OPTIMISATION

Introduction et rappels de calculs différentiels

23 Février 2021

INSAT - GL3 - M. SFAXI

Motivation

On s'intéresse essentiellement à la résolution des problèmes de type :

$$\inf_{x\in K} f(x) \qquad \text{ou} \qquad \sup_{x\in K} f(x)$$

où $K \subset \mathbb{R}^n$ et $f : K \to \mathbb{R}$ est une fonction, appelée fonction objectif.

- Si $K = \mathbb{R}^n$: on parle de problème d'optimisation **sans contrainte**.
- Si $K \subsetneq \mathbb{R}^n$: on parle de problème d'optimisation **sous contrainte**.

N.B. : Résoudre $\sup\limits_{x\in K} f(x)$ est équivalent à résoudre $\inf\limits_{x\in K} -f(x)$.

Convention : Pour indiquer que le minimum est atteint, on écrira $\min\limits_{x\in K} f(x)$.

Mission

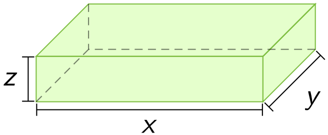
Notre mission est de répondre à un certain nombre de questions :

- Le problème $\inf_{x \in K} f(x)$ possède t-il une solution ?
- Si oui, on cherchera à la caractériser :
  - est-elle unique ?
  - la déterminer (si possible) explicitement.

  On exploitera pour cela les conditions nécessaires et/ou suffisantes d'optimalité.
- Si non, on cherchera une suite minimisante de f, i.e. une suite d'éléments de K convergeant vers l'infimum de f.
- Enfin, lorsque l'on ne sait pas déterminer explicitement la solution du problème, on mettra en place un schéma numérique adapté.

Exemple 1 (Volume maximal connaissant la surface)

On souhaite maximiser le volume d'un parallélépipède rectangle de surface égale à 6. En appelant x, y et z, les longueurs des côtés du parallélépipède, on se ramène à la formulation suivante :



$$\begin{cases} f(x,y,z) = xyz \\ xy + xz + yz = 3 \\ x \geqslant 0,\ \ y \geqslant 0,\ \ z \geqslant 0. \end{cases}$$

Il s'agit donc d'un problème d'optimisation dans $\mathbb{R}^3$ sous contrainte :

$$\sup_{(x,y,z)\in\mathcal{K}} f(x,y,z)$$

où $\quad \mathcal{K} = \{(x,y,z) \in \mathbb{R}^3,\ x \geqslant 0,\ y \geqslant 0,\ z \geqslant 0 \text{ et } xy + xz + yz = 3\}.$

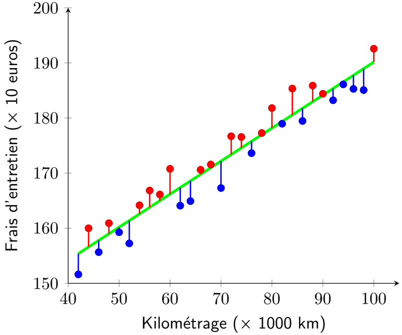
Exemple 2 (La régression linéaire)

On cherche une relation linéaire entre deux variables x et y, où x est la variable déterministe et y la variable aléatoire expliquée. Typiquement, y représente les frais d'entretien d'une voiture que l'on cherche à expliquer en fonction du kilométrage x.

Le modèle s'écrit : $\quad y = \alpha + \beta x + \varepsilon$,
où ε est une perturbation aléatoire centrée et de variance σ^2 inconnue.

À partir de n observations (x_i, y_i), on cherche à estimer α et β en minimisant la somme des résidus :

$$\inf_{(\alpha,\beta)\in\mathbb{R}^2} \sum_{i=1}^{n} (\alpha + \beta x_i - y_i)^2.$$



Exemple 2 (La régression linéaire)

Cela correspond à chercher (α, β) de sorte que la variance σ^2 soit minimale.
On cherche donc à minimiser (sur $\mathcal{K} = \mathbb{R}^2$) la fonction :

$$f(\alpha, \beta) = \left\| A \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - b \right\|^2$$

$$\text{où} \quad A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \in \mathcal{M}_{n,2}(\mathbb{R}) \quad \text{et} \quad b = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n.$$

