Outils d'optimisation pour les sciences des données et de la décision

22 septembre 2023

Aujourd'hui :

- → Présentation du cours
- → Introduction à l'optimisation

] 8h30 - 10h

- → TD d'application 10h15 - 11h45

Logistique

* Page du cours :

https://www.lamsade.dauphine.fr/~croyer/coursODD.html

* Mon mail :

clement.royer@lamsade.dauphine.fr

* Organisation :

8 séances de 3h

22/09, 29/09, 06/10, 13/10

20/10, 27/10, 03/11, 10/11

* Notation :

  - Examen (60%) 23/11 10h-12h

    1 feuille A4 de notes recto verso autorisée

  - Projet (40%)

    Notebook Python "à trous", individuel

    Date de rendu à définir (~ décembre)

INTRODUCTION À L'OPTIMISATION

↳ <u>Définition informelle</u> : Optimiser = Prendre la meilleure décision parmi un ensemble d'alternatives

↳ Mathématiquement : Calculer le maximum ou le minimum d'une certaine fonction numérique

↳ Applications : Finance, Réseaux d'énergie, Sciences des données, Travaux à Dauphine

① Anatomie d'un problème d'optimisation

Un problème d'optimisation est un objet mathématique de la forme

$$(P) \quad \boxed{\underset{w \in \mathbb{R}^d}{\text{minimiser}} \; f(w) \quad \text{s.c.} \quad w \in \mathcal{F}}$$

Variables de décision

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix} \in \mathbb{R}^d \qquad w_i \in \mathbb{R} \; \forall i$$

"sous les contraintes"

Contraintes $\mathcal{F} \subseteq \mathbb{R}^d$ ensemble réalisable

fonction objectif : $f : \mathbb{R}^d \to \mathbb{R}$, $w \mapsto f(w)$

minimiser ⇒ les meilleures valeurs sont les + petites

<u>Définitions</u>

- (P) est un problème sans contraintes si $\mathcal{F} = \mathbb{R}^d$ auquel cas on écrit
  $$\underset{w \in \mathbb{R}^d}{\text{minimiser}} \; f(w)$$
  Lorsque $\mathcal{F} \neq \mathbb{R}^d$, on dit que (P) est un problème avec contraintes
- (P) est dit réalisable si $\exists w \in \mathcal{F}$ $(\mathcal{F} \neq \emptyset)$. Sinon il est dit irréalisable
- L'ensemble des solutions de (P) est noté
  $$\underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \; f(w) \quad \text{s.c.} \quad w \in \mathcal{F} \;\;) \subseteq \mathbb{R}^d$$

On dit que $w \in \mathbb{R}^d$ est une solution de (P) si $w \in F$ et $\forall v \in F, \ f(w) \leq f(v)$

- La valeur optimale de (P) est notée

$$\left.\begin{array}{l}\min\limits_{w \in \mathbb{R}^d} f(w) \ \text{s.c.} \ w \in F\end{array}\right) \in \mathbb{R} \cup \{-\infty, +\infty\}$$

NB : $\underset{w \in \mathbb{R}^d}{\text{minimiser}} \ f(w)$ s.c. $w \in F$ est équivalent à $\underset{w \in \mathbb{R}^d}{\text{maximiser}} \ -f(w)$ s.c. $w \in F$ car

$$\underset{w \in \mathbb{R}^d}{\text{argmin}} \ f(w) \ \text{s.c.} \ w \in F = \underset{w \in \mathbb{R}^d}{\text{argmax}} \ -f(w) \ \text{s.c.} \ w \in F$$

② Solutions d'un problème d'optimisation

↳ On se concentre sur les problèmes sans contrainte.

(P) $\underset{w \in \mathbb{R}^d}{\text{minimiser}} \ f(w)$ avec $f : \mathbb{R}^d \to \mathbb{R}$

↳ Une solution de (P) est un vecteur $w^* \in \mathbb{R}^d$ tel que $f(w^*) \leq f(w)$ $\forall w \in \mathbb{R}^d$

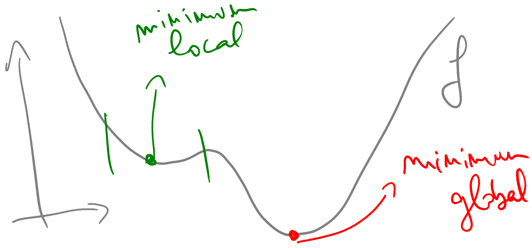
$\Rightarrow$ w^* s'appelle un minimum global de f

↳ Trouver des minima globaux via cette définition requiert d'évaluer f une infinité de fois $\Rightarrow$ pas l'approche adoptée en pratique

↳ On se contente souvent de chercher des minima locaux (des solutions locales)

Def $w^* \in \mathbb{R}^d$ est un minimum local de f si $\forall w \in \mathbb{R}^d$ tel que $\|w - w^*\|$ est suffisamment petite, alors $f(w^*) \leq f(w)$.

