Optimisation & Analyse convexe
Séance 1 : Existence d'un minimum. Convexité

Exercice 1 Soit $A \in \mathbb{R}^{n\times n}$ une matrice symétrique, et b un vecteur de $\mathbb{R}^n$. On considère la fonction :

$$f : x \in \mathbb{R}^n \longmapsto \frac{1}{2}(x, Ax) - (b, x).$$

1. Soient $\lambda_{\min}$ et $\lambda_{\max}$ la plus petite et la plus grande valeur propre de A. Montrer que

$$\forall x \in \mathbb{R}^n \qquad \lambda_{\min}\|x\|_2^2 \leq (Ax, x) \leq \lambda_{\max}\|x\|_2^2.$$

2. Montrer que f est dérivable sur $\mathbb{R}^n$. Calculer $df(x)$ et $\nabla f(x)$, pour $x \in \mathbb{R}^n$.
3. Supposons maintenant que A est symétrique positive. Montrer que dans ce cas, A est inversible si et seulement A est définie positive.

Corrigé 1

1. La matice A est symétrique, toutes ses valeurs propres λ_i sont réelles (eventuellement nulles), et elle admet une base de vecteurs propres orthonormés $\mathcal{B} = \{v_1, v_2, \cdots, v_n\}$:

$$Av_i = \lambda_i v_i \quad \text{pour } i = 1, 2, \cdots, n \quad \text{et} \quad (v_i, v_j) = \delta_{ij} \quad \text{pour } i, j = 1, 2, \cdots, n.$$

Pour tout vecteur $x \in \mathbb{R}^n$, on a donc

$$x = \sum_{i=1}^n x_i v_i, \quad Ax = \sum_{i=1}^n x_i A v_i = \sum_{i=1}^n \lambda_i x_i v_i \quad \text{et} \quad (Ax, x) = \sum_{i=1}^n \lambda_i x_i^2$$

soit finalement

$$\min_i \lambda_i \sum_{i=1}^n x_i^2 \leq (Ax, x) \leq \max_i \lambda_i \sum_{i=1}^n x_i^2.$$

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

$$\begin{aligned} f(x+h) &= \frac{1}{2}(A(x+h), x+h) - (b, x+h) \\ &= \frac{1}{2}(Ax, x) + (Ax, h) + \frac{1}{2}(Ah, h) - (b, x) - (b, h) \\ &= f(x) + (Ax - b, h) + \frac{1}{2}(Ah, h). \end{aligned}$$

Si on pose $\varepsilon(h) = \frac{1}{2}(Ah, h)/\|h\|_2$, alors

$$|(Ah, h)| \leq \|A\|\|h\|_2^2 \Rightarrow \lim_{h\to 0} |\varepsilon(h)| = 0.$$

Soit

$$df(x).h = (Ax - b, h) \quad \text{et} \quad \nabla f(x) = Ax - b.$$

Commentaire 1 : Dans le cas général (A non symétrique) :

$$\begin{aligned} &f(x) = \frac{1}{2}(x, \frac{(A + A^T)}{2} x) - (b, x); \\ &df(x).h = (\frac{1}{2}(A + A^T)x - b, h) \quad \text{et} \quad \nabla f(x) = \frac{1}{2}(A + A^T)x - b. \end{aligned}$$

Commentaire 2 : Il est clair que la fonction f est deux fois différentiable. Pour tout $x \in \mathbb{R}^n$, on a :

$$df(x+k).h = (Ax-b,h) + (Ak,h), \quad \forall h,k \in \mathbb{R}^n.$$

Comme A est supposée symétrique, on conclut que $d^2f(x).(h,k) = (Ah,k)$ et $\nabla^2 f(x) = A$.

