

# Quelques rappels de calcul différentiel, Convexité

Dans ce premier chapitre nous allons présenter quelques notions fondamentales de calcul différentiel, de l'analyse convexe, de l'optimisation. Ces notions sont utiles pour démontrer les résultats dans les chapitres suivants

## Différentiabilité, gradient, matrice Hessienne

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

2- On note  $e_1, e_2, \dots, e_n$  les éléments de la base canonique de  $\mathbb{R}^n$ , où  $e_i$  est le vecteur de  $\mathbb{R}^n$  donné par :

$$(e_i)_j = \delta_{ij} = \begin{cases} 0 & \text{si } j \neq i \\ 1 & \text{si } j = i \end{cases}, \text{ pour } j \in \{1, 2, \dots, n\}.$$

(symbole de Kronecker).

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

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

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

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

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

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

6- Si  $a, b \in \mathbb{R}^n$  on note  $[a, b]$  le segment de droite reliant les deux points  $a, b$ , le sous-ensemble de  $\mathbb{R}^n$  donné par

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

Si  $a, b \in \mathbb{R}$  avec  $a < b$  alors on retrouve le fait que  $[a, b]$  désigne l'intervalle des nombres  $x \in \mathbb{R}$  tels que  $a \leq x \leq b$ .

7- Rappelons aussi l'inégalité de Cauchy-Schwarz :

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

Notion de la dérivée partielle, gradient, matrice Hessienne

Soit  $\Omega \subset \mathbb{R}^n$  un ouvert et  $f : \Omega \rightarrow \mathbb{R}$  est une fonction.

1- On dit que  $f$  est de classe  $C^m$  sur  $\Omega$  ( $f \in C^m(\Omega)$ ) si toutes les dérivées partielles jusqu'à l'ordre  $m$  existent et sont continues.

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

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

(c'est la dérivée partielle de  $f$  en  $x$  de direction  $x_i$ ).

3- Pour tout  $x \in \Omega$  on note (quand  $\exists$ )

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

(la Jacobienne de  $f$  en  $x$ ). On appelle gradient de  $f$  au point  $x$  le vecteur colonne

$$\nabla f(x) = (J_f(x))^T = \left( \frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)^T$$

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

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

(c'est la dérivée directionnelle de  $f$  en  $x$  de direction  $h$ ) où on a noté  $\frac{\partial f}{\partial h}(x)$ .

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

b)  $\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x)$ .

Nous rappelons aussi la formule :

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

5- Gradient de la composée: Soient les deux fonctions  $f : \Omega \rightarrow \mathbb{R}$  et  $g : U \rightarrow \mathbb{R}$  et supposons qu'on deux ouverts  $\Omega \subset \mathbb{R}^n$  et  $U \subset \mathbb{R}$  avec en plus  $f(\Omega) \subset U$  (on peut alors définir  $g \circ f : \Omega \rightarrow \mathbb{R}$ ). Supposons que  $f, g$  sont de classe  $C^1$ . Alors  $g \circ f$  est aussi de classe  $C^1(\Omega)$  avec en plus

$$\nabla(g \circ f)(x) = g'(f(x)) \nabla f(x) \quad \forall x \in \Omega.$$

6- Si les dérivées de  $f$  possèdent à leur tour des dérivées partielle, on dit que  $f$  admet des dérivées partielle d'ordre 2, on pose :

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

c'est à dire :

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}.$$

$\nabla^2 f$  s'appelle le hessien de  $f$  et on note par  $H$  ou  $\nabla^2 f$ .

Si  $f$  est une fonction de classe  $C^2$  (admet des dérivées partielles d'ordre 2 continues), le hessien de  $f$  est une matrice symétrique  $H(x)$  d'ordre  $n \times n$ .

7- Soit  $H(x)$  la matrice hessienne de la fonction  $f$ , si :

a) La matrice  $H(x)$  est dite semi-définie positive (SDP) ssi :

$$\forall y \in \mathbb{R}^n, y^t H(x) y \geq 0.$$

où si les vecteurs propres de la matrice  $H(x)$  sont positives.

b) La matrice  $H(x)$  est dite définie positive (DP) ssi :

$$\forall x \in \mathbb{R}^n \setminus \{0\}, y^t H(x) y > 0.$$

c-à-d les vecteurs propres de la matrice  $H(x)$  sont strictement positives.

