# Mathématiques pour l'intelligence artificielle (MIA)

M2 IAAA, AMU, 2021-2022

Thomas Schatz

# Quatrième cours

- DM1 en ligne à rendre pour le 6 Octobre (en main propre ou par email: thomas.schatz@univ-amu.fr
- Scribes : Jean-Eudes Ayilo et Etienne Roux (rendu = fichier .tex)
- Aujourd'hui
  - Optimisation et optimisation sous contraintes

# 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**

# Jacobienne, gradient

$$f : U \subset \mathbf{R}^n \to \mathbf{R}^m, \text{ de classe } \mathrm{C}^1$$

**Dérivée partielle** $f : x \mapsto (f_1(x), f_2(x), \ldots, f_m(x))$

$$\frac{\partial f_i}{\partial x_j} : (a_1, \ldots, a_n) \mapsto \lim_{h \to 0} \frac{f_i(a_1, \ldots, a_{j-1}, a_j + h, a_{j+1}, \ldots a_n) - f_i(a_1, \ldots, a_j, \ldots, a_n)}{h}$$

**Matrice Jacobienne (transposée du gradient si m=1)**

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

**“Chain rule”**

$$J_{f \circ g}(\mathbf{a}) = J_f(g(\mathbf{a})) J_g(\mathbf{a}),$$

# Optimisation sans contraintes

$$f : \mathbf{R}^n \mapsto \mathbf{R}$$

$$\min_x f(x) \; ?$$

$x^*$ est un minimum global de $f$ ssi pour tout $x \in \mathbf{R}^n$, $f(x^*) \leq f(x)$

$x^*$ est un minimum local de $f$ ssi il existe un ensemble ouvert $U \subset \mathbf{R}^n$ contenant $x^*$ tel que pour tout $x \in U$, $f(x^*) \leq f(x)$

# Conditions d’optimalité

## Conditions nécessaires, premier ordre

Si $x^*$ est un minimum local et $f$ est de classe $\mathcal{C}^1$ sur un ouvert $U \subset \mathbf{R}^n$ contenant $x$, alors $\nabla f(x^*) = 0$

# Normes de matrices

A matrix norm is a function $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ that has the following properties:

- $\|A\| \geq 0$ for any $A \in \mathbb{R}^{m\times n}$, and $\|A\| = 0$ if and only if $A = 0$
- $\|\alpha A\| = |\alpha|\|A\|$ for any $m \times n$ matrix $A$ and scalar $\alpha$
- $\|A + B\| \leq \|A\| + \|B\|$ for any $m \times n$ matrices $A$ and $B$

