Module A3
Fonctions convexes différentiables

Dans ce module, sauf mention contraire, $\mathcal{X}$ et $\mathcal{Y}$ désignent des espaces de Hilbert, munis chacun d'un produit scalaire noté $\langle \cdot, \cdot \rangle$ et de norme associée notée $\| \cdot \|$, tandis que E désigne un espace euclidien, que l'on identifiera à $\mathbb{R}^n$ avec $n \in \mathbb{N}^*$, muni du produit scalaire usuel noté $\langle \cdot, \cdot \rangle$ également et de norme associée la norme euclidienne, notée $\| \cdot \|_2$.

En optimisation continue, les fonctions convexes différentiables occupent une place importante. L'objectif de ce module est de rappeler la définition et les premières propriétés des fonctions convexes, puis de donner des caractérisations dans le cas différentiable. On introduira enfin une classe de fonctions dites fortement convexes, qui présentent des propriétés particulièrement intéressantes en optimisation numérique.

1 Fonctions convexes différentiables

1.1 Définition et exemples

Définition 1 (Fonction convexe)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. On dit que f est *convexe* si

$$\forall\, (x_1, x_2) \in \mathcal{X}^2, \forall \lambda \in [\,0\,;1\,], \quad f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

> La définition ci-dessus est en réalité celle des fonctions convexes définies sur tout l'espace $\mathcal{X}$. La définition usuelle d'une fonction convexe tient compte de son domaine de définition qui doit être **convexe** (cf. **B6 : Projection sur un convexe**).

Exemple

Fonctions affines. Soit $a \in \mathcal{X}$ et $b \in \mathbb{R}$. Considérons la fonction

$$f : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \langle a, x \rangle + b \end{array} \right.$$

Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. Puisque

$$\langle a, \lambda\, x_1 + (1-\lambda)\, x_2 \rangle + b = \lambda \left(\langle a, x_1 \rangle + b \right) + (1-\lambda) \left(\langle a, x_2 \rangle + b \right)$$

on en déduit que f est convexe.

Géométriquement, dans le cas d'une fonction réelle ($\mathcal{X} = \mathbb{R}$), cela s'interprète de la manière suivante : la portion du graphe de f dont les abscisses sont comprises entre $x_1 \in \mathcal{X}$ et $x_2 \in \mathcal{X}$:

$$\left\{ \left(x, f(x)\right) \in \mathcal{X} \times \mathbb{R} \mid x_1 \leq x \leq x_2 \right\}$$

est située *en-dessous* du segment reliant $\big(x_1, f(x_1)\big)$ et $(x_2, f(x_2))$: on dit que

L'arc est en-dessous de la corde.

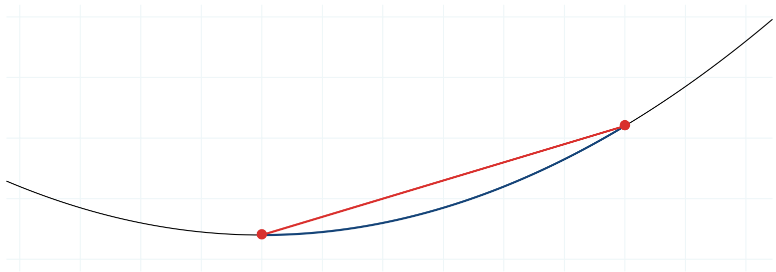

FIGURE 1 – Exemple du graphe d'une fonction convexe. En rouge la corde, en bleu l'arc.

Certaines opérations préservent la convexité des fonctions.

Proposition 1 (Combinaison linéaire)

Soient $f_1 : \mathcal{X} \to \mathbb{R}$ et $f_2 : \mathcal{X} \to \mathbb{R}$ deux fonctions convexes. Alors pour tout $\alpha_1 \geq 0$ et $\alpha_2 \geq 0$, la fonction $\alpha_1\, f_1 + \alpha_2\, f_2$ est convexe.

REMARQUE : Attention au signe des coefficients α_i !

DÉMONSTRATION : Laissée en exercice.

Proposition 2 (Composition par une application linéaire)

Soient $f : \mathcal{X} \to \mathbb{R}$ une fonction convexe et $A : \mathcal{Y} \to \mathcal{X}$ une application affine. Alors la fonction $f \circ A : \mathcal{Y} \to \mathbb{R}$ est convexe.

DÉMONSTRATION : Il suffit d'utiliser la linéarité partielle de A et de remarquer que $b = \lambda\, b + (1-\lambda)\, b$ pour tout $b \in \mathbb{R}$ et tout $\lambda \in [\,0\,;1\,]$. ■

La proposition 2 suggère que toutes les compositions de fonctions convexes ne sont pas nécessairement convexes. Il suffit de considérer par exemple la fonction $-\exp$. De manière générale, si les fonctions en jeu ne sont pas affines, seule la composition externe par une fonction convexe croissante d'une fonction convexe est assurée d'être convexe, comme on peut le voir dans la proposition suivante.

Proposition 3 (Composition par une fonction convexe croissante)

Soiten $f : \mathcal{X} \to \mathbb{R}$ une fonction convexe et $g : \mathbb{R} \to \mathbb{R}$ une fonction convexe **croissante**. Alors $g \circ f$ est convexe.

DÉMONSTRATION : On a pour tout $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$

$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

par convexité de f ; par croissance de g, on a alors

$$g \circ f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq g\big(\lambda\, f(x_1) + (1-\lambda)\, f(x_2)\big)$$

On conclut à l'aide de la convexité de g. ■

1.2 Caractérisation des fonctions convexes différentiables

Pour les fonctions différentiables sur un espace de Hilbert, la convexité est caractérisée grâce à son gradient :

Proposition 4 (Caractérisation des fonctions convexes différentiables)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction **différentiable**. Alors on a équivalence entre les énoncés suivants :

(i) f est convexe ;

(ii) $$\forall\, (x_1, x_2) \in \mathcal{X}^2, \qquad f(x_2) \geq f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle$$

(iii) ∇f est monotone, c'est-à-dire

$$\forall\, (x_1, x_2) \in \mathcal{X}^2, \qquad \langle \nabla f(x_2) - \nabla f(x_1), x_2 - x_1 \rangle \geq 0$$

> Cette caractérisation (ainsi que celles des fonctions strictement convexes et fortement convexes énoncées plus loin dans ce module) restent valables lorsque les fonctions sont définies et différentiables sur un ouvert $U \subset \mathcal{X}$ **convexe**.

