# Fondements du Machine Learning

Clément Royer

Licence 3 Informatique et Mathématiques pour la Décision et les Données

7 septembre 2020

# Premier contact

- Nous sommes en distanciel au moins jusqu'en octobre...
- Réunion de rentrée tout à l'heure;
- Pas de présentiel avant un moment.

# Premier contact

- Nous sommes en distanciel au moins jusqu'en octobre...
- Réunion de rentrée tout à l'heure;
- Pas de présentiel avant un moment.

**Présentez-vous !**

Dans le chat :

- Nom+Prénom;
- Université/Établissement en 2019-2020;
- Ville de naissance (ou votre ville préférée).

# Quelques informations

## Enseignant : Clément Royer


- Né à Pau (Pyrénées-Atlantiques);
- Maître de conférences à Dauphine (MIDO) depuis 2019;
- Chercheur au LAMSADE en optimisation pour les sciences des données;
- Email : `clement.royer@dauphine.psl.eu`
- Lien vers ma page personnelle

# Contact et liens utiles

## Liens utiles

- Page du cours :
  https://www.lamsade.dauphine.fr/∼ croyer/coursFML.html
- Polycopié (mis à jour au fil de l'eau).
- Espace My Course dédié :
  *A3IIMD_20/21_Fondements du machine learning_Clément ROYER*
- Équipe dédiée sur Microsoft Teams;

## Utilisation des ressources

- Teams pour les séances;
- Documents accessibles via ma page personnelle.

# Organisation du cours

## Telle qu'annoncée

- Cours :
  - Lundi 10h15-11h45;
  - Dernier cours le 23/11, pas de cours les 26/10 et 09/11;
- TD :
  - Mardi 13h45-15h15;
  - Premier TD le 15/09, dernier le 24/11;

## Évaluation

- Note=0.3*CC + 0.7*E;
- Contrôle continu : En présentiel si possible, devoir maison sinon.
- Examen : En janvier.

# Organisation

## En temps normal

- Ponctualité;
- Discussions;
- **Retours**.

## Ces temps-ci

- **Coupez votre micro** quand vous n'intervenez pas ou que vous arrivez en retard;
- Échangez via le chat/les canaux dédiés.

# Fonctionnement distanciel

## En temps normal

- Cours le lundi;
- TD le mardi.

# Fonctionnement distanciel

## En temps normal

- Cours le lundi;
- TD le mardi.

## Ces temps-ci

- **Polycopié disponible en avance de phase**;
- Chaque semaine, un sujet, avec cours + exercices;
  - Idéalement, cours/preuves lundi, TD le mardi;
  - Mais le fonctionnement est flexible!

# Objectifs du cours

- Présenter des modèles simples d'**analyse de données**...
- ...très fréquents en apprentissage machine *(machine learning)*.
- **Outils :** algèbre linéaire, probabilités/statistiques.

# Contenus thématiques

- Introduction et motivation;
- Algèbre linéaire et modèles linéaires :
  - Décomposition en valeurs propres/valeurs singulières;
  - Moindres carrés linéaires.
- Outils statistiques :
  - Statistique unidimensionnelle;
  - Compromis biais-variance.
- **Régression linéaire**:
  - Simple, multiple;
  - Régularisation.
- Réduction de dimension:
  - **Analyse en composantes principales**;
  - Applications.

# Terminologie

## Différents domaines

- Analyse de données *(data analysis)*;
- Fouille de données *(data mining)*;
- Apprentissage [machine/automatique/artificiel] *(machine learning)*;
- Intelligence artificielle *(AI)*;
- Masses de données *(Big Data)*;
- etc.

# Terminologie

## Différents domaines

- Analyse de données *(data analysis)*;
- Fouille de données *(data mining)*;
- Apprentissage [machine/automatique/artificiel] *(machine learning)*;
- Intelligence artificielle *(AI)*;
- Masses de données *(Big Data)*;
- etc.

Sciences des données *(data science)*

# Machine learning pour ce cours

Ensemble de **techniques** qui visent à extraire de l'information d'un jeu de **données**.

# Machine learning pour ce cours

Ensemble de **techniques** qui visent à extraire de l'information d'un jeu de **données**.

**Principe 1: Grande quantité de données**

- Pour avoir de l'information à extraire;
- Pour être représentatif.

# Machine learning pour ce cours

Ensemble de **techniques** qui visent à extraire de l'information d'un jeu de **données**.