$\forall v \in \mathbb{R}^d, \; \|v\| = \sqrt{\sum_{i=1}^{d} v_i^2}$ avec $v = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix}$



Tout minimum global est local, mais l'inverse n'est pas vraie en général.

↳ En général, trouver des minima locaux est aussi difficile que de trouver des minima globaux ! (Toujours besoin de regarder une infinité de valeurs)

⇒ Mais plus facile dans certains cas

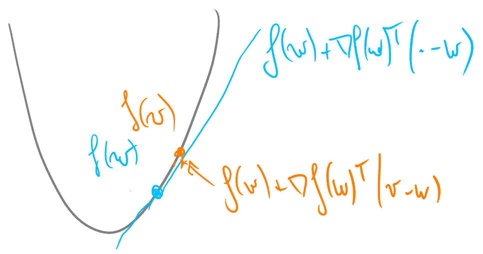
⇒ Possible de caractériser des minima locaux via des conditions vérifiables en temps fini.

(3) Conditions d'optimalité

↳ On considère toujours (P) (sans contraintes) et on suppose que f est de classe C^1 ($f \in C^1$), c'est-à-dire que f est continûment différentiable

⇒ $\forall w \in \mathbb{R}^d$, $\exists \nabla f(w) \in \mathbb{R}^d$ tel que $\forall v \in \mathbb{R}^d$, $\|v-w\|$ est suffisamment petite, alors

$$f(v) \approx f(w) + \nabla f(w)^T (v-w)$$



$\forall (u,v) \in (\mathbb{R}^d)^2$,

$u^T v = \sum_{i=1}^{d} u_i v_i$ $\quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}$

$u^T u = \|u\|^2$

$u^T = [u_1 \; \dots \; u_d]$

↳ L'approximation ci-dessus permet d'étudier le comportement local de f autour d'un point sans évaluer à nouveau la fonction !

⇒ Requiert seulement $f(w)$ et $\nabla f(w) \in \mathbb{R}^d$ que l'on appelle le gradient de f en w

Théorème : Si f est C^1 et $\bar{w} \in \mathbb{R}^d$, alors

$$[\bar{w} \text{ est un minimum local de } f] \Rightarrow \boxed{\|\nabla f(\bar{w})\| = 0}$$

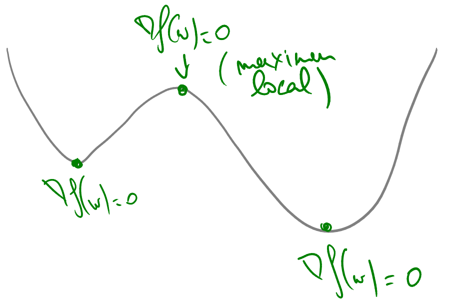
$\left(\|v\| = 0 \Leftrightarrow v = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}\right)$

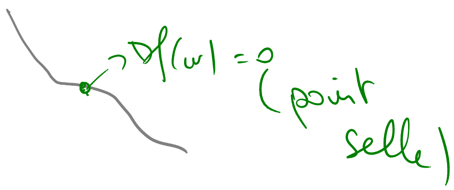
Condition d'optimalité à l'ordre 1

⇒ C'est une condition nécessaire ($\Rightarrow$) mais pas suffisante.

⇒ Il peut exister des vecteurs $w \in \mathbb{R}^d$ tels que $\|\nabla f(w)\| = 0$ qui ne soient pas des minima locaux

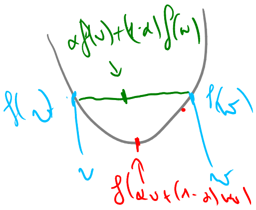
⇒ On parle alors de points stationnaires à l'ordre 1



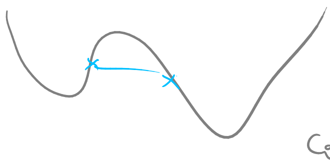


Déf : Une fonction $f : \mathbb{R}^d \to \mathbb{R}$ est dite convexe si

$$\forall (v,w) \in (\mathbb{R}^d)^2,\ \forall \alpha \in [0,1],\quad f(\alpha v + (1-\alpha) w) \leq \alpha f(v) + (1-\alpha) f(w)$$



