Mathématiques pour l’intelligence artificielle

M2 informatique parcours IAAA, Aix-Marseille Université

Thomas Schatz
Mardi 10 septembre 2024

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 points : scribe pour une séance
  - 2 points : exercices à rendre (correction d'une sélection aléatoire)
  - 7 points : partiel (1er Octobre, 2h)
  - 10 points : terminal (15 Octobre, 2h)
- Questions ?

Plan du cours (prévisionnel)

1. Optimisation : optimisation sous contraintes, dualité

2. Statistique : cadre général, calculs de biais et variance d'un estimateur

Partiel

3. Applications à l'apprentissage automatique

Terminal

Optimisation

Optimisation

$$f : \mathcal{D} = \mathrm{dom}(f) \subset \mathbf{R}^n \to \mathbf{R}$$

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

- Minimisation d'une fonction sur un domaine
  - Domaine “ouvert” ou non ?
  - Domaine et fonction convexes ?
  - Fonction lisse ou non ?

Optimisation

1. Optimisation sans contraintes

2. Optimisation sous contraintes

3. Analyse convexe et dualité

4. Analyse des algorithmes d'optimisation

Définitions

$$f : \mathcal{D} = \mathrm{dom}(f) \subset \mathbf{R}^n \to \mathbf{R}$$

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

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

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

Notions de base de topologie

Dans $\mathbf{R}^n$, La boule fermée $B_f(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x - y\|_2 \leq \epsilon$

Dans $\mathbf{R}^n$, La boule ouverte $B_o(x, \epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x - y\|_2 < \epsilon$

U est un ouvert de $\mathbf{R}^n$ ssi $U \subset \mathbf{R}^n$ et pour tout x dans U, il existe $\epsilon > 0$, tel que $B_f(x, \epsilon) \subset U$.

U est un fermé de $\mathbf{R}^n$ ssi son complément dans $\mathbf{R}^n$ est un ouvert de $\mathbf{R}^n$

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)

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

"Chain rule"

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

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 $\nabla f(x^*) = 0$

Analyse

Hessienne

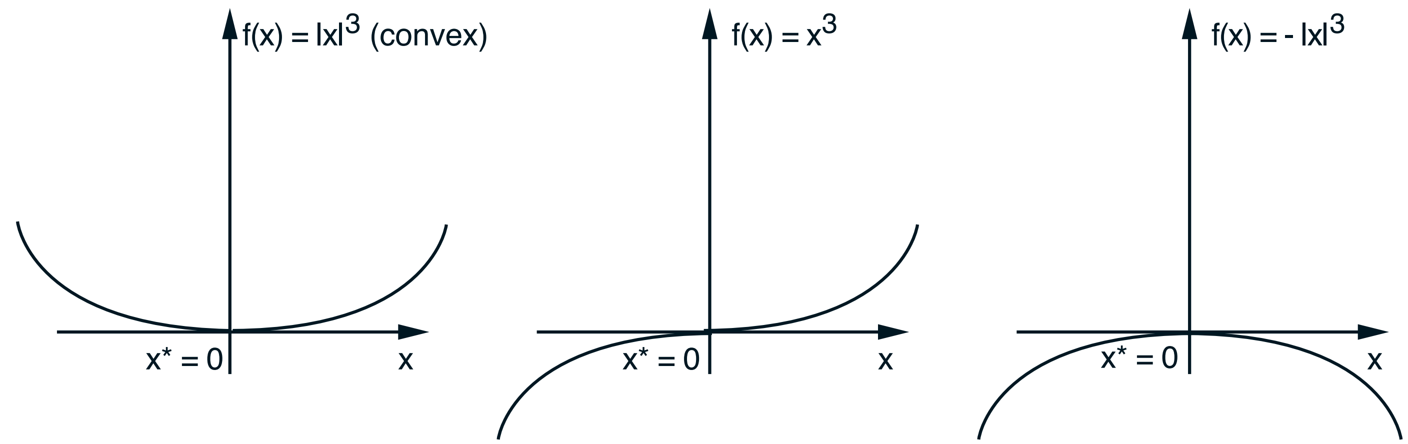
$$f : U \subset \mathbf{R}^n \mapsto \mathbf{R}$$

$$\nabla^2 f = H(f) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial {x_1}^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial {x_2}^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial {x_n}^2} \end{bmatrix}$$

