# Mathématiques pour l’intelligence artificielle

M2 informatique parcours IAAA, Aix-Marseille Université

Thomas Schatz

Lundi 4 septembre 2023

# Présentations

- A propos de moi: Thomas Schatz, maître de conférences, recherche à l'intersection entre apprentissage non-supervisé de représentations en IA et études du développement de la perception chez le bébé en sciences cognitives

- A propos de vous: nom, prénom, formation en mathématique après le bac, ce que vous envisagez de faire après le M2?

# Questions pratiques

- Calendrier des cours sur le site IAAA + ENT AMU
- Questions, etc. mattermost M2 IAAA, canal MIA
- Matériel de cours sur mon site web (https://thomas.schatz.cogserver.net/teaching/)
- Evaluation:
  - 1 point : scribe pour une séance
  - 5 points : moyennes des exercices à rendre en début de chaque séance (information sur le site du cours)
  - 14 points : terminal lors de la dernière séance (en deux parties notées séparément pour des raisons techniques), durée TBD
  - Questions ?

# Pourquoi ce cours?

# Plan du cours (prévisionnel)

Introduction

1. Notions de bases sur les preuves

2. Algèbre linéaire

3. Optimisation

4. Probabilités

5. Statistique

Discussion

# Preuves mathématiques

Méthode générale

1. Formaliser les hypothèses et le résultat à prouver
   - Identifier les objets mathématiques impliqués et les nommer
   - Appliquer les définitions formelles des notions impliquées
2. Montrer que les hypothèses impliquent logiquement le résultat
   - Utiliser uniquement les théorèmes établis (et les axiomes de la théorie dans laquelle on travaille, par exemple les axiomes usuels de la théorie des ensembles) et les règles du raisonnement logique
   - Trouver les idées de preuve en réfléchissant à partir d'exemple particuliers, par analogie etc.

Exemple: prouver que la somme de deux nombres impairs est paire

# Différents types de preuves

- Nombre limité de motifs de preuves basiques
  - Preuve directe, par disjonction des cas, par récurrence, par l'absurde, par contraposition, par contre-exemple (exercice 1.1)
- Composés hiérarchiquement pour construire des preuves plus complexes (exercice 1.2)
- Motifs qui reviennent souvent, souvent associés à un domaine donné, e.g. arithmétique, théorie des ensembles, algèbre linéaire, analyse, probabilités… (exercice 1.3)

# Language logico-mathématique

Connecteurs logiques, quantificateurs, théorie des ensembles

$\wedge, \vee, \neg, \Rightarrow, \Leftrightarrow$ $\forall, \exists, \exists!$ $x \in E, E \subset F, E \cup F, E \cap F, E \setminus F$

# Références


- Devlin, K. J. (2012). *Introduction to mathematical thinking*.


- PDF et plus de références sur la page du cours

# Plan du cours (prévisionnel)

Introduction

1. Notions de bases sur les preuves

2. Algèbre linéaire

3. Optimisation

4. Probabilités

5. Statistique

Discussion

# Algèbre linéaire

# Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires
2. Matrices
3. Angles et orthogonalité
4. Structure des applications linéaires

# Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires

2. Matrices

3. Angles et orthogonalité

4. Structure des applications linéaires

# Objets mathématiques

- Types d'objets en mathématiques nombres entiers, fractionnels, réels, complexes, fonctions, ensembles, distributions de probabilité…
- Comme en informatique mais souvent plus idéalisé (par exemple nombre réels vs nombre IEEE floating point)
- Qu'est-ce qu'une fonction en mathématiques ?

# Fonction en mathématiques

- Qu'est-ce qu'une fonction en mathématiques ?

$f = (E, F, G) \qquad G \subset E \times F$

Pour tout $x \in E$, il existe un unique $y \in F$, tel que $(x, y) \in G$

- Notation $\quad f : \left| \begin{array}{ccc} E & \to & F \\ x & \mapsto & f(x) \end{array} \right.$

- Exemples

$$+ : \left| \begin{array}{ccc} \mathbf{R}^2 & \to & \mathbf{R} \\ x, y & \mapsto & x + y \end{array} \right. \qquad \exp : \left| \begin{array}{ccc} \mathbf{R} & \to & \mathbf{R} \\ x & \mapsto e^x & \end{array} \right.$$

$$\nabla : \left| \begin{array}{ccc} \mathcal{C}^1(\mathbf{R}^n, \mathbf{R}) & \to & \mathcal{C}^0(\mathbf{R}^n, \mathbf{R}^n) \\ f & \mapsto & \nabla f \end{array} \right.$$

