OPTIMISATION MATHÉMATIQUE

Yassine Laguel

OPTIMISATION MATHÉMATIQUE

Yassine Laguel

Mail : yassine.laguel@univ-cotedazur.fr

QU'EST-CE QUE L'OPTIMISATION ?

- Définition [Wikipédia]

L'**optimisation** est une branche des mathématiques cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes qui consistent à minimiser ou maximiser une fonction sur un ensemble.

Fonction objective

$$\min_{x \in \mathcal{C}} f(x)$$

Contraintes

Modélisation

Analyse théorique

Résolution Pratique

Implémentation

Mail : yassine.laguel@univ-cotedazur.fr

APPLICATIONS

- Exemple 1: le transport

- Exemple 2 : le recrutement

On retrouve aussi l'optimisation dans l'ingénierie, la finance, le marketing, la politique, etc...

Mail : yassine.laguel@univ-cotedazur.fr

CONTENU DU COURS

- Objectif
  - Introduire l'optimisation convexe sous contraintes.
    - Savoir reconnaître les propriétés de régularité d'un problème
    - Savoir résoudre analytiquement ou numériquement un problème d'optimisation numérique.
  - Ce cours est porté essentiellement sur les aspects théoriques. On ne parlera pas en particulier des aspects modélisation et implémentation.
- Progression prévue

Semaine 1	Semaine 2	Semaine 3	Semaine 4
Rappels de Calcul Différentiel	L'algorithme du gradient	Introduction à l'optimisation sous contraintes	Dualité Lagrangienne
Rappels sur la Convexité	Zoom sur les quadratiques	Théorème des extrémas liés	

Mail : yassine.laguel@univ-cotedazur.fr

EVALUATION

- Un examen final le Mercredi 06/12 de 8h à 10h
  - L'examen dure 2 heures
  - Connaître toutes les définitions et propriétés du cours avec leur preuves.
  - Savoir refaire les exercices vus en cours et TDs

- Points bonus de participation
  - Interrogation en début de cours ou TDs
  - Porte sur des définitions ou proprietés du cours.
  - Bonus de 0 à 2 points

Mail : yassine.laguel@univ-cotedazur.fr

Dans ce cours, sauf mention contraire, $\mathcal{U}$ désigne un ouvert de $\mathbb{R}^d$.

I - Rappels de calcul différentiel

Definition : [Différentielle]

Soit $f : \mathcal{U} \to \mathbb{R}^n$ et $a \in \mathbb{R}^d$. On dit que f est différentiable au point $a \in \mathcal{U}$ s'il existe une application linéaire $L \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$ et $R : \mathbb{R}^d \to \mathbb{R}^n$ tels que

- Pour tout $h \in \mathbb{R}^d$ tel que $a+h \in \mathcal{U}$.

$$(*) \quad f(a+h) = f(a) + L(h) + R(h)$$

- $R(h) = o(\|h\|)$ lorsque h tend vers 0 dans E, c'est à dire $\lim\limits_{\substack{h \to 0 \\ h \neq 0}} \frac{\|R(h)\|}{\|h\|} = 0$.

Remarques :
- Si une telle application L existe, elle est unique : on l'appelle la différentielle de f en a.
- On note souvent cette application $L = Df(a)$ ou $f'(a)$ ou $df(a)$ ou $D_a f$, etc...

La notation gardée dans ce cours.

- On notera que $(*)$ peut aussi s'écrire

$$\begin{aligned} f(a+h) - f(a) &= Df(a)h + o(\|h\|_E) \\ &= Df(a)h + \|h\|\,\varepsilon(h) \end{aligned}$$

où $\varepsilon : \mathbb{R}^d \to \mathbb{R}^n$ satisfait $\varepsilon(h) \xrightarrow[h \to 0]{} 0$.

On a dans cette réécriture "la grande idée du calcul différentiel" :

$$\text{accroissement de la fonction} = \text{terme linéaire par rapport à l'accroissement de la variable} + \text{petit terme correctif}$$

$$\text{calcul différentiel} = \text{algèbre linéaire} + \text{majoration de normes.}$$

Pour plus de rappels voir Chap. 2 du petit guide d'introduction au calcul différentiel de F. Rouvière.

- En dimension finie, $Df(a)$ peut se représenter comme une matrice :
  - si $d = n = 1$, $Df(a) \in \mathbb{R}$
  - si $n = 1$ $Df(a) \in \mathbb{R}^d$ est un vecteur.

On définit le gradient $\nabla f(a)$ comme la transposée de $Df(a)$:

$$\begin{aligned} f(a+h) &= f(a) + Df(a).h + \|h\|\varepsilon(h) \\ &= f(a) + \nabla f(a)^T h + \|h\|\varepsilon(h) \\ &= f(a) + \langle \nabla f(a), h \rangle + \|h\|\varepsilon(h) \end{aligned}$$

Le vecteur $\nabla f(a)$ est constitué des dérivées directionnelles de f en a :

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

- Dans le cas général, on a :

$$Df(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \vdots & & \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_d} \end{pmatrix}$$

- On dit que f est de classe $\mathcal{C}^1$ sur un ouvert $\Omega \subset \mathcal{U}$ si $Df : \begin{array}{l} \mathbb{R}^d \to \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n) \\ a \mapsto Df(a) \end{array}$ est définie et continue sur Ω.

