Optimisation convexe — Cours

Ivan Lejeune

28 janvier 2025

Table des matières

Chapitre 1 — Optimisation en dimension finie

Les rappels qui suivent sont fournis afin d'essayer, dans la mesure du possible, de regrouper l'ensemble des pré-requis nécessaires pour la suite. Aussi, certaines définitions sont rappelées de manière sommaire, et les résultats parfois non re-démontrés. Tous ces résultats sont très classiques et leur preuve facilement accessible.

1.1 Quelques notations et définitions

Notation 1.1.1. Pour tous $x, y \in \mathbb{R}^n$ on note de manière équivalente $x \cdot y$ ou (x, y) le produit scalaire de x et y, qui est donné par

$$x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

Notation 1.1.2. Pour tout $x \in \mathbb{R}^n$, on note par $|x|$ la norme euclidienne de x, donnée par

$$|x| = \sqrt{x \cdot x}$$

Notation 1.1.3. Pour tout $a \in \mathbb{R}^n$ et $r \in \mathbb{R}_+^*$ on note $B(a, r)$ la boule ouverte de centre a et rayon r, donnée par

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

On note $\overline{B}(a, r)$ la boule fermée de centre a et rayon r, donnée comme l'adhérence de $B(a, r)$.

Notation 1.1.4. Pour tous $a, b \in \mathbb{R}^n$, on note $[a, b]$ le sous-ensemble de $\mathbb{R}^n$ défini par

$$[a, b] = \{(1 - t)a + tb,\ t \in [0, 1]\}.$$

L'ensemble $[a, b]$ est aussi appelé *segment reliant a à b*.

1.2 Extremum local, global

Définition 1.2.1. Extremum

Soit $U \subset \mathbb{R}^n$, $a \in U$ et $f : U \to \mathbb{R}$:

1. on dit que a est un minimum global (ou absolu) de f sur U si $f(x) \geq f(a), \forall x \in U$,
2. on dit que a est le minimum global strict de f sur U si $f(x) > f(a), \forall x \in U \setminus \{a\}$,
3. on dit que a est un minimum local (ou relatif) de f sur U si il existe un voisinage $V \subset \mathbb{R}^n$ de a tel que $f(x) \geq f(a), \forall x \in V \cap U$,
4. on dit que a est un maximum global (respectivement local) de f sur U si a est un minimum global (respectivement local) de $-f$ sur U,
5. on dit que a est un extremum global (respectivement local) de f sur U si a est : soit un minimum global (respectivement local) de f sur U, soit un maximum global (respectivement local) de f sur U.

Dans la suite, nous étudions donc uniquement la question de la minimisation d'une fonction f : pour la maximisation de f, il suffit d'étudier la minimisation de la fonction $-f$.

1.3 Un peu de calcul différentiel

Les notions de calcul différentiel nécessaires pour suivre cette U.E. sont souvent encore mal assimilées au semestre 6 de licence. Les rappels qui suivent correspondent au "minimum vital" et n'ont pas vocation à remplacer un travail approfondi du calcul différentiel.

Soit $U \subset \mathbb{R}^n$ un ouvert et $f : U \to \mathbb{R}$.

Notation 1.3.1. On dit que f est de classe C^k sur U, noté $f \in C^k(U; \mathbb{R})$, si toutes les dérivées partielles jusqu'à l'ordre k existent et sont continues.

Notation 1.3.2. Pour tous $x \in U$, et $i \in \{1, \cdots, n\}$, on note (quand c'est défini)

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0} \frac{1}{t}(f(x + te_i) - f(x)),$$

la i^{ie} dérivée partielle de f en x.

Notation 1.3.3. Pour tous $x, h \in U$, on note (quand c'est défini)

$$f'(x)(h) \text{ ou de façon équivalente } f'(x) \cdot h$$

la dérivée (ou différentielle) de f en x évaluée dans la direction h et on rappelle que $f'(x) \in L(U, \mathbb{R})$.

Notation 1.3.4. Pour tout $x \in U$, on note (quand c'est défini)

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_n}(x) \right) \in \mathbb{R}^n$$

le gradient de f en x et on a $f'(x) \cdot h = (\nabla f(x), h)$.

