

Questions?