## Principe 1: Grande quantité de données

- Pour avoir de l'information à extraire;
- Pour être représentatif.

## Principe 2: Utilisation d'algorithmes

- Traitement systématique et efficace;
- Théorie mathématique+implémentation.

# Pourquoi s'intéresser aux données ?

## Essentiel pour les entreprises

- Modèle économique des GAFA(M);
- Service gratuit mais valeur dans l'exploitation des données.

# Pourquoi s'intéresser aux données ?

## Essentiel pour les entreprises

- Modèle économique des GAFA(M);
- Service gratuit mais valeur dans l'exploitation des données.

## Important pour la recherche

- Quantité massive de données générées en biologie, médecine,...
- Difficultés mathématiques et informatiques.

# Pourquoi s'intéresser aux données ?

## Essentiel pour les entreprises

- Modèle économique des GAFA(M);
- Service gratuit mais valeur dans l'exploitation des données.

## Important pour la recherche

- Quantité massive de données générées en biologie, médecine,...
- Difficultés mathématiques et informatiques.

## Approches guidées par les données

*Ou data-driven, drivées par les données, etc.*

- Pallie le manque de modèles formels;
- Pourrait remplacer la modélisation à terme.

# Exemples: systèmes de recommendation

**But** : Suggérer du contenu en se basant sur les préférences.

- Séries/Films/Vidéos (Netflix, Youtube);
- Produits commerciaux (Ebay, Amazon);
- Hôtels/Restaurants/Voyages (Google, Facebook).

# Exemples: systèmes de recommendation

**But** : Suggérer du contenu en se basant sur les préférences.

- Séries/Films/Vidéos (Netflix, Youtube);
- Produits commerciaux (Ebay, Amazon);
- Hôtels/Restaurants/Voyages (Google, Facebook).

**Outil : Une matrice des avis (Clients/Produits).**

# Exemples: systèmes de recommendation

**But** : Suggérer du contenu en se basant sur les préférences.

- Séries/Films/Vidéos (Netflix, Youtube);
- Produits commerciaux (Ebay, Amazon);
- Hôtels/Restaurants/Voyages (Google, Facebook).

**Outil : Une matrice des avis (Clients/Produits).**

## Les grandes questions

Comment suggérer du contenu pertinent, et donc...

1. Quels sont les éléments principaux de nos préférences ?
2. Comment gérer un grand nombre d'avis ?
3. Les avis reflètent-ils vraiment la réalité ?

## Quid du cours ?

### Fondamentaux de l'apprentissage (*machine learning*)

- Des modèles linéaires de l'information dans les données;
- Les données sont vues comme des réalisations de variables aléatoires (typiquement gaussiennes).

# Quid du cours ?

## Fondamentaux de l'apprentissage (*machine learning*)

- Des modèles linéaires de l'information dans les données;
- Les données sont vues comme des réalisations de variables aléatoires (typiquement gaussiennes).

## Solide sur les fondamentaux

- Souvent efficace en pratique;
- Très souvent le premier cas considéré en recherche;
- Utilise des savoirs fondamentaux en algèbre linéaire, statistiques (et optimisation).

# Deux modèles d'apprentissage

## Modèle fonctionnel

- Distribution (de probabilité) des données connue;
- Développement de modèles ad hoc;
- Apprentissage supervisé.

$\Rightarrow$ **Régression linéaire.**

# Deux modèles d'apprentissage

## Modèle fonctionnel

- Distribution (de probabilité) des données connue;
- Développement de modèles ad hoc;
- Apprentissage supervisé.

$\Rightarrow$ **Régression linéaire.**

## Modèle prédictif

- Pas de distribution connue;
- Extraction d'information des données;
- Fréquent en apprentissage non supervisé.

$\Rightarrow$ **Analyse en composante principales.**

# En bref

## Machine learning/Apprentissage

- Décrire le comportement de données;
- Prédire les propriétés de données futures.

## Notre approche

- Le cas linéaire : souvent efficace et populaire en pratique.
- Le modèle linéaire : permet de présenter les enjeux et les outils fondamentaux.

# Sommaire

# Sommaire

# Objectifs

## Objectif 1 : Fixer les notations

- Différentes écoles : dauphinoise, toulousaine, anglo-saxonne,...
- Un seul cours !