Notation 1.3.5. Notez que dans certains ouvrage, $f'(x)$ et $\nabla f(x)$ sont assimilés à la Jacobienne de f en x. Retenez juste que, dans le cas qui nous concerne ici, toutes ces notations sont équivalentes.

Notation 1.3.6. Pour tous $x, h \in U$, on note (quand c'est défini)

$$\frac{\partial f}{\partial h}(x) \coloneqq \lim_{t \to 0} \frac{1}{t} (f(x + th) - f(x)) = g'(0),$$

la dérivée directionnelle de f en x de direction h, où on a noté $g(t) = f(x + th)$. On a alors :

$$\frac{\partial f}{\partial h}(x) = f'(x)(h) = (\nabla f(x), h) \, .$$

Notation 1.3.7. Pour tous $x \in U$, on note (quand c'est défini) $\nabla^2 f(x) \in M_n(\mathbb{R})$ la matrice hessienne de f en x, qui est définie par :

$$\left(\nabla^2 f(x)\right)_{ij} \coloneqq \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \ \forall i, j = 1, \cdots, n.$$

Notez que le Théorème de Schwarz nous assure, lorsque f est de régularité C^2, que $\nabla^2 f(x)$ est symétrique. Notez aussi que cette matrice peut-être assimilée à la dérivée seconde $f''(x) \in L(U;\, L(U; \mathbb{R}))$ ou encore la forme bilinéaire $f''(x) \in L(U \times U; \mathbb{R}))$.

Proposition 1.3.8. Gradient d'une composée
Soit $U \subset \mathbb{R}^n$ et $\Omega \subset \mathbb{R}$ ouverts. On suppose que $f \in \mathcal{C}^1(U; \mathbb{R})$ et $g \in \mathcal{C}^1(\Omega; \mathbb{R})$, avec de plus $f(U) \subset \Omega$. Alors $g \circ f$ est de classe C^1 et on a :

$$\nabla(g \circ f)(x) = g'(f(x)) \nabla f(x), \forall x \in U.$$

Proposition 1.3.9. Lien entre ∇f et $\nabla^2 f$
On a :
$$\nabla^2 f(x)h = \nabla\left(\nabla f(x), h\right)\ \forall x \in U,\ \forall h \in \mathbb{R}^n.$$

Exemple 1.3.10. Soit $a \in \mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}$ une forme linéaire définie par
$$\forall x \in \mathbb{R}^n,\ f(x) = (a, x)\,,$$
alors on a $\nabla f(x) = a$ et $\nabla^2 f(x) = 0$.

Exemple 1.3.11. Soit $A \in M_n(\mathbb{R})$ et $f : \mathbb{R}^n \to \mathbb{R}$ une forme quadratique définie par
$$\forall x \in \mathbb{R}^n,\ f(x) = (Ax, x)\,,$$
alors on a $f\nabla f(x) = (A + A^t)x$ et $\nabla^2 f(x) = A + A^t$. Si de plus on a $A \in S_n(\mathbb{R})$, alors $\nabla\left(Ax, x\right) = 2Ax$ et $\nabla^2\left(Ax, x\right) = 2A$

Exemple 1.3.12. Soit $B \in \mathcal{B}(E \times E, \mathbb{R})$ une application bilinéaire sur un espace vectoriel normé E de dimension finie. Alors B est différentiable et on a pour tout $(x, y) \in E^2$, $(h, k) \in E^2$:
$$B'(x, y) \cdot (h, k) = B(x, k) + B(h, y).$$
En effet, on a :
$$B(x + h, y + k) = B(x, y) + B(x, k) + B(h, y) + B(h, k) = B(x, y) + \mathcal{L}(h, k) + o(|(|h, k)|)|,$$
et on vérifie que $\mathcal{L}$ est linéaire.