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 $\nabla f(x^*) = 0$



Conditions nécessaires, second ordre

Si x^* est un minimum local et f est de classe $\mathcal{C}^2$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, alors $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ est semi-définie positive

Conditions suffisantes, second ordre

Si f est de classe $\mathcal{C}^2$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, si $\nabla f(x^*) = 0$ et si $\nabla^2 f(x^*)$ est définie positive, alors x^* est un minimum local.

Existence de minimum

Théorème de Weierstrass

Si f est définie et continue sur un sous-ensemble fermé et borné de $\mathbf{R}^n$, alors f admet un minimum global.

Théorème

Si f est définie et continue sur $\mathbf{R}^n$ et *coercive* (i.e. $f(x) \to +\infty$ when $\|x\| \to +\infty$, alors f admet un minimum global sur tout sous-ensemble fermé de $\mathbf{R}^n$.

Théorème

Si f est convexe et minorée, alors f admet un minimum global.

Convexité

L'ensemble X est convexe ssi pour tout x, y dans X, le segment $[x, y] := \{tx + (1-t)y \mid t \in [0,1]\}$ est inclus dans X

Une fonction $f : X \subset \mathbf{R}^n \to R$ est convexe ssi X est convexe et pour tout x, y dans X et pour tout $t \in [0,1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Si l'inégalité est stricte pour $t \in]0,1[$, la fonction est dite *strictement* convexe.

Unicité du minimum

Si f est une fonction *strictement* convexe définie sur un ensemble convexe $X \subset \mathbf{R}^n$, alors f admet au plus un minimum global.

Si f est une fonction convexe définie sur un ensemble convexe $X \subset \mathbf{R}^n$, alors tout minimum local de f est aussi un minimum global de f. Si X est ouvert, alors $\nabla f(x^*) = 0$ équivaut à x^* est un minimum global de f

Convexité

Reconnaître une fonction convexe

- les fonctions linéaires sont convexes
- les combinaisons linéaires positives de fonctions convexes sont convexes
- le maximum (point par point) d'un ensemble de fonctions convexes est convexe
- …

Si f est de classe C^2, f est convexe si et seulement si $\nabla^2 f$ est semi-définie positive sur l'intérieur de X. Si $\nabla^2 f$ est définie positive sur l'intérieur de X, f est strictement convexe. $(X = \mathrm{dom} f)$

Optimisation

1. Optimisation sans contraintes
2. Optimisation sous contraintes
3. Analyse convexe et dualité
4. Analyse des algorithmes d’optimisation

Optimisation sous contraintes

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

$$\Omega = \{x \in \mathbf{R}^n \mid \text{pour tout } i \in \mathcal{E},\ c_i(x) = 0,\ \text{pour tout } j \in \mathcal{I},\ c_j(x) \geq 0\}$$