## Objectif 2 : Parler le même langage

- Lever toute ambiguïté dès le départ...
- ...pour se rapprocher de notations usuelles (recherche+industrie).

# Des histoires de notation

## Notations en algèbre linéaire

- Coefficients/Scalaires : $\alpha, \beta, \gamma$;
- Vecteurs/Variables : $x, y, z$;
- Matrices : $A, B, C$.

# Des histoires de notation

## Notations en algèbre linéaire

- Coefficients/Scalaires : $\alpha, \beta, \gamma$;
- Vecteurs/Variables : $x, y, z$;
- Matrices : $A, B, C$.

## Notations en statistiques

- Coefficients/Scalaires : $a, b, c, \dots, \alpha, \beta, \gamma$;
- Vecteurs/Variables : $\alpha, \beta, \mu$;
- Matrices : $X, Y, Z, \dots$.

# Les notations pour ce cours

## Conventions

- Scalaires : $a, b, c, \dots, \alpha, \beta, \gamma$;
- Vecteurs : $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \dots, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$;
- Matrices : $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$.

## Distinction

- Partie algèbre linéaire : $\boldsymbol{x}, \boldsymbol{y}$ génériques;
- Partie statistiques et seq : $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Y}$ représenteront les données.

# Notations vectorielles

- $\mathbb{R}^n$ : ensemble des vecteurs à $n \geq 1$ composantes réelles;
- Par convention, $\boldsymbol{x} \in \mathbb{R}^n$ est un vecteur colonne, et on note $\boldsymbol{x}^{\mathrm{T}}$ le vecteur ligne correspondant.
- Pour tout $\boldsymbol{x} \in \mathbb{R}^n$ et tout $i \in \{1, \dots, n\}$, on notera $x_i \in \mathbb{R}$ sa $i$-ème coordonnée (dans la base canonique de $\mathbb{R}^n$) $\Rightarrow$ $\boldsymbol{x} = [x_i]_{1 \leq i \leq n}$.

# Notations vectorielles

- $\mathbb{R}^n$ : ensemble des vecteurs à $n \geq 1$ composantes réelles;
- Par convention, $\boldsymbol{x} \in \mathbb{R}^n$ est un vecteur colonne, et on note $\boldsymbol{x}^\mathrm{T}$ le vecteur ligne correspondant.
- Pour tout $\boldsymbol{x} \in \mathbb{R}^n$ et tout $i \in \{1, \dots, n\}$, on notera $x_i \in \mathbb{R}$ sa $i$-ème coordonnée (dans la base canonique de $\mathbb{R}^n$) $\Rightarrow$ $\boldsymbol{x} = [x_i]_{1 \leq i \leq n}$.

## Structure d'espace vectoriel normé

- *Addition dans* $\mathbb{R}^n$ *:* $\boldsymbol{x} + \boldsymbol{y} := [x_i + y_i]_{1 \leq i \leq n}$;
- *Multiplication par un réel :*

$$\lambda \boldsymbol{x} \overset{n}{=} \lambda \cdot \boldsymbol{x} := [\lambda x_i]_{1 \leq i \leq n} \,;$$

- *Norme euclidienne :*

$$\|\boldsymbol{x}\| := \sqrt{\sum_{i=1}^{n} x_i^2}.$$

# Notations vectorielles (suite)

## Produit scalaire sur $\mathbb{R}^n$

Le produit scalaire induit par la norme euclidienne est défini pour tous vecteurs $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ par

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} := \sum_{i=1}^{n} x_i y_i.$$

C'est une forme bilinéaire symétrique définie positive (NB : $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$).

# Notations vectorielles (suite)

## Produit scalaire sur $\mathbb{R}^n$

Le produit scalaire induit par la norme euclidienne est défini pour tous vecteurs $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ par

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} := \sum_{i=1}^{n} x_i y_i.$$

C'est une forme bilinéaire symétrique définie positive (NB : $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$).

## Sous-espace engendré

Soient $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ $p$ vecteurs de $\mathbb{R}^n$. Le *sous-espace engendré par les vecteurs* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ est le sous-espace vectoriel

$$\mathrm{vect}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_p) := \left\{ \boldsymbol{x} = \sum_{i=1}^{p} \alpha_i \boldsymbol{x}_i \;\middle|\; \alpha_i \in \mathbb{R}^n \ \forall i \right\}.$$

Ce sous-espace est de dimension au plus $\min\{n, p\}$.