Démonstration : On démontre successivement les différentes implications.

- (i) ⇒ (ii). Supposons que f est convexe. Soit $\lambda \in \,]\,0\,;1\,[$ et $(x_1, x_2) \in \mathcal{X}^2$ tels que $x_1 \neq x_2$. La définition de la convexité s'écrit

$$f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

soit, en développant les produits et en réarrangeant les termes obtenus,

$$f\big(x_2 + \lambda\,(x_1 - x_2)\big) - f(x_2) \leq \lambda\,\big(f(x_1) - f(x_2)\big)$$

Divisons par $\lambda > 0$. On obtient

$$\frac{f\big(x_2 + \lambda\,(x_1 - x_2)\big) - f(x_2)}{\lambda} \leq f(x_1) - f(x_2)$$

Puisque f est différentiable en x_2, la limite suivante existe et vaut

$$\lim_{\lambda \to 0} \frac{f\big(x_2 + \lambda\,(x_1 - x_2)\big) - f(x_2)}{\lambda} = f'(x_2; x_1 - x_2) = \langle \nabla f(x_2), x_1 - x_2 \rangle$$

Ainsi, on peut passer à la limite lorsque λ tend vers 0 dans l'inégalité précédente, ce qui donne

$$\langle \nabla f(x_2), x_1 - x_2 \rangle \leq f(x_1) - f(x_2)$$

- (ii) ⇒ (iii). On suppose que (ii) est vérifiée. Soit $(x_1, x_2) \in \mathcal{X}^2$. En particulier, l'inégalité (ii) est vraie pour (x_1, x_2) et (x_2, x_1). En sommant les deux inégalités obtenues, il en découle que

$$f(x_1) + f(x_2) \geq f(x_1) + f(x_2) + \langle \nabla f(x_1) - \nabla f(x_2), x_2 - x_1 \rangle$$

En simplifiant cette relation, on obtient le résultat désiré.

- (iii) $\Rightarrow$ (ii). On suppose que (iii) est vérifiée. On définit

$$g : \begin{cases} [\,0\,;1\,] & \to \mathbb{R} \\ t & \mapsto f\big(x_1 + t(x_2 - x_1)\big) \end{cases}$$

La fonction g est continue sur $[\,0\,;1\,]$ (par composition) et dérivable sur $]\,0\,;1\,[$ (par composition) de dérivée

$$\forall\, t \in]\,0\,;1\,[\qquad g'(t) = \langle \nabla f(x_1 + t(x_2 - x_1)), x_2 - x_1\rangle$$

D'après le théorème des accroissements finis il existe $t_0 \in]\,0\,;1\,[$ tel que

$$f\big(x_1 + (x_2 - x_1)\big) - f(x_1) = \big\langle \nabla f\big(x_1 + t_0\,(x_2 - x_1)\big), x_2 - x_1\big\rangle$$

Le membre de gauche vaut $f(x_2) - f(x_1)$, tandis que le membre de droite peut être minoré ; en effet, l'hypothèse (iii) aux points $x_1 + t_0\,(x_2 - x_1)$ et x_1 s'écrit

$$\big\langle \nabla f\big(x_1 + t_0\,(x_2 - x_1)\big) - \nabla f(x_1), \big(x_1 + t_0\,(x_2 - x_1)\big) - x_1\big\rangle \geq 0$$

ce qui donne après simplification la minoration suivante :

$$\big\langle \nabla f\big(x_1 + t_0\,(x_2 - x_1)\big), t_0\,(x_2 - x_1)\big\rangle \geq \big\langle \nabla f(x_1), t_0\,(x_2 - x_1)\big\rangle$$

Il suffit alors de simplifier par $t_0 > 0$ (grâce à la bilinéarité du produit scalaire) pour obtenir le résultat désiré.

- (ii) $\Rightarrow$ (i). On suppose que (ii) est vérifiée. Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. Posons $x_\lambda = (1-\lambda)x_1 + \lambda\, x_2$. Appliquons l'hypothèse (ii) aux points (x_1, x_λ) et (x_2, x_λ) :

$$f(x_1) \geq f(x_\lambda) + \langle \nabla f(x_\lambda), x_1 - x_\lambda\rangle \qquad \text{et} \qquad f(x_2) \geq f(x_\lambda) + \langle \nabla f(x_\lambda), x_2 - x_\lambda\rangle$$

On remarque que $x_1 - x_\lambda = \lambda(x_1 - x_2)$ et $x_2 - x_\lambda = (1-\lambda)(x_2 - x_1)$. Multiplions alors la première inégalité par $(1-\lambda)$ et la seconde par λ, les deux facteurs étant positifs. En sommant ces deux inégalités, on obtient alors, après réarrangement, que

$$f(x_\lambda) \leq (1-\lambda)\, f(x_1) + \lambda\, f(x_2)$$

d'où l'on déduit que f est convexe. ■

Dans la démonstration ci-dessus, seules les dérivées partielles sont utilisées. Aussi, on peut remplacer dans l'énoncé de la proposition 4 l'hypothèse de différentiabilité par l'existence de dérivées partielles suivant toutes les directions ou par la GATEAUX-différentiabilité. Dans ce dernier cas, l'énoncé (ii) s'écrit

$$\forall\, (x_1, x_2) \in \mathcal{X}^2, \qquad f(x_2) \geq f(x_1) + \langle \nabla_{\mathrm{G}} f(x_1), x_2 - x_1\rangle$$

et l'énoncé (iii) s'écrit

$$\forall\, (x_1, x_2) \in \mathcal{X}^2, \qquad \langle \nabla_{\mathrm{G}} f(x_2) - \nabla_{\mathrm{G}} f(x_1), x_2 - x_1\rangle \geq 0$$

Notons que, lorsque $\mathcal{X} = \mathbb{R}$, l'énoncé (iii) s'écrit

$$\forall\, (x_1, x_2) \in \mathcal{X}^2, \qquad \big(f'(x_2) - f'(x_1)\big)(x_2 - x_1) \geq 0$$

Autrement dit, si $x_1 \leq x_2$, alors $f'(x_1) \leq f'(x_2)$: la **monotonie** de ∇f est donc équivalente à la **croissance** de la dérivée f'. Par ailleurs, l'énoncé (ii) équivaut au fait que, pour tout $a \in \mathbb{R}$,

$$\forall\, x \in \mathbb{R}, \qquad f(x) \geq f(a) + f'(a)\,(x - a)$$

Le membre de droite donne l'équation d'une droite, passant par le point $(a, f(a))$ et de pente $f'(a)$: il s'agit de la tangente au graphe de f au point $(a, f(a))$. Ainsi, le graphe de f reste au-dessus de ses tangentes.