Convexe



Non convexe

Th) Soit $f : \mathbb{R}^d \to \mathbb{R}$ convexe. Alors :

a) Tout minimum local de f est un minimum global.

b) Si $f \in C^1$, alors $\forall \bar{w} \in \mathbb{R}^d$,

$[\|\nabla f(\bar{w})\| = 0] \iff \bar{w}$ est un minimum global

$\iff \bar{w}$ est un minimum local

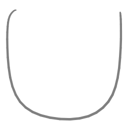
$\hookrightarrow$ Les fonctions C^1 et convexes sont particulièrement adaptées à la minimisation.

Def : $f : \mathbb{R}^d \to \mathbb{R}$ est μ-fortement convexe ($\mu > 0$) si

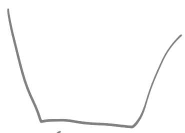
$\forall (v,w) \in (\mathbb{R}^d)^2$, $\forall \alpha \in [0,1]$,

$$f(\alpha v + (1-\alpha)w) \leq \alpha f(v) + (1-\alpha) f(w) - \frac{\mu}{2}\alpha(1-\alpha)\|v-w\|^2$$

Convexe pas fortement convexe



fortement convexe



Convexe pas fortement convexe

Th) Si $f : \mathbb{R}^d \to \mathbb{R}$ est μ-fortement convexe, alors elle possède un unique minimum global, qui est l'unique vecteur $w^* \in \mathbb{R}^d$ tel que $\nabla f(w^*) = 0$

Extension aux fonctions de classe C^2

↳ Déf : f est C^2 (deux fois continûment différentiable) si $\forall w \in \mathbb{R}^d$, $\exists \nabla f(w) \in \mathbb{R}^d$

et $\exists \nabla^2 f(w) \in \mathbb{R}^{d \times d}$

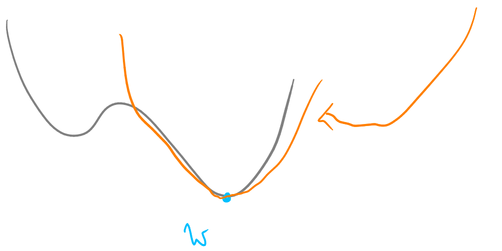
$\mathbb{R}^{m \times n}$: matrices à m lignes et n colonnes

$A = \begin{bmatrix} A_{11} \cdots A_{1n} \\ A_{m1} \cdots A_{mn} \end{bmatrix}$

tels que $\forall v \in \mathbb{R}^d$, $\|v-w\|$ suffisamment petite,

alors

$$f(v) \approx f(w) + \nabla f(w)^T (v-w) + \frac{1}{2} (v-w)^T \underbrace{\overbrace{\nabla^2 f(w)}^{\in \mathbb{R}^{d\times d}} \overbrace{(v-w)}^{\in \mathbb{R}^d}}_{\text{produit matrice vecteur } \in \mathbb{R}^d}$$



$\nabla^2 f(w)$: matrice Hessienne

Propriétés des fonctions C^2

Soit $f \in C^2$ et $\bar{w} \in \mathbb{R}^d$

- Condition nécessaire d'optimalité à l'ordre 2

$[\bar{w}$ minimum local de $f] \Rightarrow [\|\nabla f(\bar{w})\| = 0$ et $\nabla^2 f(\bar{w}) \succeq 0]$

- Condition suffisante d'optimalité à l'ordre 2

$\begin{bmatrix} \|\nabla f(\bar{w})\| = 0 \\ \text{et } \nabla^2 f(\bar{w}) \succ 0 \end{bmatrix} \Rightarrow [\bar{w}$ minimum local de $f]$

- $[f$ est convexe$] \Leftrightarrow [\nabla^2 f(w) \succeq 0 \quad \forall w \in \mathbb{R}^d]$
- $[f$ est μ-fortement convexe$] \Leftrightarrow [\nabla^2 f(w) \succeq \mu I \quad \forall w \in \mathbb{R}^d]$

$A \succeq 0$ "semi-définie positive" $\forall v \in \mathbb{R}^d$, $\underbrace{v^T A v}_{\in \mathbb{R}} \geq 0$

$A \succeq \mu I$ (identité $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$) $\Leftrightarrow$ $A - \mu I \succeq 0$

$A \succ 0$ "définie positive" : $\forall v \in \mathbb{R}^d \setminus \{0\}$, $v^T A v > 0$

NB : Ne s'applique que pour des matrices symétriques