Il s'agit d'un problème d'optimisation sans contrainte :

$$\inf_{X=(\alpha,\beta)\in\mathbb{R}^2} \|AX - b\|^2.$$

BOÎTE À OUTILS

Produit scalaire et norme dans $\mathbb{R}^n$

- Le produit scalaire de deux éléments x et y de $\mathbb{R}^n$ est le nombre réel

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i .$$

- La norme euclidienne associée d'un vecteur x de $\mathbb{R}^n$ est le nombre positif

$$\|x\| = \sqrt{\langle x, x\rangle} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} .$$

Inégalité de Cauchy-Schwarz : Soient x et y deux éléments de $\mathbb{R}^n$. On a :

$$|\langle x, y\rangle| \leqslant \|x\|\,\|y\|.$$

Produit scalaire et norme dans $\mathbb{R}^n$

Rappelons que le produit scalaire vérifie les propriétés :

- $\forall (x,y) \in [\mathbb{R}^n]^2, \quad \langle x, y\rangle = \langle y, x\rangle,$
- $\forall (x,y,z) \in [\mathbb{R}^n]^3, \; \forall(\alpha,\beta) \in \mathbb{R}^2, \quad \langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle,$
- $\forall x \in \mathbb{R}^n, \quad \langle x, x\rangle \geqslant 0 \quad$ et $\quad \langle x, x\rangle = 0 \iff x = 0.$

Une norme vérifie les propriétés :

- $\|x\| = 0 \iff x = 0.$
- $\forall \lambda \in \mathbb{R}, \; \forall x \in \mathbb{R}^n, \quad \|\lambda x\| = |\lambda|\|x\|,$
- $\forall (x,y) \in [\mathbb{R}^n]^2, \quad \|x+y\| \leqslant \|x\| + \|y\|,$

Ouvert, fermé et compact de $\mathbb{R}^n$

On appelle **boule ouverte** de centre $a \in \mathbb{R}^n$ et de rayon $r > 0$ l'ensemble

$$\mathbb{B}(a, r) = \{x \in \mathbb{R}^n,\ \|x - a\| < r\}.$$

- Un ensemble $\mathcal{O}$ est dit **ouvert** si, pour tout point $x \in \mathcal{O}$, on peut trouver une boule ouverte de centre x, incluse dans l'ensemble $\mathcal{O}$.
- Un ensemble $\mathcal{F}$ est dit **fermé** si son complémentaire est un ouvert.

N.B. : ▪ $\mathbb{R}^n$ et $\varnothing$ sont les seules parties de $\mathbb{R}^n$ à la fois ouvertes et fermées.

▪ Un ensemble qui n'est pas fermé ne veut pas dire qu'il est ouvert.

Ouvert, fermé et compact de $\mathbb{R}^n$

Un ensemble $\mathcal{B}$ est dit **borné** s'il est contenu dans une certaine boule.

Un ensemble $\mathcal{C}$ est dit **compact** s'il est à la fois fermé et borné.

Exemple : Revenons à l'exemple du parallélépipède. On voulait maximiser son volume $f(x,y,z) = xyz$ sous la contrainte :

$$\mathcal{K} = \{(x,y,z) \in \mathbb{R}^3,\ x \geqslant 0,\ y \geqslant 0,\ z \geqslant 0 \text{ et } xy + xz + yz = 3\}.$$

Déterminons sa nature topolgique.

- $\mathcal{K}$ est un fermé de $\mathbb{R}^3$. En effet :

$$\begin{aligned}\mathcal{K} &= \{(x,y,z)\in\mathbb{R}^3,\ x\geqslant 0,\ y\geqslant 0,\ z\geqslant 0\}\cap\{(x,y,z)\in\mathbb{R}^3, xy+xz+yz=3\}\\ &= (\mathbb{R}_+)^3\cap\varphi^{-1}(\{3\}),\end{aligned}$$

où $\varphi : (x,y,z)\in\mathbb{R}^3\mapsto xy+xz+yz\in\mathbb{R}$ laquelle est continue (car polynomiale). L'ensemble $\varphi^{-1}(\{3\})$ est donc un fermé de $\mathbb{R}^3$. Par conséquent, $\mathcal{K}$ est fermé.