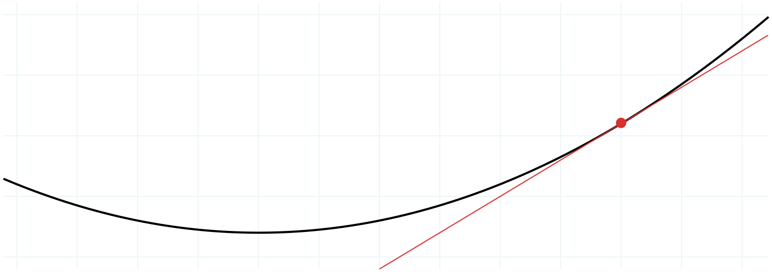

FIGURE 2 – Exemple du graphe d'une fonction convexe dérivable. En rouge une tangente au graphe.

1.3 Fonctions strictement convexes

Lorsque $x_1 = x_2$ ou $\lambda \in \{0, 1\}$, toutes les fonctions f satisfont la relation

$$f(\lambda\, x_1 + (1-\lambda)\, x_2) = \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

Ainsi, pour vérifier qu'une fonction f est convexe, il suffit de démontrer que, pour tous $x_1 \neq x_2$ et tout $\lambda \in\,]\,0\,;1\,[$,

$$f(\lambda\, x_1 + (1-\lambda)\, x_2) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

Pour les fonctions linéaires (ou affines), pour tous $x_1 \neq x_2$ et tout $\lambda \in\,]\,0\,;1\,[$, on a même

$$f(\lambda\, x_1 + (1-\lambda)\, x_2) = \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

Géométriquement, dans le cas réel $\mathcal{X} = \mathbb{R}$, cela signifie que le graphe (l'arc) se confond avec la corde quel que soit le couple d'abscisses considérées (x_1, x_2). Si l'on souhaite imposer un arc **strictement** en-dessous de la corde pour tout couple (x_1, x_2), on obtient une fonction **strictement** convexe :

Définition 2 (Fonction strictement convexe)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. On dit que f est *strictement convexe* si

$$\forall\, x_1 \neq x_2 \in \mathcal{X}, \forall \lambda \in\,]\,0\,;1\,[, \quad f(\lambda\, x_1 + (1-\lambda)\, x_2) < \lambda\, f(x_1) + (1-\lambda)\, f(x_2)$$

REMARQUE : D'après ce qui précède, il faut veiller à ce que l'inégalité stricte soit bien vérifiée en deux points distincts $x_1 \neq x_2$ et que $\lambda \notin \{0, 1\}$!

Proposition 5

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction strictement convexe. Alors f est convexe.

DÉMONSTRATION : Laissée en exercice.

EXERCICE

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction **strictement** convexe et $g : \mathbb{R} \to \mathbb{R}$ une fonction convexe **strictement croissante**. Montrer que $g \circ f$ est **strictement** convexe.

Dans le cas différentiable, la stricte convexité peut elle aussi être caractérisée à l'aide du gradient :

Proposition 6

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction *différentiable.* Alors on a équivalence entre les énoncés suivants :

(i) f est strictement convexe

(ii) $$\forall\, x_1 \neq x_2 \in \mathcal{X}, \qquad f(x_2) > f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle$$

(iii) ∇f est strictement monotone, c'est-à-dire

$$\forall\, x_1 \neq x_2 \in \mathcal{X}, \qquad \langle \nabla f(x_2) - \nabla f(x_1), x_2 - x_1 \rangle > 0$$

DÉMONSTRATION : Résultat partiellement admis : l'implication (i) $\Rightarrow$ (ii) est admise, car le passage à la limite dans la preuve du cas convexe ne permet pas de préserver l'inégalité large.

- (ii) $\Rightarrow$ (iii)**.** Il suffit de sommer l'inégalité (ii) écrite pour le couple (x_1, x_2) avec $x_1 \neq x_2$ avec l'inégalité (ii) écrite pour le couple (x_2, x_1).
- (iii) $\Rightarrow$ (ii)**.** On suit la même démonstration que pour le cas convexe, en remarquant que $x_1 \neq x_2$ et que $t_0 \neq 0$, de sorte que toutes les inégalités sont strictes.
- (ii) $\Rightarrow$ (i)**.** À nouveau, la même démonstration que dans le cas convexe suffit, en remplaçant chaque inégalité large en inégalité stricte.

REMARQUE : Si $\mathcal{X} = \mathbb{R}$, alors l'énoncé (iii) est équivalente à la *stricte croissance* de la dérivée de f. Dans les énoncés (ii) et (iii), il faut bien veiller à prendre $x_1 \neq x_2$!

2 Fonctions convexes sur un espace euclidien

Dans cette section, sauf mention contraire, on considère $E = \mathbb{R}^n$ avec $n \in \mathbb{N}^*$.

En dimension finie, les fonctions convexes présentent des propriétés très intéressantes : elles sont nécessairement continues ; la différentiabilité de FRÉCHET se confond avec celle de GATEAUX ; lorsqu'elles sont deux fois différentiables, la convexité peut-être caractérisée grâce à la matrice hessienne.

2.1 Continuité

Commençons par montrer que toute fonction convexe (définie sur tout l'espace ambiant) est continue :

Proposition 7

Soit $f : E \to \mathbb{R}$ une fonction convexe. Alors f est continue.

> Il est essentiel d'être en dimension finie pour démontrer ce résultat. Pour les fonctions convexes définies sur un ensemble convexe, on peut montrer que f est continue sur l'**intérieur** de son domaine de définition.