Quelques exemples importants :

1- Si  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  est une fonction constante alors

$$\nabla f = \nabla^2 f = 0$$

2- Soit  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  définie par

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

où  $a \in \mathbb{R}^n$  est un vecteur donné (c'est à dire,  $f$  est une fonction linéaire). Alors on calcule facilement :  $\frac{\partial f}{\partial x_k} = a_k$ , donc

$$\nabla f = a$$

le gradient est constant. Ceci nous donne

$$\nabla^2 f = 0.$$

3- Soit  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  donnée par

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

où  $A \in M_n(\mathbb{R})$  est une matrice carrée, réelle, de taille  $n$  (c'est à dire,  $f$  est une

fonction quadratique associée à la matrice  $A$  ). Alors on détermine le gradient :

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

et la matrice hessienne :

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

4- En particulier, si  $A$  est symétrique (c'est à dire  $A = A^T$  ). Alors

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

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

## Développement de Taylor

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

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

a) Formule de Taylor à l'ordre 1 avec reste intégral

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

b) Formule de Taylor - Maclaurin à l'ordre 1

$$\text{il existe } \theta \in [0,1] \text{ tel que } f(a+h) = f(a) + \langle \nabla f(a+\theta h), h \rangle.$$

c) Formule de Taylor - Young à l'ordre 1

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

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

a) Formule de Taylor à l'ordre 2 avec reste intégral

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

b) Formule de Taylor - Maclaurin à l'ordre 2 il existe  $\theta \in [0,1]$  telque

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle \nabla^2 f(a+\theta h)h, h \rangle.$$

c) Formule de Taylor - Young à l'ordre 2

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

**Remarque :** Dans les formules précédentes, la notation  $o(\|h\|^k)$  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 ).

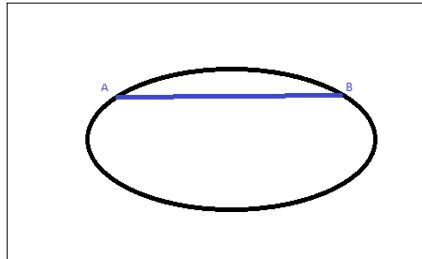
## Éléments d'analyse convexe

Dans cette section, on présente un rappel des notions fondamentales de l'analyse convexe qui serviront d'appuis pour la suite

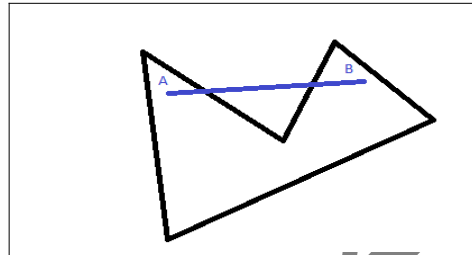
## Ensembles convexes

- Un ensemble  $C$  de  $\mathbb{R}^n$  est dit convexe si

$$\lambda x + (1 - \lambda)y \in C, \forall x, y \in C, \forall \lambda \in [0, 1].$$



Un ensemble convexe



Un ensemble non convexe

-  $C$  est dit affine si

$$\lambda x + (1 - \lambda)y \in C, \forall x, y \in C, \forall \lambda \in \mathbb{R}.$$

-  $C$  est un polyèdre convexe s'il est de la forme:

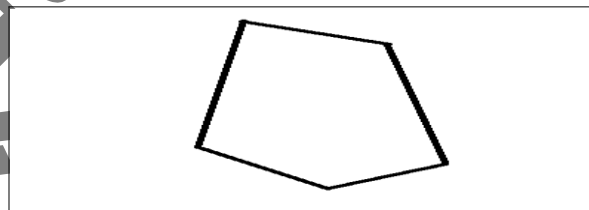
$$C = \{x \in \mathbb{R}^n : A_i^t x \leq b_i, i = 1, \dots, m\}$$

où  $A_i$  est un vecteur non nul de  $\mathbb{R}^n$  et  $b_i$  un scalaire pour  $i = 1, \dots, m$ .