# Fonction linéaire

- Objet central de l'algèbre linéaire (du point de vue conceptuel)
- Idée générale: fonction qui préserve les combinaisons linéaires
- Définition formelle

$$f : \left| \begin{array}{ccc} E & \to & F \\ x & \mapsto & f(x) \end{array} \right. \quad \text{est une fonction linéaire si et seulement si}$$

$E$ et $F$ sont des espaces vectoriels sur un même corps $K$ et pour toute paire de scalaires $(\alpha, \beta) \in K^2$ et toute paire de vecteurs $(u, v) \in E^2$,

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$$

# Espace vectoriel

## Définition

$$(E, K, +, .) \quad + : \left| \begin{array}{ccc} E^2 & \to & E \\ x, y & \mapsto & x + y \end{array} \right. \qquad . : \left| \begin{array}{ccc} K \times E & \to & E \\ \alpha, x & \mapsto & \alpha . x \end{array} \right.$$

est un espace vectoriel, si et seulement si $K$ est un corps et

Il existe $0_E \in E$, tel que
Pour tout $(u, v, w) \in E^3$,
Pour tout $(\alpha, \beta) \in K^2$
Il existe $u^{-1} \in E$, tel que

| | |
|---|---|
| $u + (v + w) = (u + v) + w$ | associativité |
| $u + v = v + u$ | commutativité |
| $u + 0_E = u$ | élément neutre |
| $u + u^{-1} = 0_E$ | inverse |
| $\alpha.(\beta.u) = (\alpha\beta).u$ | compatibilité |
| $(\alpha + \beta)u = \alpha.u + \beta.u$ | distributivité |
| $\alpha(u + v) = \alpha.u + \alpha.v$ | distributivité |
| $1_K.u = u$ | identité multiplicative |

# Espace vectoriel

Exemples

- espaces Euclidiens $\mathbf{R}^n$
- espaces Hermitiens $\mathbf{C}^n$
- espace des fonctions continues sur $[a, b] \subset \mathbf{R}$

à valeurs réelles $\mathcal{C}^0([a, b], \mathbf{R})$

# Fonction linéaire

- Objet central de l'algèbre linéaire (du point de vue conceptuel)
- Idée générale: fonction qui préserve les combinaisons linéaires
- Définition formelle

$$f : \left| \begin{array}{ccc} E & \to & F \\ x & \mapsto & f(x) \end{array} \right.$$ est une fonction linéaire si et seulement si

$E$ et $F$ sont des espaces vectoriels sur un même corps $K$ et pour toute paire de scalaires $(\alpha,\, \beta) \in K^2$ et toute paire de vecteurs $(u,\, v) \in E^2$,

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$$

# Notions élémentaires sur les espaces vectoriels

**Définition (Vect)**

Espace vectoriel engendré par une famille de vecteurs $(v_i)_{i\in I} \in E^I$:

$$\mathrm{Vect}((v_i)_{i\in I}) := \left\{ \sum_{i\in I} \lambda_i v_i \mid (\lambda_i)_{i\in I} \in K^I, \text{with finitely many } \lambda_i \text{ different from zero} \right\}$$

**Définition (indépendance linéaire)**

$v \in E$ est *linéairement indépendant* de $(v_i)_{i\in I} \in E^I$ si et seulement si $v$ n'appartient pas à $\mathrm{Vect}((v_i)_{i\in I})$.

**Définition (famille libre)**

Une famille de vecteurs $(v_i)_{i\in I} \in E^I$ est dite *libre* si et seulement si pour tout $j \in I$, $v_j$ est linéairement indépendant de $(v_i)_{i\in I\setminus\{j\}}$.

**Définition (base)**

$(v_i)_{i\in I}$ est une base de $E$ si et seulement si $(v_i)_{i\in I}$ est une famille libre de $E$ telle que $E = Vect((v_i)_{i\in I})$ (i.e. la famille $(v_i)_{i\in I}$ engendre $E$).

**Définition (dimension finie)**

Un espace vectoriel $E$ est dit de dimension finie si et seulement si il existe une base de $E$ de formée d'un nombre fini de vecteurs.

**Théorème (dimension)**

Si $E$ est un espace vectoriel de dimension finie, il existe un unique entier naturel $n$, la dimension de $E$, tel que toute base de $E$ est formée de exactement $n$ vecteurs.

# Caractérisation de l'indépendance linéaire

### Théorème

Soit $(v_i)_{i\in I} \in E^I$ est une famille de vecteurs de l'espace vectoriel $E$.