# Notations matricielles

- $\mathbb{R}^{m \times n}$ : ensemble des matrices à $m$ lignes, $n$ colonnes à coefficients réels (on supposera toujours $m \geq 1$ et $n \geq 1$).
- NB : $\mathbb{R}^{m \times 1} \simeq \mathbb{R}^m$.

## Coefficients, lignes et colonnes

Pour $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, on utilisera

- $[\boldsymbol{A}]_{ij}$ pour le coefficient $(i,j)$ de $\boldsymbol{A}$;
- $\boldsymbol{a}_i^{\mathrm{T}}$ pour la $i$-ème ligne de $\boldsymbol{A}$;
- OU $\boldsymbol{a}_j$ pour la $j$-ème colonne de $\boldsymbol{A}$.

Les notations suivantes seront équivalentes à $\boldsymbol{A}$ :

$$[\boldsymbol{A}_{ij}]_{\substack{1 \leq i \leq m, \\ 1 \leq j \leq n}} \quad \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \cdots \\ \boldsymbol{a}_m^{\mathrm{T}} \end{bmatrix} \quad \text{OU} \quad [\boldsymbol{a}_1 \cdots \boldsymbol{a}_n].$$

# Transposée, symétrie

## Définitions

Soit $\boldsymbol{A} = [\boldsymbol{A}_{ij}] \in \mathbb{R}^{m \times n}$ une matrice à $m$ lignes et $n$ colonnes.

- La *transposée de* $\boldsymbol{A}$, notée $\boldsymbol{A}^{\mathrm{T}}$, est la matrice à $n$ lignes et $m$ colonnes telle que

$$\forall i = 1, \dots, m, \ \forall j = 1, \dots, n, \qquad \left[\boldsymbol{A}^{\mathrm{T}}\right]_{ij} = \boldsymbol{A}_{ji}.$$

# Transposée, symétrie

## Définitions

Soit $\boldsymbol{A} = [\boldsymbol{A}_{ij}] \in \mathbb{R}^{m \times n}$ une matrice à $m$ lignes et $n$ colonnes.

- La *transposée de* $\boldsymbol{A}$, notée $\boldsymbol{A}^{\mathrm{T}}$, est la matrice à $n$ lignes et $m$ colonnes telle que

$$\forall i = 1, \dots, m, \ \forall j = 1, \dots, n, \qquad \left[\boldsymbol{A}^{\mathrm{T}}\right]_{ij} = \boldsymbol{A}_{ji}.$$

## Cas des matrices carrées

- $\boldsymbol{A}^{\mathrm{T}} \in \mathbb{R}^{n \times n}$;
- $\boldsymbol{A}$ est dite *symétrique* si $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$.

Ex) Matrices diagonales, de covariance, d'adjacence, etc.

# Noyau, image, rang

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

- Le **noyau** (*kernel/null space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\ker(\boldsymbol{A}) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}_{\mathbb{R}^m}\}.$$

- L'**image** (*range space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\operatorname{Im}(\boldsymbol{A}) := \{\boldsymbol{y} \in \mathbb{R}^m \mid \exists \boldsymbol{x} \in \mathbb{R}^n, \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}.$$

- Le **rang** (*rank*) de $\boldsymbol{A}$, noté $\operatorname{rang}(\boldsymbol{A})$, est la dimension du sous-espace vectoriel $\operatorname{Im}(\boldsymbol{A})$. On a $\operatorname{rang}(\boldsymbol{A}) \leq \min\{m, n\}$.

# Noyau, image, rang

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

- Le **noyau** (*kernel/null space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\ker(\boldsymbol{A}) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \quad \boldsymbol{A}\boldsymbol{x} = \mathbf{0}_{\mathbb{R}^m}\}.$$

- L'**image** (*range space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\mathrm{Im}(\boldsymbol{A}) := \{\boldsymbol{y} \in \mathbb{R}^m \mid \exists \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}.$$

- Le **rang** (*rank*) de $\boldsymbol{A}$, noté $\mathrm{rang}(\boldsymbol{A})$, est la dimension du sous-espace vectoriel $\mathrm{Im}(\boldsymbol{A})$. On a $\mathrm{rang}(\boldsymbol{A}) \le \min\{m, n\}$.

