Notes d'optimisation en dimension finie

Notes d'optimisation en dimension finie

1-Optimisation dans $\mathbb{R}$

1.1-Extremums et fonctions continues

Soient K un partie non vide de $\mathbb{R}$, $f: K \longrightarrow \mathbb{R}$ une fonction

$$(\mathcal{P}) \ \inf_{x\in K} f(x)$$

un problème.

Définition 1.1. Le problème $(\mathcal{P})$ est appelé problème de minimisation avec contraintes ou lié ou local, si $K \subsetneq \mathbb{R}$. $(\mathcal{P})$ est appelé un problème de minimisation sans contraintes ou libre ou global si $K = \mathbb{R}$.

Question: Sous quelles conditions sur K et f, le problème de minimisation $(\mathcal{P})$ admet une solution minimale i.e. un élément $x^* \in \mathbb{R}$ tel que $f(x^*) \leq f(x)$ pour tout $x \in \mathbb{R}$.?

Rappelons un résultat bien connu d'optimalité, i.e. un résultat d'existence de la solution de $(\mathcal{P})$, dont la démonstration est semblable à celle du théorème 1.2 ci dessous.

Théorème 1.1.(Weirstrass) Si $K = [a, b]$ et f est continue sur K. Le problème avec contrainte $(\mathcal{P})$ admet une solution minimale.

Définition 1.2.

$a)$ On dit que la fonction f est coercive sur K, si $f(x) \longrightarrow +\infty$ quand $|x| \longrightarrow +\infty$.

$b)$ La fonction $f: K \longrightarrow \mathbb{R} \cup \{+\infty\}$ est propre s'il existe $x_0 \in \mathbb{R}$ tel que $f(x_0) < +\infty$.

Il est clair que, s'il existe $\alpha > 0$ tel que $f(x) \geq \alpha|x|$, pour tout $x \in K$, f est coercive.

Un second résultat d'optimalité est:

Théorème 1.2. Si la fonction $f: \mathbb{R} \longrightarrow \mathbb{R} \cup \{+\infty\}$ est propre, coercive et continue. Le problème global $(\mathcal{P})$ admet au moins une solution minimale.

Preuve. Soit $m = \inf_{x\in\mathbb{R}} f(x)$, puisque il existe $x_0 \in \mathbb{R}$ tel que $f(x_0) < +\infty$, alors $m \in \mathbb{R}$ et pour tout $n \in \mathbb{N}^*$, il existe une suite $\{x_n\}$ dans $\mathbb{R}$ tel que: $m \leq f(x_n) < m + \frac{1}{n}$ et donc $\lim_{n\longrightarrow+\infty} f(x_n) = m$. La suite $\{x_n\}$ est bornée, si non $\forall n \in \mathbb{N}^*$, il existe une sous suite $\{x_{\varphi(n)}\}$ de la suite $\{x_n\}$ vérifiant $|x_{\varphi(n)}| > n$, donc $|x_{\varphi(n)}| \longrightarrow +\infty$, comme J est coercive sur $\mathbb{R}$. En tend que sous-suite de la suite $f(x_n)$, on a $f(x_{\varphi(n)}) \longrightarrow +\infty$, contradiction. $\{x_n\}$ étant bornée, elle admet une sous-suite $\{x_{\sigma(n)}\}$ convergente vers $x^* \in \mathbb{R}$. Par continuité de $f(x^*) = m \leq f(x)$, pour tout $x \in \mathbb{R}$.

1.2-Extremums et dérivés premières

On note par $I(x, r)$ l'intervalle ouvert de centre $x \in \mathbb{R}$ et de rayon $r > 0$. Soient E un partie non vide de $\mathbb{R}$, et $f: E \longrightarrow \mathbb{R}$ une fonction.

Définition 1.3. Un point d'accumulation $x_0 \in E$ est dit:

$a)$ Un minimum local (respectivement un minimum global) de f dans E, s'il existe $\delta > 0$ tel que: $f(x_0) \leq f(x)$, $\forall x \in I(x_0, \delta)$, (respectivement $\forall x \in E$). Quand x_0 existe, on note par $f_{min}(x_0)$ la valeur minimale de f.

$b)$ Un maximum local (respectivement un maximum global) de f dans E, s'il existe $\delta > 0$ tel que: $f(x) \leq f(x_0)$, $\forall x \in I(x_0, \delta)$ (respectivement $\forall x \in E$). Quand x_0 existe, on note par $f_{max}(x_0)$ la valeur maximale de f.

$c)$ Un extremum local (respectivement un extremum global), s'il est minimum local ou maximum Local (respectivement s'il est minimum global ou maximum global).

La valeur $f(x_0)$ est dite selon le cas: la valeur minimale locale ou maximale locale (respectivement minimale globale ou. maximale globale).

$d)$ Un point selle s'il n'est pas un extremum.

Proposition 1.1. Si $x_0 \in E$ est un extremum local ou global et si $f'(x_0)$ (la dérivée de f au point x_0) existe. Alors $f'(x_0) = 0$.

Remarque 1.1.

a) Si f est dérivable sur E, les solutions de l'équation $f'(x) = 0$ sont appelées points critiques de f, ils sont selon la forme de la courbe, des points stationnaires ou des points cols de f.

b) Il existe des fonctions dont les dérivées sont nulles en x_0, mais x_0 n'est pas un extremum.

Exemple 1.1. La fonction $f(x) = x^3$ sur $\mathbb{R}$ est telle que $f'(0) = 0$ qui n'est pas un extremum, car la fonction f change de signe au voisinage de 0.

Proposition 1.2. Soit f une fonction dérivable sur E. Alors:

a) Si $f'(x) = 0$, alors $f(x) = \ cte$ (constante).

b) Si, $f'(x) \geq 0$ (respectivement $f'(x) > 0$), alors f est croissante (respectivement f est strictement croissante).

c) Si, $f'(x) \leq 0$ (respectivement $f'(x) < 0$), alors f est décroissante (respectivement f est strictement décroissante).

d) L'équation de la tangente au point $\left(x_0, f(x_0)\right)$ de la courbe de f s'écrit: $f(x) = f'(x_0)(x - x_0) + f(x_0)$, $f'(x_0)$ est donc la pente de cette tangente. Donc si $x_0 \in E$ est un extremum, la tangente au point $\left(x_0, f(x_0)\right)$ est parallèle à l'axe des x. Si admet la dérivée à droite $f'_d(x_0)$ $\big($respectivement la dérivée à gauche $f'_g(x_0)$) au point $x_0\big)$. On a une demi-tangente à droite (respectivement une demi $-$ tangente à gauche) au point $\left(x_0, f(x_0)\right)$ d'équation $f(x) = f'_d(x_0)(x - x_0) + f(x_0)$ $\Big($respectivement $f(x) = f'_g(x_0)(x - x_0) +$ $fx0$.

Définition 1.4. On dit que la fonction f est différentiable en x_0, s'il existe $a \in \mathbb{R}$ et une fonction $\alpha : E \longrightarrow \mathbb{R}$, $h \longmapsto \alpha(h)$ vérifiant $\alpha(h) \longrightarrow 0$, quant $h \longrightarrow 0$ $(h \neq 0)$ tel que: $f(x_0 +$ $h - fx0 = ah + \alpha hh$.

Remarque 1.2. Il y'a équivalences entre la dérivabilité de f au point x_0 et sa différentiabilité en x_0. Dans le cas où $f'(x_0)$ existe, $a = f'(x_0)$.

Définition 1.5. Si f est différentielle au point $x_0 \in E$. La différentiable de f au point $x_0 \in E$ est une forme linéaire (continue) notée df_{x_0} définie de $\mathbb{R}$ dans $\mathbb{R}$, par $df_{x_0}(h) = f'(x_0)h$ pour tout $h \in \mathbb{R}$. Lorsque $f(x) = x$, $f'(x) = 1$ pour tout $x \in E$. On a donc, pou tout $x \in E$ et pour tout $h \in \mathbb{R}$, $dx_x(h) = h$ ou simplement $dx(h) = h$, d'où $df_{x_0}(h) = f'(x_0)dx(h)$, pour tout $h \in \mathbb{R}$, ce qui donne $df_{x_0} = f'(x_0)dx$. Si f est différentiable en tout point $x \in E$, $df_x = df(x) = f'(x)dx$ ou $f'(x) = \frac{df(x)}{dx}$.

Remarque 1.3.

a) Une fonction dérivable en $x_0 \in E$ est continue en x_0.

b) Une fonction peut avoir un extremum en un point sans être dérivable en ce point.

Donnons un résultat d'existence d'un extremum d'une fonction non nécessairement dérivable en x_0.

Théorème 1.3. Soit f une fonction dérivable au voisinage de x_0 sauf au point x_0 et elle est continue en x_0. Alors:

a) Si au voisinage de x_0, $f'(x)$ est négative à gauche de x_0 et elle est positive à droite de x_0, f admet un minimum local en x_0.

b) Si au voisinage de x_0, $f'(x)$ est positive à gauche de x_0 et elle est négative à droite de x_0, f admet un maximum local en x_0.

c) Si au voisinage de x_0, $f'(x)$ garde un signe constant à droite et à gauche de x_0, f n'a pas d'extremums local en x_0.

Preuve. $a)$ Appliquons Lagrange à f sur $[x, x_0]$ $(x < x_0)$, il existe $\alpha \in]x, x_0[$ tel que: $f(x)$-$f(x_0) = f'(\alpha)(x - x_0)$, comme $f'(\alpha) \leq 0$ car $f'(x)$ est négative à gauche de x_0 et comme $x - x_0 < 0$, alors $f(x) - f(x_0) \geq 0$, donc x_0 est un minimum local de f. Si maintenant $x > x_0$, il existe $\beta \in]x_0, x[$ tel que $f(x) - f(x_0) = f'(\beta)(x - x_0) \geq 0$ car $f'(x)$ est positive à droite de x_0, donc x_0 est un minimum local de f. La démonstration de $b)$ se fait de la même manière. $c)$ L'égalité $f(x) - f(x_0) = f'(\alpha)(x - x_0)$ de $a)$ change de signe selon le signe de $f'(\alpha)$, donc f n'a pas d'extremums local en x_0.

Exemple 1.2.

$a)$ La fonction $f(x) = |x|$ sur $\mathbb{R}$ et telle que $f(0) = 0 \leq f(x) = |x|$ sur $\mathbb{R}$, alors que $f'(0)$ n'existe pas. Il est claire que $f'(x) = \pm 1$ change de signe au voisinage de 0 qui est un minimum globale.

$b)$ La fonction $f(x) = \sqrt[3]{x^2}$ sur $\mathbb{R}$ est continue sur $\mathbb{R}$ et elle est dérivable sur $\mathbb{R}^*$ et $f'(x) = \frac{2}{3\sqrt[3]{x}}$.. Donc $f'(x) > 0$ si $x > 0$ et $f'(x) < 0$ si $x < 0$ d'où f admet un minimum global en 0.

$c)$ La fonction

$$f(x) = \begin{cases} \dfrac{x}{1 + e^{\frac{1}{x}}} & \text{si } x \neq 0; \\ 0 & \text{si } x = 0, \end{cases}$$

est continue sur $\mathbb{R}$ et elle est dérivable sur $\mathbb{R}^*$, $f'(x) = \frac{1+\left(1+\frac{1}{x}\right)e^{\frac{1}{x}}}{\left(1+e^{\frac{1}{x}}\right)^2}$, $f_d'(0) = 0 \neq 1 = f_g'(0)$.

Comme $f'(x) \geq 0$ à droite et à gauche de 0, f n'a pas d'extremum en 0.

1.3-Extremums et dérivées d'ordre supérieur

Soit $f \in C^n([a, b])$ i.e. la fonction f est n fois dérivable et sa dérivée d'ordre n, notée $f^{(n)}$ est continue sur $[a, b]$, telle que la fonction $f^{(n)}$ es dérivable sur $]a, b[$ et soit $x_0 \in]a, b[$. On suppose que pour tout $t \in [0,1]$ et pour tout $h \in \mathbb{R}$, $a + th \in [a, b]$.

Théorème 1.4. (Développement de Taylor). Sous les conditions ci-dessus. Pour tout $x \in [a, b]$ $(x \neq x_0)$ et pour tout $p \in \mathbb{N}^*$, il existe $\theta \in]0,1[$ tel que:

$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_{n+1}(x)$. (TG)

où $R_{n+1}(x) = \frac{f^{(n+1)}(x_0+\theta(x-x_0))}{n!p}(x - x_0)^{n+1}(1 - \theta)^{n-p+1}$. (RTG)

La formule (TG) s'appelle le développement à l'ordre n de Taylor, avec reste généralisé, de la fonction f au point x_0. En posant $x - x_0 = h$, (TG) s'écrit:

$f(x_0 + h) = f(x_0) + \frac{f'(x_0)}{1!}h + \frac{f''(x_0)}{2!}h^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}h^n + R_{n+1}(h)$ où $R_{n+1}(h) = \frac{f^{(n+1)}(x_0+\theta h)}{n!p}h^{n+1}(1 - \theta)^{n-p+1}$.

Remarque 1.4.

$a)$ Pour $p = 1$, on obtient la formule de Taylor-Cauchy (TC), dont le reste est $R_{n+1}(h) = \frac{f^{(n+1)}(x_0+\theta h)}{n!}h^{n+1}(1 - \theta)^n$ (RTC)

$b)$ Pour $p = n + 1$, on obtient la formule de Taylor-Lagrange (TL), dont le reste est $R_{n+1}(h) = \frac{f^{(n+1)}(x_0+\theta h)}{(n+1)!}h^{n+1}$ (RTL)

Théorème 1.5 (développement de Taylor − Péano). Si $f \in C^{n-1}([a, b])$ et si $f^{(n-1)}$ est dérivable en $x_0 \in]a, b[$, on a la formule de Taylor-Péano suivante:

$f(x_0+h)=f(x_0)+\frac{f'(x_0)}{1!}h+\frac{f''(x_0)}{2!}h^2+\cdots+\frac{f^{(n-1)}(x_0)}{(n-1)!}h^{n-1}+R_{n+1}(h)$ où $R_{n+1}(h)=o(h^n)$ i.e $R_{n+1}(h)$ est négligeable devant h^n autrement dit $\lim_{h\to 0}\frac{R_{n+1}(h)}{h^n}$=0.

Théorème 1.6 (développement de Taylor avec reste intégrale). Si $f\in C^{n+1}([a,b])$, on obtient la formule de Taylor avec reste intégrale suivante:

$f(x_0+h)=f(x_0)+\frac{f'(x_0)}{1!}h+\frac{f''(x_0)}{2!}h^2+\cdots+\frac{f^{(n)}(x_0)}{n!}h^n+R_{n+1}(h)$ où $R_{n+1}(h)=\frac{1}{n!}\int_{x_0}^{x_0+h}f^{(n+1)}(s)(x_0+h-s)^n\,ds$. En faisant un premier changement de variable en posant $u=x_0+h-s$, $R_{n+1}(h)=\frac{1}{n!}\int_0^h f^{(n+1)}(x_0+h-u)(u)^n\,ds$ et par un deuxième changement de variable $h-u=th$, $R_{n+1}(h)=\frac{1}{n!}\int_0^1(1-t)^n f^{(n+1)}(x_0+ht)\,h^{n+1}dt$

Remarque 1.5.

$a)$ Dans la cas où $x_0=0$ dans les différentes formules précédentes, on obtient la formule de Taylor-Maclaurin (TM) avec restes de Taylor-Maclaurin (RTM).