- S'il existe une famille $(\lambda_i)_{i\in I} \in K^I$ de scalaires avec au plus un nombre fini de $\lambda_i$ différents de 0 et telle que :

  1. il existe $i \in I$, tel que $\lambda_i \neq 0$
  2. $\sum_{i\in I} \lambda_i v_i = 0$

  alors $(v_i)_{i\in I}$ est une famille liée.

- Réciproquement, si pour toute famille $(\lambda_i)_{i\in I} \in K^I$ de scalaires avec au plus un nombre fini de $\lambda_i$ différents de 0, on a:

$$\left(\sum_{i\in I} \lambda_i v_i = 0\right) \Rightarrow \left(\text{ pour tout } i \in I,\ \lambda_i = 0\right),$$

  alors $(v_i)_{i\in I}$ est une famille libre.

# Décomposition sur une base

**Définition (décomposition sur une base)**

Soit une base $V = (v_i)_{i \in I}$ de E et $\Lambda = (\lambda_i)_{i \in I} \in K^I$ une famille de scalaires telle que au plus un nombre fini des $\lambda_i$ soient différents de 0.

On dit que $\Lambda$ est une décomposition de $v \in E$ sur la base $V$ si et seulement si $v = \sum_{i \in I} \lambda_i v_i$.

**Théorème (existence et unicité de la décomposition)**

Etant donné une base $V = (v_i)_{i \in I}$ de $E$, tout vecteur $v \in E$ peut-être décomposé sur cette base et cette décomposition est unique.

Toute base peut donc servir de système de coordonnées pour E

# Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires
2. Matrices
3. Angles et orthogonalité
4. Structure des applications linéaires

# Matrice

- Objet central de l'algèbre linéaire (en pratique)
- Idée générale: simplement un tableau de nombres en deux dimensions

$$\mathrm{M} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np} \end{pmatrix}$$

# Matrice d'une fonction linéaire

Multiplication matrice/vecteur colonne: $(Mv)_i := \sum_{j=1}^{n} m_{ij} v_i$

Multiplication matrice-matrice: $(AB)_{ij} := \sum_{k=1}^{n} a_{ik} b_{kj}$

# Propriétés du produit matriciel

$$A \times (B \times C) = (A \times B) \times C$$

$$A \times (B + C) = A \times B + A \times C$$

$$(A + B) \times C = A \times C + B \times C$$

$$c(A \times B) = (cA) \times B = A \times (cB)$$

# Transposée

Transposée de A : obtenue en échangeant les lignes et les colonnes de A

$$(A+B)^{\mathsf{T}} = A^{\mathsf{T}} + B^{\mathsf{T}}, \qquad (\alpha A)^{\mathsf{T}} = \alpha A^{\mathsf{T}}$$

$$(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}$$

# Matrice

- Objet central de l'algèbre linéaire (en pratique)
- Idée générale: simplement un tableau de nombres en deux dimensions

$$M = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np} \end{pmatrix}$$

# Matrice

- Objet central de l'algèbre linéaire (en pratique)
- Idée générale: simplement un tableau de nombres en deux dimensions
- Pourquoi est-ce utile ?

$$
M = \begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\
a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\
\vdots & \vdots & & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\
\vdots & \vdots & & \vdots & & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np}
\end{pmatrix}
$$

# Matrice

- Objet central de l'algèbre linéaire (en pratique)
- Idée générale: simplement un tableau de nombres en deux dimensions
- Pourquoi est-ce utile ?
    - Permet de représenter les fonctions linéaires en dimensions finie de manière simple à appréhender et pratique pour les calculs

$$\mathrm{M} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np} \end{pmatrix}$$

# Matrice d'une fonction linéaire

$E$, $F$ espaces vectoriels de dimension finie
$V = (v_1, ..., v_p)$ base de $E$
$W = (w_1, ..., w_n)$ base de $F$
$f : E \to F$, fonction linéaire
$M = M(f, V, W)$ matrice de $f$ de $V$ vers $W$.

$$
\mathrm{M} = \begin{array}{cc}
\begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\
a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\
\vdots & \vdots & & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\
\vdots & \vdots & & \vdots & & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np}
\end{pmatrix}
&
\begin{array}{l}
\leftarrow w_1 \\
\leftarrow w_2 \\
\\
\leftarrow w_i \\
\\
\leftarrow w_n
\end{array}
\\
\begin{array}{cccccc}
\uparrow & \uparrow & & \uparrow & & \uparrow \\
f(v_1) & f(v_2) & & f(v_j) & & f(v_p)
\end{array}
&
\end{array}
$$