Pour démontrer ce résultat, on va établir successivement deux lemmes (qui permettent par ailleurs d'obtenir des résultats intermédiaires dans le cas de la dimension infinie, car on n'y suppose pas que $\mathcal{X}$ est euclidien). On va suivre la démarche suivante :

- on montre d'abord que si f est convexe et localement bornée, alors f est localement lipschitzienne (lemme 1) ;
- ensuite, on montre que si f est convexe, alors elle est nécessairement localement minorée, de sorte qu'il suffit qu'elle soit localement majorée pour être localement lipschitzienne (lemme 2) ;
- enfin, on montre que, si f est convexe sur un espace euclidien, alors elle est nécessairement localement majorée ; ainsi, elle est localement lipschitzienne et donc continue (proposition 7).

Lemme 1

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction convexe. Soit $a \in \mathcal{X}$. On suppose qu'il existe $\delta > 0$ et $M > 0$ tels que

$$\forall\, x \in \overline{\mathcal{B}}(a,\delta), \qquad |f(x)| \le M$$

Alors f est $(4\,M/\delta)$-lipschitzienne sur $\overline{\mathcal{B}}(a,\delta/2)$.

Démonstration : Soit $x, y \in \overline{\mathcal{B}}(a,\delta/2)$. On introduit le point suivant :

$$z = y - \frac{\delta}{2}\,\frac{x-y}{\|x-y\|}$$

On commence par noter que $z \in \overline{\mathcal{B}}(a,\delta)$; en effet, l'inégalité triangulaire assure que

$$\|z-a\| \le \|y-a\| + \frac{\delta}{2}\,\frac{\|x-y\|}{\|x-y\|} \le \|y-a\| + \frac{\delta}{2} \le \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

Par ailleurs, $y \in [\,x\,;z\,]$; plus précisément, on peut vérifier que

$$y = \frac{\delta}{2\,\|x-y\|+\delta}\,x + \left(1 - \frac{\delta}{2\,\|x-y\|+\delta}\right) z$$

Par convexité, on obtient donc que

$$f(y) \le \frac{\delta}{2\,\|x-y\|+\delta}\,f(x) + \left(1 - \frac{\delta}{2\,\|x-y\|+\delta}\right) f(z)$$

soit
$$f(y) - f(x) \le \left(1 - \frac{\delta}{2\,\|x-y\|+\delta}\right) (f(z) - f(x))$$

Puisque f est bornée sur $\overline{\mathcal{B}}(a,\delta)$, il s'ensuit que

$$f(y) - f(x) \le \left(1 - \frac{\delta}{2\,\|x-y\|+\delta}\right) 2\,M = \frac{4\,M}{2\,\|x-y\|+\delta}\,\|x-y\| \le \frac{4\,M}{\delta}\,\|x-y\|$$

En inversant le rôle de x et de y, on démontre une inégalité analogue qui permet de majorer $|f(x) - f(y)|$. ■

On supprime maintenant l'hypothèse de minoration locale :

Lemme 2

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction convexe. Soit $a \in \mathcal{X}$. On suppose qu'il existe $\delta > 0$ et $M > 0$ tels que

$$\forall\, x \in \overline{\mathcal{B}}(a, \delta), \qquad f(x) \leq M$$

Alors f est lipschitzienne sur $\overline{\mathcal{B}}(a, \delta/4)$.

Démonstration : D'après le lemme 1, il suffit de montrer que f est localement minorée. Pour cela, on remarque que, si $x \in \overline{\mathcal{B}}(a, \delta/2)$, alors $2\,a - x \in \overline{\mathcal{B}}(a, \delta/2)$. En effet,

$$\|2\,a - x - a\| = \|a - x\| \leq \frac{\delta}{2}$$

Par ailleurs, $a \in [\,x\,; 2\,a - x\,]$ car

$$a = \frac{1}{2}\,x + \frac{1}{2}\,(2\,a - x)$$

Il s'ensuit que, si $x \in \overline{\mathcal{B}}(a, \delta/2)$, alors la convexité de f assure que

$$f(a) \leq \frac{1}{2}\,f(x) + \frac{1}{2}\,f(2\,a - x) \qquad \text{soit} \qquad f(x) \geq 2\,f(a) - f(2\,a - x) \geq 2\,f(a) - M$$

On en déduit que f est minorée sur $\overline{\mathcal{B}}(a, \delta/2)$, donc bornée sur $\overline{\mathcal{B}}(a, \delta/2)$. Le lemme 1 assure donc que f est lipschitzienne sur $\overline{\mathcal{B}}(a, \delta/4)$. ■

On peut à présent démontrer la proposition 7 :

Démonstration : D'après ce qui précède, il suffit de montrer que toute fonction convexe est localement majorée. C'est pour ce résultat que la dimension finie est requise. Soit $a = (a_i)_{1 \leq i \leq n} \in E$. Soit $\delta > 0$ et

$$\mathcal{R} = \prod_{i=1}^{n} [\,a_i - \delta\,; a_i + \delta\,]$$

On note que $a \in \mathcal{R}$ et que $\mathcal{R}$ est un convexe. Posons

$$M = \max\left\{ f((a_i + \varepsilon_i \delta)_{1 \leq i \leq n}) \mid \forall\, i \in [\![\,1\,;\,n\,]\!]\,, \varepsilon_i \in \{1, -1\} \right\}$$

la plus grande valeur prise par f sur les sommets de $\mathcal{R}$. Soit $x \in \mathcal{R}$. Par définition de $\mathcal{R}$, il existe $\lambda_1, \ldots, \lambda_n \in [\,0\,; 1\,]$ tels que

$$x = \big(\lambda_i\,(a_i - \delta) + (1 - \lambda_i)\,(a_i + \delta)\big)_{1 \leq i \leq n}$$

On a donc

$$f(x) \leq \lambda_1 \underbrace{f(a_1 - \delta, x_2, \ldots, x_n)}_{\substack{\leq \lambda_2\, f(a_1 - \delta, a_2 - \delta, \ldots, x_n) \\ + (1 - \lambda_2)\, f(a_1 - \delta, a_2 + \delta, \ldots, x_n)}} + (1 - \lambda_1) \underbrace{f(a_1 + \delta, x_2, \ldots, x_n)}_{\substack{\leq \lambda_2\, f(a_1 + \delta, a_2 - \delta, \ldots, x_n) \\ + (1 - \lambda_2)\, f(a_1 + \delta, a_2 + \delta, \ldots, x_n)}}$$

