Chapitre 3

Rappel sur le calcul différentiel

3.0.1 Quelques Notations

1 . Pour tout $n \in \mathbb{N}^*$, $\mathbb{R}^n$ désinge l'espace **euclidien** $\mathbb{R} \times \mathbb{R} \times ......\mathbb{R}$ ("produit n fois") . En général un vecteur $x \in \mathbb{R}^n$ sera noté $x = (x_1, x_2, ...x_n)^T$ (vecteur colonne).

2 . On note $e_1, e_2, ...e_n$ les éléments de la **base canonique** de $\mathbb{R}^n$, ou e_i est le vecteur de $\mathbb{R}^n$ donnè par :

$$(e_i)_j = \delta_{ij} = \begin{cases} 0 & \text{si } j \neq i \\ 1 & \text{si } j = i \end{cases} \quad , \quad \forall i, j = 1, 2....n \tag{1.1}$$

(δ_{ij} = symboles de **Kronecker**) .

3 . Pour tous $x, y \in \mathbb{R}^n$ on note par $\langle x, y \rangle \in \mathbb{R}$ le **produit scalaire** de x et y, qui est donné par

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

Deux vecteurs $x, y \in \mathbb{R}^n$ sont **orthogonaux** (on notera $x \perp y$) si $\langle x, y \rangle = 0$.

4 . Pour tout $x \in \mathbb{R}^n$ on note par $\|x\| \geq 0$ la **norme euclidienne** de x, donné par

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Rappellons les **propriétés d'une norme (**donc aussi de la norme euclidienne) :

i) $\|\lambda x\| = |\lambda| \|x\| \quad \forall\lambda \in \mathbb{R}, \forall x \in \mathbb{R}^n$

ii) $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in \mathbb{R}^n$

iii) $\|0\| = 0$ et $\|x\| > 0$ si $x \in \mathbb{R}^n - \{0\}$.

5 . Pour tous $x \in \mathbb{R}^n$ et $r > 0$ on notera par $B(x, r)$ la **boule ouverte** du centre x et rayon r, donnée par

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

6 . Si $\{x^{(k)}\}_{k\in\mathbb{N}}$ est une suite dans $\mathbb{R}^n$ et x est un élément de $\mathbb{R}^n$ on dit que $x^{(k)}$ **converge** vers x (notée $x^{(k)} \to x$) si $\left\|x^{(k)} - x\right\| \to 0$.

Rappellons que nous avons : $x^{(k)} \to x$ si et seulement si $x_i^{(k)} \to x_i$ en $\mathbb{R}$ ou $x_i^{(k)}$(respectivement x_i) est la i-éme composante de $x^{(k)}$ (respectivement x).

7 . Soit $U \subset \mathbb{R}^n$.

- On définit **l'intérieur** de U comme l'ensemble des éléments $x \in U$ pour lesquels il existe $r > 0$ tel que $B(x, r) \subset U$.

- On dit que U est **ouvert** si $\forall x \in U \ \exists r > 0$ tel que $B(x, r) \subset U$.

-On dit que U est **fermé** si pour tout suite $\{x^{(k)}\} \subset U$ tel que $x^{(k)} \to x \in \mathbb{R}^n$ on a $x \in U$.

8 .Si $a, b \in \mathbb{R}^n$ on note $[a, b]$ le sous-ensemble de $\mathbb{R}^n$ donné par

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

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

Remarques :

$\cdot$ $[a, b] = [b, a]$.

$\cdot$Si $a, b \in \mathbb{R}$ avec $a < b$ on retrouve la notation $[a, b]$ pour l'intervalle des nombres $x \in \mathbb{R}$ tels que $a \leq x \leq b$.

9 . Rappellons aussi l'inégalité de Cauchy-Schwarz :

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \quad \forall x, y \in \mathbb{R}^n.$$

3.0.2 Quelques rappels sur le calcul diffèrentiel

On considére dans cette partie m et n deux nombres de $\mathbb{N}^*$ (trés souvent, on aura $m = 1$).

1 . Soit U un sous-ensemble de $\mathbb{R}^n$ et $f : U \to \mathbb{R}^m$.

On dit que f est **continue** en $x \in U$ si $f\left(x^{(k)}\right) \to f(x)$ pour toute suite $x^{(k)} \subset U$ telle que $x^{(k)} \to x$.

On dit que f est continue sur U si f est continue en tout point $x \in U$.

Remarque : Si $f = (f_1, f_2, ...f_m)$ avec $f_1, f_2, ...f_m : U \to \mathbb{R}$ alors f est continue en $x \in U$ si et seulement si $f_1, f_2, ...f_m$ sont continues en x.

Pour tous les poins suivants on va supposer que Ω est un ouvert de $\mathbb{R}^n$ et f est une fonction définie de $\Omega \to \mathbb{R}^m$.