### Théorème du rang

Pour toute matrice $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, on a

$$\dim(\ker(\boldsymbol{A})) + \mathrm{rang}(\boldsymbol{A}) = n.$$

# Matrices inversibles, définies, positives

Une matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ est dite *inversible* s'il existe $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ telle que $\boldsymbol{BA} = \boldsymbol{AB} = \mathbf{I}_n$, où $\mathbf{I}_n$ est la matrice identité de $\mathbb{R}^{n \times n}$.
La matrice $\boldsymbol{B}$ est alors unique : elle est appelée l'*inverse de* $\boldsymbol{A}$ et se note $\boldsymbol{A}^{-1}$.

Une matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ est dite *semi-définie positive* si

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} \geq 0.$$

Elle est dite *définie positive* lorsque $\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} > 0$ pour tout vecteur $\boldsymbol{x}$ non nul.

# Sommaire

# Valeurs propres et vecteurs propres

## Definition

Soit $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. On dit que $\lambda \in \mathbb{R}$ est une *valeur propre de* $\boldsymbol{A}$ si

$$\exists \boldsymbol{v} \in \mathbb{R}^n, \boldsymbol{v} \neq \vec{0}, \qquad \boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

Le vecteur $\boldsymbol{v}$ est alors un *vecteur propre associé à la valeur propre* $\lambda$.
L'ensemble des valeurs propres de $\boldsymbol{A}$ s'appelle le *spectre*.

# Valeurs propres et vecteurs propres

## Definition

Soit $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. On dit que $\lambda \in \mathbb{R}$ est une *valeur propre de* $\boldsymbol{A}$ si

$$\exists \boldsymbol{v} \in \mathbb{R}^n, \boldsymbol{v} \neq \vec{0}, \qquad \boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

Le vecteur $\boldsymbol{v}$ est alors un *vecteur propre associé à la valeur propre* $\lambda$. L'ensemble des valeurs propres de $\boldsymbol{A}$ s'appelle le *spectre*.

- Toute matrice de $\mathbb{R}^{n \times n}$ possède $n$ valeurs propres complexes, mais pas nécessairement $n$ valeurs propres réelles.
- Les valeurs propres réelles d'une matrice semi-définie positive (resp. définie positive) sont positives (resp. strictement positives).
- Le noyau de $\boldsymbol{A}$ est engendré par les vecteurs propres associés à la valeur propre 0.

# Décomposition en valeurs propres

**Théorème**

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ symétrique admet une décomposition dite **spectrale** de la forme :

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1},$$

avec

- $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ matrice orthogonale ($\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$), dont les colonnes $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n$ forment une *base orthonormée* de vecteurs propres.
- $\boldsymbol{\Lambda}$ matrice diagonale, qui contient les $n$ valeurs propres de $\boldsymbol{A}$ $\lambda_1, \dots, \lambda_n$ sur la diagonale.

# Décomposition en valeurs propres

**Théorème**

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ symétrique admet une décomposition dite **spectrale** de la forme :

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1},$$

avec

- $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ matrice orthogonale ($\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$), dont les colonnes $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n$ forment une *base orthonormée* de vecteurs propres.
- $\boldsymbol{\Lambda}$ matrice diagonale, qui contient les $n$ valeurs propres de $\boldsymbol{A}$ $\lambda_1, \dots, \lambda_n$ sur la diagonale.

- Il n'y a pas unicité de la décomposition spectrale;
- Aux permutations près, l'ensemble des valeurs propres est unique.

# Sur la décomposition spectrale

## Importance de la décomposition

Chaque valeur propre $\lambda_i$ caractérise l'effet de $\boldsymbol{A}$ sur le vecteur $\boldsymbol{p}_i$ :

- $|\lambda_i| >> 1 \;\Rightarrow\; \|\boldsymbol{A}\boldsymbol{p}_i\| >> \|\boldsymbol{p}_i\|$;
- $|\lambda_i| << 1 \;\Rightarrow\; \|\boldsymbol{A}\boldsymbol{p}_i\| << \|\boldsymbol{p}_i\|$.

## Interprétation géométrique

L'action de la matrice $\boldsymbol{A}$ sur un vecteur $\boldsymbol{x}$

- Allonge les composantes de $\boldsymbol{x}$ dans la base des vecteurs propres associées aux plus grandes valeurs propres en valeur absolue;
- Réduit les composantes de $\boldsymbol{x}$ associées aux valeurs propres les plus faibles en valeur absolue;
- Cas extrême : si $\ker(\boldsymbol{A}) \neq \{\mathbf{0}\}$, toute composante de $\boldsymbol{x}$ selon un vecteur du noyau est réduite à zéro par $\boldsymbol{A}$.