Théorème [fondamental de l'analyse]

Soit $f : \mathbb{R}^d \to \mathbb{R}^n$ de classe $\mathcal{C}_1$ sur $\mathbb{R}^d$.

Alors pour tout $x, h \in \mathbb{R}^d$,

$$f(x+h) = f(x) + \int_0^1 Df(x+th).(h) \; dt$$

Remarques :
- La preuve de ce théorème se base sur le théorème des accroissements finis, qui découle du théorème de Rolle.
- Cette formule se généralise aux fonctions de classe $\mathcal{C}^k$, c'est la formule de Taylor :

$$f(x+h) = f(x) + Df(x)(h) + \frac{1}{2}D^2f(x)(h,h) + \dots + \frac{1}{k!}D^kf(x)(h,\dots,h) + \left\{ \begin{array}{l} o(\|h\|^k) \\ \displaystyle\int_0^1 \frac{(1-t)^k}{k!} D^{k+1}f(x+th)(\underbrace{h,\dots,h,h}_{k+1 \text{ fois}})\,dt \end{array}\right.$$

Notons qu'une intégration par parties donne

$$\int_0^1 \frac{(1-t)^k}{k!} D^{k+1}f(x+th)(\underbrace{h,\dots,h,h}_{k+1 \text{ fois}})\,dt$$
$$= \left[\frac{-(1-t)^{k+1}}{(k+1)!} D^{k+1}f(x+th)(h,\dots,h,h) \right]_0^1$$
$$+ \int_0^1 \frac{(1-t)^{k+1}}{(k+1)!} D^{k+2}f(x+th)(\underbrace{h,\dots,h,h}_{k+2 \text{ fois}})\,dt$$

ce qui permet de prouver cette formule par récurrence.

- Dans le cas particulier $k=2$, $n=1$, on a la formule:

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \underbrace{\int_0^1 (1-t) D^2f(x+th)(h,h)\,dt}$$
$$= h^t \underbrace{\nabla^2 f(x+th)}_{\text{Hessienne de } f \text{ en } h.} h$$

La matrice associée $\left[\nabla^2 f(x)_{ij}\right]_{\substack{1\le i\le d \\ 1\le j\le d}}$ satisfait

$$\nabla^2 f(x)_{ij} = \frac{\partial f(x)}{\partial x_i \partial x_j}.$$

C'est une matrice symmétrique.

Proposition: [Différentiation d'une composée]

Soient $\mathcal{U}$ et V des ouverts respectifs de $\mathbb{R}^d$ et $\mathbb{R}^n$.

Soient $f: \mathcal{U} \to \mathbb{R}^n$ et $g: V \to \mathbb{R}^m$ telles que $f(\mathcal{U}) \subset V$.

Si f et derivable en $a \in \mathcal{U}$ et g est derivable en $b = f(a)$. Alors $g \circ f$ est derivable en a et

$$D(g\circ f)(a) = Dg(f(a)) \circ Df(a)$$

Remarque: Dans le cas $m=1$, on a:

$$\nabla(g\circ f)(a) = \left(\mathrm{Jac}\, g(f(a)) \cdot \mathrm{Jac}\, f(a)\right)^T$$
$$= \left(\nabla g(f(a))^T\, \mathrm{Jac}\, f(a)\right)^T$$
$$= \mathrm{Jac}\, f(a)^T \cdot \nabla g(f(a)).$$

Exercices • Calculer la différentielle de $x \mapsto \|x\|^2$.

Exercice:

Soient $A \in \mathcal{M}_{n\times d}(\mathbb{R})$, $b \in \mathbb{R}^n$ et $f: \mathbb{R}^d \to \mathbb{R}$, $x \mapsto \|b - Ax\|^2$.

Calculer le gradient de f.

Solution: Pour $x, h \in \mathbb{R}^d$, on a:

$$f(x+h) = \|b - A(x+h)\|^2$$
$$= \|b - Ax\|^2 - 2\langle b - Ax, h\rangle + \|Ah\|^2$$
$$= f(x) + \langle -2(b-Ax), h\rangle + \underbrace{\|Ah\|^2}_{\substack{\le \|h\|^2 \cdot \|\|A\|\|^2 \\ = o(\|h\|)}}$$

Ainsi, f est differentiable en x

II - Notions d'extrema

Définition : [minimum global, minimum local]

Soit $f : \mathcal{U} \mapsto \mathbb{R}$ et $x \in \mathcal{U}$.