2 . Pour tout $x \in \Omega$ et $h \in \mathbb{R}^n$ on note (quand $\exists$)

$$\frac{\partial f}{\partial h}(x) = \lim_{t \to 0} \frac{1}{t}\left[f(x + th) - f(x)\right],$$

(c'est la **dérivée directionnelle** de f en x dans la direction h).

Remarques :

$i)$ $\frac{\partial f}{\partial 0}(x) = 0$.

$ii)$ Si $f = (f_1, f_2, ...f_n)^T \in \mathbb{R}^n$ avec $f_1, f_2, ...f_m : \Omega \to \mathbb{R}$ alors

$$\frac{\partial f}{\partial h}(x) = \left(\frac{\partial f_1}{\partial h}(x), \frac{\partial f_2}{\partial h}(x), ...\frac{\partial f_m}{\partial h}(x)\right)^T$$

3 . Pour tout $x \in \Omega$ et tout $i \in \{1, 2, ..., n\}$ on note (quand $\exists$)

$$\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x) = \lim_{t \to 0} \frac{1}{t}\left[f(x + te_i) - f(x)\right]$$

(c'est la **dérivée partielle** de f en x par rapport à la variable x_i)

En particulier, si $n = 1$ on note $f^{'}(x) = \frac{\partial f}{\partial x_1}(x) = \lim_{t \to 0} \frac{1}{t}\left[f(x+t) - f(x)\right] = \lim_{y \to x} \frac{1}{y-x}\left[f(y) - f(x)\right].$

4 . Pour tout $x \in \Omega$ on note (quand $\exists$) $J_f(x)$ =la **matrice Jacobienne** de f en x qui est un élément de $M_{m,n}(\mathbb{R})$ définie par

$$(J_f(x))_{ij} = \frac{\partial f_i}{\partial x_j}(x) \in \mathbb{R} \quad \forall i = 1, ...m, \forall j = 1, ...n.$$

Le **gradient** de f en x est défini comme la transposée de la matrice Jacobienne de f en x :

$$\nabla f(x) = (J_f(x))^T \in M_{m,n}(\mathbb{R}).$$

Remarque importante : Dans le cas particulier $m = 1$(donc $f : \Omega \to \mathbb{R}$) alors en considérant tout élément de $M_{n,1}$comme un vecteur colonne de $\mathbb{R}^n$, on va dire que $\nabla f(x)$ est le vecteur colonne

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}\frac{\partial f}{\partial x_2}, ...\frac{\partial f}{\partial x_n}\right)^T \in \mathbb{R}^n.$$

Rappellons la formule :

$$\frac{\partial f}{\partial h}(x) = \langle \nabla f(x), h \rangle \quad \forall x \in \Omega \ \forall h \in \mathbb{R}^n.$$

5 . Si $f : \Omega \to \mathbb{R}$ (ici $m = 1$) on dit qu'un point $x \in \Omega$ un **point critique** pour la fonction f si $\nabla f(x) = 0$.

6 . Pour tout $x \in \Omega$ et $i, j \in \{1, 2, ...n\}$ on note (quand $\exists$)

$$\frac{\partial^2 f}{\partial x_i \partial x_J}(x) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x) \in \mathbb{R}^m,$$

la dérivée partielle à l'ordre 2.

Notation : pour $i = j$ on écrira $\frac{\partial^2 f}{\partial^2 x_i}(x)$ à la place de $\frac{\partial^2 f}{\partial x_i \partial x_i}(x)$.

7 . Dans le cas $m = 1$ on note pour tout $x \in \Omega$ (quand $\exists$) $\nabla^2 f(x) =$ la matrice carrée $\in M_n(\mathbb{R})$ donnée par

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

($\nabla^2 f(x)$ s'appelle aussi **la matrice Hessienne** de f en x).

8 . On dit que f est de classe C^p sur Ω (on notera $f \in C^p(\Omega)$) pour $p = 1$ ou $p = 2$ si les dérivées partielles des f jusqu'à l'ordre p existent et sont continues sur Ω. Par extension on dit que f est de classe C^0 sur Ω si f est continue sur Ω.

9 . On a le Théoréme de Schwarz : si $f \in C^2(\Omega)$ alors

$$\frac{\partial^2 f}{\partial x_i \partial xj}(x) = \frac{\partial^2 f}{\partial xj \partial xi}(x) \quad \forall x \in \Omega \;\; \forall i, j = 1, 2, ...n.$$

(c'est à dire, la matrice $\nabla^2 f(x)$ est symmétrique.