En réitérant le processus sur chaque composante, on obtient

$$f(x) \leq \sum_{i=1}^{n^2} \Lambda_i\, f(a_1 \pm \delta, \ldots, a_n \pm \delta)$$

avec $\Lambda_i \geq 0$. Puisque chaque terme $f(a_1 \pm \delta, \ldots, a_n \pm \delta)$ est majoré par M, et que la somme est finie, on en déduit que f est majorée sur $\mathcal{R}$. Puisqu'il existe une boule fermée centrée en a et contenue dans $\mathcal{R}$, on en déduit que f y est majorée. Le lemme 2 assure donc que f y est lipschitzienne, donc continue. ■

2.2 Différentiabilité

On vient de voir dans le paragraphe précédent qu'une fonction convexe sur un espace euclidien est nécessairement continue. Qu'en est-il de la différentiabilité ? En général, une fonction convexe n'est pas différentiable ; il suffit de considérer la fonction valeur absolue, qui est convexe mais non différentiable (dérivable) en 0. En revanche, la différentiabilité de Gateaux coïncide avec celle de Fréchet :

Proposition 8

Soit $f : E \to \mathbb{R}$ une fonction convexe. Soit $a \in E$. Alors f est Fréchet-différentiable en a si et seulement si f est Gateaux-différentiable en a.

Démonstration : Il suffit de démontrer que Gateaux-différentiable implique Fréchet-différentiable puisque la réciproque est toujours vraie.

- Soit $\varepsilon > 0$. En dimension finie, la sphère unité $\mathcal{S}$ est compacte ; donc on peut considérer un recouvrement fini de boules de rayon ε :
$$\mathcal{S} \subset \bigcup_{i=1}^{r} \mathcal{B}(v_i, \varepsilon)$$
Autrement dit, pour tout $v \in \mathcal{S}$, il existe $i \in [\![1 ; r]\!]$ tel que $\|v - v_i\|_2 \leq \varepsilon$.
- Commençons par remarquer que, si f est différentiable, alors sa différentielle coïncide avec sa différentielle au sens de Gateaux. Soit $h \in E$. Ainsi, il existe $v \in \mathcal{S}$ tel que $h = \|h\|_2\, v$ et $i \in [\![1 ; r]\!]$ tel que $\|v - v_i\|_2 \leq \varepsilon$. Par ailleurs,
$$\begin{aligned} f(a+h) - f(a) - f'(a;h) = & \, f(a + \|h\|_2\, v_i) - f(a) - \|h\|_2\, f'(a; v_i) \\ & + f(a + \|h\|_2\, v) - f(a + \|h\|_2\, v_i) \\ & + \|h\|_2 \left(f'(a; v_i) - f'(a; v) \right) \end{aligned}$$
L'inégalité triangulaire permet d'obtenir la majoration suivante :
$$\begin{aligned} |f(a+h) - f(a) - f'(a;h)| \leq & \, |f(a + \|h\|_2\, v_i) - f(a) - \|h\|_2\, f'(a; v_i)| \\ & + |f(a + \|h\|_2\, v) - f(a + \|h\|_2\, v_i)| \\ & + \|h\|_2\, |f'(a; v_i) - f'(a; v)| \end{aligned}$$
- Puisque f est convexe, on a montré dans la démonstration de la proposition 7 que f est localement lipschitzienne ; autrement dit, il existe $M > 0$ tel que
$$\begin{aligned} |f(a + \|h\|_2\, v) - f(a + \|h\|_2\, v_i)| & \leq M \, \|(a + \|h\|_2\, v) - (a + \|h\|_2\, v_i)\|_2 \\ & \leq M \, \|h\|_2 \, \|v - v_i\|_2 \\ & \leq M \, \|h\|_2 \, \varepsilon \end{aligned}$$
- Puisque f est Gateaux-différentiable en a, la différentielle de Gateaux de f en a
$$d_G f(a) : v \mapsto f'(a; v)$$
est linéaire et continue (par définition) ; elle est en particulier lipschitzienne, et il existe $M' > 0$ tel que
$$|f'(a; v_i) - f'(a; v)| \leq M' \, \|v_i - v\|_2 \leq M' \, \varepsilon$$
- Puisque f est Gateaux-différentiable en a, pour tout $i \in [\![1 ; r]\!]$, il existe $\delta_i > 0$ tel que pour tout $h \in \mathcal{B}(0, \delta_i)$
$$\left| \frac{f(a + \|h\|_2\, v_i) - f(a) - \|h\|_2\, f'(a; v_i)}{\|h\|_2} \right| \leq \varepsilon$$

Soit $\delta = \min\{\delta_i\}_{1\leq i\leq r}$. On a donc pour tout $i \in [\![1 ; r]\!]$

$$\forall\, h \in \mathcal{B}(0,\delta), \qquad \big|f(a+\|h\|_2\, v_i) - f(a) - \|h\|_2\, f'(a; v_i)\big| \leq \varepsilon\, \|h\|_2$$

- Finalement, on a démontré que

$$|f(a+h) - f(a) - f'(a;h)| \leq (1+M+M')\, \|h\|_2\, \varepsilon$$

Autrement dit, f est différentiable en a. ■

On termine en signalant que, dans le cas réel ($E = \mathbb{R}$), même lorsqu'une fonction convexe n'est pas dérivable, elle admet tout de même des dérivées à gauche et à droite. Pour cela, on commence par énoncer le lemme des trois cordes :

Lemme 3 (des trois cordes)

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Soit $(a,b,c) \in \mathbb{R}^3$ tels que $a < b < c$. Alors on a la double inégalité suivante :

$$\frac{f(b)-f(a)}{b-a} \leq \frac{f(c)-f(a)}{c-a} \leq \frac{f(c)-f(b)}{c-b}$$

Démonstration :

Soit $x \in \mathbb{R}$. On pose $\qquad \forall\, y \neq x, \qquad g_x(y) = \dfrac{f(y)-f(x)}{y-x}$

- **Croissance de g_x sur $]\,x\,;+\infty\,[$.** Soit $y \geq y' > x$. Puisque y' est un point du segment $[\,x\,;y\,]$, il peut s'écrire comme combinaison convexe de x et y. Plus précisément, si on pose

$$y' = \lambda\, x + (1-\lambda)\, y$$

alors on obtient $\qquad \lambda = \dfrac{y-y'}{y-x}$

Puisque λ est le rapport entre deux nombres strictement positifs différents et que le numérateur est strictement plus petit que le dénominateur, il s'ensuit que $\lambda \in \,]\,0\,;1\,[$. la définition de la convexité de f assure alors que

$$f(y') \leq \lambda\, f(x) + (1-\lambda)\, f(y) \qquad \text{soit} \qquad f(y') - f(x) \leq (1-\lambda)\,(f(y) - f(x))$$

Or, puisque $\qquad 1-\lambda = \dfrac{y-x-(y-y')}{y-x} = \dfrac{y'-x}{y-x}$

on obtient finalement que $g_x(y') \leq g_x(y)$.