3. Supposons que A soit inversible. Soit w tel que $(Aw,w) = 0$. On va montrer que w est en fait égal à zéro. Soient $d \in \mathbb{R}^n$ $(d \neq 0)$ et $\lambda > 0$. Comme A est positive et symétrique :
$$0 \leq (A\{w+\lambda\, d\}, w+\lambda\, d) = 2\lambda(Aw,d) + \lambda^2(Ad,d).$$
En mettant λ en facteur, puis en faisant tendre λ vers 0 dans le facteur restant, on obtient $(Aw,d) \geq 0$. Comme c'est valable pour toute direction ($d \in \mathbb{R}^n$), on en déduit que $Aw = 0$, ce qui conduit enfin à $w = 0$ par hypothèse sur A.
La réciproque est aisée et classique. En effet, si A est définie-positive, $Aw = 0$ entraîne que $(Aw,w) = 0$, et donc que $w = 0$: par conséquent, A est inversible.

Exercice 2 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = \dfrac{1}{2}(x,Ax) - (b,x) + c$ avec A une matrice $n \times n$ symétrique, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

1. Montrer que f admet un minimum sur tout compact K de $\mathbb{R}^n$.
2. Montrer que f est convexe ssi A est positive[1].
3. Montrer que f est stritement convexe ssi A est définie positive.
4. Montrer que si A est définie positive, alors f est "*infinie à l'infini*"et admet un minimum unique sur tout fermé convexe $K \subset \mathbb{R}^n$.
5. Montrer que, s'ils existent, les minima de f vérifient la relation $Ax = b$.
En déduire que si A est non positive, alors f n'admet pas de minimum sur $\mathbb{R}^n$.

Corrigé 2

1. f étant continue, elle atteint ses minima sur tout compact K.
2. Rappelons que la fonction f est convexe si et seulement si :
$$\forall x,y \in \mathbb{R}^n, (\nabla f(y) - \nabla f(x), y-x) \geq 0.$$
D'autre part, on a $\nabla f(x) = Ax - b$. Il vient alors :
$$\begin{aligned} f \text{ est convexe} \quad &\Leftrightarrow \quad ((Ay-b)-(Ax-b), y-x) \geq 0 \quad \forall y,x \in \mathbb{R}^n \\ &\Leftrightarrow \quad (A(y-x), y-x) \geq 0 \quad \forall y,x \in \mathbb{R}^n \\ &\Leftrightarrow \quad (Az,z) \geq 0 \quad \forall z \in \mathbb{R}^n \\ &\Leftrightarrow \quad A \text{ est positive.} \end{aligned}$$
3. La stricte convexité de f est equivalente à :
$$\forall x \neq y \in \mathbb{R}^n, (\nabla f(y) - \nabla f(x), y-x) > 0.$$
4. Supposons que la matrice A est symétrique définie positive. Donc d'après l'exercice 1, on peut minorer la foncion f par :
$$f(x) \geq \frac{1}{2}\lambda_{\min}(A)\|x\|_2^2 - \|b\|_2\|x\|_2 + c,$$

[1] c'est à dire $(Av,v) \geq 0$ pour tout $v \in \mathbb{R}^n$. On parle aussi dans certains livres de semi-définie positive

avec $\lambda_{\min}(A) > 0$ et $\lim\limits_{z\to+\infty} \frac{1}{2}\lambda_{\min}(A)z^2 - \|b\|_2 z + c = +\infty$. On en déduit alors que :

$$\lim_{x\in K, \|x\|\to+\infty} f(x) = +\infty;$$

pour tout ensemble fermé K de $\mathbb{R}^n$. f est alors "infinie à l'infini" et strictement convexe, elle admet donc un minimum unique sur tout fermé convexe K.

5. Supposons que f atteint le minimum en $x \in \mathbb{R}^n$, alors il existe V un voisinage de x tel que :

$$f(y) \geq f(x), \forall y \in V.$$

En particulier, $f(x+\lambda d) - f(x) \geq 0$ pour tout $d \in \mathbb{R}^n$ et pour tout $\lambda > 0$ assez petit tel que $x + \lambda d \in V$. En divisant par λ et en faisant tendre λ vers 0 par valeurs supérieures, on en déduit que x satisfait :

$$\nabla f(x).d = (Ax - b, d) \geq 0, \quad \forall d \in \mathbb{R}^n.$$

Ou encore, puisque d parcourt $\mathbb{R}^n$,

$$Ax = b.$$

D'autre part, si x existe alors

$$f(x+\lambda d) - f(x) = \lambda(Ax-b,d) + \frac{\lambda^2}{2}(Ad,d) = \frac{\lambda^2}{2}(Ad,d) \geq 0,$$

pour tout $d \in \mathbb{R}^n$ (et pour tout $\lambda > 0$, assez petit tel que $x + \lambda d$ reste dans le voisinage V de x). Donc A est necessairement positive.