10 . (Lien entre ∇, J_f et ∇^2) : Si $f : \Omega \to \mathbb{R}$ est de classe C^2 alors

$$\nabla^2 f(x) = J_{\nabla f}(x) = \nabla J_f(x) \;\; \forall x \in \Omega$$

(la matrice Hessienne de f est le Jacobien du gradient de f ou le gradient de la Jacobienne de f).

11 . (Composition) Soient $\Omega \subset \mathbb{R}^n, U \subset \mathbb{R}^m$ avec Ω, U ouverts $f : \Omega \to \mathbb{R}^m, g : U \to \mathbb{R}^p$ avec $p \in \mathbb{N}^*$ et $f(\Omega) \subset U$. Considérons la fonction composée $g \circ f : \Omega \to \mathbb{R}^p$

$i)$ Si f et g sont continues alors $g \circ f$ est continue.

$i)$ Si f et g sont de classe C^1 alors $g \circ f$ est de classe C^1 et on a l'égalité matricielle

$$J_{g \circ f}(x) = J_g(f(x)) \cdot J_f(x) \quad \forall x \in \Omega.$$

Conséquences :

$i)$ Si $m = p = 1$ alors

$$\nabla (g \circ f)(x) = g^{'}(f(x)) \nabla f(x).$$

$i)$ Si $n = p = 1$ alors

$$(g \circ f)^{'}(x) = \Big\langle \nabla g(f(x)) f^{'}(x) \Big\rangle.$$

Proposition 1.1. Nous avons

$$\nabla^2 f(x) h = \nabla \langle \nabla f(x), h \rangle \;\; \forall x \in \Omega, \forall h \in \mathbb{R}^n$$

ou le premier gradient dans le membre de droite de l'égalité est considéré par rapport à la variable x.

Démonstration. On a :

$$\frac{\partial}{\partial x_i} \langle \nabla f(x), h \rangle = \frac{\partial}{\partial x_i} \left(\sum_{j=1}^{n} \frac{\partial f}{\partial xj}(x) h_j \right) = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i xj}(x) h_j = \left(\nabla^2 f(x) h \right)_i.$$

3.0.3 Quelques exemples importants :

1 . Si $f : \mathbb{R}^n \to \mathbb{R}^m$ est une fonction **constante** alors $\nabla f = 0$ et $J_f = 0$. On a aussi évidement $\nabla^2 f = 0$ dans le cas $m = 1$.

2 .Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ définie par

$$f(x) = Ax \quad \forall x \in \mathbb{R}^n,$$

ou $A \in M_{m,n}(\mathbb{R})$ est une matrice donné (c'est à dire, f est une fonction **linéaire).** Il est facile de voir qu'on a

$$J_f(x) = A \;\; \forall x \in \mathbb{R}^n,$$

(la matrice Jacobienne est constante).

Dans le cas particulier $m = 1$ une fonction linéaire générale peut être écrite sous la forme

$$f(x) = \langle a, x\rangle \quad \forall x \in \mathbb{R}^n,$$

ou $a \in \mathbb{R}^n$ est un vecteur donné. Il est clair alors que

$$\nabla f = a,$$

et

$$\nabla^2 f = 0$$

3 . Soit $f : \mathbb{R}^n \to \mathbb{R}$ donnée par

$$f(x) = \langle Ax, x\rangle \quad \forall x \in \mathbb{R}^n,$$

où $A \in M_n(\mathbb{R})$ est un matrice carrée, réelle, de taille n (c'est à dire, f est la **forme quadratique** associée à la matrice A). Alors pour un $p \in \{1, 2, ...; n\}$ fixè, on peut écrire

$$f(x) = \sum_{i,j=1}^{n} A_{ji}x_i x_j = A_{pp}x_p^2 + \sum_{j=1, j\neq p}^{n} A_{pj}x_p x_j + \sum_{i=1, i\neq p}^{n} A_{ip}x_i x_p + \sum_{i,j=1, i\neq p, j\neq p}^{n} A_{ij}x_i x_j$$

ce qui nous donne

$$\frac{\partial f}{\partial x_p} = 2A_{pp}x_p + \sum_{j=1, j\neq p}^{n} A_{pj}x_j + \sum_{i=1, i\neq p}^{n} A_{ip}x_i = \sum_{j=1}^{n} A_{pj}x_j + \sum_{i=1}^{n} A_{ip}x_i = (Ax)_p + \left(A^T x\right)_p.$$

Nous avons donc obtenu :

$$\nabla f(x) = \left(A + A^T\right) x, \quad \forall x \in \mathbb{R}^n.$$

En utilisant la formule $\nabla^2 f = J_{\nabla f}$ on déduit

$$\nabla^2 f(x) = A + A^T, \quad \forall x \in \mathbb{R}^n,$$

(donc la hessienne de f est constante).