$b)$ Les restes dans les différentes formules de Taylor sont négligeables. Montrons par exemple que: $R_{n+1}(h)=\frac{f^{(n+1)}(\theta h)}{(n+1)!}h^{n+1}\longrightarrow 0$ quand $n\longrightarrow\infty$, puisque, $0\le|R_{n+1}(h)|=\left|\frac{f^{(n+1)}(\theta h)}{(n+1)!}h^{n+1}\right|=\left|f^{(n+1)}(\theta h)\right|\frac{1}{(n+1)!}|h|^{n+1}\le M\frac{1}{(n+1)!}|h|^{n+1}$ $\left(\left|f^{(n+1)}(\theta h)\right|\text{ est supposée bornée}\right)$. Ce qui permet d'obtenir une meilleure approximation de $f(x)$ en la remplaçant par un polynôme.

On a vu dans la proposition 1.1 que: si au voisinage de x_0,une fonction est dérivable au point x_0 et admet x_0 comme extremum local en x_0, sa dérivée en x_0 est nulle.

Question: Parmi les solutions de l'équation $f'(x)=0$, existe-il des extremums de f?
La réponse est donnée par le théorème suivant:

Théorème 1.7. Si, au voisinage de x_0, une fonction f est de classe C^n, $n\in\mathbb{N}^*$ et telle que: $f'(x_0)=f''(x_0)=\cdots=f^{(n-1)}(x_0)=0$ et $f^{(n)}(x_0)\neq 0$. Alors, x_0 est un extremum local ssi n est pair.et:

$a)$ x_0 est un minimum local si $f^{(n)}(x_0)>0$.

$b)$ x_0 est un maximum local si $f^{(n)}(x_0)<0$.

Preuve. La formule (TL) s'écrit: $f(x_0+h)-f(x_0)=\frac{f^{(n)}(x_0+\theta h)}{n!}h^n$ où $0<\theta<1$.
Supposons que n est pair et que $f^{(n)}(x_0)>0$, par continuité de $f^{(n)}$, on a $\lim_{h\to 0}f^{(n)}(x_0+$ *θh=fnx0>0*, donc pour *12fnx0*, il existe *δ>0* tel que pout tout *h∈I0,δ* on a $f^{(n)}(x_0+\theta h)>f^{(n)}(x_0)-\frac{1}{2}f^{(n)}(x_0)=f^{(n)}(x_0)>0$, donc $f(x_0+h)-f(x_0)>0$ ou bien $f(x)-f(x_0)>0$ pour tout $x\in I(x_0,\delta)$ i.e. x_0 est un minimum local. Même argumentation pour le maximum local.

Remarque 1.6.

$a)$ Dans le cas où n est impair $f(x_0+h)-f(x_0)$ change de signe avec h on a ni maximum ni minimum. On a un point d'inflexion si $f^{(n)}(x_0)>0$ i.e. la courbe traverse la tangente au point $(x_0,f(x_0))$.

$b)$ Dans le cas où n est paire et $f^{(n)}(x_0)=0$, on peut rien dire. Il se peut qu'on a un extremum ou un point selle. Comme le montre les exemples suivants: la fonction $f(x)=x^4$ a pour $f'(0)=f''(0)=0$ et $f^{(4)}(0)=24$, donc 0 est un minimum, alors que la fonction $f(x)=x^3$ a pour $f'(0)=f''(0)=0$, comme $f'''(x)=6x$ et $f'''(0)=6>0$, 0 est un point d'infliction, donc se n'est pas un extremum.

Exemple 1.3.

$a)$ La fonction $f(x)=x^3-3x-4$ a pour points stationnaires ($x=0$ et $x=2$).

Comme pour $f''(0) = -6$, 0 est un maximum local de f, $f_{max}(0) = -4$.et comme $f''(2) = 6$, 2 est un minimum local de f, $f_{min}(2) = -8$.

b) La fonction $f(x) = (x - x_0)^n$ est telle que: $f'(x_0) = \cdots = f^{(n-1)}(x_0) = 0$ et $f^{(n)}(x_0) = n! > 0$, donc x_0 est un minimum local de f si n est pair et un point d'inflexion si n est impair.

1.4-Extremums et fonctions convexe

Définition 1.6.

a) Une partie non vide K de $\mathbb{R}$ est dite convexe si $\forall t \in [0,1]$ et $\forall x, y \in K$, la combinaison convexe $tx + (1-t)y \in K$.

b) Une fonction f définie sur une partie convexe K de $\mathbb{R}$ est dite convexe, (respectivement strictement convexe) sur K, si $\forall t \in [0,1]$ et $\forall x, y \in K$, $f[tx +$ $1-ty \leq tfx+1-tfy$ respectivement $ftx+1-ty<tfx+1-tfy$.

c) Une fonction f définie sur un convexe K est dite concave, si la fonction $(-f)$ est convexe sur K.

Exemple 1.4.

a) Les singletons, tous les intervalles de $\mathbb{R}$ sont des convexes et les sous-espaces vectoriels sont des convexes.

b) Les fonctions $f(x) = x^2$ et $g(x) = |x|$ définies de $\mathbb{R}$ dans $\mathbb{R}_+$ sont deux fonctions convexes.

Théorème 1.8. Soient I un intervalle ouvert de $\mathbb{R}$, $x_0 \in I$ et f une fonction, définie et convexe sur I. Alors, f est dérivable à droite et à gauche au point x_0 et f est continue au point x_0.

Preuve. Démontrons d'abord que, la fonction $h(x) = \frac{f(x)-f(x_0)}{x-x_0}$ est croissante sur $I \setminus (x_0)$,. soient x et y dans $I \setminus (x_0)$, tel que $x < x_0 < y$ (même résonnement si $x_0 < x < y$ ou $x < y<x0$, on a $0<t=y-x0y-x<1$, $1-t=x0-xy-x$ et $x0=tx+1-ty$, comme f est convexe, $f(x_0) \leq \left(\frac{y-x_0}{y-x}\right) f(x) + \left(\frac{x_0-x}{y-x}\right) f(y) \Leftrightarrow f(x_0)x_0 + f(x_0)(y-x) \leq f(x_0)x_0 + f(x)(y-x_0) + f(y)(x_0-x) \Leftrightarrow \big(f(x) - f(x_0)\big)(y-x_0) \leq \big(f(y) - f(x_0)\big)(x-x_0) \Leftrightarrow h(x) = \frac{f(x)-f(x_0)}{x-x_0} \leq h(y) = \frac{f(y)-f(x_0)}{y-x_0}$, donc g est croissante sur $I \setminus (x_0)$. Pour la dérivabilité, puisque g est croissante sur $[a, b] \setminus (x_0)$ où $a, b \in I$, donc elle est bornée sur tout intervalle intérieur à $]a, b[$, donc $\lim_{x\to x_0^-} h(x) = \max_{\{x\in I, x<x_0\}} h(x) = f'_g(x_0)$ et $\lim_{x\to x_0^+} h(x) = \min_{\{x\in I, x>x_0\}} h(x) = f'_d(x_0)$,.donc f est dérivable à droite et à gauche au point x_0, ce qui conduit à la continuité de f à droite et à gauche de x_0.

En permutant x_0 et x dans l'inégalité de convexité ci-dessus on aura, $\lim_{x\to x_0^+} f(x) = f(x_0) = \lim_{x\to x_0^+} f(x)$, f est donc continue en x_0.

Proposition 1.4. Soit f une fonction, définie et dérivable sur un intervalle I de $\mathbb{R}$. Alors, f est convexe ssi la dérivée f' est croissante sur I.

Preuve. Si f est convexe sur I. Pour tout x dans l'intérieur de I, $f'_g(x)$ et $f'_d(x)$ existent, donc si $x < y < z$ on a $f'_d(x) \leq f'_d(x)(y)$ car $\frac{f(y)-f(x)}{y-x} \leq \frac{f(z)-f(y)}{z-y}$. La fonction $x \mapsto f'_d(x)(x)$ est croissante $\big($respectivement la fonction $x \mapsto f'_g(x)$ est croissante$\big)$. Comme par hypothèse, $f'(x)$ existe sur I, alors $f'(x) = f'_d(x)(x) = f'_g(x)$ et f' est croissante. Réciproquement, soient $t \in [0,1]$, $x, y \in I$ $(x < y)$ et $z = tx + (1-t)y \in I$ qui est convexe. Appliquant le théorème de Lagrange pour f sur les deux intervalles $[x, z]$ et $[z, y]$, ils existent $\alpha \in]x, z[$ et $\beta \in]z, y[$ tel que: $f(x) = f(z) + f'(\alpha)(x - z)$ et $f(y) = f(z) + f'(\beta)(y - z)$. En remarquant que, $x-z=(t-1)(y-x)$, $y-z=t(y-x)$ et en multipliant la première égalité par t et la seconde par $(1-t)$, on a $tf(x)=tf(z)+tf^{'}(\alpha)(t-1)(y-x)$ et $(1-t)f(y)=(1-t)f(z)+t(1-t)f^{'}(\beta)(y-x)$. D'où $tf(x)+(1-t)f(y)=f(tx+(1-t)y)+\ t(t-1)f^{'}(\alpha)(y-x)+t(1-t)f^{'}(\beta)(y-x)=f(tx+(1-t)y)+t(1-t)(y-x)[f^{'}(\beta)-f^{'}(\alpha)]$, comme $t(1-t)(y-x)[f^{'}(\beta)-f^{'}(\alpha)]\geq 0$ car $f^{'}$ est croissante donc, $f(tx+(1-t)y)\leq tf(x)+(1-t)\,f(y)$ et f est convexe sur I.

Corollaire 1.1. Le courbe d'une fonction f, qui est convexe et dérivable sur un intervalle I de $\mathbb{R}$, est située au-dessus de chacune de ses tangentes i.e. au point x_0 on a l'inégalité de convexité suivante: $f(x)\geq\ f'(x_0)(x-x_0)+f(x_0)$ pour tout $x\in I$.

Preuve. Soit $(x_0,f(x_0))$ un point de la tangente à la courbe de f, dont l'équation est $y=f'(x_0)(x-x_0)+f(x_0)$. Alors $f(x)-y=f(x)-[f'(x_0)(x-x_0)+f(x_0)]=[f(x)-f(x_0)]-[f'(x_0)(x-x_0)]$. Le théorème de Lagrange, assure l'existence d'un point $c\in\,]x_0,x[$, tel que $f(x)-f(x_0)=f'(c)(x-x_0)$, donc $f(x)-y=[f'(c)-f'(x_0)](x-x0$, comme d'après la proposition 1.4, f'est croissante on a $fx-y\geq 0$ i.e. la courbe de f est située au-dessus de la tangente à la courbe de f au point $(x_0,f(x_0))$ ou bien $f(x)\geq f'(x_0)(x-x_0)+f(x_0)$ pour tout $x\in I$.

Corollaire 1.2. Soit f une fonction convexe et dérivable sur un intervalle I de $\mathbb{R}$. Alors, si x_0 est un minimum local, il est minimum global.

Preuve. Puisque $f(x)\geq\ f^{'}(x_0)(x-x_0)+f(x_0)$ pour tout $x\in I$ et $f'(x_0)=0$, alors $f(x)\geq\ f(x_0)$ pour tout $x\in I$.

Il est immédiat que:

Corollaire 1.3. si, f est une fonction convexe et dérivable sur un intervalle I de $\mathbb{R}$ et si, $x_0\in I$ satisfait $f'(x_0)=0$, alors x_0 est un minimum de f sur I.

Corollaire 1.4. Si f est deux fois dérivable sur un intervalle I de $\mathbb{R}$. Alors, f est convexe ssi $f''\geq 0$.

Preuve. Puisque $f''\geq 0$ sur I, alors la dérivée f' est croissante sur I, d'après la proposition 1.3 f est convexe sur I. Réciproquement, puisque d'après (TY) appliquée au point x arbitraire dans I, on a $f(x+h)=f(x)+f'(x)h+\frac{1}{2}f''(x)h^2+h^2\varepsilon(h)$ avec $\varepsilon(h)\longrightarrow 0$ quand $h\longrightarrow 0$ et d'après le corollaire 1.1, $f(x+h)-f(x)-f'(x)h\geq 0$, alors $\frac{1}{2}f''(x)+\varepsilon(h)\geq 0$ quand $h\longrightarrow 0$, $f''(x)\geq 0$.

Remarque 1.7.

a) La somme de deux fonctions convexe est une fonction convexe et le produit d'une fonction convexe par un nombre réel positif est une fonction convexe.

b) Le produit de deux fonctions convexes n'est pas toujours convexe. En effet les deux fonctions $f(x)=x$ et $g(x)=x^2$ sont convexes sur $\mathbb{R}$, alors que $h(x)=f(x)g(x)=x^3$ n'est pas convexe sur $\mathbb{R}$, car $h''(x)=6x$ est alternative.

Proposition 1.5. Une fonction continue sur un intervalle I de $\mathbb{R}$ est convexe sur I ssi pour tout $,y\in I, f\left(\frac{x+y}{2}\right)\leq\frac{1}{2}\big(f(x)+f(y)\big)$.

Preuve. Soient $,b\in I$ $(a<b)$, $z\in[a,b]$ et $[a_1,b_1]$ l'un des deux intervalles $\left[a,\frac{a+b}{2}\right]$ et $\left[\frac{a+b}{2},b\right]$ contenant z. Par itération, on obtient une suite d'intervalles $[a_n,b_n]$ de longueur $\frac{b-a}{2^n}$ contenant z. Les deux suites $\{a_n\}$ et $\{b_n\}$ convergent vers z et prennent les formes: $a_n=a+k\left(\frac{b-a}{2^n}\right)=\left(1-\frac{k}{2^n}\right)a+\frac{k}{2^n}b$ et $b_n=a_n+\frac{b-a}{2^n}=\left(1-\frac{k+1}{2^n}\right)a+\frac{k+1}{2^n}b$ où $1\leq k\leq 2^n-1$, $n\in\mathbb{N}^*$. Par l'inégalité de convexité, on a $f(a_n)\leq\left(1-\frac{k}{2^n}\right)f(a)+\frac{k}{2^n}\,f(b)$ et $f(b_n)\leq\left(1-\frac{k+1}{2^n}\right)f(a)+\frac{k+1}{2^n}f(b)$, $\forall n\in\mathbb{N}^*$ et $1\leq k\leq 2^n-1$. Comme $a_n\longrightarrow z=(1-t)a+tb$

où $t = \lim_{n \longrightarrow +\infty} \frac{k}{2^n} \in [a, b]$. Par continuité de f en z, on a $f(z) \leq (1 - t)f(a) + t\, f(b)$ ce qui donne la convexité de f.

Donnons en fin un résultat d'unicité de la solution optimale du problème $(\mathcal{P})$.

Corollaire 1.5. Si la fonction $f: I \longrightarrow \mathbb{R}$ est strictement convexe. La solution optimale du problème $(\mathcal{P})$, quand elle existe est unique.

Preuve. Supposons que $(\mathcal{P})$ admet deux solutions optimales distincts x^* et y^*. Puisque f est strictement convexe. Alors $f\left(\frac{x^*+y^*}{2}\right) < \frac{1}{2}\big(f(x^*) + f(y^*)\big)$, comme $f(x^*) = f(y^*)$ donc $f\left(\frac{x^*+y^*}{2}\right) < f(x^*)$ contradiction.