- $\mathcal{K}$ n'est pas borné (donc non compact). En effet :

Remarquons que, pour tout $x>0$, le vecteur $(x,\frac{3}{x},0)$ appartient à $\mathcal{K}$. Alors que :

$$\left\|(x,\frac{3}{x},0)\right\|^2 = x^2+\frac{9}{x^2}\xrightarrow[x\to+\infty]{}+\infty.$$

CALCUL DIFFÉRENTIEL

Différentiabilité

Définition : Soit $\mathcal{D}$ un ouvert de $\mathbb{R}^n$ et a un point de $\mathcal{D}$. On dit qu'une application $f : \mathcal{D} \to \mathbb{R}^m$ est **différentiable** en a s'il existe $df_a \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ (continue), telle que

$$f(a+h) = f(a) + df_a(h) + o(\|h\|).$$

Par définition, l'application df_a est continue. Il n'en est pas nécessairement de même de l'application

$$df : \begin{array}{ccc} \mathcal{D} & \to & \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m) \\ a & \mapsto & df_a \end{array}.$$

Si c'est le cas, on dira que f est de **classe** $\mathscr{C}^1$ au voisinage de a.

Différentiabilité

- Si f et g sont différentiables en a, et si λ est réel constant, alors $f+g$ et λf sont différentiables en a et on a : $d(f+g)_a = df_a + dg_a, \quad d(\lambda f)_a = \lambda\, df_a$.
- Si f est constante, alors elle différentiable en tout point a et on a : $df_a = 0$.
- Si f est linéaire, alors elle est différentiable en tout point a et on a : $df_a = f$.

Proposition : Soient $\mathcal{U}$ un ouvert de $\mathbb{R}^n$, $\mathcal{V}$ un ouvert de $\mathbb{R}^p$, $f : \mathcal{U} \to \mathbb{R}^p$ et $g : \mathcal{V} \to \mathbb{R}^p$ deux applications telles que $f(U) \subset \mathcal{V}$. Si f est différentiable en $a \in \mathcal{U}$ et g différentiable en $f(a) \in \mathcal{V}$, alors $g \circ f$ est différentiable en a de différentielle :

$$d(g \circ f)_a = dg_{f(a)} \circ df_a.$$

Calcul de différentielle

Exemple 1 : Montrons que l'application $f : x \mapsto \|x\|^2$ est différentiable en tout point a de $\mathbb{R}^n$ et calculons sa différentielle.

Pour tout $h \in \mathbb{R}^n$, on a :

$$f(a+h) = \|a+h\|^2 = \langle a+h, a+h\rangle = \underbrace{\|a\|^2}_{f(a)} + \underbrace{2\langle a, h\rangle}_{df_a(h)} + \underbrace{\|h\|^2}_{o(\|h\|)}.$$

f est donc différentiable en a (donc sur tout $\mathbb{R}^n$) de différentielle : $df_a(\cdot) = 2\langle a, \cdot\rangle$.

EXERCICE : Montrer que l'application norme $\Phi : x \in \mathbb{R}^n \mapsto \|x\|$ est différentiable en tout point a de $\mathbb{R}^n \setminus \{0\}$ et déterminer sa différentielle.

Exemple 2 : Calculons la différentielle au point $a = (2,3)$ de l'application f définie sur $\mathbb{R}^2$ par $f(x,y) = (x - 2xy, y^2)$. Pour tout $(h,k) \in \mathbb{R}^2$, on a :

$$\begin{aligned} f\big((2,3) + (h,k)\big) &= f(2+h, 3+k) = (-10 - 5h - 4k - 2hk, 9 + 6k + k^2) \\ &= \underbrace{(-10,9)}_{f(2,3)} + \underbrace{(-5h - 4k, 6k)}_{df_{(2,3)}(h,k)} + \underbrace{(-2hk, k^2)}_{o(\|(h,k)\|)\ ?}. \end{aligned}$$

Pour $(h,k) \neq (0,0)$, on a :

$$\frac{\|(-2hk, k^2)\|}{\|(h,k)\|} = \frac{|k|\sqrt{4h^2 + k^2}}{\sqrt{h^2 + k^2}} \leqslant 2|k| \xrightarrow[(h,k)\to(0,0)]{} 0.$$

f est donc différentiable en $(2,3)$ de différentielle : $df_{(2,3)} : (h,k) \mapsto (-5h - 4k, 6k)$.

Calcul pratique de différentielle

Si l'on a au préalable démontré que f est différentiable en a , alors, on peut écrire pour tout vecteur unitaire $v \in \mathbb{R}^n$: $f(a+tv) = f(a) + t\, df_a(v) + o(t)$, ou encore

$$df_a(v) = \lim_{\substack{t \to 0 \\ t \in \mathbb{R}}} \frac{f(a+tv) - f(a)}{t}.$$

Cette limite s'appelle **dérivée directionnelle** de f en a suivant le vecteur v. Elle est souvent notée $D_v f(a)$ ou $f'_v(a)$.