Remarque : En particulier, si A est **symmétrique (** c'est à dire $A = A^T$) alors

$$\nabla \langle Ax, x \rangle = 2Ax, \quad \forall x \in \mathbb{R}^n.$$

$$\nabla^2 \langle Ax, x \rangle = 2A, \quad \forall x \in \mathbb{R}^n.$$

3.0.4 Rappel formules de Taylor

Proposition 1.2

Soit $\Omega \subset \mathbb{R}^n$ ouvert, $f : \Omega \to \mathbb{R}$, $a \in \Omega$ et $h \in \mathbb{R}^n$ tels que $[a, a+h] \subset \Omega$. Alors :

1 . Si $f \in C^1(\Omega)$ alors

$i)$ $f(a+h) = f(a) + \int_0^1 \langle \nabla f(a+th), h \rangle \, dt$

(formule de Taylor à l'ordre 1 avec reste intégral).

$ii)$ $f(a+h) = f(a) + \langle \nabla f(a+\theta h), h \rangle$ avec $0 < \theta < 1$

(formule de Taylor- Maclaurin à l'ordre 1)

$iii)$ $f(a+h) = f(a) + \langle \nabla f(a), h \rangle + o(\|h\|)$

(formule de Taylor- Young à l'ordre 1).

2 . Si $f \in C^2(\Omega)$ alors

$i)$ $f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \int_0^1 (1-t) \left\langle \nabla^2 f(a+th) h, h \right\rangle dt$

(formule de Taylor à l'ordre 2 avec reste intégral).

$ii)$ $f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \left\langle \nabla^2 f(a+\theta h) h, h \right\rangle$ avec $0 < \theta < 1$

(formule de Taylor- Maclaurin à l'ordre 2)

iii) $f(a+h) = f(a) + \langle \nabla f(a), h\rangle + \frac{1}{2}\left\langle \nabla^2 f(a)\, h, h\right\rangle + o\left(\|h\|^2\right)$

(formule de Taylor- Young à l'ordre 2).

Remarque : Dans la proposition précédente la notation $o\left(\|h\|^k\right)$ pour $k \in \mathbb{N}^*$ signifie une expression qui tend vers 0 plus vite que $\|h\|^k$ (c'est à dire, si on la divise par $\|h\|^k$, le résultat tend vers 0 quand $\|h\|$ tend vers 0).

3.0.5 Quelque rappels sur le matrices carrées réelles

Soit $n \in \mathbb{N}^*$ et $A \in M_n(\mathbb{R})$ une matrice carrée réelle.

1 . Soit $\mathbb{C}$: l'ensemble des nombres complexes. On rappelle que $\lambda \in \mathbb{C}$ est une **valeur propre** de A s'il existe $x \in \mathbb{C}^n$ avec $x \neq 0$ tel que $Ax = \lambda x$; on appelle x **vecteur propre** de A associé à la valeur propre λ.

2 . On dit que la matrice A est **semi-définie positive** si $\langle Ax, x\rangle \geq 0,\ \forall x \in \mathbb{R}^n$.

On dit que A est **définie positive** si $\langle Ax, x\rangle > 0,\ \forall x \in \mathbb{R}^n$ avec $x \neq 0$.

3 . Rappellons que si A est symétrique alors toutes les valeurs propres de A sont réelles ; en plus il existe n vecteurs propres de A appartenant à $\mathbb{R}^n$ formant une base orthonormée en $\mathbb{R}^n$.

4 . Supposons que la matrice A est symétrique. Alors

$$\langle Ah, h\rangle \geq \lambda_{\min} \|h\|^2,\ \forall h \in \mathbb{R}^n,$$

où $\lambda_{\min} \in \mathbb{R}$ est la plus petite valeur propre de A.

Rémarquons que l'inégalité précédente devient égalité si h est un vecteur propre associé à la valeur propre $\lambda_{\min}$.

5 . Supposons que A est symètrique. Alors A est semi-définie positive si et seulement si $\lambda_{\min} \geq 0$ et A est définie positive si et seulement si $\lambda_{\min} > 0$.

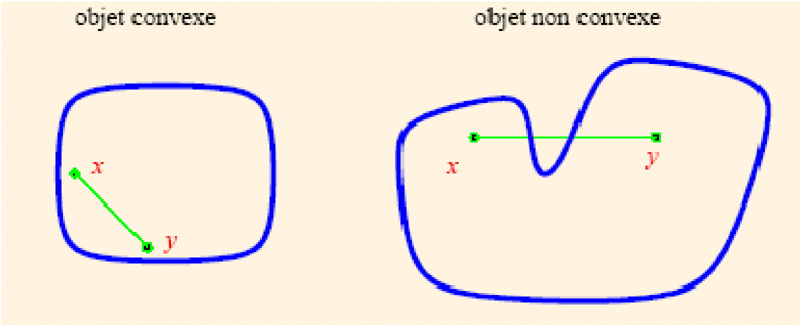
6 . **Abréviation :** La notation SDP pour une matrice carrée rélle signifie "matrice symétrique et définie positive" (elle ne signifie pas "matrice semi-définie positive" !).