$C$  peut s'écrire sous la forme matricielle suivante :

$$C = \{x \in \mathbb{R}^n / Ax \leq b\},$$

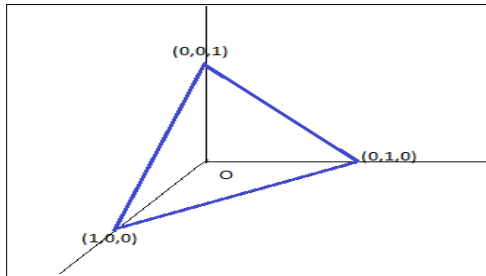
où  $A$  est une matrice de  $\mathbb{R}^{m \times n}$  et  $b$  un vecteur de  $\mathbb{R}^m$ .



Un polyèdre convexe

-  $S_n$  est un (n-simplexe) s'il est de la forme :

$$S_n = \left\{ x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1 \right\}$$



$S_n$  est un (3-simplexe)

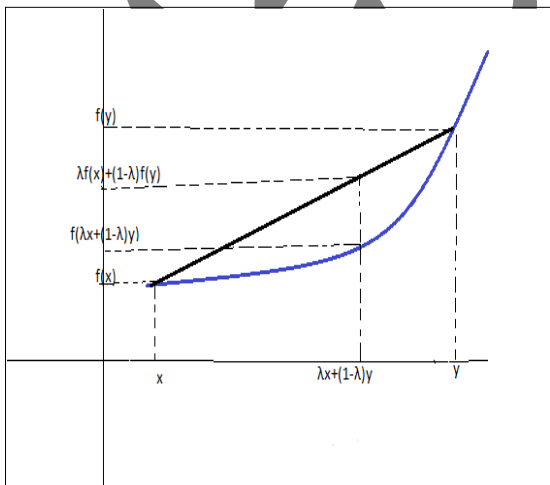
- Un point  $x \in S_n$  est dit extrémal (ou sommet de  $S_n$ ) si l'on a :  
 $\forall t \in [0,1], \forall (y,z) \in S_n^2 : x = (1-t)y + tz \Rightarrow x = y = z$

### Fonctions convexes :

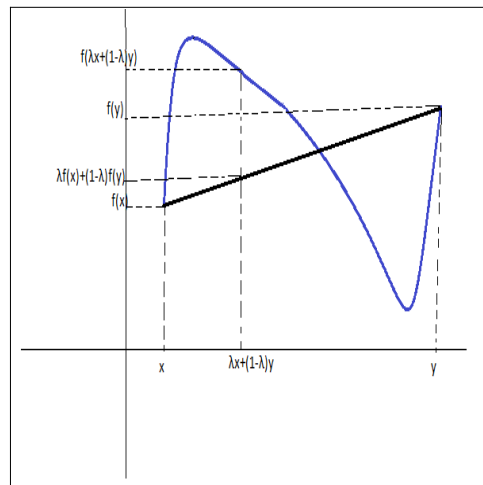
Soit  $f : C \rightarrow \mathbb{R}$  une fonction et  $C$  un ensemble convexe de  $\mathbb{R}^n$ .

- $f$  est dite convexe sur  $C$  si l'inégalité suivante est satisfaite :  
 $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \forall \lambda \in [0,1], \forall x, y \in C.$

- $f$  est dite strictement convexe sur  $C$  si :  
 $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y), \forall \lambda \in ]0,1[, \forall x, y \in C \text{ et } x \neq y.$



Une fonction convexe



Une fonction non convexe

- $f$  est dite mid-convexe sur  $C$  si :

$$\forall x, y \in C, f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}.$$

- $f$  est dite quasi-convexe sur  $C$  si :

$$f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y)), \forall \lambda \in [0, 1], \forall x, y \in C.$$

- $f$  est dite fortement convexe sur  $C$  s'il existe  $\alpha > 0$ , tel que :

$$\forall \lambda \in ]0, 1[, \forall x, y \in C \text{ et } x \neq y,$$

on a:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{1}{2} \alpha \lambda(1-\lambda) \|x-y\|^2.$$

(on dit aussi que  $f$  est  $\alpha$ -convexe).