En particulier, si $v = e_j$ (le j-ième vecteur de la b.c. de $\mathbb{R}^n$), la dérivée directionnelle $f'_{e_j}(a)$ est appelée la **dérivée partielle** de f par rapport à la j-ième variable x_j. Ainsi :

$$\frac{\partial f}{\partial x_j}(a) := f'_{e_j}(a) = \lim_{t \to 0} \frac{f(a+te_j) - f(a)}{t} = df_a(e_j) \in \mathbb{R}^m.$$

Jacobienne

Soit $\mathcal{D}$ un ouvert de $\mathbb{R}^n$ et $f = (f_1, \ldots, f_m) : \mathcal{D} \to \mathbb{R}^m$. Si f est différentiable en a, alors toutes ses dérivées partielles existent et df_a est l'application linéaire représentée par une matrice de taille $m \times n$ appelée **matrice jacobienne** de f en a :

$$Jf(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix}$$

Retour à l'exemple 2 : On voulait déterminer $df_{(2,3)}$ de $f : (x, y) \mapsto (x - 2xy, y^2)$.

$$Jf(x,y) = \begin{bmatrix} 1-2y & -2x \\ 0 & 2y \end{bmatrix} \implies Jf(2,3) = \begin{bmatrix} -5 & -4 \\ 0 & 6 \end{bmatrix}$$

Ainsi : $df_{(2,3)}(h,k) = Jf(2,3) \cdot (h,k)^T = (-5h - 4k, 6k)$.

Gradient

On se placera dorénavant dans le cas particulier d'une fonction $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, avec $\mathcal{D}$ ouvert de $\mathbb{R}^n$. Soit $a \in \mathcal{D}$.

Supposons que f est différentiable en a. Sa matrice jacobienne existe et est réduite à une ligne. Sa transposée est donc une matrice de taille $n \times 1$ que l'on confond avec un vecteur de $\mathbb{R}^n$. Ce vecteur est appelé **gradient** de f en a et est noté $\nabla f(a)$, i.e. :

$$\nabla f(a) := \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right).$$

Dès lors, la différentielle de f en a s'écrit :

$$df_a(h) = Jf(a) \cdot h = \langle \nabla f(a), h \rangle, \quad \text{pour tout } h \in \mathbb{R}^n.$$

Différentiabilité d'ordre 2

Si $f : \mathcal{D} \to \mathbb{R}$ est différentiable sur $\mathcal{D}$ (ouvert de $\mathbb{R}^n$). On a alors une application

$$df : \mathcal{D} \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}).$$

Si df est différentiable en a, alors $d(df)_a$ appartient à $\mathscr{L}\left(\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n, \mathbb{R})\right) \overset{\Psi}{\simeq} \mathscr{B}(\mathbb{R}^n)$ l'espace des formes bilinéaires (continues) de $\mathbb{R}^n$ dans $\mathbb{R}$.

On dit que f est deux fois différentiable en a si et seulement si df est différentiable en a. On appelle alors **différentielle seconde** en a l'application :

$$d^2 f_a = \Psi\big(d(df)_a\big) \in \mathscr{B}(\mathbb{R}^n).$$

Si de plus, l'application $d^2 f : a \mapsto d^2 f_a$ est continue, on dira que f est de **classe $\mathscr{C}^2$**.

Hessienne

Soit $\mathcal{D}$ un ouvert de $\mathbb{R}^n$ et $f : \mathcal{D} \to \mathbb{R}$. Si f est deux fois différentiable en $a \in \mathcal{D}$, alors toutes ses dérivées partielles secondes existent et $d^2 f_a$ est l'application bilinéaire représentée par une matrice de taille $n \times n$ appelée **matrice hessienne** de f en a :

$$\operatorname{Hess} f(a) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{bmatrix}$$

On notera que, si f est deux fois différentiable, en vertu du théorème de Schwarz, $\operatorname{Hess} f(a)$ est symétrique réelle (se rappeler du contre-exemple de Peano lorsque la fonction n'est pas deux fois différentiable).