3.0.6 Convexité

Fonctions convexes, strictement convexes, fortement convexes

Définition. Un ensemble $U \subset \mathbb{R}^n$ est dit **convexe** si

$\forall x, y \in U$ on a $[x, y] \subset U$ (quelque soit deux points dans U, tout le segment qui les unit est dans U).



Définition. Soit $U \subset \mathbb{R}^n$ un ensemble **convexe** et $f : U \to \mathbb{R}$ une fonction.

1 . On dit que f est **convexe** sur U si

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x), \quad \forall x, y \in U, \ \forall t \in [0,1]$$

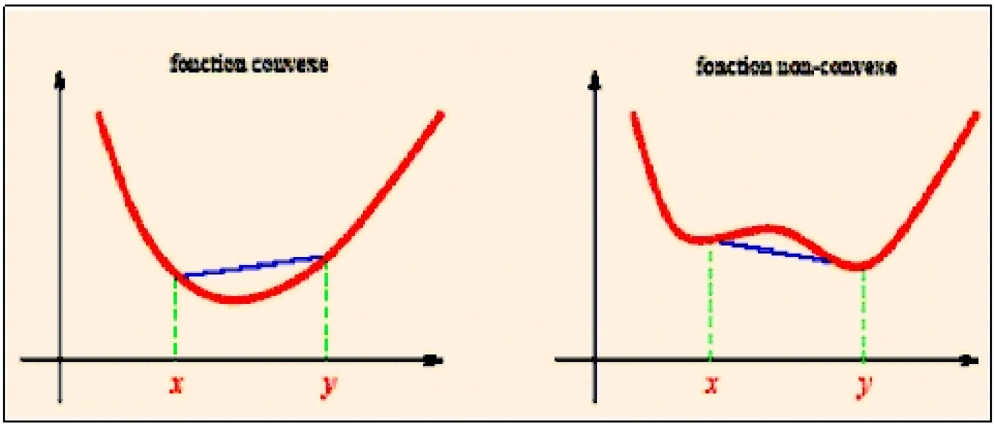
2On dit que f est **strictement convexe** sur U si

$$f(ty + (1-t)x) < tf(y) + (1-t)f(x), \quad \forall x, y \in U \quad \text{avec } x \neq y, \ \forall t \in]0,1[$$

3 . On dit que f est **fortement convexe** sur U s'il existe $\alpha \succ 0$ tel que

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x) - \alpha t(1-t)\|y - x\|^2, \quad \forall x, y \in U \quad , \ \forall t \in [0,1]$$

4 . On dit que f est **concave** (respectivement **strictement concave,** respectivement **fortement concave**) si $-f$ est convexe(respectivement strictement convexe,respectivement fortement convexe).



Remarque : Il est facile de voir qu'on a : fortement convexe $\Longrightarrow$ strictement convexe $\Longrightarrow$ convexe.

Les réciproques ne sont pas vraies en général ; par exemple une application affine $f(x) = Ax + b$ est convexe (et aussi concave) mais elle n'est pas strictement convexe(ni strictement concave) donc elle n'est pas fortement convexe(ni fortement concave). On a le résultat utile suivant :

Proposition 1.3 Soit $U \subset \mathbb{R}^n$ un ensemble convexe, $p \in \mathbb{N}^*, f_1, f_2, ..., f_p : U \longrightarrow \mathbb{R}$ des fonctions convexes et $\gamma_1, \gamma_2, ..., \gamma_n$ des constantes strictement positives.

Posons $f = \gamma_1 f_1 + \gamma_2 f_2 + ... + \gamma_P f_P$. Alors on a :

1 . La fonction f est convexe (donc toute combinaison linéaire avec des coefficients strictement positifs de fonctions convexes est convexe).

2 . Si au moins l'une des fonctions $f_1, ..., f_p$ est strictement convexe alors f est strictement convexe.

3 . Si au moins l'une des fonctions $f_1, ..., f_p$ est fortement convexe alors f est fortement convexe.

Il est en général difficile de vérifier la convexité d'une fonction en utilisant uniquement la définition (essayez avec $f(x) = x^2$ ou avec $f(x) = x^4$!) Les propositions suivantes donnent des critéres de convexité, convexité stricte et convexité forte, plus faciles à utiliser que les définitions respectives.