2-Optimisation dans $\mathbb{R}^n$

2.1-Dérivabilité des fonctions à plusieurs variables

Soient Ω un ouvert de $\mathbb{R}^n$, $a = (a_1, \dots, a_i, \dots, a_n) \in \Omega$, $h = (h_1, \dots, h_i, \dots, h_n) \in \mathbb{R}^n$ tel que $a + h \in \Omega$, $\|\quad\|$ une norme de $\mathbb{R}^n$, $\langle x, y\rangle \coloneqq x^T y$ où x^T est le transposé de x, le produit scalaire de $x, y \in \mathbb{R}^n$ et $f: x = (x_1, \dots, x_i, \dots, x_n) \in \Omega \longmapsto f(x) \in \mathbb{R}$ une fonction. Rappelons que:

■ $i)$ $\forall x, y, z \in \mathbb{R}^n$ et $\forall \lambda \in \mathbb{R}$, $\lambda\langle x, y\rangle = \langle \lambda x, y\rangle = \langle x, \lambda y\rangle$, $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ et $\langle x, x\rangle = \|x\|^2$ (la norme Euclidienne dans $\mathbb{R}^n$)

$ii)$ $\forall x, y \in \mathbb{R}^n$, $\langle x, y\rangle = \langle y, x\rangle$, $\|x + y\|^2 + \|x - y\|^2 = 2[\|x\|^2 + \|y\|^2]$ (identité du parallélograme) et $|\langle x, y\rangle| \leq \|x\|\|y\|$ (inégalité de Cauchy Schwarz).

$iii)$ Si, $a \in \mathbb{R}^n$ est fixé et $\langle x, a\rangle = 0$, pour tout $x \in \mathbb{R}^n$, alors $a = 0$.

■ Pour toute forme linéaire $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, il existe un unique élément $c \in \mathbb{R}^n$ tel que $f(x) = \langle c, x\rangle$.

■ f est dérivable au point a, si ses dérivées partielles au point a, i.e. les $\lim_{h_i \longrightarrow 0} \frac{f(a_1, \dots, a_i + h_i, \dots, a_n) - f(a_1, \dots, a_i, \dots, a_n)}{h_i}$ existent pour tout $i \in \{1, \dots, n\}$, autrement dit pour tout $i \in \{1, \dots, n\}$, les fonctions d'une seule variable dites fonctions partielles $\varphi_i(x_i) = f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n)$ sont dérivables aux points a_i. On note pour tout $i \in \{1, \dots, n\}$, $\frac{\partial f}{\partial x_i}(a)$ ou $f'_{x_i}(a)$ ou simplement $\partial_i f(a)$ la dérivée partielle de f par rapport à la composante x_i.

■ La dérivée de f au point a quand elle existe, est:
$f'(a) = \big(\partial_1 f(a), \dots, \partial_i f(a), \dots, \partial_n f(a)\big) = (\partial_1 f, \dots, \partial_i f, \dots, \partial_n f)_a$ et le gradient de f au point a noté $\nabla f(a)$, est définie par $\nabla f(a) = \langle f'(a), e\rangle = \sum_1^n \partial_i f(a) e_i$ où $e = (e_1, \dots, e_i, \dots, e_n) \in \mathbb{R}^n$, avec $|e_i|$=1, pour tout $i \in \{1, \dots, n\}$. Si f est dérivable en tout point a de Ω, la fonction dérivée $f': \Omega \longrightarrow \mathbb{R}^n$ est donnée par
$f'(x) = \big(\partial_1 f(x), \dots, \partial_i f(x), \dots, \partial_n f(x)\big)$ elle est dans la plus par des cas notée comme le gradient de f.

■ La dérivée directionnelle de f au point a suivant la direction du vecteur $u \in \mathbb{R}^n$ où $a + tu \in \Omega$, $t \in \mathbb{R}$, est la $\lim_{t \longrightarrow 0} \frac{f(a+tu) - f(a)}{t}$ quand elle existe, on la note par $\frac{\partial f}{\partial u}(a)$ ou $\partial_u f(x)$ ou $f'_u(a)$. Dans le cas où u est un vecteur unitaire normal i.e. $\|u\| = 1$, on obtient la dérivée normal de f au point a notée $f'_n(a)$ ou $\frac{\partial f}{\partial n}(a)$.

■ Notons que, si les dérivées directionnelles de f au point a existent dans toutes les directions, la dérivée $\frac{\partial f}{\partial x_i}(a)$ est obtenu à partir de la dérivée directionnelle de f au point a suivant la direction du vecteur $e_i = (0, \dots, 1, \dots, 0)$, (1 se trouve à la position i) de la base canonique $\{e_i, 1 \leq i \leq n\}$ de $\mathbb{R}^n$ i.e. $f'_{x_i}(a) = f'_{e_i}(a)$ pour tout $i \in \{1, \dots, n\}$.

Remarque 2.1.
$a)$ Une fonction dérivable en un point de $\mathbb{R}^n$ n'est pas toujours continue en ce point.
$b)$ Il se peut que les dérivées partielles d'une fonction en un point de $\mathbb{R}^n$ existent, mais les dérivées directionnelles dans toutes les sens n'existent pas.

Exemple 2.1. Pour $(x, y) \in \mathbb{R}^2$, la fonction $f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{si } x^2 + y^2 \neq 0; \\ 0 & \text{si } x^2 + y^2 = 0, \end{cases}$

a pour dérivée au point $(0,0)$, $f'(0,0) = \left(f'_x(0,0), f'_y(0,0)\right) = (0,0)$. Alors que f n'est pas continue en $(0,0)$. Car la suite $\left\{\left(\frac{1}{n}, \frac{1}{n}\right)\right\}$ converge vers $(0,0)$, mais $f\left(\frac{1}{n}, \frac{1}{n}\right) \longrightarrow \frac{1}{2} \neq 0 = f(0,0)$

Exemple 2.2. Pour $(x, y) \in \mathbb{R}^2$, la fonction $f(x,y) = \begin{cases} x + y & \text{si } xy \neq 0; \\ 0 & \text{si } xy = 0, \end{cases}$ a pour dérivée au point $(0,0)$, $f'(0,0) = \left(f'_x(0,0), f'_y(0,0)\right) = (1,1)$, mais pour n'importe quel vecteur $u = (u_1, u_2) \neq (0,0)$ on a $\lim_{t\to 0} \frac{f(tu_1,tu_2)-f(0,0)}{t} = \lim\limits_{t\to 0} \frac{1}{t} = \infty$, donc la dérivée directionnelle de f au point $(0,0)$, dans la direction u n'existe pas.

Exemple 2.3. Pour $(x, y) \in \mathbb{R}^2$, la fonction $f(x,y) = \begin{cases} \frac{x^2y}{x^4+y^2} & \text{si } (x,y) \neq 0; \\ 0 & \text{si } (x,y) = 0, \end{cases}$ a pour dérivée directionnelle au point $(0,0)$, dans la direction $u = (u_1, u_2)$ où $u_2 \neq 0$, la $\lim_{t\to 0} \frac{f(tu_1,tu_2)-f(0,0)}{t} = \lim_{t\to 0} \frac{{u_1}^2 u_2}{t^2{u_1}^4+{u_2}^2} = \frac{{u_1}^2}{u_2}$. Donc pour $u = e_2 = (0,1)$, $f'_y(0,0) = 0$ et $f'_x(0,0) = \lim_{h\to 0} \frac{f(h,0)-f(0,0)}{h} = \lim_{h\to 0} \frac{0}{h} = 0$, alors que la dérivée dans la direction $u = e_2 = (1,0)$ n'existe pas.

■ Si pour tout $x \in \Omega$, $f(x) = \left(f_1(x), \dots, f_j(x), \dots, f_m(x)\right) \in \mathbb{R}^m$, $m \in \mathbb{N}^*$ et si $f_1{}'(a), \dots, f_j{}'(a), \dots, f_m{}'(a)$ existent, alors: $f'(a) = \begin{pmatrix} f_1{}' \\ \vdots \\ f_j{}' \\ \vdots \\ f_m{}' \end{pmatrix}_{\text{a}} = \begin{pmatrix} \partial_1 f_1 \dots \partial_i f_1 \dots \partial_n f_1 \\ \vdots \\ \partial_1 f_j \dots \partial_i f_j \dots \partial_n f_n \\ \vdots \\ \partial_1 f_m \dots \partial_i f_m \dots \partial_n f_m \end{pmatrix}_{\text{a}} = \left(\partial_i f_j(a)\right)_{\substack{1\le i\le n \\ 1\le j\le m}}$. $f'(a)$ est une matrice m lignes et n colonnes.

Théorème 2.1. Si $f'(a)$ existe et si g est une fonction dérivable au point $f(a) = b$ d'un ouvert $U \subset f(\Omega) \subset \mathbb{R}^m$ et pour tout $x \in \Omega$,
$h(x) = (g \circ f)(x) = g\left(f_1(x), \dots, f_j(x), \dots, f_m(x)\right) = \left(h_1, \dots, h_k, \dots, h_p\right)_x \in \mathbb{R}^p$, $p \in \mathbb{N}^*$.
Alors, la fonction h est dérivable au point a et $h'(a) = [(g \circ f)(a)]' = g'(f(a)).f'(a) = g'(b)).f'(a)$.
L'écriture matricielle de $h'(a)$ est:

$$h'(a)=\begin{pmatrix}h'_1\\ \vdots\\ h'_k\\ \vdots\\ h'_p\end{pmatrix}_a=\begin{pmatrix}\partial_1 h_1 \dots \partial_i h_1 \dots \partial_n h_1\\ \vdots\\ \partial_1 h_k \dots \partial_i h_k \dots \partial_n h_k\\ \vdots\\ \partial_1 h_p \dots \partial_i h_p \dots \partial_n h_p\end{pmatrix}_a$$

$$=\begin{pmatrix}\partial_1 g_1 \dots \partial_j g_1 \dots \partial_m g_1\\ \vdots\\ \partial_1 g_k \dots \partial_j g_k \dots \partial_m g_k\\ \vdots\\ \partial_1 g_p \dots \partial_k g_p \dots \partial_m g_p\end{pmatrix}_b\begin{pmatrix}\partial_1 f_1 \dots \partial_i f_1 \dots \partial_n f_1\\ \vdots\\ \partial_1 f_j \dots \partial_i f_j \dots \partial_n f_n\\ \vdots\\ \partial_1 f_m \dots \partial_i f_m \dots \partial_n f_m\end{pmatrix}_a$$

Ce qui revient à écrire, pout tout $i\in\{1,\dots,n\}$ et pour tout $\in\{1,\dots,p\}$, $\partial_i h_k(a)=\sum_{j=1}^{m}\partial_j g_k(b).\partial_i f_j(a)$. Les dérivées $f'(a)$, $g'(b)$ et $h'(a)$ sont des matrices dites Jacobiennes et s'écrivent respectivement: $Jh(a)$, $Jg(b)$ et $Jf(a)$ et $Jh(a)=Jg(b).Jf(a)$. Dans le cas où $n=m$, $Jf(a)$ quand elle existe est une matrice carrée, son déterminant est appelé le Jacobien de f au point a et on le note: $det\big(Jf(a)\big)$ ou $\frac{D(f_1,\dots,f_i,\dots,f_n)}{D(x_1,\dots,x_i,\dots,x_n)}$ ou $\frac{\partial(f_1,\dots,f_i,\dots,f_n)}{\partial(x_1,\dots,x_i,\dots,x_n)}$.

Exemple 2.4. Calculer le Jacobien de $Jf(\rho,\theta,\varphi)$ où $\rho\in\mathbb{R}_+^*$, $\theta\in]0,2\pi[$ et $\varphi\in]0,\pi[$ et $f(\rho,\theta,\varphi)=(\rho cos\theta sin\varphi,\rho sin\theta sin\varphi,\rho cos\varphi)=(f_1,f_2,f_3)_{(\rho,\theta,\varphi)}$.

$$Jf(\rho,\theta,\varphi)=\begin{pmatrix}f_1{}'\\ f_2{}'\\ f_3{}'\end{pmatrix}_{(\rho,\theta,\varphi)}=\begin{pmatrix}\partial_1 f_1 & \partial_2 f_1 & \partial_3 f_1\\ \partial_1 f_2 & \partial_2 f_2 & \partial_3 f_2\\ \partial_1 f_3 & \partial_2 f_3 & \partial_3 f_3\end{pmatrix}_{(\rho,\theta,\varphi)}$$

$$=\begin{pmatrix}cos\theta sin\varphi & -\rho sin\theta sin\varphi & \rho cos\theta cos\varphi\\ sin\theta sin\varphi & \rho cos\theta sin\varphi & \rho sin\theta cos\varphi\\ cos\varphi & 0 & -\rho sin\varphi\end{pmatrix}\qquad\begin{pmatrix}cos\theta sin\varphi & -\rho sin\theta sin\varphi & \rho cos\theta cos\varphi\\ sin\theta sin\varphi & \rho cos\theta sin\varphi & \rho sin\theta cos\varphi\\ cos\varphi & 0 & -\rho sin\varphi\end{pmatrix}$$

Donc

$$det(Jf(\rho,\theta,\varphi)=\rho^2cos\varphi\begin{vmatrix}-sin\theta sin\varphi & cos\theta cos\varphi\\ cos\theta sin\varphi & sin\theta cos\varphi\end{vmatrix}-\rho^2 sin\varphi\begin{vmatrix}cos\theta sin\varphi & -sin\theta sin\varphi\\ sin\theta sin\varphi & cos\theta sin\varphi\end{vmatrix}$$

$$=-\rho^2(sin^2\theta+cos^2\theta)sin\varphi cos^2\varphi-\rho^2(cos^2\theta+sin^2\theta)sin\varphi\, sin^2\varphi$$

$$=-\rho^2 sin\varphi(sin^2\varphi+cos^2\varphi)=-\rho^2 sin\varphi.$$

Exemple 2.5. soit $f:\mathbb{R}^2\longrightarrow\mathbb{R}^2$, définie par:

$f(x,y)=(x+y,x-y)=(f_1,f_2)_{(x,y)}$ et soit $g:\mathbb{R}^2\longrightarrow\mathbb{R}$, définie par, $g(x,y)=sin(x^2-y^2)$, pour tout $(x,y)\in\mathbb{R}^2$. Alors:

$$Jf(x,y)=\begin{pmatrix}\partial_1 f_1 & \partial_2 f_1\\ \partial_1 f_2 & \partial_2 f_2\end{pmatrix}_{(x,y)}=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},$$

$Jg(x,y)=(\partial_1 g \quad \partial_2 g)_{(x,y)}=\big(2xcos(x^2-y^2)\quad -2ycos(x^2-y^2)\big)$. La fonction composée $g\circ f:\mathbb{R}^2\longrightarrow\mathbb{R}$ est définie par $(g\circ f)(x,y)=g\big(f(x,y)\big)=g(x+y,x-y)=sin((x+y)^2-(x-y)^2)=sin(4xy)$. Donc,

$J(g\circ f)(x,y)=(\partial_1(g\circ f)\quad \partial_2(g\circ f))_{(x,y)}=(4ycos(4xy),\quad 4xcos(4xy))=4cos(4xy)(y,\quad x)$ et

$Jg\big(f(x,y)\big)=(\partial_1 g,\quad \partial_2 g)_{f(x,y)}=$

$\big(2f_1(x,y)cos\big(f_1^2(x,y)-f_2^2(x,y)\big),\quad -2f_2(x,y)cos\big(f_1^2(x,y)-f_2^2(x,y)\big)\big)=$

$\big(2(x+y)cos(4xy),\quad -2(x-y)cos(4xy)\big)=2cos(4xy)(x+y,y-x)$

En appliquant d'une part, la dérivée d'une fonction composée

$J(g\circ f)(x,y)=Jg\big(f(x,y)\big).Jf(x,y)$ on obtient le même résultat, en effet: $2cos(4xy)[(x+$ *y,y−x111−1=2cos4xy2y2x=4cos4xyy,x.*

2.2-Différentiabilité

Définition 2.4. On dit que la fonction f est différentiable au point $a \in \Omega$, s'il existe $A = (A_1, \dots, A_i, \dots, A_n) \in \mathbb{R}^n$ et une fonction $\alpha: \mathbb{R}^n \longrightarrow \mathbb{R}$, $h \longmapsto \alpha(h)$ vérifiant $\alpha(h) \longrightarrow 0$, quant $h \longrightarrow 0$ $(h \neq 0)$ tel que: $\lim_{h \longrightarrow 0} \frac{f(a+h)-f(a)-\langle A,h\rangle}{\|h\|} = 0$ ou bien

$$f(a+h) - f(a) = \langle A, h\rangle + \|h\|\alpha = \sum_1^n A_i h_i + \|h\|\alpha(h) \qquad (*)$$

sachant que $a + h \in \Omega$.