# Sommaire

# Le cas des matrices rectangulaires

## Notion de valeur propre

Soit $\mathbf{A} \in \mathbb{R}^{m \times n}$. On peut parler :

- des valeurs propres de $\mathbf{A}^{\mathrm{T}}\mathbf{A} \in \mathbb{R}^{n \times n}$;
- des valeurs propres de $\mathbf{A}\mathbf{A}^{\mathrm{T}} \in \mathbb{R}^{m \times m}$.

Peut-on s'en servir pour obtenir une décomposition de $\mathbf{A}$ ?

# Le cas des matrices rectangulaires

## Notion de valeur propre

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. On peut parler :

- des valeurs propres de $\boldsymbol{A}^\mathrm{T}\boldsymbol{A} \in \mathbb{R}^{n \times n}$;
- des valeurs propres de $\boldsymbol{A}\boldsymbol{A}^\mathrm{T} \in \mathbb{R}^{m \times m}$.

Peut-on s'en servir pour obtenir une décomposition de $\boldsymbol{A}$ ?

## Observations concernant $\boldsymbol{A}^\mathrm{T}\boldsymbol{A}$

- $\boldsymbol{A}^\mathrm{T}\boldsymbol{A}$ est semi-définie positive;
- $\boldsymbol{A}^\mathrm{T}\boldsymbol{A}$ est symétrique.
- $\ker(\boldsymbol{A}^\mathrm{T}\boldsymbol{A}) = \ker(\boldsymbol{A})$;
- $\mathrm{Im}(\boldsymbol{A}^\mathrm{T}\boldsymbol{A}) = \mathrm{Im}(\boldsymbol{A}^\mathrm{T})$;
- $\mathrm{rang}(\boldsymbol{A}^T\boldsymbol{A}) = \mathrm{rang}(\boldsymbol{A})$.

(On a des résultats similaires pour $\boldsymbol{A}\boldsymbol{A}^\mathrm{T}$.)

# Décomposition en valeurs singulières

## Théorème

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ admet une **décomposition en valeurs singulières** (*SVD*) de la forme

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}.$$

- $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ est orthogonale;
- $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ est orthogonale;
- $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ est diagonale par blocs, avec des coefficients nuls sauf ceux de la diagonale $\{[\boldsymbol{\Sigma}]_{ii}\}_i$ qui sont positifs (ou nuls). Ces éléments, notés $\{\sigma_i\}$, s'appellent les **valeurs singulières de $\boldsymbol{A}$**.

### Théorème

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ admet une **décomposition en valeurs singulières** (*SVD*) de la forme

$$\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^{\mathrm{T}}.$$

- $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ est orthogonale;
- $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ est orthogonale;
- $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ est diagonale par blocs, avec des coefficients nuls sauf ceux de la diagonale $\{[\boldsymbol{\Sigma}]_{ii}\}_i$ qui sont positifs (ou nuls). Ces éléments, notés $\{\sigma_i\}$, s'appellent les **valeurs singulières de $\boldsymbol{A}$**.

- $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^{\mathrm{T}} \Leftrightarrow \boldsymbol{U}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}^{\mathrm{T}} = \boldsymbol{\Sigma}$;
- Une *SVD* n'est pas définie de façon unique mais ses valeurs singulières le sont.

# Pour la suite

## Cours

- La décomposition en valeurs singulières (SVD) : outil fondamental;
- Applications.

## TD

- Quelques exercices d'algèbre linéaire;
- Preuve constructive de la SVD.

**Exercice (I-1 polycopié)**

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Démontrer les propriétés suivantes sans utiliser la décomposition en valeurs singulières :

a) $\ker(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \ker(\boldsymbol{A})$;

b) $\mathrm{rang}(\boldsymbol{A}^{\mathrm{T}}) = \mathrm{rang}(\boldsymbol{A})$;

c) $\mathrm{rang}(\boldsymbol{A}^{T}\boldsymbol{A}) = \mathrm{rang}(\boldsymbol{A})$, et en déduire que $\mathrm{Im}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathrm{Im}(\boldsymbol{A}^{\mathrm{T}})$.

Comment les démontrer en utilisant maintenant la SVD ?