Proposition 1.4. (caractérisation de la convexité)

Soit $\Omega \subset \mathbb{R}^n$ ouvert, $U \subset \Omega$ avec U convexe et $f : \Omega \to \mathbb{R}$ une fonction de classe C^1. Alors

$a)$ les 3 propositions suivantes sont équivalentes :

1 . f est convexe sur U

2 .

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in U$$

3 . ∇f est **monotone sur** U, c'est à dire

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0, \quad \forall x, y \in U.$$

$b)$ Si de plus f est de classe C^2sur Ω alors f est convexe sur U si et seulement si

$$\left\langle \nabla^2 f(x)(y - x), y - x \right\rangle \geq 0, \quad \forall x, y \in U \tag{1.2}$$

Démonstration $a)$ On montre ici l'équivalence entre $1)$, $2)$ et $3)$.

$1) \Longrightarrow 2)$: Supposons f convexe ; la définition de la convexité peut s'écrire

$$f(x + t(y - x)) - f(x) \leq t[f(y) - f(x)]$$

En fixant x, y en divisant par t et en faisant t tendre vers 0 (ce qui est possible car $t \in [0, 1]$) on obtient $2)$

$2) \Longrightarrow 3)$: De$2)$ on déduit

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in U$$

et aussi (en inversant x et y) :

$$f(x) > f(y) + \langle \nabla f(y), x - y \rangle, \quad \forall x, y \in U$$

En faisant la somme de ces 2 inégalités on obtient 3).

$3) \Longrightarrow 1)$: Soient $x, y \in U$ fixés. On introduit la fonction $g : I \to \mathbb{R}$ définie par

$$t \in I \to g(t) = f(ty + (1-t)x) \in \mathbb{R}$$

où I est un intervalle ouvert qui contient $[0,1]$ .Il est facile de voir que g est de classe C^1 et on a

$$g^{'}(t) = \langle \nabla f(ty + (1-t)x), y - x \rangle \quad \forall t \in I.$$

Soient $t_1, t_2 \in [0,1]$ avec $t_1 \prec t_2$. Alors

$$\begin{aligned} g^{'}(t_2) - g^{'}(t_1) &= \langle \nabla f(x + t_2(y-x)) - \nabla f(x + t_1(y-x)), y - x \rangle = \\ & \langle \nabla f(x + t_2(y-x)) - \nabla f(x + t_1(y-x)), (t_2 - t_1)(y-x) \rangle \frac{1}{t_2 - t_1}. \end{aligned}$$

Par hypothése 3) le dernier term de l'égalité précédente est ≥ 0, ce qui montre que la fonction $g^{'}$ est une fonction croissante. On déduit alors que g est une fonction convexe sur $[0,1]$, ce qui nous donne pour tout $t \in [0,1]$:

$$g(t \cdot 1 + (1-t) \cdot 0) \leq tg(1) + (1-t)g(0)$$

c'est à dire

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x)$$

donc f est convexe.

b) On suppose ici $f \in C^2(\Omega)$.

"$\Longrightarrow$" Supposons que f est convexe et montrons (1.2). Soit $h \in \mathbb{R}^n$ fixé et notons $g : \Omega \to \mathbb{R}$ la fonction donnée par $g(x) = \langle \nabla f(x), h \rangle \ \ \forall x \in \Omega$. Nous avons en utilisant aussi la Proposition 1.1 :

$$\left\langle \nabla^2 f(x)\, h, h \right\rangle = \langle \nabla g(x), h \rangle = \frac{\partial g}{\partial h}(x) = \lim_{t \to 0} \frac{\langle \nabla f(x+th), h \rangle - \langle \nabla f(x), h \rangle}{t}$$

ce qui nous donne

$$\left\langle \nabla^2 f(x)\, h, h \right\rangle = \lim_{t \to 0} \frac{\langle \nabla f(x+th) - \nabla f(x), th \rangle}{t^2}$$

Considérons maintenant $x, y \in U$ arbitraires et $h = y - x$. Comme $x + t(y-x) \in U \quad \forall t \in [0,1]$, de l'égalité précédente on déduit à l'aide de la monotonie de ∇f que $\left\langle \nabla^2 f(x)\, h, h \right\rangle \geq 0$, c'est à dire (1.2) ;

"$\Longleftarrow$" Supposons maintenant que (1.2) est satisfaite et montrons que f est convexe. Soient $x, y \in U$ fixées arbitraires, et considérons la fonction $g_1 : \Omega \to \mathbb{R}$ donnée par $g_1(z) = \langle \nabla f(z), x - y \rangle \quad \forall z \in \Omega$. Alors

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = g_1(x) - g_1(y) = \langle \nabla g_1(y + \theta(x-y)), x - y \rangle$$

avec $\theta \in]0,1[$ (on a utilisé l'une des formules de Taylor).