Remarque 2.2.

a) La partie linéaire $\langle A, h\rangle$ dans $(*)$ est notée $df_a(h)$ et, elle est appelée la différentielle de f au point a. $df_a(h)$ quand elle existe est unique.

b) En remplaçant pour tout $i \in \{1, \dots, n\}$, h_i par dx_i et donc $h = x - a$ par dx, il est parfois commode d'écrire $df_a(h) = df_a(dx) = \sum_1^n A_i dx_i$.

c) La fonction affine $h \mapsto f(a) + df_a(h)$ réalise quand h est assez petit la première approximation de f an point a.

d) Une fonction différentiable en un point est continue en ce point.

e) Si f est différentiable en tout point $x \in \Omega$, la différentielle de f s'écrit $df(dx) = f'(x)dx$ ou simplement $df = f'(x)dx$.

Remarque 2.3. En dimension 3, f est différentiable au point $a = (\alpha, \beta, \gamma) \in \Omega \subset \mathbb{R}^2$ signifie géométriquement, qu'il existe un plan π tangent à la courbe $z = f(x, y)$ au point a, d'équation $z = \partial_1 f(a)(x - \alpha) + \partial_2 f(a)(y - \beta) + cte$, le vecteur normal à π a pour composante $(\partial_1 f(a), \partial_2 f(a), -1)$.

Proposition 2.1.

a) Une fonction f différentiable en un point a est dérivable en a et $\partial_i f(a) = A_i$ pour tout $i \in \{1, \dots, n\}$.

b) Une fonction f différentiable en un point a admet les dérivées suivant tout vecteur $u = (u_1, \dots, u_i, \dots, u_n) \in \mathbb{R}^n$ au point a et $\partial_u f(a) = \sum_1^n u_i\, \partial_i f(a) = df_a(u)$.

Preuve. Montrons b). Puisque $f(a + h) - f(a) = \sum_1^n u_i \partial_i f(a) + \alpha(h)\,\|h\|$, où $\alpha(h) \longrightarrow 0$ quand $h \longrightarrow 0$. Donc pout $t \in \mathbb{R}^*$, et $h = tu$ on a $f(a + tu) - f(a) = t \sum_1^n u_i \partial_i f(a) +$ *tαtuu* et donc *fa+tu−fat=1nuiðifa+ttαtuu*, quand *t⟶0* on a le résultat.

Proposition 2.2. Une fonction f dérivable, de dérivée continue au point a, est différentiable en ce point et $df_a(h) = \langle f'(a), h\rangle$ pour tout $h \in \mathbb{R}^n$.

Exemple 2.6.

a) Si, il existe une constante $c \in \mathbb{R}^n$ tel que $f(x) = c$ alors $df = 0$. En effet, pour tout x, h dans $\mathbb{R}^n$, $f(x + h) - f(x) = 0 = df(h) + 0(h)\|h\|$ (où $\alpha(h) = 0(h)$ est la fonction nulle), donc $df(h) = \langle f'(x), h\rangle = 0$, pour tout h dans $\mathbb{R}^n$ d'où $f'(x) = 0$ pour tout x dans $\mathbb{R}^n$, alors $f' = 0$.

b) La différentielle d'une forme linéaire f définie sur $\mathbb{R}^n$ est elle même. En effet, pour tout x, h dans $\mathbb{R}^n$, $f(x + h) - f(x) = f(h) + 0(h)\|h\| = df(h) + \alpha(h)\|h\|$, donc $f(h) = df(h) = \langle f'(x), h\rangle$ pour tout x, h dans $\mathbb{R}^n$. Comme, il existe $c \in \mathbb{R}^n$ tel $f(h) = \langle c, h\rangle$, donc $\langle c - f'(x), h\rangle = 0$ pour tout h dans $\mathbb{R}^n$ d'où $f'(x) = c$ pour tout x dans $\mathbb{R}^n$ donc $f' = c$.

c) La différentielle de la fonction affine $f(x) = \langle a, x\rangle + b$ où $x, a \in \mathbb{R}^n$ et $b \in \mathbb{R}$ est $df(h) = \langle a, h\rangle$ pour tout h dans $\mathbb{R}^n$. En effet, $f(x + h) - f(x) = \langle a, x + h\rangle + b - \langle a, x\rangle - b = \langle a, h\rangle = \langle a, h\rangle + 0(h)\|h\|$, pour tout h dans $\mathbb{R}^n$. Comme $df(h) = \langle f'(x), h\rangle = \langle a, h\rangle$ pour tout x dans $\mathbb{R}^n$, alors $f'(x) = a$, pour tout x dans $\mathbb{R}^n$, d'où $f' = c$.

d) Si, pour tout $x \in \mathbb{R}^n$, $f(x) = \langle ax, x\rangle$ où $a \in \mathbb{R}$, alors $df(h) = 2a\langle x, h\rangle$ pour tout $h \in \mathbb{R}^n$. En effet, $f(x + h) - f(x) = a\langle x + h, x + h\rangle - a\langle x, x\rangle = a[\langle x, x\rangle + 2\langle x, h\rangle +$

$\langle h,h\rangle] - a\langle x,x\rangle = 2a\langle x,h\rangle + \|h\|(a\|h\|) = df(h) + \|h\|\alpha(h)$ pour tout $x, y \in \mathbb{R}^n$ où $\alpha(h) = a\|h\| \longrightarrow 0$, quand $h \longrightarrow 0$.

Exemple 2.7. La différentielle en $(0,0) \in \mathbb{R}^2$ de la fonction

$$f(x,y) = \begin{cases} (x^2+y^2)sin\left(\frac{1}{x^2+y^2}\right) & \text{si } (x,y) \neq (0,0); \\ 0 & \text{si } (x,y) = (0,0), \end{cases}$$

est la forme linéaire nulle. En effet, $f\big((0,0) + (h_1,h_2)\big) - f\big((0,0)\big) = f(h_1,h_2) = \left({h_1}^2 + {h_2}^2\right)sin\left(\frac{1}{{h_1}^2+{h_2}^2}\right) = \left[\|h\|sin\left(\frac{1}{\|h\|^2}\right)\right]\|h\| = \alpha(h)\|h\|$, où $\|h\| = \sqrt{{h_1}^2+{h_2}^2}$, et $\alpha(h) = \|h\|sin\left(\frac{1}{\|h\|^2}\right) \longrightarrow 0$ quand $h \longrightarrow 0$. En déduit que $\partial_1 f\big((0,0)\big) = \partial_2 f\big((0,0)\big) = 0$.

Exemple 2.8. La différentielle en $(0,0) \in \mathbb{R}^2$ de la fonction

$$f(x,y) = \begin{cases} x + y + \frac{xy}{\sqrt{x^2+y^2}} & \text{si } (x,y) \neq (0,0); \\ 0 & \text{si } (x,y) = (0,0), \end{cases}$$

n'existe pas. En effet si f est différentiable en $(0,0)$ elle est dérivable en $(0,0)$, et $f\big((0,0) +$ *h1,h2−f0,0=fh1,h2=h1+h2+h1h2h12+h22=∂1f0,0h1+∂2f0,0h2+oh=h1+h2+αhh,* d'où $\frac{h_1h_2}{\sqrt{{h_1}^2+{h_2}^2}} = \frac{h_1h_2}{\|h\|} = \alpha(h)\|h\|$, en déduit que $\frac{h_1h_2}{\|h\|^2} = \alpha(h) \longrightarrow 0$ quand $h \longrightarrow 0$ impossible, puisque la fonction $g(h_1,h_2) = \frac{h_1h_2}{\|h\|^2}$ n'a pas de limite quand $(h_1,h_2) \longrightarrow (0,0)$.

2.3-Extremums d'une fonction à plusieurs variables

En prenant $a \in K$ un partie non vide de $\mathbb{R}^n$, $f\colon K \longrightarrow \mathbb{R}$ et $B(a,r) = \{x \in \mathbb{R}^n,\ \|x-a\| <$ *r a∈K*, la boule ouverte dans *ℝn*, de centre *a* et de rayon *r>0* au lieu de *]x0,r*, la norme au lieu de la valeur absolue $|\quad|$. Les définitions et la terminologie relatives aux problème $(\mathcal{P})$, aux extremums et à la coercivité de f dans le cas d'une fonction à une seule variables restent valables dans le cas d'une fonction à plusieurs variables. On a alors:

■ Si f est continue sur un compact K, le problème $(\mathcal{P})$ admet un solution minimale. Résultat équivalent au théorème 1.1.

■ Si la fonction $f\colon \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ est propre, coercive et continue. Le problème global $(\mathcal{P})$ admet une solution minimale. Résultat équivalent au théorème 1.2.

■ Si, la fonction $f\colon \Omega \longrightarrow \mathbb{R}$ est dérivable au point $a \in \Omega$, et a est un extremum local de f, alors $f'(a) = (\partial_1 f, \ldots, \partial_i f, \ldots, \partial_n f)_a = 0$. La réciproque est fausse comme le montre l'exemple suivant :

Exemple 2.9. la fonction $f(x,y) = xy$ sur $\mathbb{R}^2$ et dérivable sur $\mathbb{R}^2$ et $f'(x,y) = (\partial_1 f, \partial_2 f)_{(x,y)} = (0,0) \Leftrightarrow (y,x) = (0,0) \Leftrightarrow (y = 0 \text{ et } x = 0)$, donc $f'(0,0) = 0$ alors que $f(x,y)$ change de signe au point $B\left((0,0), \sqrt{3}\right)$ $\Big(\big(f(-1,1)\big) < f(0,0) < f(1,1)\Big)$.

Exemple 2.10. $(0,0)$ est un maximum global de la fonction $f(x,y) = 1 - x^2 - y^2$, $\forall (x,y) \in \mathbb{R}^2$. En effet, $f'(0,0) = (\partial_1 f, \partial_2 f)_{(0,0)} = (0,0)$ donc $(0,0)$ est un point stationnaire de f et $f(x,y) - f(0,0) = -(x^2+y^2) \leq 0$, $\forall (x,y) \in \mathbb{R}^2$, $f_{max}(0,0) = 1$.

Définition 2.5. Si, la fonction $f\colon \Omega \longrightarrow \mathbb{R}$ est dérivable sur Ω et, $f'(x) = (\partial_1 f, \ldots, \partial_i f, \ldots, \partial_n f)_x$ est dérivable au point $a \in \Omega$, i.e. pour tout $i \in \{1, \ldots, n\}$ la fonction dérivée partielle $\partial_i f$ est dérivable au point a. La fonction f est deux fois dérivable au point a et la dérivée seconde de f au point a s'écrit

$$f''(a) = \begin{pmatrix} (\partial_1 f)' \\ \vdots \\ (\partial_i f)' \\ \vdots \\ (\partial_n f)' \end{pmatrix}_a$$

où pour tout $i, j \in \{1, \ldots, n\}$ $\left(\partial_j f\right)'(a) = \left(\partial_{j1} f, \ldots, \partial_{ji} f, \ldots, \partial_{jn} f\right)_a$ avec $\partial_{ji} f(a) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a) = f''_{x_j x_i}(a)$. Il vient que

$$f''(a) = \begin{pmatrix} \partial_{11} f \cdots \partial_{1i} f \cdots \partial_{1n} f \\ \partial_{21} f \cdots \partial_{2i} f \cdots \partial_{2n} f \\ \vdots \\ \partial_{n1} f \cdots \partial_{ni} f \cdots \partial_{nn} f \end{pmatrix}$$

est une matrice carré d'ordre n.

Définition 2.6. Si, la dérivée d'ordre $(\alpha - 1)$ de f, existe en tout point de Ω, et si cette dérivée admet une dérivée au point a, on dit que f admet une dérivée d'ordre α au point a. Dans ce cas la dérivée partielle d'ordre α au point a par rapport à $x_{i_1}, \ldots, x_{i_\alpha}$ s'écrit: $\frac{\partial^\alpha f}{\partial x_{i_1} \ldots \partial x_{i_\alpha}}(a)$ ou $\partial_{i_1 \ldots i_\alpha} f(a)$. Dans la cas où les indices ne sont pas égaux, on dit qu'on a une dérivée partielle mixte d'ordre α.

Remarque 2.4. Le nombre de dérivées partielles quand elles existent est n^α où α est l'ordre de la dérivée partielle.

Exemple 2.11. Les dérivées partielles mixtes d'ordre $\alpha = 2$ au point $(0,0)$ de la fonction

$$f(x,y) = \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} & \text{si } (x,y) \neq (0,0); \\ 0 & \text{si } (x,y) = (0,0), \end{cases}$$

sont

$$\partial_1 f(x,y) = \begin{cases} y\dfrac{(x^4 - y^4 + 4x^2y^2)}{(x^2 + y^2)^2} & \text{si } (x,y) \neq (0,0); \\ 0 & \text{si } (x,y) = (0,0), \end{cases}$$

et

$$\partial_2 f(x,y) = \begin{cases} -x\dfrac{(y^4 - x^4 + 4x^2y^2)}{(x^2 + y^2)^2} & \text{si } (x,y) \neq (0,0); \\ 0 & \text{si } (x,y) = (0,0), \end{cases}$$

donc

$\partial_{12} f(0,0) = \lim_{h \to 0} \frac{\partial_1 f(0,h) - \partial_1 f(0,0)}{h} = -1$ et $\partial_{21} f(0,0) = 1$.

Comme $\partial_{12} f(0,0) \neq \partial_{21} f(0,0)$ l'ordre dans la dérivation est important.

Exemple 2.12. Les dérivées partielles à l'ordre 2 de la fonction $f(x,y) = Arctg\left(\frac{x}{y}\right)$, $y \neq 0$. sont: $\partial_1 f(x,y) = y\frac{1}{x^2+y^2}$, $\partial_2 f(x,y) = -x\frac{1}{x^2+y^2}$, $\partial_{11} f(x,y) = -2xy\frac{1}{(x^2+y^2)^2}$, $\partial_{22} f(x,y) = 2xy\frac{1}{(x^2+y^2)^2}$ et $\partial_{12} f(x,y) = \partial_{21} f(x,y) = \frac{x^2-y^2}{(x^2+y^2)^2}$, ici les dérivées mixtes sont égales.

Définition 2.7. On dit que f est différentiable au point a, jusqu'à l'ordre α, si ca dérivée d'ordre $\alpha - 1$ existe et elle est différentiable au point a. Dans ce cas, $d^\alpha f = \mathrm{d}(d^{\alpha-1} f) = \cdots = d^{\alpha-1}(df)$ où $d^\alpha f$ dénote la différentielle d'ordre α de f au point a.