- **Croissance de g_x sur $]\,-\infty\,;x\,[$.** Il suffit de réécrire les mêmes calculs que dans le point précédent, en considérant cette fois $x < y' \leq y$.
- On applique maintenant les deux résultats précédents aux points $a < b < c$. On a d'une part

$$g_a(b) \leq g_a(c) \qquad \text{soit} \qquad \frac{f(b)-f(a)}{b-a} \leq \frac{f(c)-f(a)}{c-a}$$

et d'autre part

$$g_c(a) \leq g_c(b) \qquad \text{soit} \qquad \frac{f(a)-f(c)}{a-c} \leq \frac{f(b)-f(c)}{b-c} = \frac{f(c)-f(b)}{c-b} \;■$$

Proposition 9 (Dérivabilité à droite et à gauche)

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction convexe et $x \in \mathbb{R}$. Alors

- f admet une *dérivée à droite* en x, c'est-à-dire que la limite suivante existe et est finie
$$f'_+(x) = \lim_{\substack{t\to 0\\ t>0}} \frac{f(x+t)-f(x)}{t}$$
- f admet une *dérivée à gauche* en x, c'est-à-dire que la limite suivante existe et est finie
$$f'_-(x) = \lim_{\substack{t\to 0\\ t>0}} \frac{f(x)-f(x-t)}{t}$$
- et on a de plus $\qquad f'_-(x) \leq f'_+(x)$

Démonstration :

- Soit $t > 0$. En appliquant le lemme des trois pentes aux points $a = x - t$, $b = x$ et $c = x + t$, on obtient en particulier que
$$\frac{f(x)-f(x-t)}{t} \leq \frac{f(x+t)-f(x)}{t}$$
Ainsi, si les limites lorsque t tend vers 0 existent, alors on a bien l'inégalité annoncée dans le troisième point.
- *On reprend les notations de la preuve du lemme des trois pentes.* On rappelle que g_x est une fonction croissante sur $]\,x\,;+\infty\,[$. Il s'ensuit que la fonction
$$g : \left\{ \begin{array}{ccc}]\,0\,;+\infty\,[& \to & \mathbb{R} \\ t & \mapsto & g_x(x+t) \end{array} \right.$$
est croissante. Montrons que g est minorée. Le lemme des trois pentes appliqué à $(a,b,c) = (x-1, x, x+t)$ assure en effet que
$$\forall\, t > 0, \qquad f(x) - f(x-1) \leq g(t)$$
Ainsi, g admet une limite à droite en 0, ce qui équivaut à la dérivabilité à droite de f en 0. Un raisonnement analogue démontre la dérivabilité à gauche de f en 0. ■

2.3 Différentiabilité d'ordre deux

Si $\mathcal{X} = \mathbb{R}$ et f est dérivable, on a vu plus haut que f est convexe si et seulement f' est croissante. Il en découle que, si f est deux fois dérivable (et donc que f' est dérivable), que f est convexe si et seulement $f''(x)$ est positif pour tout $x \in \mathcal{X}$. La généralisation de cette remarque pour le cas non réel est donnée par la proposition suivante :

Proposition 10 (Caractérisation des fonctions convexes deux fois différentiables)

Soit $f : E \to \mathbb{R}$ une fonction *deux fois différentiable*. Alors on a équivalence entre les énoncés suivants :

(i) f est convexe ;

(ii) $\mathrm{Hess} f(x)$ est semi-définie positive en tout point $x \in E$.

> Rappel : on dit qu'une matrice carrée A de taille n est *semi-définie positive* si, pour tout $x \in \mathbb{R}^n$, on a
> $$\langle A\,x, x\rangle \geq 0$$
> On peut montrer que A est semi-définie positive si et seulement si ses valeurs propres sont positives ou nulles. Dans le cas du plan ($n = 2$), puisque la trace (resp. le déterminant) de A donne la somme (resp. le produit) de ses valeurs propres, on en déduit que A est semi-définie positive si et seulement si
> $$\mathrm{tr}\,(A) \geq 0 \qquad \text{et} \qquad \det(A) \geq 0$$
> **Attention :** cette dernière caractérisation n'est valable qu'en dimension 2 !

> Le caractère semi-défini positif d'une matrice est **une** façon de généraliser la notion de positivité pour une matrice. C'est la raison pour laquelle certains auteurs parlent plus simplement de *matrices positives*. Il ne s'agit pas de la seule généralisation possible : une manière plus simple (et plus naturelle) de définir la "positivité" d'une matrice est de considérer les matrices dont les coefficients sont positifs. 'A cause de cette ambigüité (et parce que l'expression *matrices positives* risque plus sûrement d'évoquer une matrice à coefficients positifs qu'une matrice dont les valeurs propres sont positives), nous choisissons dans ce cours d'utiliser l'expression *matrice semi-définie positive*.

Démonstration : Grâce à la proposition 4, il suffit de démontrer que (ii) équivaut à la monotonie du gradient, c'est-à-dire à

$$\forall\,(x_1, x_2) \in E^2, \quad \langle \nabla f(x_2) - \nabla f(x_1), x_2 - x_1\rangle \geq 0$$

Démontrons séparément les deux sens de équivalence.