Exemple 1.3.13. Soit $A \in M_n(\mathbb{R})$, on considère l'application suivante :
$$f\ :\ \mathcal{GL}_n(\mathbb{R}) \ni M \mapsto M^{-1} \in \mathcal{GL}_n(\mathbb{R}).$$
Alors f est différentiable et pour tout $M \in \mathcal{GL}_n(\mathbb{R})$, $H \in \mathcal{GL}_n(\mathbb{R})$, on a :
$$f'(M) \cdot H = -M^{-1}HM^{-1}.$$
En effet,
$$f(M + H) = (M + H)^{-1} = \left(M(I_n + M^{-1}H)\right)^{-1} = (I_n + M^{-1}H)^{-1}M^{-1},$$
d'où
$$f(M + H) = \left((I_n - M^{-1}H + o(|H||)|\right)M^{-1} = f(M) - M^{-1}HM^{-1} + o(|H||)|,$$
et on vérifie sans peine que, pour $M \in \mathcal{GL}_n(\mathbb{R})$, l'application $\mathcal{L} : \mathcal{GL}_n(\mathbb{R}) \ni H \mapsto -M^{-1}HM^{-1} = \mathcal{L}(H)$ est linéaire.

Théorème 1.3.14. Théorème fondamental de l'analyse
Soit $U \subset \mathbb{R}^n$ ouvert. On suppose que $f \in \mathcal{C}^1(U; \mathbb{R})$. Alors $\forall (x, y) \in U^2$, tels que $\forall t \in [0, 1],\ x + t(y - x) \in U$, on a :
$$f(y) = f(x) + \int_0^1 f'(x + t(y - x)) \cdot (y - x)\, dt.$$

ou encore, en posant $y = x + h$ et utilisant la notation vectorielle :

$$f(x+h) = f(x) + \int_0^1 (\nabla f(x+th), h)\, dt.$$

Démonstration. On considère

$$\phi : \left| \begin{array}{ccc} [0,1] & \longrightarrow & \mathbb{R} \\ t & \longmapsto & f\left(x + t(y-x)\right), \end{array} \right|$$

Par construction, ϕ est de régularité C^1 et on a

$$\phi'(t) = f'\left(x + t(y-x)\right) \cdot (y-x).$$

En appliquant le TFA pour les fonctions $C^1(\mathbb{R};\mathbb{R})$, on a

$$\phi(1) = \phi(0) + \int_0^1 \phi'(s)\, ds,$$

si bien que

$$f(y) = f(x) + \int_0^1 f'(x + t(y-x)) \cdot (y-x)\, dt.$$

Notez que cette formule est également la formule de Taylor à l'ordre 1 avec reste intégral. $\square$

Proposition 1.3.15. Formules de Taylor-Young
Soit $U \subset \mathbb{R}^n$ ouvert. On suppose que $f \in \mathcal{C}^2(U;\mathbb{R})$. Alors $\forall x \in U$, il existe un voisinage $V \in U$ de x tels que $\forall y = x + h \in V$, on ait :

$$f(x+h) = f(x) + f'(x) \cdot h + o(|h|) \quad \text{(ordre 1)},$$

et

$$f(x+h) = f(x) + f'(x) \cdot h + f''(x) \cdot (h,h) + o(|h|^2) \quad \text{(ordre 2)},$$

ou encore en notation matricielle :

$$f(x+h) = f(x) + (\nabla f(x), h) + \left(\nabla^2 f(x) h, h\right) + o(|h|^2).$$

Démonstration. ⋯....
Rappelons que la notation $o(|h|^k)$ pour $k \in \mathbb{N}^*$, signifie une expression qui tend vers 0 plus vite que $|h|^k$: si on la divise par $|h|^k$, le résultat tend toujours vers 0 quand $|h|$ tend vers 0. $\square$

Proposition 1.3.16. Formule de Taylor-Lagrange d'ordre 1
Soit $U \subset \mathbb{R}^n$ ouvert. On suppose que $f \in \mathcal{C}^1(U;\mathbb{R})$. Alors $\forall x \in U$, il existe un voisinage $V \in U$ de x et $0 < \theta < 1$ tels que $\forall y = x + h \in V$, tels que :

$$f(x+h) = f(x) + (\nabla f(x + \theta h), h).$$

Démonstration. ⋯ $\square$