# Mathématiques pour l'intelligence artificielle (MIA)

M2 IAAA, AMU, 2021-2022

Thomas Schatz

# Questions pratiques

- Calendrier des cours sur ADE
- Questions, etc.: sur le mattermost du M2 IAAA
- Matériel de cours sur ametice (à venir)
- Evaluation: 3 DMs notés (un toutes les 3 séances, 3 semaines pour le faire, n'attendez pas le dernier moment pour commencer) + scribe pour une séance
- Questions ?

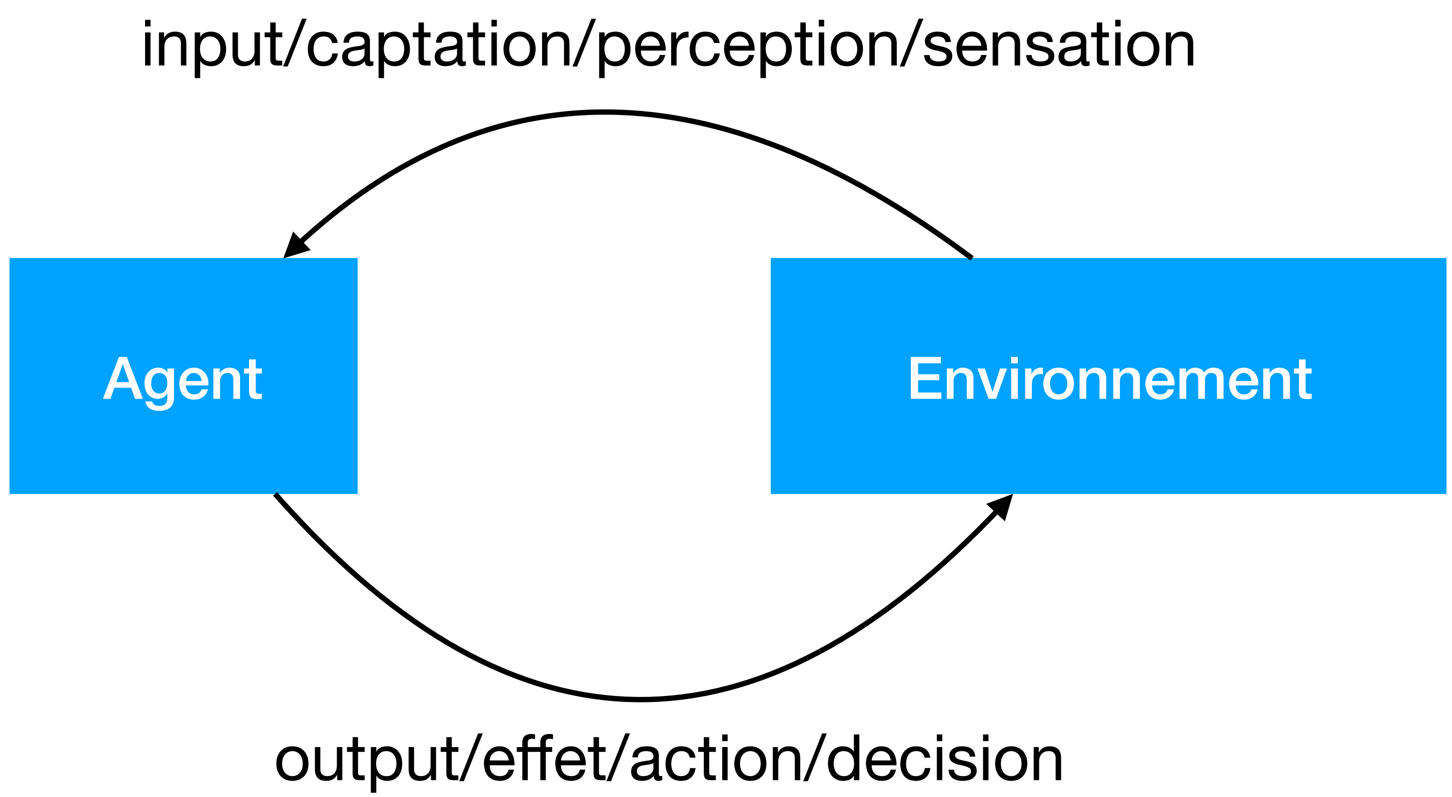
# Partie 1 — Introduction générale

- Pourquoi êtes vous ici ?

- Pourquoi êtes vous ici ?

- Quelles mathématiques en intelligence artificielle et pourquoi ?