Si  $f$  est une fonction continue sur un convexe  $C$ , on a :

- $f$  est convexe sur  $C$  si et seulement si  $f$  est mid-convexe sur  $C$ .
- $f$  est  $\alpha$ -convexe sur  $C$  si et seulement si :

$$\forall x, y \in C, f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} - \frac{\alpha}{8} \|x-y\|^2.$$

## Caractérisation d'une fonction convexe différentiable

Avant de donner les théorèmes de caractérisation d'une fonction convexe différentiable, nous allons donner une lemme très importante pour la démonstration des théorèmes de caractérisation:

**Lemme :** Soient  $I$  est un intervalle ouvert, et  $\varphi$  est une fonction de classe  $C^1$  sur  $I$ , Alors  $\varphi$  est une fonction convexe sur  $I$  ssi  $\varphi'$  est une fonction croissante sur  $I$ .

### Démonstration :

On montre que  $\varphi$  est une fonction convexe sur  $I \Leftrightarrow \varphi'$  est une fonction croissante sur  $I$ .

On montre d'abord  $\varphi$  est une fonction convexe sur  $I \Rightarrow \varphi'$  est une fonction croissante sur  $I$ . Soient  $\lambda \in ]0, 1[$  et  $(x, y) \in I \times I$ , tel que  $y > x$ , et  $z = \lambda y + (1-\lambda)x$ . On a:

$$\varphi(z) = \varphi(\lambda y + (1-\lambda)x) = \varphi(x + \lambda(y-x)) \leq \lambda \varphi(y) + (1-\lambda)\varphi(x)$$

donc

$$\frac{\varphi(z)-\varphi(x)}{\lambda} \leq \varphi(y)-\varphi(x)$$

$\Leftrightarrow$

$$\frac{\varphi(z)-\varphi(x)}{z-x} \leq \frac{\varphi(y)-\varphi(x)}{y-x}$$

autre part

$$\frac{\varphi(y)-\varphi(z)}{1-\lambda} \geq \varphi(y)-\varphi(x)$$

$\Leftrightarrow$

$$\frac{\varphi(y)-\varphi(z)}{y-z} \geq \frac{\varphi(y)-\varphi(x)}{y-x}$$

Alors

$$\frac{\varphi(z)-\varphi(x)}{z-x} \leq \frac{\varphi(y)-\varphi(x)}{y-x} \leq \frac{\varphi(y)-\varphi(z)}{y-z}, \quad \forall (x, y, z) \in I^3 \text{ et } y > z > x$$

donc

$$\frac{\varphi(z)-\varphi(x)}{z-x} \leq \frac{\varphi(y)-\varphi(x)}{y-x}, \quad \forall (x, y, z) \in I^3 \text{ et } y > z > x$$

$$\frac{\varphi(y)-\varphi(x)}{y-x} \leq \frac{\varphi(y)-\varphi(t)}{y-t}, \quad \forall (x, y, z) \in I^3 \text{ et } y > t > x$$

on passant à la limites pour  $z$  tend vers  $x$ , et  $t$  tend vers  $y$ , on a

$$\varphi'(x) \leq \varphi'(y), \quad \forall (x, y) \in I^2 \text{ et } y > x$$

donc  $\varphi'$  est une fonction croissante sur  $I$ .

### Réciproquement

Soient  $t \in [0,1]$  et  $(x, y) \in I \times I$ , on considère la fonction  $\psi$  définie par

$$\psi(t) = t\varphi(y) + (1-t)\varphi(x) - \varphi(x+t(y-x))$$

est une fonction de classe  $C^1([0,1])$  car  $\varphi$  est une fonction de classe  $C^1(I)$  et on a

$$\psi'(t) = \varphi(y) - \varphi(x) - (y-x)\varphi'(x+t(y-x))$$

soient  $(t_1, t_2) \in [0,1]^2$  tel que  $t_1 \neq t_2$  suppose que  $t_1 > t_2$  et d'après  $\varphi'$  est une fonction croissante, on a

$$\begin{aligned} \psi'(t_1) - \psi'(t_2) &= (y-x)[\varphi'(x+t_2(y-x)) - \varphi'(x+t_1(y-x))] \\ &= \frac{1}{(t_2-t_1)}[\varphi'(x+t_2(y-x)) - \varphi'(x+t_1(y-x))][(t_2-t_1)(y-x)] \\ &\leq 0 \end{aligned}$$

donc  $\psi'$  décroissante par rapport à  $t$  sur  $[0,1]$ .