Théoreme 2.1. Si la dérivée d'ordre α de f existe et que cette dérivée est continue en a, alors f est différentiable au point a à l'ordre α.
Théoreme 2.2. (Schwarz généralisé). Si f est α-fois différentiable au point a. Alors, les dérivées partielles mixtes d'ordre α au point a sont égales i.e.

$$\frac{\partial^\alpha f}{\partial x_{i_1}\dots\partial x_{i_k}\partial x_{i_{k+1}}\dots\partial x_{i_\alpha}}(a)=\frac{\partial^\alpha f}{\partial x_{i_1}\dots\partial x_{i_{k+1}}\partial x_{i_k}\dots\partial x_{i_\alpha}}(a)=\frac{\partial^\alpha f}{\partial x_1^{\alpha_1}\dots\partial x_m^{\alpha_m}}(a),$$

où $\alpha_1+\dots+\alpha_m=\alpha$, $(0<\alpha_i<\alpha)$.
■ Pour $(x,y)\in\mathbb{R}^2$ et $\alpha=2$, si les dérivées partielles seconde existent et elles sont continues au point a, alors $\frac{\partial^2 f}{\partial x\partial y}(a)=\frac{\partial^2 f}{\partial y\partial x}(a)$, (théorème de Schwarz) et les dérivées partielles d'ordre deux au point a s'écrivent $\frac{\partial^2 f}{\partial x^{\alpha_1}\partial y^{\alpha_2}}(a)$ avec $\alpha_1+\alpha_2=2$, $(0<\alpha_i,<\alpha, i=1,2)$.
■ Si f est deux fois différentielle au point a,
$d^2 f_a=\mathrm{d}(df)=\mathrm{d}(\sum_{i=1}^n\partial_i f(x)\,dx_i)=\sum_{i=1}^n d(\partial_i f(x)dx_i)=\sum_{i=1}^n\big[d\big(\partial_i f(x)\big)dx_i+$ $\partial ifxddxi$, comme $ddxi=j=1n\partial jdxidxj=0$, car $\partial jdxi=0$ et $d\partial ifx=j=1n\partial ijfxdxj$ pour tout $i\in\{1,\dots,n\}$, alors: $d^2f_a(x)=\sum_{i=1}^n\sum_{j=1}^n\partial_{ij}f(a)dx_j\,dx_i$, avec $\partial_{ij}f(a)=\partial_{ji}f(a)$. La dérivée seconde $f''(a)$ est une matrice symétrique d'ordre n appelée la Hissienne de f au point a notée $Hf(a)$ et $f''(a)\in L\big(\mathbb{R}^n,L(\mathbb{R}^n,\mathbb{R})\big)$ qui est isomorphe à l'ensemble des formes bilinéaire définies sur $\mathbb{R}^n$, noté $L_2(\mathbb{R}^n,\mathbb{R})$, $\langle f''(a)h,k\rangle=\langle h,f''(a)k\rangle=\langle f''(a)k,h\rangle=$ $f''(a)(h,k)$ en tend qu'élément de $L_2(\mathbb{R}^n,\mathbb{R})$, pour tout $h,k\in\mathbb{R}^n$. L'écriture matricielle de d^2f_a, en fonction de la dérivée seconde $f''(a)$ ou $Hf(a)$ est: $d^2f_a(\,h\,)=\langle f''(a)h,h\rangle=$ $\langle Hf(a)h,h\rangle$, pout tout $h\in\mathbb{R}^n$. Puisque $Hf(h)$ est symétrique, il existe une base de vecteurs propres orthonormale $\{u_1,\dots,u_n\}$ dont les valeurs propres associées, $\{\lambda_1,\dots,\lambda_n\}$ sont des éléments de $\mathbb{R}_+^*$ tel que $d^2f(h)=\langle f''(x)h,h\rangle=\langle Ah,h\rangle=\sum_1^n\lambda_i\,h_ih_j\langle u_i,u_j\rangle=\sum_1^n\lambda_i\,{h_i}^2$, pour tout $h\in\mathbb{R}^n$. Si $D_{Hf(a)}$ est la matrice diagonale de $Hf(a)$, son déterminant (déterminant de Hesse) est: $d\acute{e}t\big(D_{Hf(a)}\big)=\lambda_1\lambda_2\dots\lambda_n=d\acute{e}t\big(Hf(a)\big)$ (produit des éléments diagonaux de $D_{Hf(a)}$). La trace de la matrice $Hf(a)$ est $Tra\big(Hf(a)\big)=a_{11}+a_{22}+\dots+a_{nn}=\lambda_1+\lambda_2+\dots+\lambda_n$. Les deux nombres $d\acute{e}t\big(D_{Hf(a)}\big)$ et $Tra\big(Hf(a)\big)$ sont invariant par le changement de base.

■ Si, on désigne par convention la différentielle de f par $df=(dx_1\partial_1+\dots+dx_i\partial_i+\dots+$ $dxn\partial n1f$, alors

$$d^2f=(dx_1\partial_1+\dots+dx_i\partial_i+\dots+dx_n\partial_n)^2f.$$

Par récurrence, quand $d^\alpha f$ existe en tout point de Ω, elle s'écrit.

$$\begin{aligned}d^\alpha f&=(dx_1\partial_1+\dots+dx_i\partial_i+\dots+dx_n\partial_n)^\alpha f\\&=\sum_{i_1=1}^n\sum_{i_2=1}^n\dots\sum_{i_\alpha=1}^n\frac{\partial^\alpha f}{\partial x_{i_1}\partial x_{i_2}\dots\partial x_{i_\alpha}}dx_{i_1}dx_{i_2}\dots dx_{i_\alpha}\\&=\sum_{\alpha_1+\dots+\alpha_n=\alpha}\frac{(\alpha_1+\dots+\alpha_n)!}{\alpha_1!\dots\alpha_n!}\times\frac{\partial^\alpha f}{\partial x_1^{\alpha_1}\dots\partial x_m^{\alpha_m}}(dx_1)^{\alpha_1}\dots(dx_n)^{\alpha_n},\end{aligned}$$

où on a utilisé le polynôme de Newton généralisé

$$(a_1+\dots+a_n)^\alpha=\sum_{\alpha_1+\dots+\alpha_n=\alpha}\frac{\alpha!}{\alpha_1!\dots\alpha_n!}{a_1}^{\alpha_1}\dots{a_n}^{\alpha_n},$$

avec, pour tout $\in\{1,\dots,n\}$, $\alpha_i\in\mathbb{N}$, $(0<\alpha_i<\alpha)$ et $\alpha_1+\dots+\alpha_m=\alpha$.

Exemple 2.13. Calculer $df(x,y)$ et $d^2f(x,y)$ de la fonction $f(x,y)=xlny+ylnx$, $x,y\in\mathbb{R}_+^*$, et déduire $df_{(1,1)}$ et $d^2f_{(1,1)}$.
$\partial_1f(x,y)=lny+\frac{y}{x}$, $\partial_2f(x,y)=lnx+\frac{x}{y}$, $\partial_{11}f(x,y)=-\frac{y}{x^2}$, $\partial_{12}f(x,y)=\frac{1}{y}+\frac{1}{x}=$ $\partial_{21}f(x,y)$ et $\partial_{22}f(x,y)=-\frac{x}{y^2}$.
Alors,

$$df=(dx\partial_1+dy\partial_2)f=\left(lny+\frac{y}{x}\right)dx+\left(lnx+\frac{x}{y}\right)dy.$$

Au point (1,1),

$$df_{(1,1)}=dx+dy.$$

$$d^2f(x,y)=(dx\partial_1+dy\partial_2)^2f=-\frac{y}{x^2}\,(dx)^2+2\left(\frac{1}{y}+\frac{1}{x}\right)dxdy-\frac{x}{y^2}(dy)^2.$$

Au point (1,1),

$$d^2f_{(1,1)}(x,y)=-(dx)^2+4dxdy-(dy)^2.$$

2.4-Développements de Taylor

Soit $f\in C^{\alpha}(\Omega)$, telle que la fonction $f^{(\alpha)}$ est dérivable sur Ω et soit $a\in\Omega$. On suppose que pour tout $t\in[0,1]$ et pour tout $h\in\mathbb{R}$, $a+th\in\Omega$. Comme pour une fonction à une seule variable et avec les mêmes remarques on a:

Théorème 2.3. (développement de Taylor − Lagrange). Sous les conditions ci-dessus. Pour tout $x\in\Omega$ $(x\neq a)$ et pour tout $p\in\mathbb{N}^*$, il existe $\theta\in]0,1[$ tel que: En posant $x-a=h$, (TG) s'écrit:

$f(a+h)=f(a)+\frac{1}{1!}df_a(h)+\frac{1}{2!}d^2f_a(h)+\cdots+\frac{1}{\alpha!}d^{\alpha}f_a(h)+\frac{1}{(\alpha+1)!}d^{\alpha+1}f_{a+\theta h}(h)$ où $x-a=dx=h$. (TL)

Théorème 2.4 (développement de Taylor − Péano). Si $f\in C^{\alpha-1}(\Omega)$ et si $f^{(\alpha-1)}$ est dérivable en $a\in\Omega$, on a la formule de Taylor-Péano suivante:

$f(a+h)=f(a)+\frac{1}{1!}df_a(h)+\frac{1}{2!}d^2f_a(h)+\cdots+\frac{1}{(\alpha-1)!}d^{\alpha-1}f_a(h)+o(\|h\|^{\alpha})$ (TP).

$o(\|h\|^{\alpha})$ s'écrit aussi $\|h\|^{\alpha}\alpha(h)$ avec $\alpha(h)\longrightarrow 0$ quand $h\longrightarrow 0$ et (TP) s'appelle Taylor Young (TP).

Théorème 2.5 (développement de Taylor avec reste intégrale). Si $f\in C^{\alpha+1}(\Omega)$, on obtient la formule de Taylor avec reste intégrale suivante:

$f(a+h)=$
$f(a)+\frac{1}{1!}df_a(h)+\frac{1}{2!}d^2f_a(h)+\cdots+\frac{1}{\alpha!}d^{\alpha}f_a(h)+\frac{1}{\alpha!}\int_0^1(1-t)^{\alpha}d^{\alpha+1}f_{a+th}(h)\,dt$ (TI).

Exemple 2.14. Ecrire (TL) à l'ordre 2 en $(0,0)$ de la fonction à deux variable $f(x,y)=e^x siny$. On doit calculer pour $h,k\in\mathbb{R}$, $df_a(h,k)$, $d^2f_a(h,k)$ et $d^3f_{a+\theta h}(h,k)$. On a alors:

$$\begin{aligned}f(h,k)&=f(h,k)+\frac{1}{1!}df_{(0,0)}(h,k)+\frac{1}{2!}d^2f_{(0,0)}(h,k)+\frac{1}{3!}d^3f_{a+\theta h}(h,k)\\&=\frac{1}{1!}\left(h\partial_1f_{(0,0)}+k\partial_2f_{(0,0)}\right)^1+\frac{1}{2!}\left(h\partial_1f_{(0,0)}+k\partial_2f_{(0,0)}\right)^2\\&+\frac{1}{3!}\left(h\partial_1f_{(\theta h,\theta k)}+k\partial_2f_{(\theta h,\theta k)}\right)^3\\&=\left(h\partial_1f_{(0,0)}+k\partial_2f_{(0,0)}\right)+\frac{1}{2!}\left(h^2\partial_{11}f_{(0,0)}+2hk\partial_{12}f_{(0,0)}+k^2\partial_{22}f_{(0,0)}\right)\\&+\frac{1}{3!}\left(h^3\partial_{111}f_{(\theta h,\theta k)}+3h^2k\partial_{112}f_{(\theta h,\theta k)}+3k^2h\partial_{122}f_{(\theta h,\theta k)}\right.\\&\left.+k^3\partial_{222}f_{(\theta h,\theta k)}\right)\\&=k+hk+\frac{1}{3!}e^{\theta h}(h^3sin\theta k+3h^2kcos\theta k-3k^2hsin\theta k-k^3cos\theta k)\end{aligned}$$

2.5-Conditions suffisantes d'existence d'extremums locaux

Une fonction q: $\mathbb{R}^n\longrightarrow\mathbb{R}$ est une forme quadratique, si elle est de la forme,
$q(h)=\sum_1^n\sum_1^n a_{ij}\,h_ih_j$, pour tout $h=(h_1,\dots,h_i,\dots,h_n)\in\mathbb{R}^n$
où $a_{ij}\in\mathbb{R}$, pour tout $i,j\in\{1,\dots,n\}$. Evidement q est une forme linéaire sur $\mathbb{R}^n$.

Définition 2.8. Une forme quadratique q est dite:

a) définie positive si $\mathrm{q}(h) > 0$, $\forall h \in \mathbb{R}^n$ et $h \neq 0$.

b) semi-définie positive si $\mathrm{q}(h) \geq 0$, $\forall h \in \mathbb{R}^n$.

c) définie négative si $\mathrm{q}(h) < 0$, $\forall h \in \mathbb{R}^n$ et $h \neq 0$.

d) semi-définie négative si $q(h) \leq 0$, $\forall h \in \mathbb{R}^n$

c) alternée si elle est définie positive pour certaines valeurs de h et elle est définie négative pour certaines valeurs de h.

Proposition 2.3. Si q est alternée, il existe $h', h'' \in \mathbb{R}^n$ dont $\|h'\| = \|h''\| = 1$, tel que $\mathrm{q}(h') > 0$ et $\mathrm{q}(h'') < 0$

Preuve. Puisque q est alternée, il existe $t', t'' \in \mathbb{R}^n$ non nuls, tel que $\mathrm{q}(t') > 0$ et $\mathrm{q}(t'') < 0$. Alors $h' = \frac{t'}{\|t'\|}$, $h'' = \frac{t''}{\|t''\|} \in \mathbb{R}^n$, $\|h'\| = \|h''\| = 1$, $\mathrm{q}(h') = \frac{1}{\|t'\|}\mathrm{q}(t') > 0$ et $\mathrm{q}(h'') = \frac{1}{\|t''\|}\mathrm{q}(t'')$.

■ La matrice de la forme quadratique q est une matrice carrée de la forme

$$A = \begin{pmatrix} a_{11} & a_{12} \dots a_{1i} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2i} \dots a_{2n} \\ & \vdots \\ a_{n1} & a_{n2} \dots a_{ni} \dots a_{nn} \end{pmatrix} = (a_{ij})_{\substack{1\leq i\leq n,\\ 1\leq j\leq n}}$$

La matrice A est symétrique si pour tout $i, j \in \{1, \dots, n\}$ $a_{ij} = a_{ji}$, i.e. $A = A^T$ où A^T est la transposé de A.

■ Dans le cas où $A = A^T$, A est diagonalisable. Les mineurs principaux de A sont définis par:

$$A_1 = a_{11}, A_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, A_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{22} \\ a_{11} & a_{12} & a_{13} \end{vmatrix}, \dots, A_n = \begin{vmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ & \vdots \\ a_{n1} & a_{n2} \dots a_{nn} \end{vmatrix}$$

■ Si a est un point stationnaires, d'une fonction f deux fois différentielle en a, pour déterminer la nature de a, on peut utiliser les valeurs propres de la matrice $Hf(a)$, comme suit:

a) On calcule les valeurs propres λ de $Hf(a)$ en résolvant l'équation de son polynôme caractéristique suivante: $dét(Hf(a) - \lambda I) = 0$ (déterminant de la matrice $Hf(a) - \lambda I$ est nul où I est la matrice identité). Alors

i) si la matrice $Hf(a)$ admet n valeurs propres positives et différentes, a est un minimum.

ii) si la matrice $Hf(a)$ admet n valeurs propres négatives et différentes, a est un maximum.

iii) si la matrice $Hf(a)$ admet n valeurs propres différentes qui ne sont pas de même signe, a est un point selle.

b) **En dimension deux**,. Pour déterminer le signe des valeurs propre de la matrice $Hf(a)$, il suffit de calculer le déterminant et la trace de $Hf(a)$.

Dans le cas où $\boldsymbol{dét}\big(\boldsymbol{Hf}(\boldsymbol{a})\big) > 0$:

i) si la $tr\big(Hf(a)\big) > 0$, les deux valeurs propres sont différentes et positives, a est un minimum.

ii) si la $tr\big(Hf(a)\big) < 0$, les deux valeurs propres sont différentes et négatives, a est un maximum.

Dans le cas où $\boldsymbol{dét}\big(\boldsymbol{Hf}(\boldsymbol{a})\big) < 0$, les deux valeurs propres sont différentes et ne sont pas de même signe, a est un point selle.