$$\|A\| = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$$

$$\|A\|_2 = \sigma_1$$

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}.$$

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

# Plan du cours

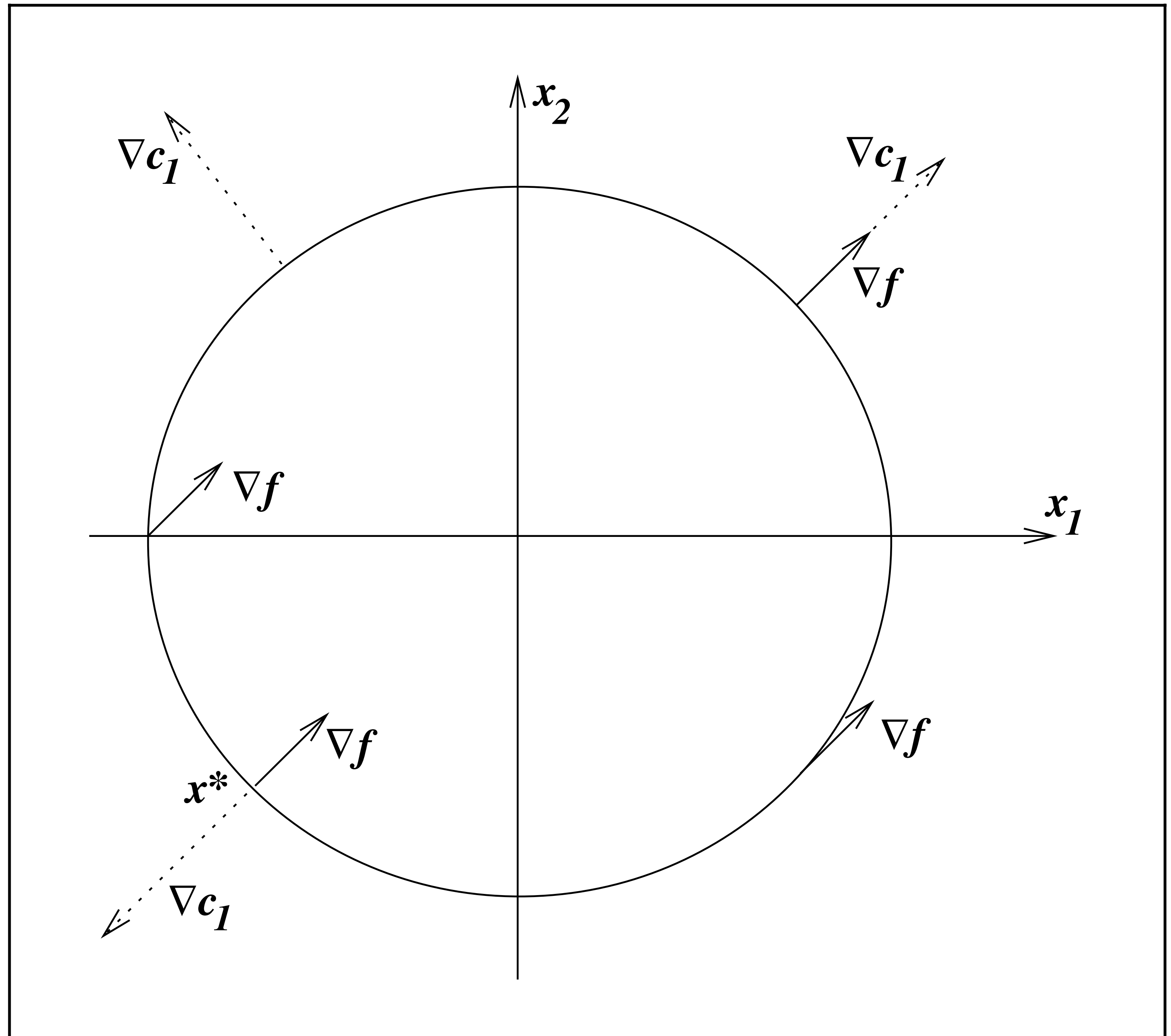
1. Introduction générale
2. Preuves (revue)
3. Algèbre linéaire (revue)
4. Optimisation (revue)
5. Optimisation sous contraintes
6. Probabilités (revue)
7. Statistique
8. **Théorie de l'apprentissage**
9. **Optimisation pour l'apprentissage**

# Optimisation sous contraintes

$$\min_{x \in \mathbf{R}^n} f(x) \text{ subject to } \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}$$