Exercice 3

1. La composée de deux fonctions convexes est-elle convexe ?
2. Examiner, parmi les fonctions suivantes, lesquelles sont convexes, strictement convexes, $\alpha-$convexes :

$$f(x) = -\ln(x), \quad g(x) = x^4 - x^2, \quad h(x) = x^\beta \ (\beta \in \mathbb{N}), \quad \ell(x) = |x|^\beta \ (\beta > 1)?$$

3. Soit $\|\cdot\|$ une norme de $\mathbb{R}^n$. Montrer que $x \in \mathbb{R}^n \longmapsto \|x\|$ et $x \in \mathbb{R}^n \longmapsto \|x\|^2$ sont convexes.

Corrigé 3

1. La composée de deux fonctions convexes n'est pas forcément convexe. En effet, les fonctions $g : x \longmapsto e^{-x}$ et $f : x \longmapsto x^2$ sont bien convexes sur R, et pourtant $g \circ f : x \longmapsto e^{-x^2}$ n'est pas convexe sur $\mathbb{R}$!

   Par contre, on peut facilement vérifier le résultat suivant : Soit K une partie convexe de $\mathbb{R}$.

   > Soient $f : \mathbb{R}^n \longrightarrow K$ une fonction convexe et $g : K \longrightarrow \mathbb{R}$ une fonction convexe croissante. Alors $g \circ f$ est convexe.

2. La fonction f est strictement convexe (sur $\mathbb{R}_+^*$), mais n'est pas $\alpha-$convexe.
La fonction g n'est pas convexe sur $\mathbb{R}$, mais est convexe sur les intervalles $]-\infty, -\frac{\sqrt{6}}{6}]$ et $[\frac{\sqrt{6}}{6}, +\infty[$.
Pour $\beta \in \{0,1\}$, la fonction h est convexe. Pour $\beta \geq 2$, h n'est convexe sur $\mathbb{R}$ que si β est paire. Dans le cas $\beta = 2$, h est α-convexe avec $\alpha = 2$.
La fonction ℓ se décompose en $s \circ r$ où s et r sont définies par :
$$s : x \in \mathbb{R}^+ \longmapsto x^\beta, \quad r : x \in \mathbb{R} \longmapsto |x|.$$
Il est facile de vérifier que r est convexe, et s est strictement convexe et croissante sur $\mathbb{R}^+$. On conclut alors que pour $\beta > 1$, ℓ est strictement convexe sur $\mathbb{R}$.

Exercice 4 Soit f de $\mathbb{R}^2$ dans $\mathbb{R}^2$ définie par :
$$f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{sinon.} \end{cases}$$

1. Montrer que f est continue en $(0,0)$.
2. Montrer que f n'est pas différentiable en $(0,0)$.
3. Est-elle Gâteaux-différentiable ?

Corrigé 4

1. De l'inégalité $2|xy| \leq x^2+y^2$, on obtient la majoration de $|f(x,y)|$ par :
$$|f(x,y)| \leq \frac{1}{2}\sqrt{x^2+y^2}, \quad \forall x,y \in \mathbb{R}.$$
On en déduit alors que
$$\lim_{(x,y)\to 0} f(x,y) = 0 = f(0,0)$$
ce qui prouve la continuité de f en $(0,0)$.
2. Pour que f soit différentiable en $(0,0)$ il faut et il suffit qu'il existe une forme linéaire $df(0,0)$ tel que :
$$f(x,y) = f(0,0) + df(0,0).(x,y) + \|(x,y)\|\varepsilon(x,y); \qquad \text{avec} \lim_{(x,y)\to(0,0)} \varepsilon(x,y) = 0.$$
Or si on fixe $y = 0$, on obient :
$$f(x,0) = 0 = f(0,0) + 0, \quad \forall x \in \mathbb{R}.$$
Donc si f est différentiable en $(0,0)$, alors $df(0,0).(x,0) = 0$ pour tout $x \in \mathbb{R}$, et de même on pourrait établir que $df(0,0).(0,y) = 0$. Ce qui entraînerait que $df(0,0) \equiv 0$.
Maintenant, si on prend $x = y \neq 0$, alors on obtient :
$$f(x,x) = \frac{x^2}{\sqrt{2x^2}} = \frac{|x|}{\sqrt{2}},$$
et
$$\lim_{(x,x)\to 0} \frac{f(x,x) - f(0,0) - 0.(x,x)}{\|(x,x)\|} \neq 0.$$
Donc la forme linéaire nulle n'est pas la différentielle de f en $(0,0)$ et f n'est pas différentiable en ce point.