Autre part  $\psi(1) = \psi(0) = 0$  d'après le théorème de Rolle, il existe  $a \in ]0,1[$  tel que  $\psi'(a) = 0$ . Alors  $\psi(t) \geq 0$  sur  $[0,1]$  donc

$$(1-t)\varphi(x) + t\varphi(y) - \varphi(x+t(y-x)) \geq 0, \quad \forall t \in [0,1]$$

La convexité de  $\varphi$  s'en déduit.



Maintenant nous allons donner la théorème si la fonction différentiable de première ordre de la convexité

**Théorème** : Soit  $C$  est un ouvert convexe de  $\mathbb{R}^n$  et  $f : C \rightarrow \mathbb{R}$ ,  $f \in C^1(C)$

. Les propositions suivantes sont équivalences:

- 1-  $f$  est une fonction convexe sur  $C$
- 2-  $f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle, \forall x, y \in C$ .
- 3-  $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0, \forall x, y \in C$ .

### Démonstration :

On montre que  $((1) \Leftrightarrow (2))$

On montre d'abord  $((1) \Rightarrow (2))$ . Soient  $\lambda \in [0, 1]$  et  $(x, y) \in C \times C$ , on a la dérivée directionnelle

$$\lim_{\lambda \rightarrow 0} \frac{f(x + \lambda(y-x)) - f(x)}{\lambda} = \langle \nabla f(x), (y-x) \rangle, \text{ (car } f \in C^1(C))$$

Par convexité de  $f$ , on a de plus

$$\begin{aligned} f(x + \lambda(y-x)) &= f((1-\lambda)x + \lambda y) \\ &\leq (1-\lambda)f(x) + \lambda f(y) \end{aligned}$$

donc

$$\frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x)$$

En passant à la limite quand  $\lambda \rightarrow 0$ , on obtient

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

Montrons maintenant  $((2) \Rightarrow (1))$ . Soient  $(x, y) \in C \times C$  et  $z = \lambda x + (1-\lambda)y$ . On peut écrire:

$$\begin{aligned} f(x) - f(z) &\geq \langle \nabla f(z), x - z \rangle \\ f(y) - f(z) &\geq \langle \nabla f(z), y - z \rangle \end{aligned}$$

Soit encore :

$$\begin{aligned} f(x) - f(\lambda x + (1-\lambda)y) &\geq \langle \nabla f(z), (1-\lambda)(x-y) \rangle \\ f(y) - f(\lambda x + (1-\lambda)y) &\geq \langle \nabla f(z), \lambda(y-x) \rangle \end{aligned}$$

En multipliant la première inégalité par  $\lambda$ , la seconde par  $(1-\lambda)$ , on a

$$\begin{aligned} \lambda(f(x) - f(\lambda x + (1-\lambda)y)) &\geq \lambda \langle \nabla f(z), (1-\lambda)(x-y) \rangle \\ (1-\lambda)(f(y) - f(\lambda x + (1-\lambda)y)) &\geq (1-\lambda) \langle \nabla f(z), \lambda(y-x) \rangle \end{aligned}$$

puis en additionnant les deux, on obtient

$$\lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y) \geq 0$$

c'est l'inégalité de convexité. Alors  $((1) \Leftrightarrow (2))$ .

On montre que  $((1) \Leftrightarrow (3))$

On montre d'abord  $((1) \Rightarrow (3))$ . Soient  $t \in \mathbb{R}$  et  $(x, y) \in C \times C$ , on a

$$\varphi : \mathbb{R} \rightarrow \mathbb{R} \text{ définie par } t \mapsto \varphi(t) = f(x + t(y-x))$$

est une fonction convexe de classe  $C^1(\mathbb{R})$  car  $f$  est une fonction convexe de classe  $C^1(C)$

