Optimisation - Résumé

October 18, 2023

THEVENET Louis

Table des matières

1. Rappels

1.1. Différentielle d'une composée

Théorème 1.1.1: f, g telles que $g \circ f$ soit dérivable en $x \in \Omega$, on a :

$$\forall h \in E, (g \circ f)'(x). \cdot h = g'(f(x)) \times (f'(x) \cdot h)$$

1.2. Gradient

Définition 1.2.1: $a \in \Omega$, $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ doublement dérivable sur Ω:

$$\nabla f(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}$$

$$\nabla^2 f(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \frac{\partial^2 f}{\partial x_n \partial x_2}(a) & \dots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{pmatrix}$$

Voir[1]

1.3. Un autre truc

Proposition 1.3.1:

$$\forall h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in \mathbb{R}_n : f'(a) \cdot h = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a) \cdot h_k$$

$$k = f'(a) \cdot h \Leftrightarrow \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_p \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \dots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial x_1}(a) & \frac{\partial f_p}{\partial x_2}(a) & \dots & \frac{\partial f_p}{\partial x_n(a)} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix}$$

1.4. Convexité

Théorème 1.4.1: f dérivable sur $D_0 \subset \Omega$ un convexe :

$$f \text{ est convexe} \Longleftrightarrow \forall x, y \in D_0, f(y) - f(x) \geq f'(x)(y - x)$$
$$f \text{ est strictement convexe} \Longleftrightarrow \forall x, y \in D_0, x \neq y, f(y) - f(x) > f'(x)(y - x)$$
$$f \text{ est uniformément convexe} \Longleftrightarrow \forall x, y \in D_0, f(y) - f(x) \geq f'(x)(y - x) + c\|y - x\|_E^2$$

$$\begin{aligned} f \text{ est convexe} &\Longleftrightarrow \forall x \in D_0 : f''(x) \text{ est semi-définie postivive} \\ &\Longleftrightarrow \forall x \in D_0 : \nabla^2 f(x) \text{ semi-def. pos.} \end{aligned}$$

$$\forall x \in D_0 : f''(x) \text{ ou } \nabla^2 f(x) \text{ est définie postivive} \Rightarrow f \text{ est strictement convexe}$$

2. Définitions

Définition 2.1 : Problème d'optimisation avec contraintes C

On cherche à minimiser une fonctionnelle f sur un ensemble $C \subset \mathbb{R}^n$, on note ce problème :

$$(P)\begin{cases} \min(f(x)) \\ x \in C \subset \mathbb{R}^n \end{cases}$$

Le problème est
- Non différentiable si f est non dérivable
- Convexe si f et C sont convexes

3. Existence de solutions

3.1. Problème avec contraintes C

Théorème 3.1.1 : Soit (P) un problème d'opti. sous contraintes C

- (P) admet une solution si
  - Si f est continue
  - C est un compact non vide
- (P) admet une solution si
  - $f : \mathbb{R}^n \to \mathbb{R}$ continue et **0-coercive**
  - C est un fermé non vide

3.2. Cas convexe

Théorème 3.2.1 : Ici C est un convexe de E EVN

- il existe au plus une solution si
  - f est **strictement** convexe à valeurs réelles
- tout minimum local sur C est global sur C si
  - f est convexe à valeurs réelles

4. Condition nécessaire et suffisante

4.1. Premier ordre

4.1.1. Cas sans contrainte

Proposition 4.1.1.1:

Si

- f à valeurs réelles, définie sur un ouvert
- x^* minimum local de f
- f dérivable en x^*.

Alors $f'(x^*) = 0$

4.1.2. Cas f convexe sur C

Proposition 4.1.2.1:

- $\forall y \in C, f'(x^*)(y - x) \geq 0$ si
  - f convexe sur un ouvert convexe C
  - x^* minimum local sur C
  - f dérivable en x^*
- Si f est dérivable et convexe en tout point de C, ces conditions sont **équivalentes** :
  1. x^* minimum local sur C
  2. x^* minimum global sur C
  3. $\forall y \in C, f'(x^*)(y - x) \geq 0$

(la condition devient suffisante)

4.2. Second ordre

4.2.1. Condition nécessaire

Théorème 4.2.1.1:

- x^* minimum local de f
- f deux fois dérivable en x^*.

Alors $f''(x^*)$ est **semi-définie positive**

4.2.2. Condition suffisante

Théorème 4.2.2.1:

- Si
  - x^* point critique de f
  - f deux fois dérivable en x^*
  - $f''(x^*)$ uniformément définie positive

  Alors x^* est un **minimum local** de f
- Si
  - f deux fois dérivable sur Ω et $\exists B(x^*, \eta) \mid f''(x)$ est **semi-définie positive**
  - $f'(x^*) = 0$

  Alors x^* est un **minimum local** de f

5. Problèmes aux moindres carrés

Définition 5.1: Problème aux moindres carrés

C'est un problème d'optimisation **sans contrainte** où f est de la forme suivante :

$$f(\beta) = \frac{1}{2}\|r(\beta)\|^2 = \frac{1}{2}(r(\beta) \mid r(\beta)) = \frac{1}{2}\sum_{i=1}^{n} r_i^2(\beta)$$

Le problème est dit **aux moindres carrés linéaires** si r est affine :

$$r : \begin{cases} \mathbb{R}^p \longrightarrow \mathbb{R}^n \\ \beta \quad \mapsto y - X\beta \end{cases}$$

où X est de taille (n, p) et $y \in \mathbb{R}^n$

5.1. Méthode de Newton

Avec

$$f(\beta) = \frac{1}{2}\|r(\beta)\|^2 = \frac{1}{2}\sum_{k=1}^{n} r_i^2(\beta)$$

$$\begin{aligned} \nabla f(\beta) &= \sum_i r_i(\beta)\nabla r_i(\beta) = J_r(\beta)^T r(\beta) \\ \nabla^2 f(\beta) &= \sum_i r_i(\beta)\nabla^2 r_i(\beta) + \sum_i \nabla r_i(\beta)\nabla r_i(\beta)^T \\ &= S(\beta) + J_r(\beta)^T J_r(\beta) \end{aligned}$$

Algo de Newton :

- Initialisation
  - choisir $\beta^{(0)} \in \mathbb{R}^n$
  - choisir $\varepsilon > 0$ et `MaxIter`

    - $k := 0$
- Corps
    - Répéter
        - $\beta^{(k+1)} := \beta^{(k)} - \left[S\left(\beta^{(k)}\right) + J_r\left(\beta^{(k)}\right)^T J_r\left(\beta^{(k)}\right)\right]^{-1} J_r\left(\beta^{(k)}\right)^T r\left(\beta^{(k)}\right)$
        - $k := k + 1$
    - Jusqu'à $\left(\left\|f\left(\beta^{(k)}\right)\right\| < \varepsilon\left(\left\|f\left(\beta^{(0)}\right) + 1\right\|\right)\right)$ `ou` $(k = \texttt{MaxIter})$

5.2. Méthode de Gauß-Newton

La formule de récurrence pour l'algo de Gauß-Newton change :

$$\beta^{(k+1)} := \beta^{(k)} - \left[J_r\left(\beta^{(k)}\right)^T J_r\left(\beta^{(k)}\right)\right]^{-1} J_r\left(\beta^{(k)}\right)^T r\left(\beta^{(k)}\right)$$

Bibliographie

[1] Carmen Cincotti, "La jacobienne contre l'hessienne contre le gradient," 2022. [Online]. Available: https://carmencincotti.com/fr/2022-08-15/la-jacobienne-contre-la-hessienne-contre-le-gradient/