Dans le cas où $\boldsymbol{dét}\big(\boldsymbol{Hf}(\boldsymbol{a})\big) = \mathbf{0}$, on peut rien dire, il faut aller aux différentielles d'ordre supérieure à deux pour déterminer la nature de a.

Exemple 2.15. Etude des extremums des deux fonctions suivantes: la fonction $f(x,y) = x^3 + y^3 - 3mxy$ où $x, y \in \mathbb{R}^n$ et $m \in \mathbb{R}^*$ et la fonction $g(x,y) = 2x^2 + (x+1)y^2$ où $x, y \in \mathbb{R}^n$
Pour chercher les extremums de f.

$i)$ On cherche les points critiques de f, en résolvant l'équation $f'(x) = 0$, donc on résous le système

$$\begin{cases} \partial_1 f(x,y) = 0; \\ \partial_2 f(x,y) = 0. \end{cases}$$

Ce système est équivalent à

$$\begin{cases} x^2 - my = 0; \\ y^2 - mx = 0. \end{cases}$$

Les ponts critiques sont: $a = (0,0)$ et $b_m = (m, m)$.

$ii)$ On calcule $tra\big(Hf(a)\big)$ et $tra\big(Hf(b_m)\big)$. Puisque $\partial_{11} f(x,y) = 6x$, $\partial_{22} f(x,y) = 6y$ alors $\partial_{11} f(a) = \partial_{22} f(a) = 0$, comme $\partial_{12} f(a) = -m = \partial_{21} f(a)$, alors $dét\big(Hf(a)\big) = -m^2 < 0$, les deux valeurs propres sont non nulles et de signe différents, donc $a = (0,0)$ est un point selle. Puisque $\partial_{11} f(b_m) = \partial_{22} f(b_m) = 2m$, comme $\partial_{12} f(b_m) = -m = \partial_{21} f(b_m)$, alors $dét\big(Hf(v)\big) = 3m^2 > 0$. Si, $m > 0$ la $tra\big(Hf(b_m)\big) = 4\,m > 0$, les deux valeurs propres sont différentes et positives, $b_m = (m, m)$ est un minimum local. Si $m < 0$, la $tra\big(Hf(b_m)\big) = 4\,m < 0$ les deux valeurs propres distinctes sont négatives et b_m est un maximum local.

Pour chercher les extremums de g. Comme précédemment, on résous le système

$$\begin{cases} \partial_1 f(x,y) = 0; \\ \partial_2 f(x,y) = 0. \end{cases}$$

Donc

$$\begin{cases} 4x + y^2 = 0; \\ 2y(x+1) = 0. \end{cases}$$

Les points critiques sont $a = (0,0)$, $b = (-1,2)$ et $c = (-1,-2)$.

Comme $\partial_{11} f(x,y) = 4$, $\partial_{22} f(x,y) = 2(x+1)$ alors $\partial_{11} f(a) = 4$ et $\partial_{22} f(a) = 2$, et puisque $\partial_{12} f(x,y) = \partial_{21} f(x,y) = 2y$, alors $\partial_{12} f(a) = \partial_{21} f(a) = 0$, donc $dét\big(Hf(a)\big) = 8 > 0$, comme $tra\big(Hf(a)\big) = 6 > 0$, les deux valeurs propres sont non nulles, distinctes et elles sont positives donc $a = (0,0)$ est un minimum local. Pour le point $b = (-1,2)$, $\partial_{11} f(b) = 4$, $\partial_{22} f(b) = 0$, et $\partial_{12} f(b) = \partial_{21} f(b) = 4$, donc $dét\big(Hf(b)\big) = -16 < 0$, les deux valeurs propres sont non nulles, distinctes et elles sont de signe différents, b est un point selle. Pour le point $c = (-1,-2)$, $\partial_{11} f(c) = 4$, $\partial_{22} f(c) = 0$, et $\partial_{12} f(c) = \partial_{21} f(c) = -4$, donc $dét\big(Hf(c)\big) = -16 < 0$, c est un point selle.

Théorème 2.6. (Sylvester).

$a)$ q est définie positive ssi $A_i > 0$. pour tout $i \in \{1, \ldots, n\}$

$b)$ q est définie négative ssi $(-1)^i A_i > 0 \Leftrightarrow (A_1 < 0,\ A_2 > 0,\ A_3 < 0,\ A_4 > 0, \ldots)$, pour tout $i \in \{1, \ldots, n\}$.

Théorème 2.7. Si, au voisinage de a, une fonction f est de classe C^α, $\alpha \in \mathbb{N}^*$ et si, $f'(a) = \cdots = f^{(\alpha-1)}(a) = 0$ et $f^{(\alpha)}(a) \neq 0$. Alors, a est un extremum local strict ssi α est pair.

$a)$ si $d^\alpha f_a > 0$, a est un minimum local strict.

$b)$ si $d^\alpha f_a < 0$. a est un maximum local strict.

$c)$ si $d^\alpha f_a$ est alternée au voisinage de a. a n'est pas un extremum.

Preuve. Il suffit de faire la démonstration pour $\alpha = 2$. Pour démontrer $b)$ on utilise la même argumentation pour la démonstration de $a)$, il suffit donc de démontrer $a)$ Comme $df_a = 0$ car $\partial_i f(a) = 0$, $\forall i \in \{1, \ldots, n\}$, la formule de (TY) à l'ordre deux de f au point a s'écrit: $f(x) - f(a) = \frac{1}{2!}\sum_{i=1}^n \sum_{j=1}^n \partial_{ij} f(a) dx_j \sum_{i=1}^n \sum_{j=1}^n \partial_{ij} f(a) dx_j\, dx_i + \rho^2 \alpha(\rho)$, où $dx_i = x_i -$

a_i, $\forall i \in \{1,\dots,n\}$. Posons pour $i \in \{1,\dots,n\}$, $h_i = \frac{dx_i}{\rho}$, où $\rho = \|x-a\|$, alors $|h_i| = \frac{1}{\rho}|dx_i| \le 1$, $\forall i \in \{1,\dots,n\}$, et $\|h\|^2 = 1$, donc $f(x)-f(a) = \frac{\rho^2}{2!}\sum_{i=1}^n \sum_{j=1}^n \partial_{ij} f(a) k_j\, h_i + \rho^2\alpha(\rho) = \rho^2\left[\frac{1}{2!}d^2f_a(h)+\alpha(\rho)\right]$. Comme $d^2f_a > 0$ et elle est continue sur la boule unité fermé $\tilde{B}(0,1)$ qui est compact, d'après Weierstrass, $d^2f_a(h)$ atteint son minimum $m > 0$ sur $\tilde{B}(0,1)$ et comme $\alpha(\rho) \longrightarrow 0$ quand $\rho \longrightarrow 0$, il existe $\delta > 0$ tel que pour tout ρ vérifiant $0 < \|\rho\| < \delta$, on a $|\alpha(\rho)| < \frac{m}{2}$. En conclusion $d^2f_a(h)+\alpha(\rho) > \frac{1}{2}m - \frac{1}{2}m = 0$ et donc $f(x)-f(a) > 0$ pour tout $x \in B(a,\delta)$, a est donc un minimum local de f. c) Puisque d^2f_a est alternée, d'après la proposition 2.3, il existe h', $h'' \in \mathbb{R}^n$ dont $\|h'\| = \|h''\| = 1$, tel que $d^2f_a(h') > 0$ et $d^2f_a(h'') < 0$. Posons pour tout $\varepsilon > 0$ et pour tout $i \in \{1,\dots,n\}$, $dx_i' = \varepsilon h_i'$ et $dx_i'' = \varepsilon h_i''$, $\|x'-a\| = \|x''-a\| = \varepsilon$ i.e. $x', x'' \in \tilde{B}(a,\varepsilon)$. Appliquant (TY) à l'ordre deux de f au points (x',a) et (x'',a), on a: $f(x')-f(a) = \varepsilon^2\left[\frac{1}{2}d^2f_a(h')+\alpha(\varepsilon)\right]$ et $f(x'')-f(a) = \varepsilon^2\left[\frac{1}{2}d^2f_a(h'')+\alpha(\varepsilon)\right]$, comme $d^2f_a(h') > 0$ et $d^2f_a(h'') < 0$ et $\alpha(\varepsilon) \longrightarrow 0$ quand $\rho \longrightarrow 0$, on a $f(x')-f(a) \ge 0$ et $f(x'')-f(a) \le 0$. Conclusion $\forall\varepsilon > 0$, il existe x' et x'' dans $\tilde{B}(a,\varepsilon) \subset B(a,2\varepsilon)$ tel que $f(x') \ge f(a) \ge f(x'')$ donc a n'est pas un extremum de f.

Exemple 2. 16. Soit la fonction $(x,y) = x^3+y^3$, $(x,y) \in \mathbb{R}^2$.
Les dérivées partielles premières de $f(x,y)$ sont: $\partial_1 f(x,y) = 3x^2$ et $\partial_2 f(x,y) = 3y^2$, donc $(0,0)$ est un point critique de f.
Les dérivées partielles secondes de $f(x,y)$ sont: $\partial_{11}f(x,y) = 6x$, $\partial_{22}f(x,y) = 6y$ et $\partial_{12}f(x,y) = \partial_{21}f(x,y) = 0$
La différentielle seconde est: $d^2f(h,k) = h^2\partial_{11}f(x,y) + 2hk\partial_{12}f(x,y) + k^2\partial_{22}f(x,y) = 6xh^2+6yk^2$, donc $d^2f_{(0,0)}(h,k) = 0$ et $(0,0)$ n'est pas un extremum de f car pour tout $\varepsilon > 0$ les deux points $\left(0,\sqrt[3]{\varepsilon}\right)$ et $\left(0,-\sqrt[3]{\varepsilon}\right)$ dans $B(0,\varepsilon)$ de $\mathbb{R}^2$, vérifient $f\left(0,\sqrt[3]{\varepsilon}\right) = \varepsilon > 0$ et $f\left(0,-\sqrt[3]{\varepsilon}\right) = -\varepsilon < 0$.

Exemple 2. 17. Soit la fonction $(x,y) = x^4+y^4$, $(x,y) \in \mathbb{R}^2$.
Les dérivées partielles premières de $f(x,y)$ sont: $\partial_1 f(x,y) = 4x^3$ et $\partial_2 f(x,y) = 4y^3$, donc $(0,0)$ est un point critique de f.
Les dérivées partielles secondes de $f(x,y)$ sont: $\partial_{11}f(x,y) = 12x^2$, $\partial_{22}f(x,y) = 12y^2$ et $\partial_{12}f(x,y) = \partial_{21}f(x,y) = 0$
La différentielle seconde est: $d^2f(h,k) = h^2\partial_{11}f(x,y) + 2hk\partial_{12}f(x,y) + k^2\partial_{22}f(x,y) = 12x^2h^2+12y^2k^2$, donc $d^2f_{(0,0)}(h,k) = 0$ et $(0,0)$ est un minimum global de f car $f(x,y) \ge 0 = f(0,0)$, pour tout $(x,y) \in \mathbb{R}^2$.

Remarque 2.6. Les conditions $d^2f_a \ge 0$ (respectivement $d^2f_a < 0$) sont nécessaires. En effet, si par exemple a est un minimum et $d^2f_a < 0$, la fonction $F(t) = f(a+th)$ où t est voisin de 0 dans $\mathbb{R}$, admet 0 comme minimum local de F. Alors que $F''(0) = d^2f_a < 0$, contradiction.

■ En appliquant Sylvester, pour une fonction à deux variables on ait devant quatre cas:
Cas 1. Si, $A_1 = a_{11} > 0$ et $A_2 = a_{11}a_{22} - {a_{12}}^2 > 0$, f admet un minimum local strict et f est convexe au voisinage de a.
Cas 2. Si, $A_1 = a_{11} < 0$ et $A_2 = a_{11}a_{22} - {a_{12}}^2 > 0$, f admet un maximum local strict et $(-f)$ est convexe au voisinage de a.i.e. f est concave au voisinage de a.
Cas 3. Si, $A_2 = a_{11}a_{22} - {a_{12}}^2 < 0$, f n'a pas d'extremums locaux.
Cas 4. Si, $A_2 = 0$, f peut avoir ou ne peut pas avoir des extremums locaux.

Preuve. Les cas 1-2 sont assurés par Selvester. Conçernant la cas 3, supposons que, $A_2 = a_{11}a_{22} - {a_{12}}^2 < 0$ et $a_{11} \neq 0$. Démontrons que, $d^2 f_a$ est alternée. Soient $a = (\alpha, \beta)$, $dx = x - \alpha = \rho h$, $dy = y - \beta = \rho k$, où $\rho = \|(dx, dy)\|$, $a_{11} = \partial_{11} f(a)$, $a_{12} = \partial_{12} f(a)$ et $a_{22} = \partial_{22} f(a)$, alors $d^2 f_a(dx, dy) = dx^2 a_{11} + 2dxdya_{12} + dy^2 a_{22}$, donc $d^2 f_a(h,k) = \rho^2(h^2 a_{11} + 2hka_{12} + k^2 a_{22}) = \frac{\rho^2}{a_{11}}[(ha_{11} + ka_{12})^2 + (a_{11}a_{22} - a_{12}^2)k^2]$, donc pour $(h,k) = (1{,}0)$ on a $d^2 f_a(1{,}0) = \rho^2 a_{11}$ et pour $(h', k') = \left(\frac{a_{12}}{\sqrt{a_{11}^2 + a_{12}^2}}, \frac{-a_{11}}{\sqrt{a_{11}^2 + a_{12}^2}}\right)$,

$d^2 f_a\left(\frac{a_{12}}{\sqrt{a_{11}^2 + a_{12}^2}}, \frac{-a_{12}}{\sqrt{a_{11}^2 + a_{12}^2}}\right) = \rho^2 \frac{a_{11}}{\sqrt{a_{11}^2 + a_{12}^2}}(a_{11}a_{22} - a_{12}^2)$, quelque soit le signe de a_{11}, $d^2 f_a$ est alternative. Si maintenant $a_{11} = 0$, $-{a_{12}}^2 < 0$, donc $a_{12} \neq 0$ et $d^2 f_a(h,k) = \rho^2(2hka_{12} + k^2 a_{22})$, donc $d^2 f_a(0{,}1) = \rho^2 a_{22}$ et $d^2 f_a\left(\frac{a_{22}}{\sqrt{a_{22}^2 + a_{12}^2}}, \frac{-a_{12}}{\sqrt{a_{22}^2 + a_{12}^2}}\right) =$

$\rho^2\left(-2\frac{{a_{12}}^2 a_{22}}{a_{22}^2 + a_{12}^2} + \frac{{a_{12}}^2 a_{22}}{a_{22}^2 + a_{12}^2}\right) = -\rho^2 \frac{{a_{12}}^2}{a_{22}^2 + a_{12}^2} a_{22}$, quelque soit le signe de a_{22}, $d^2 f_a$ est alternative.

Exemple 2.18. Trouver les extremums de la fonction $f(x,y) = 3xy - x^3 - y^3$, $x, y \in \mathbb{R}$.
Les dérivées partielles premières de f sont: $\partial_1 f(x,y) = 3(y - x^2)$ et $\partial_2 f(x,y) = 3(x -$ *y2*, donc *0,0* et *1,1* sont deux points critiques de *f*.
Les dérivées partielles secondes de $f(x,y)$ sont: $\partial_{11} f(x,y) = -6x$, $\partial_{22} f(x,y) = -6y$ et $\partial_{12} f(x,y) = \partial_{21} f(x,y) = 3$
La différentielle seconde est: $d^2 f(h,k) = -6xh^2 + 6hk - 6yk^2$, donc $d^2 f_{(0,0)}(h,k) = 6hk$ elle est donc alternée et $(0{,}0)$, n'est pas un extremum de f et $d^2 f_{(1,1)}(h,k) = -6h^2 + 6hk - 6k^2 = -6(h^2 - hk + k^2) = -6[(h-k)^2 + hk] = -6(h-k)^2 - 6hk \leq -6(h-k)^2 + 3(h^2 + k^2) = -3(h^2 + k^2) + 6hk = -3(h-k)^2 \leq 0$. En utilisant Selvester, $A_1 = a_{11} = -6 < 0$ et $A_2 = a_{11}a_{22} - {a_{12}}^2 = 27 > 0$ alors $(1{,}1)$ est un maximum local de f. En calculant $tr\big(Hf(1{,}1)\big) = -12 < 0$, les deux valeurs propres sont différentes et négatives, donc $(1{,}1)$ est un maximum.