- **Sens direct.** Soit $x \in E$. On suppose que le gradient est monotone. Soit $h \in E$. Par définition de la matrice hessienne, on a pour tout $t \neq 0$ voisin de 0
$$\mathrm{Hess} f(x)\,h = \frac{\nabla f(x + t\,h) - \nabla f(x) + t\,\|h\|_2\,\varepsilon(\|t\,h\|_2)}{t}$$
avec $\varepsilon(t)$ tendant vers 0 quand t tend vers 0. On en déduit donc que
$$\langle \mathrm{Hess} f(x)\,h, h\rangle = \frac{\langle \nabla f(x + t\,h) - \nabla f(x), h\rangle}{t} + \langle \|h\|_2\,\varepsilon(\|t\,h\|_2), h\rangle$$
On fait tendre t vers 0 ; le second terme s'annule, tandis que le premier est positif grâce à la monotonie du gradient.
- **Réciproque.** On suppose que la matrice hessienne est semi-définie positive. Soit $(x_1, x_2) \in E^2$. Posons
$$g : \left\{ \begin{array}{rcl} [\,0\,;1\,] & \to & \mathbb{R} \\ t & \mapsto & \langle \nabla f(x_1 + t\,(x_2 - x_1)), x_2 - x_1\rangle \end{array} \right.$$
Cette fonction est dérivable car ∇f est différentiable sur $[\,x_1\,; x_2\,]$, sa dérivée étant donnée par
$$\forall\,t \in \,]\,0\,;1\,[\,, \qquad g'(t) = \langle \mathrm{Hess} f(x_1 + t\,(x_2 - x_1))\,(x_2 - x_1), x_2 - x_1\rangle$$
Notons d'ores-et-déjà que, par hypothèse sur la matrice hessienne, $g'(t) \geq 0$ pour tout $t \in \,]\,0\,;1\,[$. On applique finalement le théorème des accroissements finis, qui implique l'existence d'un $t_0 \in \,]\,0\,;1\,[$ tel que
$$g(1) - g(0) = g'(t_0)$$

Autrement dit,

$$\langle \nabla f(x_2), x_2 - x_1 \rangle - \langle \nabla f(x_1), x_2 - x_1 \rangle \geq 0$$

ce qui est bien la monotonie du gradient. ■

Exemple

Norme au carré. On considère la fonction suivante :

$$f : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & \|x\|_2^2 \end{array} \right.$$

La fonction f est deux fois différentiable en tant que fonction polynomiale, de gradient et de matrice hessienne respectifs

$$\forall\, x \in \mathbb{R}^n, \qquad \nabla f(x) = 2\,x \quad \text{et} \quad \mathrm{Hess} f(x) = 2\,\mathrm{I}_n$$

La matrice hessienne est bien semi-définie positive, puisque

$$\forall\, x \in \mathbb{R}^n, \qquad \langle 2\,\mathrm{I}_n\, x, x \rangle = 2\,\langle x, x \rangle = 2\,\|x\|_2^2 \geq 0$$

donc la fonction f est convexe.

Pour les fonctions strictement convexes, de même que la stricte croissance d'une fonction réelle n'implique pas nécessairement la stricte positivité de sa dérivée (qui peut s'annuler ponctuellement), on n'a que la condition suffisante suivante :

Proposition 11

Soit $f : E \to \mathbb{R}$ une fonction *deux fois différentiable*. On suppose que $\mathrm{Hess} f(x)$ est définie positive en tout point $x \in E$. Alors f est strictement convexe.

> Rappel : on dit qu'une matrice carrée A de taille n est *définie positive* si, pour tout $x \in \mathbb{R}^n \setminus \{0\}$, on a
>
> $$\langle A\,x, x \rangle > 0$$
>
> Une matrice est définie positive si et seulement si ses valeurs propres sont **strictement** positives ce qui, dans le cas $n = 2$, équivaut à la stricte positivité de sa trace et de son déterminant.

Démonstration : Laissée en exercice.

Pour résumer, si $f : E \to \mathbb{R}$ est une fonction deux fois différentiable alors :

- si $\mathrm{Hess}\, f(x)$ est définie positive pour tout $x \in E$, alors f est strictement convexe ;
- s'il existe $x^0 \in E$ tel que $\mathrm{Hess}\, f(x^0)$ **n'est pas** semi-définie positive, alors f **n'est pas** convexe ;
- si $\mathrm{Hess}\, f(x)$ est semi-définie positive pour tout $x \in E$ mais qu'il existe $x^0 \in E$ tel que $\mathrm{Hess}\, f(x^0)$ **n'est pas** définie positive, alors f est convexe, mais on ne peut pas conclure directement quant à sa stricte convexité (il faut revenir à la définition de la stricte convexité ou démontrer la stricte monotonie du gradient par exemple).

3 Fonctions fortement convexes

3.1 Définition

On termine cette section en considérant une sous-classe de fonctions strictement convexes :

Définition 3 (Fonction fortement convexe)

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. Soit $\alpha > 0$. On dit que f est *fortement convexe* de module α si

$$\forall (x_1, x_2) \in \mathcal{X}^2, \forall \lambda \in [\,0\,;1\,],$$
$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2) - \frac{\alpha}{2}\, \lambda\, (1-\lambda)\, \|x_1 - x_2\|^2$$

> On dit parfois que f est α-convexe, ou encore que f est α-*elliptique*.

Voici un exemple central de fonctions fortement convexes :

EXEMPLE

Distance au carré. Soit $\mathcal{X}$ un espace de HILBERT. Soient $x^0 \in \mathcal{X}$ et $\alpha > 0$. Considérons la fonction suivante :

$$f : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & \dfrac{\alpha}{2}\, \|x - x^0\|^2 \end{array} \right.$$

Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. En remarquant que $x^0 = \lambda\, x^0 + (1-\lambda)\, x^0$ et en développant le premier carré, on obtient que

$$\frac{\alpha}{2}\, \big\|\lambda\, x_1 + (1-\lambda)\, x_2 - x^0\big\|^2 - \frac{\alpha}{2}\, \lambda\, \|x_1 - x^0\|^2 - \frac{\alpha}{2}\, (1-\lambda)\, \|x_2 - x^0\|^2$$
$$= -\frac{\alpha}{2}\, \lambda\, (1-\lambda)\big(\|x_1 - x^0\|^2 + \|x_2 - x^0\|^2 - 2\, \langle x_1 - x^0, x_2 - x^0\rangle\big)$$

En utilisant une identité remarquable pour simplifier le terme entre parenthèses dans l'expression précédente, on prouve que f est une fonction fortement convexe de module α.