Soient  $\lambda \in [0,1]$  et  $(t_1, t_2) \in \mathbb{R}^2$ , on a

$$\begin{aligned} \varphi(\lambda t_1 + (1-\lambda)t_2) &= f(x + (\lambda t_1 + (1-\lambda)t_2)(y-x)) \\ &= f(\lambda(x + t_1(y-x)) + (1-\lambda)(x + t_2(y-x))) \\ &\leq \lambda f(x + t_1(y-x)) + (1-\lambda)f(x + t_2(y-x)) \\ &= \lambda \varphi(t_1) + (1-\lambda)\varphi(t_2) \end{aligned}$$

Comme  $\varphi$  est une fonction convexe de classe  $C^1(\mathbb{R})$  alors  $\varphi'$  est une fonction croissante sur  $\mathbb{R}$  par rapport à  $t$  et on a

$$\varphi'(t) = \langle \nabla f(x + t(y-x)), (y-x) \rangle$$

donc

$$\varphi'(1) = \langle \nabla f(y), (y-x) \rangle \geq \varphi'(0) = \langle \nabla f(x), (y-x) \rangle$$

donc

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

Montrons maintenant  $((3) \Rightarrow (1))$ . Soient  $t \in \mathbb{R}$  et  $(x, y) \in C \times C$ , on a

$$g : [0, 1] \rightarrow \mathbb{R} \text{ définie par } t \mapsto g(t) = (1-t)f(x) + tf(y) - f(x + t(y-x))$$

est une fonction de classe  $C^1([0,1])$  car  $f$  est une fonction de classe  $C^1(C)$  et on a

$$g'(t) = f(y) - f(x) - \langle \nabla f(x + t(y-x)), (y-x) \rangle$$

soient  $(t_1, t_2) \in [0,1]^2$  tel que  $t_1 \neq t_2$  suppose que  $t_1 > t_2$  et d'après (3), on a

$$\begin{aligned} g'(t_1) - g'(t_2) &= \langle \nabla f(x + t_2(y-x)) - \nabla f(x + t_1(y-x)), (y-x) \rangle \\ &= \frac{1}{(t_2-t_1)} [\langle \nabla f(x + t_2(y-x)) - \nabla f(x + t_1(y-x)), (t_2-t_1)(y-x) \rangle] \\ &\leq 0 \end{aligned}$$

donc  $g'$  décroissante par rapport à  $t$  sur  $[0,1]$ .

Autre part  $g(1) = g(0) = 0$  d'après le théorème de Rolle, il existe  $a \in ]0,1[$  tel que  $g'(a) = 0$ . Alors  $g(t) \geq 0$  sur  $[0,1]$  donc

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

La convexité de  $f$  s'en déduit.

Nous allons donner la théorème si la fonction différentiable de second ordre de la convexité

**Théorème :** Soit  $C$  est un ouvert convexe de  $\mathbb{R}^n$  et  $f : C \rightarrow \mathbb{R}$ ,  $f \in C^2(C)$ .  
Les deux propositions suivantes sont équivalences:

- 1-  $f$  est une fonction convexe sur  $C$
- 2-  $\langle \nabla^2 f(x)h, h \rangle \geq 0, \forall x \in C, \forall h \in \mathbb{R}^n$ .

### Démonstration :

On montre que  $((1) \Leftrightarrow (2))$

On montre d'abord  $((1) \Rightarrow (2))$ . D'après la théorème précédente on a

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

donc pour  $y = x + th$  on a  $\langle \nabla f(x + th) - \nabla f(x), th \rangle \geq 0$ . On divise alors cette inégalité par  $t^2$  puis on fait tendre  $t$  vers 0, ce qui fournit :

$$\langle \nabla^2 f(x)h, h \rangle \geq 0, \forall x \in C, \forall h \in \mathbb{R}^n.$$

Montrons maintenant  $((2) \Rightarrow (1))$ . Soient  $t \in \mathbb{R}$  et  $(x, y) \in C \times C$ ; On applique la formule de Taylor-Mac Laurin à l'ordre deux : il existe  $\theta \in [0, 1]$  tel que

$$\begin{aligned} f(y) &= f(x + (y - x)) \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x + \theta(y - x))(y - x), (y - x) \rangle \\ &\geq f(x) + \langle \nabla f(x), y - x \rangle \end{aligned}$$

donc

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

D'après la théorème précédente on a  $f$  est une fonction convexe sur  $C$