Exemple 2.19. Trouver les extremums de la fonction $f(x,y) = \sum_1^n m_i[(x - \alpha_i)^2 +$ *y−βi2, x,y∈ℝ, α=α1,…,αi,…,αn, β=β1,…,βi,…,βn* et *m1,…,mi,…,mn* sont fixe dans *ℝ+∗.*
On a $\partial_1 f(x,y) = 2\sum_1^n m_i(x - \alpha_i)$ et $\partial_2 f(x,y) = 2\sum_1^n m_i(y - \alpha_i)$ donc $a = \left(\frac{\sum_1^n m_i \alpha_i}{\sum_1^n m_i}, \frac{\sum_1^n m_i \beta_i}{\sum_1^n m_i}\right)$ est un point critique de f. Comme $a_{11} = \partial_{11} f(a) = \partial_{22} f(a) = a_{22} = 2\sum_1^n m_i$ et $\partial_{12} f(a) = \partial_{21} f(a) = a_{12} = 0$ alors $A_1 > 0$ et $A_2 = a_{11}a_{22} - {a_{12}}^2 = 4(\sum_1^n m_i)^2 > 0$, donc a est un minimum local de f.

■ Une partie non vide K dans $\mathbb{R}^n$ est convexe si pout tout $x, y \in \mathbb{R}^n$ et pour tout $t \in [0,1]$, $tx + (1-t)y \in K$. Une fonction f définie sur un convexe K est convexe sur K, si pout tout $x, y \in \mathbb{R}^n$ et pour tout $t \in [0,1]$, $f[x + t(y - x)] \leq f(x) + t[f(y) - f(x)]$, la fonction f est strictement convexe si l'inégalité précédente est stricte. Il est facile à voir que:

a) $\{\emptyset\}$, les singletons, les boules fermées, les boules ouvertes et les sous espaces vectoriels dans $\mathbb{R}^n$ sont des convexes.

b) L'intersection quelconque de convexes est un convexe.

c) Si $\{K_n\}$ est une suite décroissante de convexes alors, $\cup_{n\in\mathbb{N}} K_n$ est un convexe.

d) Si K et K' sont deux convexes, alors $K + K'$ et λK pour tout $\lambda \in \mathbb{R}$ sont des convexes.

e) Si K est convexe alors sa fermeture $\overline{K}$ est un convexe et $K + K = 2K$.

f) Si f est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^m$ elle est convexe et si K est un convexe dans $\mathbb{R}^n$, alors $f(K)$ est un convexe dans $\mathbb{R}^m$.

g) Une fonction convexe sur un convexe K est continue sur K.

Remarque 2.5.

a) a peut être un extremum ou pas, si $d^2f_a = 0$. Dans ce cas il faux aller aux différentielles d'ordre supérieur à 2 au point a quand elles existent.

b) Dans le cas d'une fonction à deux variables si pour tout $(x,y) \in \mathbb{R}^2$, $\left(\partial_{11}\partial_{22} - \partial122\, fx,y>0\right.$ on a un minimum global et f est convexe sur $\mathbb{R}2$ et si pour tout $x,y\in\mathbb{R}2$, $\left(\partial_{11}\partial_{22} - {\partial_{12}}^2\right) f(x,y) < 0$ on a un maximum global et f est concave sur $\mathbb{R}^2$

Exemple 2.20. Cherchons les extremums de la fonction $f(x,y) = x^2\left(\frac{1}{2}y + 1\right) + \frac{y^3}{3} - 4y$, $x, y \in \mathbb{R}$.

Les points critiques de f sont $(0,2)$ et $(0,-2)$ solutions du système:

$$\begin{cases} \partial_1 f(x,y) = 2x\left(1 + \dfrac{y}{2}\right) = 0; \\ \partial_2 f(x,y) = \dfrac{x^2}{2} + y^2 - 4 = 0. \end{cases}$$

Les dérivées seconde de f sont: $\partial_{11}f(x,y) = 2 + y$, $\partial_{22}f(x,y) = 2y$ et $\partial_{12}f(x,y) = \partial_{21}f(x,y) = x$.

Au point $a = (0,2)$, on a: $A_1 = a_{11} = \partial_{11}f(a) = 4$, $a_{22} = \partial_{22}f(a) = 4$, $a_{12} = \partial_{12}f(a) = 0$, donc $A_2 = a_{11}a_{22} - {a_{12}}^2 = 16 > 0$, a est un minimum local. En utilisant les valeurs propres, $tr(Hf(a) = 8 > 0$ les deux valeurs propres sont distinctes et positives, donc a est un minimum local.

Au point $b = (0,-2)$, on a: $A_1 = a_{11} = \partial_{11}f(b) = 0$, $a_{22} = \partial_{22}f(b) = -4$, $a_{12} = \partial_{12}f(b) = 0$, donc $A_2 = a_{11}a_{22} - {a_{12}}^2 = 0$. Dans ce cas $d^2f_b(h,k) = 0$, et b est un point selle de f.

Lemme 2.1. Si, K est un convexe dans $\mathbb{R}^n$ et $f{:}K \longrightarrow \mathbb{R}$ est une fonction deux fois différentiable sur K . Alors:

a) f est convexe ssi $d^2f \geq 0$ sur K

b) Si $d^2f > 0$, alors f est strictement convexe. La réciproque est fausse.

Preuve. Il suffit de démontrer que si $d^2f \geq 0$, la fonction $F{:}\,[0,1] \longrightarrow \mathbb{R}$ définie par $F(t) = f[x + t(y - x)] - f(x) - t[f(y) - f(x)] \leq 0$, pour tout $t \in [0,1]$ et tout $x, y \in \mathbb{R}^n$. Il est clair que $F(0) = F(1) = 0$ et que $F''(t) = d^2f\big(x + t(y - x)\big) \geq 0$ pour tout $t \in [0,1]$ et tout $x, y \in \mathbb{R}^n$. Si, on suppose qu'il existe $s \in [0,1]$ tel que $F(s) > 0$ on arrive à une contradiction. En effet, puisque F est continue sur le compact $[0,1]$, d'après Weierstrass il existe t_0 dans $]0,1[$ tel que $F(t_0) \geq F(t)$ pour tout $t \in [0,1]$, donc $F'(t_0) = 0$. Comme $F'(t)$ est croissante sur $[0,1]$, F' est aussi croissante sur $[t_0, 1]$ alors $0 = F'(t_0) \leq F'(t)$, pour tout $t \in [t_0, 1]$ donc F est croissante sur $[t_0, 1]$, d'où $0 < F(s) \leq F(t_0) \leq F(1) = 0$, contradiction. La démonstration du cas où $d^2f > 0$ se fait de la même façon.

Exemple 2.21.Etudier la nature des extremums de la fonction $f(x,y) = x^2 + y^2$, $x, y \in \mathbb{R}$.

Résolution du système

$$\begin{cases} \partial_1 f(x,y) = 2x = 0; \\ \partial_2 f(x,y) = 2y = 0. \end{cases}$$

Le point $a = (0,0)$ est le seul point critique de la fonction f. Les dérivées seconde de f sont: $\partial_{11}f(x,y) = 2$, $\partial_{22}f(x,y) = 2$ et $\partial_{12}f(x,y) = \partial_{21}f(x,y) = 0$. Alors, $A_1 = a_{11} = \partial_{11}f(a) = 2$, $a_{22} = \partial_{22}f(a) = 2$, $a_{12} = \partial_{12}f(a) = 0$, donc $A_2 = a_{11}a_{22} - {a_{12}}^2 = 4 > 0$, a est un minimum local de f. Comme $d^2f(h,k) = (h,k)\begin{pmatrix} \partial_{11}f(x,y) & \partial_{12}f(x,y) \\ \partial_{12}f(x,y) & \partial_{22}f(x,y) \end{pmatrix}\begin{pmatrix} h \\ k \end{pmatrix} =$

$(h,k)\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}h\\ k\end{pmatrix} = 2(h,k)\begin{pmatrix}h\\ k\end{pmatrix} = 2(h^2+k^2) > 0$ pour tout $(h,k) \in \mathbb{R}^2$ et $(h,k) \neq (0,0)$ la fonction est strictement convexe et a est minimum globale.

3-Fonctions fortement convexes et fonctions elliptiques

3.1-Fonctions fortement convexes

Nous allons maintenant introduire une classe de fonctions très importante en optimisation.

Définition 3.1. Une fonction $f\colon K \longrightarrow \mathbb{R}$ est dite fortement convexe sur K, de paramètre $m \in \mathbb{R}_+^*$ ssi, $f[tx+(1-t)y] \leq tf(x) + (1-t)\, f(y) - \frac{m}{2}t(1-t)\|x-y\|^2$, pour tout $x, y \in K$ et pour tout $t \in [0,1]$.

Proposition 3.1. Une fonction f est fortement convexe sur K, de paramètre $m \in \mathbb{R}_+^*$ ssi la fonction $g(x) = f(x) - \frac{m}{2}\|x\|^2$ est convexe sur K.

Preuve. Pour tout $x, y \in K$ et pour tout $t \in [0,1]$, $g([tx+(1-t)y]) = f([tx+$ *1−ty−m2tx+1−ty2≤tfx+1−t* $f(y) - \frac{m}{2}[t(1-t)(\|x\|^2 - 2\langle x,y\rangle + \|y\|^2) + t^2\|x\|^2 + 2t(1-t)\langle x,y\rangle + (1-t)^2\|y\|^2] = tf(x) + (1-t)\, f(y) - \frac{m}{2}[t(1-t)\|x\|^2 - 2t(1-t)\langle x,y\rangle + t(1-t)\|y\|^2 + t^2\|x\|^2 +$ *2t1−tx,y+1−t2y2=tfx+1−t fy−m2t1−tx2+t1−ty2+t2x2+1−t2y2=tfx+1−t* $f(y) - g([tx+(1-t)y])[t\|x\|^2 + (1-t)\|y\|^2] = t\left[f(x) - \frac{m}{2}\|x\|^2\right] + (1-t)\left[f(y) -\right.$ *m2y2=tgx+1−tgy*, donc *g* est convexe. Réciproquement si, pour tout *x,y∈K* et pour tout $t \in [0,1]$, $g([tx+(1-t)y]) \leq tg(x) + (1-t)g(y)$, alors pour tout $x, y \in K$ et pour tout $t \in [0,1]$, $f([tx+(1-t)y]) - \frac{m}{2}\|tx+(1-t)y\|^2 \leq t\left[f(x) - \frac{m}{2}\|x\|^2\right] + (1-t)\left[f(y) -\right.$ *m2y2*. Donc, pour tout *x,y∈K* et pour tout *t∈0,1*, *ftx+1−ty≤tfx+1−t* $f(y) - \frac{m}{2}\|tx+(1-t)y\|^2 - \frac{m}{2}[t\|x\|^2 + (1-t)\|y\|^2] = tf(x) + (1-t)\, f(y) - \frac{m}{2}[t^2\|x\|^2 + 2t(1-t)\langle x,y\rangle + (1-t)^2\|y\|^2 + t\|x\|^2 + (1-t)\|y\|^2] = tf(x) + (1-t)\, f(y) - \frac{m}{2}[t(t+1)\|x\|^2 + 2t(1-t)\langle x,y\rangle + [(2-t)(1-t)]\|y\|^2] \leq tf(x) + (1-t)\, f(y) - \frac{m}{2}t(t-1)[\|x\|^2 + 2\langle x,y\rangle + \|y\|^2] \leq tf(x) + (1-t)\, f(y) - \frac{m}{2}t(t-$ *1x2−2x,y+y2= tfx+1−t fy−m2tt−1x−y2.*

Proposition 3.2. Si, f est une fonction fortement convexe sur K de paramètre $m \in \mathbb{R}_+^*$ et g est une fonction convexe sur K. Alors, la fonction $h = f + g$ est fortement convexe de paramètre $m \in \mathbb{R}_+^*$.

Preuve. Pour tout $x, y \in K$ et pour tout $t \in [0,1]$, $h([tx+(1-t)y]) = f([tx+$ *1−ty+gtx+1−ty≤tfx+1−t fy+tgx+1−t* $g(y) - \frac{m}{2}t(1-t)\|x-y\|^2 = t[f(x)+g(x)] + (1-t)[f(y)+g(y)] - \frac{m}{2}t(1-$ *tx−y2=thx+1−thy−m2t1−tx−y2.*

3.2-Fonctions elliptiques

Définition 3.2. Une fonction f de casse $C^1(\mathbb{R}^n)$ est dite elliptique de paramètre $\gamma \in \mathbb{R}_+^*$ si $\langle f'(y) - f'(x), y - x\rangle \geq \gamma\|y-x\|^2$, pour tout $x, y \in \mathbb{R}^n$ (γ est appelée la constante d'ellipticité de f).

Proposition 3.3. Une fonction f de casse $C^1(\mathbb{R}^n)$ est fortement convexe de paramètre $m \in \mathbb{R}_+^*$, ssi

$a)$ elle est elliptique de paramètre $m \in \mathbb{R}_+^*$.

$b)$ $f(x) \geq f(y) + \langle f'(y), x - y\rangle + \frac{m}{2}\|x-y\|^2$, pour tout , $y \in \mathbb{R}^n$, $(*)$.

Preuve. $a)$ La fonction $g(x) = f(x) - \frac{m}{2}\|x\|^2$ est convexe sur $\mathbb{R}^n$, ssi $\langle g'(x) - g'(y), x - y\geq 0$, pour tout $x,y\in\mathbb{R}n$ ssi $f'x-f'y+my-x,x-y\geq 0$, pour tout $x,y\in\mathbb{R}n$ ssi $\langle f'(x) - f'(y), x - y\rangle \geq m\langle x-y, x-y\rangle = m\|x-y\|^2$, pour tout $x, y \in \mathbb{R}^n$. On utilise l'inégalité de convexité $g(x) \geq \langle g'(y), x - y\rangle + g(y)$ pour tout $x, y \in \mathbb{R}^n$, pour démontrer que $\langle g'(x) - g'(y), x - y\rangle \geq 0$, pour tout $x, y \in \mathbb{R}^n$. $b)$, $g(x) \geq \langle g'(y), x - y\rangle + g(y)$ pour tout $x, y \in \mathbb{R}^n$ ssi $f(x) - \frac{m}{2}\|x\|^2 \geq \langle f'(y) - my, x - y\rangle + f(y) - \frac{m}{2}\|y\|^2 = \langle f'(y), x - y-my,x-y+fy-m2y2$, pour tout $x,y\in\mathbb{R}n$ ssi $f(x) \geq \langle f'(y), x - y\rangle + f(y) - m\langle y, x\rangle + \frac{m}{2}\|x\|^2 + \frac{m}{2}\|y\|^2 = \langle f'(y), x - y\rangle + f(y) + \frac{m}{2}(-2\langle y, x\rangle + \|x\|^2 + \|y\|^2) = \langle f'(y), x - y\rangle + f(y) + \frac{m}{2}\|x-y\|^2$, pour tout $x, y \in \mathbb{R}^n$.