On dit que x est un minimum global de f sur $\mathcal{U}$ si :

$$\forall x' \in \mathcal{U} \quad f(x) \leq f(x').$$

On dit que x est minimum local s'il existe un ouvert $W \subset \mathcal{U}$ tel que :

$$\begin{cases} x \in \mathcal{U} \\ \forall x' \in W, \quad f(x) \leq f(x'). \end{cases}$$

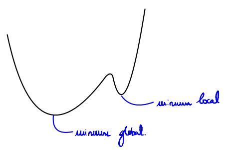
Remarque :

- Les notions de maximums locaux et globaux se définissent similairement.

  [x maximum global (resp. local) de f]

  $\Leftrightarrow$ [x minimum global (resp. local) de $-f$]

- Ensemble des minimum globaux $\subsetneq$ Ensemble des minimums locaux



- Le minimum est dit strict si on peut remplacer $\leq$ par $<$.

- On va voir dans la suite de cette partie comment relier l'existence d'un minimum aux propriétés de f.

Proposition : [Existence d'extrema et compacité]

Soit $K \subset \mathbb{R}^d$ compact et $f : K \to \mathbb{R}$ continue.
Alors f admet un minimum global sur K.

Remarque : • On en déduit que le problème $\min_{x \in K} f(x)$ admet une solution globale dès lors que f est continue et K compact.

Preuve : Soit $(x_n)_{n \geq 0}$ une suite de K telle que

$$f(x_n) \xrightarrow[n \to \infty]{} \inf F(K).$$

Par compacité de K, on peut extraire $(x_{\varphi(n)})_{n \geq 0}$ de (x_n) pour une extractrice telle que $(x_{\varphi(n)})_{n \geq 0}$ admet une limite $x \in K$. Alors, par continuité de f,

$$f(x) = \lim_{n \to +\infty} f(x_{\varphi(n)}) = \inf f(K).$$

Proposition : [fonction continue coercive]

Soit $f : \mathbb{R}^d \to \mathbb{R}$ continue telle que $f(x) \xrightarrow[\|x\| \to +\infty]{} +\infty$
Alors f admet un minimum global.

Preuve : On note que $K = \{x, f(x) \leq f(0)\}$ est

- non-vide ($0 \in K$),
- fermé en tant qu'image réciproque d'un fermé par une fonction continue.
- borné, par hypothèse de coercivité.

Donc d'après le théorème précédent, f admet un minimum global x sur K.
Par définition de K, $f(x) \leq f(0) \leq f(y)$ pour tout $y \in \mathbb{R}^d \setminus K$. Donc x est un minimum global de f sur $\mathbb{R}^d$.

Proposition :

Soit $f : \mathcal{U} \mapsto \mathbb{R}$ différentiable sur $\mathcal{U}$.

Soit $x \in \mathcal{U}$ un minimum local de $\mathcal{U}$.

Alors $\nabla f(a) = 0$.

Preuve : Soit $u \in \mathbb{R}^d$ fixé. Pour tout $t > 0$ tel que $x + tu \in \mathcal{U}$, on a :

$$f(x+tu) - f(x) = t \langle \nabla f(x), u \rangle + t \|u\| \varepsilon(tu)$$

avec $\varepsilon(h) \xrightarrow[h \to 0]{} 0$.

Ainsi, pour $t > 0$ assez petit,

$$0 \leq \frac{f(x+tu) - f(x)}{t} = \langle \nabla f(x), u \rangle + \varepsilon(tu) \xrightarrow[t \to 0]{} \langle \nabla f(x),$$

Donc, pour tout $u \in \mathbb{R}^d$, $\langle \nabla f(x), u \rangle \geqslant 0$.

En particulier, pour $u = -\nabla f(x)$, on a:

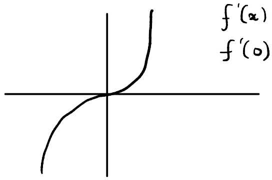
$$-\|\nabla f(x)\|^2 \geqslant 0 \quad \Rightarrow \quad \|\nabla f(x)\| = 0$$

Remarques:

- On a:

$$\{\text{minimum globaux}\} \subsetneq \{\text{minimums locaux}\} \subsetneq \{\text{points critiques}\}$$

- Un exemple simple de point critique qui n'est pas un minimum local: $x \mapsto x^3$



$f'(x) = 3x^2$

$f'(0) = 0.$

Proposition: [minimum local et dérivée seconde]

Soit $f: U \to \mathbb{R}$ deux fois dérivable en $a \in U$.

(i) Si f admet un minimum relatif en a, alors

$$\nabla^2 f(a) \succcurlyeq 0, \quad \text{i.e } h^t \nabla^2 f(a) h \geqslant 0, \text{ pour tout } h \in \mathbb{R}^d$$

(ii) Si $Df(a) = 0$ et $D^2 f(a)$ est définie positive,

$$\left(\text{i.e } \begin{cases} h^T D^2 f(a) h \geqslant 0, \ \forall h \in \mathbb{R}^d \\ h^T D^2 f(a) h = 0 \iff h = 0 \end{cases}\right)$$

alors a est un minimum local strict en x_0.