D'autre part, nous avons

$$\nabla g_1(z) = \nabla^2 f(z)(x-y)$$

et ceci nous permet de déduire, en utilisant aussi (1.2) :

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = \left\langle \nabla^2 f(y + \theta(x-y))(x-y), x - y \right\rangle \geq 0 \ .$$

Ceci nous donne la monotonie de ∇f donc la convexité de f.

Proposition 1.5. (caractérisation de la convexité stricte)

Soit $\Omega \subset \mathbb{R}^n$ ouvert, $U \subset \Omega$ avec U convexe et $f : \Omega \to \mathbb{R}$ une fonction de classe C^1. Alors

a) les 3 propositions suivantes sont équivalentes :

1 . f est strictement convexe sur U

2 .

$$f(y) \succ f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in U \text{ avec } x \neq y$$

3 . ∇f est **strictement monotone sur** U, c'est à dire

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle > 0 \quad \forall x, y \in U \text{ avec } x \neq y.$$

b) Si de plus f est de classe C^2 sur Ω alors si

$$\left\langle \nabla^2 f(x)(y - x), y - x \right\rangle > 0, \quad \forall x, y \in U \text{ avec } x \neq y \tag{1.3}$$

alors f est strictement convexe sur U.

Proposition 1.6. (caractérisation de la convexité forte)

Soit $\Omega \subset \mathbb{R}^n$ ouvert, $U \subset \Omega$ avec U convexe et $f : \Omega \to \mathbb{R}$ une fonction de classe C^1. Alors

a) les 3 propositions suivantes sont équivalentes :

1 . f est fortement convexe sur U

2 . Il existe $\beta > 0$ tel que

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \beta \|y - x\|^2 \quad \forall x, y \in U.$$

3 . Il existe $\gamma > 0$ tel que

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \gamma \|y - x\|^2 \quad \forall x, y \in U.$$

b) Si de plus f est de classe C^2 sur Ω alors f est fortement convexe sur U si et seulement si il existe $\beta > 0$ tel que

$$\left\langle \nabla^2 f(x)(y - x), y - x \right\rangle \geq \beta \|y - x\|^2 \quad \forall x, y \in U \tag{1.4}$$

Remarques :

1 . Dans la Proposition 1.6b) il n'y a pas d'équivalence entre la stricte convexitè de f et l'inégalité (1.3) de cette proposition dans le cas $f \in C^2$ (par exemple la fonction $f : \mathbb{R} \to \mathbb{R}$ donnée par $f(x) = x^4$ est strictement convexe, mais l'inégalité n'est pas satisfaite si $x = 0$).

2 . Dans le cas particulier ou l'ensemble U est ouvert alors pour les 3 propositions précédentes les inégalités (1.2) , (1.3) et (1.4) concernant la matrice Hessienne $\nabla^2 f$ s'écrivent respectivement :

$$\left\langle \nabla^2 f(x)h, h \right\rangle \geq 0 \quad \forall x \in U, \ \ \forall h \in \mathbb{R}^n \ \ \text{(pour la Proposition 1.4)},$$

$$\left\langle \nabla^2 f(x)h, h \right\rangle > 0 \quad \forall x \in U, \ \ \forall h \in \mathbb{R}^n \ \ \text{(pour la Proposition 1.5)},$$

$$\left\langle \nabla^2 f(x)h, h \right\rangle \geq \beta \left\| h \right\|^2 \quad \forall x \in U, \ \ \forall h \in \mathbb{R}^n \ \ \text{(pour la Proposition 1.6)}.$$

Pour voir ceci il suffit de remarquer que d'une part, pour tout $y \in U$ on peut considérer $h = y - x \in \mathbb{R}^n$ et d'autre part, que pour tout $h \in \mathbb{R}^n$ on peut considèrer $y = x + th \in U$ avec $t > 0$ assez petit

3 . Dans le cas particulier $n = 1$ et U un intervalle dans $\mathbb{R}$, on a $\nabla^2 f\left(x\right) = f^{''}(x)$ donc $\left\langle \nabla^2 f(x)\left(y - x\right), y - x \right\rangle = f^{''}(x)\left(y - x\right)^2$. Alors les 3 inégalités précédentes peuvent s'écrire respectivement :

$$f^{''}(x) \geq 0 \ \forall x \in U \ , \qquad f^{''}(x) > 0 \ \forall x \in U \ \ \text{et} \ \ f^{''}(x) \geq \beta \ \forall x \in U.$$

On rétrouve alors une caractérisation bien connue pour la convexitè et la convexité stricte.

Définition. On appelle **fonction elliptique** une fonction $f : \mathbb{R}^n \to \mathbb{R}$ de classe C^1 et fortement convexe.