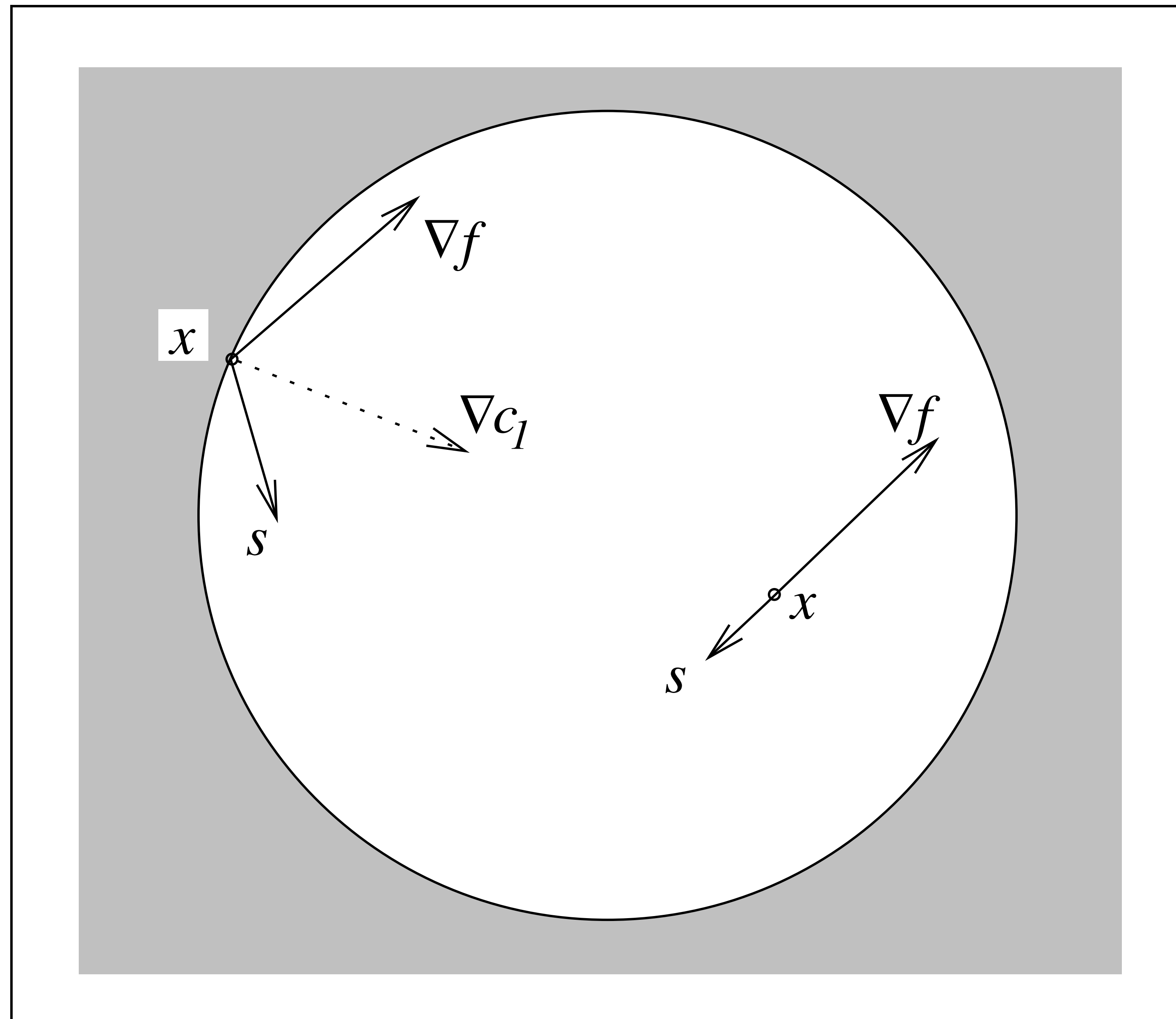
$x^*$ est une solution locale du problème ssi $x^* \in \Omega$ et il existe un ensemble ouvert $U \subset \mathbf{R}^n$ contenant $x^*$ tel que pour tout $x \in U \cap \Omega$, $f(x^*) \leq f(x)$

# Condition nécessaire d'optimalité: intuition




$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*).$$

# Condition nécessaire d'optimalité: intuition




$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*).$$

$$\lambda_1^* c_1(x^*) = 0.$$

# Condition nécessaire d'optimalité: KKT

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x).$$

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\
c_i(x^*) &= 0, \quad && \text{for all } i \in \mathcal{E}, \\
c_i(x^*) &\geq 0, \quad && \text{for all } i \in \mathcal{I}, \\
\lambda_i^* &\geq 0, \quad && \text{for all } i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, \quad && \text{for all } i \in \mathcal{E} \cup \mathcal{I}.
\end{aligned}$$

# Crédits

Certaines figures de

- Jorge Nocedal and Stephen Wright
- Wikipedia