3. Pour tout $t > 0$ et tout $(x, y) \neq (0, 0)$, on a :
$$\frac{f(tx, ty) - f(0, 0)}{t} = \frac{xy}{\sqrt{x^2 + y^2}}.$$
donc
$$\lim_{t \to 0^+} \frac{f(tx, ty) - f(0, 0)}{t} = \frac{xy}{\sqrt{x^2 + y^2}}.$$
Cette limite n'étant pas une forme linéaire, on conclut que f n'est pas Gâteaux-différentiable.

Exercice 5 Montrer que :

1. L'application $f : x \in \mathbb{R}^n \longmapsto \|x\|$ n'est jamais différentiable à l'origine.
2. L'application $f : x \in \mathbb{R}^n \longmapsto \|x\|_2$ (la norme euclidienne) est différentiable au sens de Fréchet en tout point $x \neq 0$, et
$$df(x).h = \frac{(x, h)}{\|x\|_2}, \qquad \forall h \in \mathbb{R}^n.$$
3. L'application $f : x \in \mathbb{R}^n \longmapsto \|x\|_\infty$ est dérivable en un point $a \in \mathbb{R}^n$, si et seulement si il existe un indice i_0, tel que
$$\forall i \neq i_0, \quad |a_{i_0}| > |a_i|.$$

Corrigé 5

1. On suppose que l'application $f : x \in \mathbb{R}^n \longmapsto \|x\|$ est différentiable en 0 ; alors on peut écrire
$$f(0 + h) = f(0) + (\nabla f(0), h) + \|h\|\varepsilon(h), \quad \text{avec } \lim_{h \to 0} \|\varepsilon(h)\| = 0;$$
soit encore
$$1 = \frac{\|h\|}{\|h\|} = 0 + (\nabla f(0), \frac{h}{\|h\|}) + \varepsilon(h), \quad \text{avec } \lim_{h \to 0} \|\varepsilon(h)\| = 0.$$
En changeant h par $-h$, on aussi :
$$1 = \frac{\|h\|}{\|h\|} = 0 - (\nabla f(0), \frac{h}{\|h\|}) + \varepsilon(-h).$$
Par addition, on obtient la contradiction
$$2 = \varepsilon(h) + \varepsilon(-h) \quad \text{avec } \lim_{h \to 0} \|\varepsilon(h)\| = \lim_{h \to 0} \|\varepsilon(-h)\| = 0.$$
2. Pour la norme euclidienne, on écrit pour tout $x \neq 0$
$$\|x + h\|_2^2 - \|x\|_2^2 = 2(x, h) + \|h\|_2^2.$$
On obtient
$$\|x + h\|_2 - \|x\|_2 = \frac{2(x, h) + \|h\|_2^2}{\|x + h\|_2 + \|x\|_2}.$$

Reécrivons tout d'abord la partie en $\|h\|_2^2$, sous la forme $\|h\|_2\varepsilon_1(h)$, avec

$$\varepsilon_1(h) = \frac{\|h\|_2}{\|x+h\|_2 + \|x\|_2}; \quad \text{on a } 0 \le \varepsilon_1(h) \le \frac{\|h\|_2}{\|x\|_2}, \text{ et } \lim_{h\to 0} |\varepsilon_1(h)| = 0.$$