# Matrice d’une fonction linéaire

Multiplication matrice/vecteur colonne: $(Mv)_i := \sum_{j=1}^{n} m_{ij} v_i$

Multiplication matrice-matrice: $(AB)_{ij} := \sum_{k=1}^{n} a_{ik} b_{kj}$

# An example vector space

$$\mathbb{R}_2[X] = \mathrm{Vect}(1, X, X^2)$$

$$P \in \mathbb{R}_2[X] \Leftrightarrow \big(\text{there exists } (a_0, a_1, a_2) \in \mathbb{R}^3,\ P = a_0 + a_1 X + a_2 X^2\big)$$

Formellement:

$$1 = (1, 0, 0, \dots)$$
$$X = (0, 1, 0, \dots)$$
$$X^2 = (0, 0, 1, \dots)$$

Soit $x \in \mathbb{R}$

$$e_x : \left| \begin{array}{ccc} \mathbb{R}_2[X] & \to & \mathbb{R} \\ P = a_0 + a_1 X + a_2 X^2 & \mapsto & P(x) = a_0 + a_1 x + a_2 x^2 \end{array} \right.$$

Soit $M$ une matrice à coefficients réels de taille $n$ pas $n$

$$e_M : \left| \begin{array}{ccc} \mathbb{R}_2[X] & \to & \mathcal{M}_n(\mathbb{R}) \\ P = a_0 + a_1 X + a_2 X^2 & \mapsto & P(M) = a_0 + a_1 M + a_2 M^2 \end{array} \right.$$

# Matrice d'une fonction linéaire, exemple

$V = (1, X, X^2)$ base de $\mathbb{R}_2[X]$

$W = (1, X, X^2, X^3)$ base de $\mathbb{R}_3[X]$

$$f : \left| \begin{array}{ccc} \mathbb{R}_2[X] & \to & \mathbb{R}_3[X] \\ P = a_0 + a_1 X + a_2 X^2 & \mapsto & f(P) = (a_2 + a_0)X^3 + a_1 X^2 + a_0 \end{array} \right.$$

$M(f, V, W)$?

$$\mathrm{M} = \begin{array}{c} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np} \end{pmatrix} \begin{array}{l} \leftarrow w_1 \\ \leftarrow w_2 \\ \\ \leftarrow w_i \\ \\ \leftarrow w_n \end{array} \\ \begin{array}{cccccc} \uparrow & \uparrow & & \uparrow & & \uparrow \\ f(v_1) & f(v_2) & & f(v_j) & & f(v_p) \end{array} \end{array}$$

# Matrice d’une fonction linéaire

L’évaluation en un point d’une fonction linéaire et la composition d’application linéaire se réduit (en dimension finie) à l’application “mécanique” de règles de calcul matriciel

Pas trop “mécanique” quand même:

Crimes contre les matrices
http://ee263.stanford.edu/notes/matrix_crimes.pdf

# Changement de base

Toute base peut servir de système de coordonnées pour E

Changement de base == changement de système de coordonnées

Matrice de passage de V à W: $P(V, W) = M(Id_E, V, W)$

Définition: transforme un vecteur v exprimé en terme de ses coordonnées dans V, en le même vecteur exprimé en terme de ses coordonnées dans W

Pour la trouver: les colonnes de P(V,W) sont les coordonnées des éléments de V (dans un ordre fixé) dans W

# Matrice par blocs

La matrice

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \end{bmatrix}$$

peut être partitionnée en quatre blocs

$$\mathbf{P}_{11} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathbf{P}_{12} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}, \mathbf{P}_{21} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \\ 3 & 3 \end{bmatrix}, \mathbf{P}_{22} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

On peut alors écrire la matrice par bloc comme :

$$\mathbf{P}_{\text{partitionnee}} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}.$$

# Interpretation des opérations matricielles

write $A \in \mathbf{R}^{m \times n}$ in terms of its columns:

$$A = \left[ \begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array} \right]$$

then $y = Ax$ can be written as

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

write $A$ in terms of its rows:

$$A = \left[ \begin{array}{c} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_n^T \end{array} \right]$$

then $y = Ax$ can be written as

$$y = \left[ \begin{array}{c} \tilde{a}_1^T x \\ \tilde{a}_2^T x \\ \vdots \\ \tilde{a}_m^T x \end{array} \right]$$

# Interpretation des opérations matricielles

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

$$C = \left[ \begin{array}{ccc} c_1 & \cdots & c_p \end{array} \right] = AB = \left[ \begin{array}{ccc} Ab_1 & \cdots & Ab_p \end{array} \right]$$

$$C = \left[ \begin{array}{c} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{array} \right] = AB = \left[ \begin{array}{c} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{array} \right]$$

$$C = \sum_i a_i \tilde{b}_i^T$$

# Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires
2. Matrices
3. Angles et orthogonalité
4. Structure des applications linéaires et matrices