Exemples des fonctions convexes, strictement convexes et fortement convexes

1 . Soit $f : \mathbb{R} \to \mathbb{R}$ donnée par $f(x) = x^2$. Comme $\nabla f(x) = f'(x) = 2x$ on a $\forall x, y \in \mathbb{R}$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = \left[f'(x) - f'(y) \right] \cdot (x - y) = 2(x - y)^2 \equiv \beta \|x - y\|^2,$$

avec $\beta = 2$, donc f est fortement convexe.

C'est encore plus facile si on utilise f'' : $f''(x) = 2 \equiv \beta > 0$ donc f est fortement convexe.

2 . Plus généralement, si $n = 1$ et $\Omega \subset \mathbb{R}$ intervalle ouvert :

Toute fonction $f : \Omega \to \mathbb{R}$ de classe C^2 et satisfaisant :

$$\exists \alpha > 0, \; f''(x) \geq \alpha, \quad \forall x \in \Omega$$

est une fonction fortement convexe(c'est une conséquence immédiate de la Proposition 1.6). Si $\Omega = \mathbb{R}$ alors la fonction f est elliptique.

Exemples :

$i)$ $f(x) = ax^2 + bx + c$ avec $a > 0$

$ii)$ $f(x) = x^2 + \sin(x)$ (car $f''(x) = 2 - \sin(x) \geq 1, \forall x \in \mathbb{R}$).

3 . **Le cas général** $(n \in \mathbb{N}^*)$

Soit $f : \mathbb{R}^n \to \mathbb{R}$ donnée par

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c, \quad \forall x \in \mathbb{R}^n,$$

avec $A \in M_n(\mathbb{R})$ une matrice carrée réelle et **symétrique** de taille n, avec $b \in \mathbb{R}^n$ un vecteur et $c \in \mathbb{R}$ un scalaire(on appelle encore, par abus de language, fonction(ou forme) **quadratique** une fonction de ce type). On calcule facilement :

$$\nabla f(x) = Ax - b$$

$$\nabla^2 f(x) = A$$

(donc la hessienne de f est constante).

En utilisant les rappels de la Section précédente et les Propositions 1.4, 1.5 et 1.6, on déduit :

$i)$ f est une fonction convexe $\Longleftrightarrow$ A est une matrice semi-définie positive

$ii)$ f est fortement convexe $\Longleftrightarrow$ f est strictement convexe $\Leftrightarrow$ A est une matrice définie positive .

Fonctions coercives

Définition. Soit $\Omega \subset \mathbb{R}^n$ un ensemble **non borné** et $f : \Omega \to \mathbb{R}$. On dit que f est **coercive sur** Ω si on a

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

(ceci peut aussi s'écrire : $f\left(x^k\right) \to +\infty$ pour toute suite $\left\{x^k\right\}_{k\in\mathbb{N}} \subset \Omega$ telle que $\left\|x^k\right\| \to +\infty$)

Remarque : Supposons qu'on a $\Omega_0 \subset \Omega \subset \mathbb{R}^n$ avec Ω_0 non borné et $f : \Omega \to \mathbb{R}$. Si f est coercive sur Ω alors f est coercive sur Ω_0, la réciproque n'étant pas vraie en général.

Proposition 1.7. Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $U \subset \Omega$ un ensemble convexe et non borné. Soit $f : \Omega \to \mathbb{R}$ une fonction de classe C^1sur Ω et fortement convexe sur U. Alors f est coercive sur U.

Démonstration. Ecrivons l'inégalité 2 de la Proposition 1.6 avec $x = a$ ou a est un élément fixé de U :

$$f(y) = f(a) + \langle \nabla f(a), y - a \rangle + \beta \|y - a\|^2 \quad \forall y \in U,$$

où $\beta > 0$ est un nombre fixé. D'autre part, nous avons

$$\langle \nabla f(a), y-a \rangle \geq -\|\nabla f(a)\| \cdot \|y-a\|$$

(conséquence de l'inégalité de Cauchy-Scwhwarz). Ceci donne

$$f(y) \geq f(a) - \|\nabla f(a)\| \cdot \|y-a\| + \beta \|y-a\|^2 \quad \forall y \in U.$$

En faisant $\|y\| \to +\infty$ (ce qui donne $\|y-a\| \to +\infty$) on obtient le résultat.

3.1 Existence et unicité d'un point de minimum

Théorème (Existence)

Soit $U \subset \mathbb{R}^n$ un ensemble non-vide et fermé et $f : U \to \mathbb{R}$ une fonction continue.On suppose

1.Soit U est borné,

2.Soit U est non borné et f est une fonction coercive.

Alors il existe au moins un point de minimum de f sur U .

Théorème (Unicité)

Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \to \mathbb{R}$ une fonction strictement convexe.Alors il existe au plus un point de minimum de f sur U.