Si maintenant on extrait la partie linéaire de ce qui reste, on trouve

$$\frac{2(x,h)}{\|x+h\|_2 + \|x\|_2} = \frac{(x,h)}{\|x\|_2} + \frac{(x,h)\{\|x\|_2 - \|x+h\|_2\}}{\|x\|_2\{\|x+h\|_2 + \|x\|_2\}}.$$

Ecrivons le second terme sous la forme $\|h\|_2\varepsilon_2(h)$, avec

$$\varepsilon_2(h) = \frac{(x,h)\{\|x\|_2 - \|x+h\|_2\}}{\|h\|_2\|x\|_2\{\|x+h\|_2 + \|x\|_2\}}.$$

D'après l'inégalité de Cauchy-Schwartz, on a $|(x,h)| \le \|h\|_2\|x\|_2$, et d'après l'inégalité triangulaire, on a également $|\|x\|_2 - \|x+h\|_2| \le \|h\|_2$. Ainsi

$$0 \le |\varepsilon_2(h)| \le \frac{\|h\|_2}{\|x\|_2}, \text{ et } \lim_{h\to 0} |\varepsilon_2(h)| = 0.$$

Remarque. Un raisonnement plus rapide consiste à exprimer f sous la forme $f = g \circ f_2$, avec $g : t \longmapsto \sqrt{t}$ et $f_2 : x \longmapsto \|x\|_2^2$. Dés que $x \neq 0$, on déduit du théorème de composition des différentielles que f est bien différentiable, et on a :

$$df(x).h = dg(f_2(x)).(df_2(x).h), \quad \forall h \in \mathbb{R}^n.$$

Or on a

$$dg(s).y = \frac{1}{2\sqrt{s}}y, \ \forall s \in \mathbb{R}_+^*, \qquad df_2(x).h = 2(x,h), \ \forall h \in \mathbb{R}^n.$$

On en déduit finalement que, pour $x \neq 0$, on a :

$$df(x).h = \frac{(x,h)}{\|x\|_2}, \quad \text{et } \nabla f(x) = \frac{1}{\|x\|_2}x.$$

3. On considre d'abord le cas où il existe un indice i_0 tel que $|a_{i_0}| > |a_i|$ pour tout $i \neq i_0$. On pose $\delta = |a_{i_0}| - \max_{i\neq i_0} |a_i|$; $\delta > 0$ et pour tout $h \in \mathbb{R}^n$ tel que $\|h\|_\infty \le \delta/2$, on a

$$f(a+h) - f(a) = \|a+h\|_\infty - \|a\|_\infty = |a_{i_0} + h_{i_0}| - |a_{i_0}| = h_{i_0}\text{sign}(a_{i_0}).$$

ainsi l'application f est différentiable en a et

$$df(a).h = (\nabla f(a), h) = h_{i_0}\text{sign}(a_{i_0}).$$

On considère maintenant le cas où il existe (au moins) deux indices i_0 et j_0, tels que :

$$\|a\|_\infty = |a_{i_0}| = |a_{j_0}|$$

et $a \neq 0$ (aucune norme n'est dérivable en 0, voir plus haut). Alors pour tout $h \in \mathbb{R}^n$ tel que $|h_{i_0}| < |a_{i_0}|$, et $h_i = 0$ pour tout $i \neq i_0$, on écrit

$$\|a+h\|_\infty - \|a\|_\infty = \max_i |a_i + h_i| - |a_{i_0}|.$$

A partir de là, on a $\max_i |a_i + h_i| = |a_{i_0} + h_{i_0}| = \|a\|_\infty + \|h\|_\infty$ si h_{i_0} et a_{i_0} sont de même signe, et $\max_i |a_i + h_i| = |a_{i_0} = \|a\|_\infty$ sinon. On trouve donc

$$\|a+h\|_\infty - \|a\|_\infty = \begin{cases} \|h\|_\infty & \text{si sign}(h_{i_0}) = \text{sign}(a_{i_0}), \\ 0 & \text{si sign}(h_{i_0}) = -\text{sign}(a_{i_0}). \end{cases}$$

L'application f n'est donc pas différentiable en a.