# Problème général en IA/ML (informel)




On cherche à determiner une politique d’action/décision pour l’agent qui optimise une métrique d’évaluation

Problème de recherche -> algorithmique, optimisation, optimisation continue, algèbre linéaire

La métrique d’évaluation n’est pas directement accessible
-> modélisation probabiliste, statistique, algèbre linéaire

**IA/ML: Quand ces deux problèmes (grand espace de recherche + incertitude et besoin de généraliser) se rencontrent**

# Plan du cours

1. Introduction générale
2. Preuves (revue)
3. Algèbre linéaire (revue)
4. Optimisation (revue)
5. Optimisation sous contraintes
6. Probabilités (revue)
7. Statistique
8. **Théorie de l’apprentissage**
9. **Optimisation pour l’apprentissage**

# Partie 2 — Preuves (revue)

# Différents types de preuves

1. Prouver que la somme de deux nombre entiers impairs est paire.
2. Prouver que pour tout nombre entier $n$, $n^2 + n$ est pair.
3. Montrer que pour tout entier $n$, $2^{3n+1} + 5$ est un multiple de 7.
4. Soit $a$ un nombre rationnel et $b$ un nombre irrationnel.
   Prouver que $a + b$ est irrationnel.
5. Montrer que si $n$ est un nombre entier et $n^2$ est impair,
   alors $n$ est impair.

# Références

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

Stefanowicz, A., Kyle, J., & Grove, M. (2014). *Proofs and mathematical reasoning*.

# Partie 3 —Algèbre linéaire (revue)

# Objets mathématiques

- Types d'objets en mathématiques nombres entiers, fractionnels, réels, complexes, fonctions, ensembles, distribution de probabilités…
- 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 ?

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

- 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

$$f : \left| \begin{array}{ccc} E & \to & F \\ x & \mapsto & f(x) \end{array} \right.$$ 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 $(u, v) \in E^2$,

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

# Indépendance linéaire, base, dimension…

**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.

# Décomposition sur une base

**Théorème (indépendance linéaire)**

Soit $(v_i)_{i\in I} \in E^I$ est une famille libre et $(\lambda_i)_{i\in I} \in K^I$ une famille de scalaires telle que au plus un nombre fini des $\lambda_i$ soient différents de 0. Alors $\sum_{i\in I} \lambda_i v_i = 0$ implique que pour tout $i \in I$, $\lambda_i = 0$.

**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

# Matrice

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

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

# Matrice

- 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

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

# 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 $W$.

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

# 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

# Matrice d’une fonction linéaire

Multiplication matrice/vecteur colonne: règle et interprétation

Multiplication matrice-matrice: règle et interprétation

# 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 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

# Orthogonalité

Définition: dans un espace vectoriel réel, un *produit scalaire* est une forme bilinéaire symétrique définie positive

Forme : $( \mid ) : \left| \begin{array}{ccc} E \times E & \to & \mathbf{R} \\ (v, w) & \mapsto & (v \mid w) \end{array} \right.$

Bilinéaire : linéaire en v pour w fixé et en w pour v fixé

Symétrique : $\text{Pour tout } v, w \in E^2,\ (v \mid w) = (w \mid v)$

Définie positive : $\text{Pour tout } v \in E \setminus \{0_E\}, (v \mid v) > 0$

# Orthogonalité

Norme associée au produit scalaire : $\|v\| = \sqrt{(v \mid v)}$

Exemple: produit scalaire et norme Euclidienne sur $\mathbf{R}^n$

Vecteurs orthogonaux: si et seulement si $(v \mid v) = 0$

Famille orthogonale:

$$(v_i)_{i \in I},\ \text{telle que pour tout } (i,j) \in I^2,\ i \neq j,\ (v_i \mid v_j) = 0$$

Famille orthonormale: famille orthogonale telle que

$$\text{pour tout } i \in I,\ \|v_i\| = 1$$

# Orthogonalité

Théorème : une famille de vecteurs orthogonaux est libre

Comment construire des familles orthonormales?

Gram-Schmidt procedure : étant donné $k$ vecteurs indépendants $(a_1, a_2, \ldots, a_k) \in E^k$, trouver des vecteurs orthonormaux $(q_1, q_2, \ldots, a_k)$, tels que pour tout $r \leq k$, $(q_1, q_2, \ldots, q_r)$ soit une base orthonormale de $\mathrm{Vect}(a_1, a_2, \ldots, a_r)$

- step 1a. $\tilde{q}_1 := a_1$
- step 1b. $q_1 := \tilde{q}_1 / \|\tilde{q}_1\|$ (normalize)
- step 2a. $\tilde{q}_2 := a_2 - (q_1^T a_2) q_1$ (remove $q_1$ component from $a_2$)
- step 2b. $q_2 := \tilde{q}_2 / \|\tilde{q}_2\|$ (normalize)
- step 3a. $\tilde{q}_3 := a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2$ (remove $q_1$, $q_2$ components)
- step 3b. $q_3 := \tilde{q}_3 / \|\tilde{q}_3\|$ (normalize)
- etc.

# Crédits

Certaines figures de

- Strang, Gilbert. *Linear algebra and learning from data*. Cambridge: Wellesley-Cambridge Press, 2019
- Boyd, Stephen EE263 - Introduction to Linear Dynamical Systems (https://see.stanford.edu/Course/EE263)
- Wikipedia