Calcul de différentielle seconde

Calculons la différentielle seconde de l'application $f : x \in \mathbb{R}^n \mapsto \|x\|^2 = \langle x, x \rangle$.

Rappelons que la différentielle de f est donnée par $\quad df_x(h) = df(x)(h) = 2\langle x, h \rangle$.
Pour tout $(h, k) \in \mathbb{R}^n$, la différentielle seconde de f en a est :

$$d^2 f_a(h, k) := d\big(df\big)_a(h)(k) = d\big(df(\cdot)(h)\big)_a(k) = d\Big(2\langle \cdot, h \rangle\Big)_a(k) = 2\langle k, h \rangle.$$

Retrouvons cette formule à l'aide de la matrice hessienne. Pour tout $(h, k) \in [\mathbb{R}^n]^2$:

$$d^2 f_a(h, k) = h^T \underbrace{\mathrm{Hess}\, f(a)}_{=2\, I_n} k = 2\langle h, k \rangle$$

où I_n est la matrice identité de taille n.

TRAVAUX DIRIGÉS

Exercice 1 : Montrer que la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par :

$$f(x,y) = \begin{cases} \frac{y^2}{x} & \text{si} \quad x \neq 0 \\ y & \text{si} \quad x = 0 \end{cases}$$

admet des dérivées partielles au point $(0,0)$, mais n'est pas continue en $(0,0)$.

Exercice 2 : Soit $A \in \mathcal{M}_{n,m}(\mathbb{R})$, avec $(n,m) \in [\mathbb{N}^\star]^2$.

(a) Montrer que l'application $J : \mathbb{R}^m \to \mathbb{R}$ définie par $J(X) = \|AX\|^2$ est différentiable et calculer sa différentielle.

(b) Soit $f \in \mathscr{C}^1(\mathbb{R})$. Montrer que l'application $G : \mathbb{R}^m \to \mathbb{R}$ définie par $G(X) = f(J(X))$ est différentiable et calculer sa différentielle.

Exercice 3 : Calculer les différentielles d'ordre 1 et 2 en $(0,0)$ des applications suivantes :

(a) $f(x,y) = x^3 + \sqrt{1 + x^2 + y^2}$.

(b) $g(x,y,z) = xyz\sin(xy) + 2x + 5$.

Exercice 4 : Soit f la fonction définie sur $\mathbb{R}^2$ par :

$$f(x,y) = \frac{|x|^{3/2}y}{x^2+y^2} \text{ si } (x,y) \neq (0,0) \quad \text{et} \quad f(0,0) = 0.$$

(a) f est-elle continue ?

(b) Calculer les dérivées partielles de f . Sont-elles continues ?

(c) Calculer les dérivées directionnelles (si elles existent) de f en $(0,0)$.

(d) f est-elle différentiable en $(0,0)$?

Exercice 5 : Soit f la fonction définie sur $\mathbb{R}^2$ par :

$$f(x,y) = \frac{xy}{\sqrt{x^2+y^2}} \text{ si } (x,y) \neq (0,0) \quad \text{et} \quad f(0,0) = 0.$$

(a) Étudier la différentiabilité de f.

(b) f admet-elle des dérivées partielles ?

(c) f est-elle de classe $\mathscr{C}^1$?

Exercice 6 : Reprendre les questions de l'exercice précédent avec la fonction :

$$f(x,y) = (x^4+y^4)\sin\left(\frac{1}{\sqrt{x^4+y^4}}\right) \text{ si } (x,y) \neq (0,0) \quad \text{et} \quad f(0,0) = 0.$$

Exercice 7 : Soit f la fonction définie sur $\mathbb{R}^2$ par :

$$f(x, y) = x^2 y \sin\left(\frac{y}{x}\right) \quad \text{si } (x, y) \neq (0, 0) \quad \text{et} \quad f(0, 0) = 0.$$

(a) f admet-elle des dérivées partielles sur $\mathbb{R}^2$?

(b) f est-elle de classe $\mathscr{C}^1$ sur $\mathbb{R}^2$?

(c) Calculer (si elles existent) $\frac{\partial^2 f}{\partial x \partial y}(0, 0)$ et $\frac{\partial^2 f}{\partial y \partial x}(0, 0)$.

Exercice 8 : Soient $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction différentiable et a un point de $\mathbb{R}^2$. On considère les lignes de niveau $f(x) = c$.

Montrer que $\nabla f(a)$ est orthogonal à la ligne de niveau de f passant par le point a.