3.2 Caractérisation de la forte convexité

Comme on peut le voir dans l'exemple précédent, démontrer la forte convexité d'une fonction en utilisant la définition peut être relativement ardu. Aussi, on va établir une première manière simple de caractériser la forte convexité. Commençons par faire l'observation suivante :

Proposition 12

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction fortement convexe de module α et $g : \mathcal{X} \to \mathbb{R}$ une fonction convexe. Alors $f + g$ est fortement convexe, de module α.

DÉMONSTRATION : Soit $(x_1, x_2) \in \mathcal{X}^2$ et $\lambda \in [\,0\,;1\,]$. Puisque f est fortement convexe, on a

$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2) - \frac{\alpha}{2}\, \lambda\,(1-\lambda)\, \|x_1 - x_2\|^2$$

La convexité de g permet quant à elle d'écrire

$$g\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, g(x_1) + (1-\lambda)\, g(x_2)$$

Il suffit alors de sommer les deux inégalités pour conclure. ■

La réciproque est vraie, comme le montre la proposition suivante :

Proposition 13

Soit $\mathcal{X}$ un espace de HILBERT. Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. Alors on a l'équivalence entre les énoncés suivants :

(i) la fonction f est fortement convexe de module α ;

(ii) pour tout $x^0 \in \mathcal{X}$, la fonction

$$g : \left\{ \begin{array}{ccc} \mathcal{X} & \to & \mathbb{R} \\ x & \mapsto & f(x) - \dfrac{\alpha}{2}\, \|x - x^0\|^2 \end{array} \right.$$

est une fonction convexe.

On peut réécrire l'affirmation (ii) de la proposition 13 de la manière suivante :

$$\forall\, x \in \mathcal{X}, \qquad f(x) = g(x) + \frac{\alpha}{2}\, \|x - x^0\|^2$$

avec $g : \mathcal{X} \to \mathbb{R}$ une fonction convexe. Autrement dit, toute fonction fortement convexe est somme d'une fonction convexe et d'une fonction fortement convexe de même module.

DÉMONSTRATION : Puisque la réciproque a déjà été démontrée, on se contente du sens direct. Soit $x^0 \in \mathcal{X}$. Le calcul effectué dans l'exemple précédent prouve que, pour tout $(x_1, x_2) \in \mathcal{X}^2$ et tout $\lambda \in [\,0\,;1\,]$, on a

$$\big\|\lambda\, x_1 + (1-\lambda)\, x_2 - x^0\big\|^2 - \lambda\, \|x_1 - x^0\|^2 - (1-\lambda)\, \|x_2 - x^0\|^2 = -\lambda\,(1-\lambda)\, \|x_1 - x_2\|^2$$

Par définition, f est fortement convexe de module α si et seulement si

$$\begin{aligned} &\forall\, (x_1, x_2) \in \mathcal{X}^2, \forall \lambda \in [\,0\,;1\,]\,, \\ &\qquad f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2) - \frac{\alpha}{2}\, \lambda\,(1-\lambda)\, \|x_1 - x_2\|^2 \end{aligned}$$

soit, en y ajoutant l'égalité ci-dessus multipliée par $\alpha/2$,

$$\begin{aligned} f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) - \frac{\alpha}{2}\, \big\|\lambda\, x_1 + (1-\lambda)\, x_2 - x^0\big\|^2 \\ \leq \lambda\, f(x_1) - \frac{\alpha}{2}\, \lambda\, \|x_1 - x^0\|^2 + (1-\lambda)\, f(x_2) - \frac{\alpha}{2}\, (1-\lambda)\, \|x_2 - x^0\|^2 \end{aligned}$$

qui n'est autre que la définition de la convexité appliquée à la fonction g. ■

3.3 Cas différentiable

La proposition 13 permet de caractériser les fonctions fortement convexes différentiables à l'aide de leur différentielle ; si $\mathcal{X}$ est un espace de HILBERT, alors on peut écrire :

Corollaire 1

Soit $\mathcal{X}$ un espace de HILBERT. Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction *différentiable*. Alors on a équivalence entre les énoncés suivants :

(i) f est fortement convexe de module α ;

(ii) $\forall\,(x_1, x_2) \in \mathcal{X}^2, \qquad f(x_2) \geq f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle + \dfrac{\alpha}{2}\,\|x_1 - x_2\|^2$

(iii) $\forall\,(x_1, x_2) \in \mathcal{X}^2, \qquad \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \geq \alpha\,\|x_1 - x_2\|^2$

DÉMONSTRATION : Il suffit d'appliquer la proposition 4 à la fonction

$$x \mapsto f(x) - \frac{\alpha}{2}\,\|x - x_1\|^2$$

qui est une fonction différentiable par somme. ■

De même, si $\mathcal{X} = E = \mathbb{R}^n$ est un espace euclidien, alors on peut caractériser la forte convexité à l'aide de la matrice hessienne :

Corollaire 2

Soit $f : E \to \mathbb{R}$ une fonction *deux fois différentiable*. Alors on a équivalence entre les énoncés suivants :

(i) f est fortement convexe de module α ;

(ii) pour tout point $x \in E$, les valeurs propres de la matrice $\operatorname{Hess} f(x)$ sont supérieures ou égales à α.

DÉMONSTRATION : On applique cette fois la proposition 10 à la fonction

$$g : x \mapsto f(x) - \frac{\alpha}{2}\,\|x - x_1\|_2^2$$

qui est deux fois différentiable par somme, de matrice hessienne $\operatorname{Hess} f(x) - \alpha\,\mathrm{I}_n$, qui est semi-définie positive car g est convexe. Or, puisque

$$P^{-1}\,\operatorname{Hess} f(x)\,P - \alpha\,\mathrm{I}_n = P^{-1}\,\operatorname{Hess} f(x)\,P - \alpha\,P^{-1}\,\mathrm{I}_n\,P = P^{-1}(\operatorname{Hess} f(x) - \alpha\,\mathrm{I}_n)\,P$$

il s'ensuit que les valeurs propres de $\operatorname{Hess} f(x) - \alpha\,\mathrm{I}_n$, qui sont positives ou nulles, sont exactement celles de $\operatorname{Hess} f(x)$ auxquelles on retranche α. ■