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

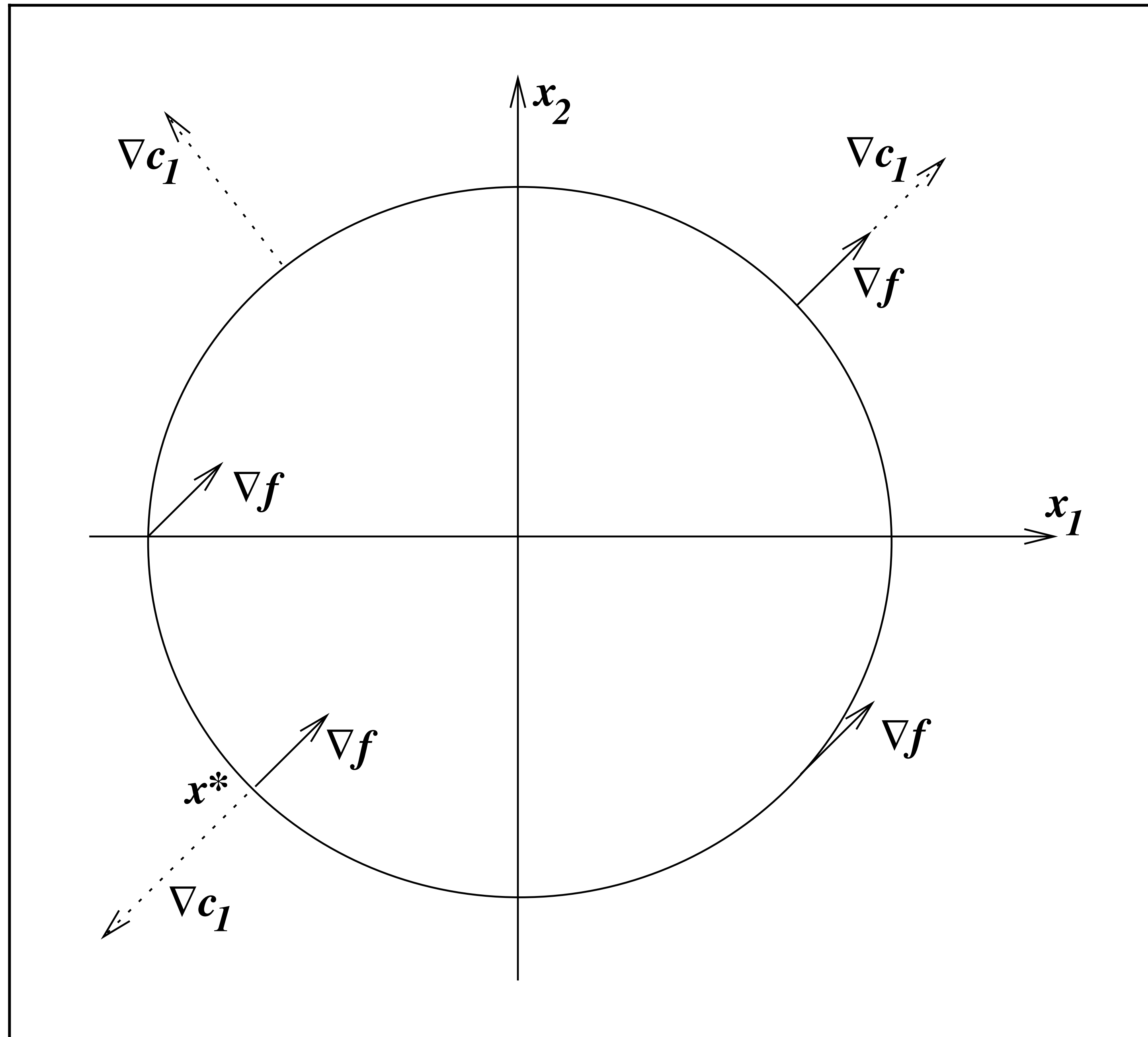
$$f : x_1, x_2 \mapsto x_1 + x_2$$

Cas d’une seule contrainte d’égalité

$$x_1^2 + x_2^2 = 2$$

Condition nécessaire d’optimalité

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



Condition nécessaire d'optimalité: intuition

$$f : x_1, x_2 \mapsto x_1 + x_2$$

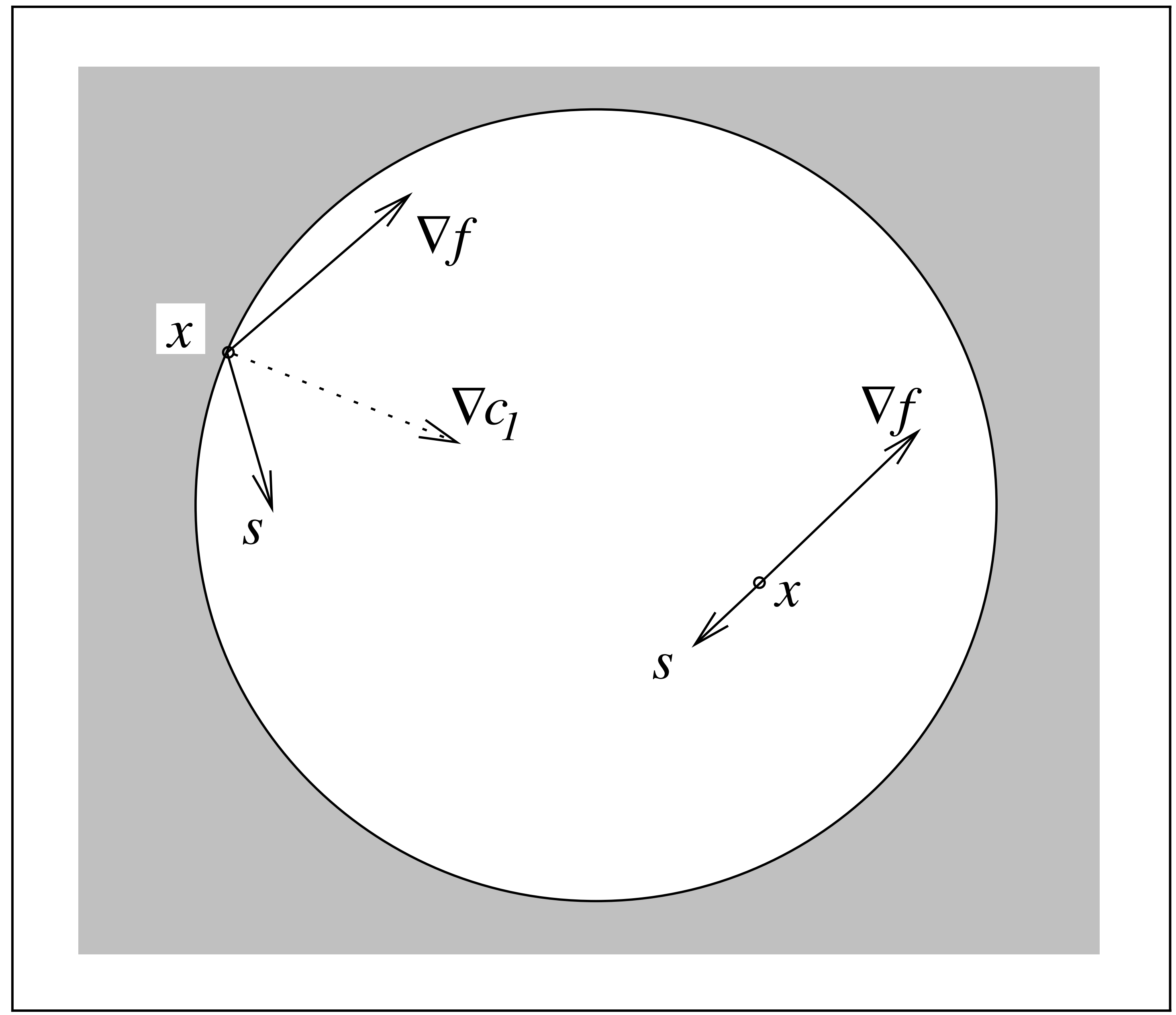
Cas d'une seule contrainte d'inégalité

$$x_1^2 + x_2^2 \leq 2$$

Condition nécessaire d'optimalité

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

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



Condition nécessaire d'optimalité: cas général

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

Conditions de Karush-Kuhn-Tucker (KKT)

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