Théorème 3.1. Si f est une fonction elliptique sur $\mathbb{R}^n$ de paramètre $m \in \mathbb{R}_+^*$. Alors:

$a)$ $m\|x-y\| \leq \|f'(x) - f'(y)\|$, pour tout, $x, y \in \mathbb{R}^n$.

$b)$ $f(y) - f(x) \geq \langle f'(x), y - x\rangle$, pour tout, $x, y \in \mathbb{R}^n$ (inégalité de convexité).

$c)$ f est strictement convexe et coercive.

$d)$ f admet un minimum unique sur $\mathbb{R}^n$.

Preuve. $a)$ de $(*)$ dans la proposition 3.3 $b)$ et de l'inégalité de Cauchy Schwarz, on a $0 \geq \langle f'(y) - f'(x), x - y\rangle + m\|x-y\|^2$, pour tout $x, y \in \mathbb{R}^n$, donc $m\|x-y\|^2 \leq \langle f'(x) - f'(y), x - y\rangle \leq \|f'(x) - f'(y)\|\|x-y\|$ pour tout $x, y \in \mathbb{R}^n$, d'où $m\|x-y\| \leq \|f'(x) - f'(y)\|$ pour tout $x, y \in \mathbb{R}^n$. $b)$ Il est clair que $f(y) - f(x) \geq \langle f'(x), y - x\rangle$ pour tout $x, y \in \mathbb{R}^n$ et donc $f(y) \geq f\big(y + t(x-y)\big) - t\langle f'\big(y + t(x-y)\big), y - x\rangle$ et $f(x) \geq f\big(y + t(x-y)\big) - (1-t)\langle f'\big(y + t(x-y)\big), y - x\rangle$ pour tout $t \in [0,1]$ et pour tout $x, y \in \mathbb{R}^n$, en multipliant la première inégalité par $(1-t)$ et la deuxième par t et en sommant les deux inégalité on a l'inégalité de convexité $(1-t)f(y) + tf(x) \geq f\big(y + t(x-y)\big)$ pour tout $t \in [0,1]$ et pour tout $x, y \in \mathbb{R}^n$ et on a une inégalité de convexité stricte de f si $x \neq y$ se qui donne $c)$. Puisque $f(y) - f(x) \geq \langle f'(x), y - x\rangle + \frac{m}{2}\|x-y\|^2$, pour tout $x, y \in \mathbb{R}^n$, par l'inégalité de Cauchy Schwarz et $(*)$ on a $f(y) \geq f(0) - \|f'(0)\|\|y\| + \frac{m}{2}\|y\|^2 = f(0) + \|y\|\left[\frac{m}{2}\|y\| - \|f'(0)\|\right]$ pour tout $y \in \mathbb{R}^n$ et donc f est coercive. $d)$ Puisque f est strictement convexe, continue et elle est coercive sur $\mathbb{R}^n$, le problème

$$(\mathcal{P})\ \inf_{x\in\mathbb{R}^n} f(x)$$

admet une solution unique, se qui achève la démonstration du théorème.

Proposition 3.4. Une fonction f de casse $C^2(\mathbb{R}^n)$ est elliptique de paramètre $m \in \mathbb{R}_+^*$ ssi $\langle f''(x)h, h\rangle = d^2f(h) \geq m\|h\|^2$ pour tout $x, h \in \mathbb{R}^n$.

Preuve. Puisque f est elliptique de paramètre $m \in \mathbb{R}_+^*$. D'après $(*)$ dans la proposition 3.3 $b)$ $f(y) - f(x) \geq \langle f'(x), y - x\rangle + \frac{m}{2}\|x-y\|^2$, pour tout $x, y \in \mathbb{R}^n$, donc $f(x+th) - f(x) \geq \langle f'(x), th\rangle + \frac{m}{2}t^2\|h\|^2$ et $f(x) - f(x+th) \geq \langle f'(x+th), -th\rangle + \frac{m}{2}t^2\|h\|^2$ pour tout $x, h \in \mathbb{R}^n$ et $t > 0$, donc $0 \geq \langle f'(x), th\rangle - \langle f'(x+th), th\rangle + mt^2\|h\|^2$ pour tout $x, h \in \mathbb{R}^n$ et $t > 0$, d'où $\frac{\langle f'(x+th), h\rangle - \langle f'(x), h\rangle}{t} \geq m\|h\|^2$ pour tout $x, h \in \mathbb{R}^n$ et $t > 0$, comme pour tout $x, h \in \mathbb{R}^n$ $f''(x)h = \lim_{t\to 0}\frac{f'(x+th) - f'(x)}{t}$, donc $d^2f(h) = \langle f''(x)h, h\rangle = \lim_{t\to 0}\frac{\langle f'(x+th) - f'(x), h\rangle}{t} \geq m\|h\|^2$ pour tout $x, h \in \mathbb{R}^n$.

Réciproquement, appliquant (TL) et au point $x \in \mathbb{R}^n$ et à l'ordre 1, pour la fonction $g(t) = \langle f'(t), y - x\rangle$ pour tout $t, x, y \in \mathbb{R}^n$. Alors, $g(t) = g(x) + \langle g'\big(x + \theta(t - x)\big), t - x\rangle$, donc $\langle f'(t), y - x\rangle = \langle f'(x), y - x\rangle + \langle f''\big(x + \theta(t - x)\big)(t - x), t - x\rangle$ pour $t = y$ on a $\langle f'(y), y - x\rangle - \langle f'(x), y - x\rangle = \langle f''\big(x + \theta(y - x)\big)(y - x), y - x\rangle \geq m\|y - x\|^2$, pour tout $x, y \in \mathbb{R}^n$ et donc $\langle f'(y) - f'(x), y - x\rangle \geq m\|y - x\|^2$, pour tout $x, y \in \mathbb{R}^n$ et f est elliptique.

De l'ellipticité, l'inégalité de Cauchy-Schawarz et la proposition 3.5, on a:

Corollaire 3.1. Une fonction $f \in C^2(K)$ est fortement convexe sur K, ssi il existe $k_1, k_2 \in \mathbb{R}_+^*$ tel que $k_1\|h\|^2 \leq d^2f(h) \leq k_2\|h\|^2$ pour tout $h \in K$.

Exemple 3.1. Soit la fonction f défine de $\mathbb{R}^n$ dans $\mathbb{R}$ par: $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ où A est une matrice carrée d'ordre n, symétrique ($A^T = A$, A^Test la matrice transposée de A) et $b \in \mathbb{R}^n$ est une constante. f est dite une fonction quadratique elliptique. Alors: pour tout $x, h \in \mathbb{R}^n$

$i)$ $f'(x) = Ax - b$ et $f''(x) = A$, $df(h) = \langle f'(x), h\rangle = \langle Ax - b, h\rangle = \langle Ax, h\rangle - \langle b, h\rangle$ et $d^2f(h) = \langle f''(x)h, h\rangle = \langle Ah, h\rangle$. En effet, $f(x + h) = \frac{1}{2}\langle A(x + h), x + h\rangle - \langle b, x + h\rangle = \frac{1}{2}[\langle Ax + Ah, x + h\rangle] - \langle b, x\rangle - \langle b, h\rangle = \frac{1}{2}[\langle Ax, x\rangle + \langle Ax, h\rangle + \langle Ah, x\rangle + \langle Ah, h\rangle] - \langle b, x\rangle - \langle b, h\rangle$, comme $A^t = A$ donc $\langle Ah, x\rangle = \langle h, A^t x\rangle = \langle h, Ax\rangle = \langle Ax, h\rangle$. Alors $f(x + h) = f(x) + [\langle Ax, h\rangle - \langle b, h\rangle] + \frac{1}{2}\langle Ah, h\rangle$, d'où $f(x + h) - f(x) = [\langle Ax, h\rangle - \langle b, h\rangle] + \left(\frac{1}{2\|h\|}\langle Ah, h\rangle\right)\|h\|$ avec $0 \leq |\alpha(h)| = \frac{1}{2\|h\|}|Ah, h| \leq \frac{1}{2\|h\|}\|Ah\|\|h\| = \frac{1}{2}\|Ah\|$, quand $h \longrightarrow 0$, $\alpha(h) \longrightarrow 0$, donc $df(h) = \langle Ax, h\rangle - \langle b, h\rangle = \langle Ax - b, h\rangle = \langle f'(x), h\rangle$ ou bien $\langle f'(x) - Ax-b,h=0$, d'où $f'x=Ax-b$ et donc $f''x=A$ et $d2fh=f''xh,h=Ah,h$.

$ii)$ f est elliptique de paramètre $m \in \mathbb{R}_+^*$. En effet, puisque A est symétrique, il existe une base de vecteurs propres orthonormale $\{u_1, \ldots, u_n\}$ dont les valeurs propres associées, $\{\lambda_1, \ldots, \lambda_n\}$ sont des éléments de $\mathbb{R}_+^*$ tel que $d^2f(h) = \langle f''(x)h, h\rangle = \langle Ah, h\rangle = \sum_1^n \lambda_i h_i h_j \langle u_i, u_j\rangle = \sum_1^n \lambda_i {h_i}^2 \geq m\|h\|^2$, pour tout $h \in \mathbb{R}^n$ où $m = \min_{1\leq i\leq n} \lambda_i$.

$iii)$ La dérivée f' de la fonction f est M −Lipschitzienne, plus précisément, il existe $M = \max_{1\leq i\leq n} \lambda_i > 0$, ($\{\lambda_1, \ldots, \lambda_n\}$ sont comme dans $ii)$) tel que $\|f'(x) - f'(y)\| \leq M\|x - y\|$ pour tout $x, y \in \mathbb{R}^n$. En effet, pour tout, $x, y \in \mathbb{R}^n$ $f'(x) - f'(y) = Ax - Ay = A(x - y) = \sum_1^n \lambda_i (x_i - y_i)u_i$, donc $\|f'(x) - f'(y)\| = \|\sum_1^n \lambda_i (x_i - y_i)u_i\| \leq \max_{1\leq i\leq n} \lambda_i (\sum_1^n |x_i - y_i|)\|u_i\| = M\|x - y\|$ où $\|u_i\| = 1, \forall i \in \{1, \ldots, n\}$ et $M = \max_{1\leq i\leq n} \lambda_i$.

Exemple 3.2. Puisque, la fonction $g(x) = \|x\|^2 = \langle x, x\rangle$ sur le convexe K est elliptique si, f est une fonction convexe sur K. Pour tout $\lambda \in \mathbb{R}_+^*$, la fonction $h(x) = f + \lambda\|x\|^2$ est fortement convexe sur K.

3.3-Résultats d'existence d'un minimum

Définition 3.3. Un point x_0 de K est un minimum local de f s'il existe un $\delta > 0$ tel que $f(x_0) \leq f(x)$ pour tout $x \in B(x_0, \delta) \cap K = B_K(x_0, \delta)$ (la boule ouverte de centre x_0 et de rayon δ dans le sous espace K).

Lemme 3.1. Pour une fonction f, convexe et différentiable sur un convexe K. Un point x_0 de K est un minimum local de f ssi $\langle \nabla f(x_0), h\rangle \geq 0$ pour tout h vérifiant $x_0 + h \in K$.

Preuve. Puisque f est différentiable au point x_0, d'après la proposition 2.1 $b)$, pour tout $h = x - x_0 \neq 0$ avec $x \in K$, la dérivée de f dans la direction du vecteur unitaire $u = \frac{h}{\|h\|}$, s'écrit $\partial_u f(x_0) = df_{x_0}(u) = \sum_1^n \partial_i f(x_0)u_i = \sum_1^n \partial_i f(x_0)\frac{h_i}{\|h\|} = \frac{1}{\|h\|}(\sum_1^n \partial_i f(x_0)h_i) = \frac{1}{\|h\|}df_{x_0}(h) = \frac{1}{\|h\|}\langle\nabla f(x_0), h\rangle$, donc $\partial_u f(x_0)\|h\| = \langle\nabla f(x_0), h\rangle = df_{x_0}(h)$. Comme x_0 est un minimum local de f, il existe $r > 0$ tel que, $f(x) \geq f(x_0)$ pour tout $x \in B_K = B(x_0, r) \cap K$ où $B(x_0, r)$ est la boule ouverte dans $\mathbb{R}^n$ de centre x_0 et de rayon r. Comme, pour tout $0 < t < r$, $x_0 + tu \in B_K$, en effet $x_0 + tu \in B(x_0, r)$ et $x_0 + tu = x_0 + \frac{t}{\|h\|}(x - x_0) = \frac{t}{\|h\|}x + \left(1 - \frac{t}{\|h\|}\right)x_0 \in K$. Alors $f(x_0 + tu) \geq f(x_0)$ et $\frac{f(x_0+tu)-f(x_0)}{t} \geq 0$ donc $\partial_u f(x_0) \geq 0$ ($\partial_u f(x_0) = 0$, si x_0 est un point intérieur de K et $\partial_u f(x_0) > 0$, si x_0 est un point frontière de K), d'où $\langle\nabla f(x_0), h\rangle \geq 0$ pour tout h vérifiant $x_0 + h \in K$. Inversement, puisque f est convexe, pour x_0 et $x_0 + h$ dans K et pout tout $t \in]0,1]$, on a $f(x_0 + th) \leq tf(x_0) + (1-t)f(x_0 + h)$, donc $f(x_0 + h) - f(x_0) \geq \frac{f(x_0+th)-f(x_0)}{t}$ et quand $t \longrightarrow 0^+$ $f(x_0 + h) - f(x_0) \geq \partial_h f(x_0) = \langle\nabla f(x_0), h\rangle \geq 0$, donc x_0 est un minimum local de f.

Remarque 3.1.

a) L'inégalité dans la lemme 3.1 est une égalité si x_0 est un point intérieur de K.

b) La convexité de f n'intervient pas dans " $\Longrightarrow$ ".

c) Si f est strictement convexe, x_0 est unique.

Théorème 3.2. Si, K est un convexe fermé de $\mathbb{R}^n$ et $f \in C^2(K)$ fortement convexe de paramètre $m \in \mathbb{R}_+^*$. Alors, f admet un minimum local unique sur K.

Preuve. Soit $x_1$un point intérieur de K, puisque f est deux fois différentiable sur K, la formule de (TL) au point x_1 s'écrit $f(x) = f(x_1) + df_{x_1} + \frac{1}{2}d^2 f_{[x_1+\theta(x-x_1)]}$, où $\theta \in]0.1[$, puisque K est convexe $x_1 + \theta(x - x_1) \in K$. Posons $h = x - x_1$, alors $df_{x_1} = \langle\nabla f(x_1), h\rangle$, donc $|df_{x_1}| = |\langle\nabla f(x_1), h\rangle| \leq \|\nabla f(x_1)\|\|h\|$, comme f est fortement convexe de paramètre $m \in \mathbb{R}_+^*$, d'après la proposition 3.4, $d^2 f_{[x_1+\theta(x-x_1)]} \geq m\|h\|^2$ et d'après $(*)$ dans la proposition 3.3 b) et l'inégalité de Cauchy schwarz $f(x) - f(x_1) \geq -\|\nabla f(x_1)\|\|h\| + \frac{m}{2}\|h\|^2 = \|h\|\left[\frac{m}{2}\|h\| - \|\nabla f(x_1)\|\right] = \left(\|h\|\frac{m}{2}\right)\left[\|h\| - \frac{2}{m}\|\nabla f(x_1)\|\right]$, pour avoir $f(x) - f(x_1) > 0$, il suffit d'avoir un $r \in \mathbb{R}_+^*$, vérifiant $\|h\| > r \geq \frac{2}{m}\|\nabla f(x_1)\|$. Donc, x_1 est un minimum local de f dans l'extérieur de la boule fermée $\tilde{B}(x_1, r)$. Soit $K_r = \mathrm{K} \cap \tilde{B}(x_1, r)$, c'est un convexe fermé et borné dans $\mathbb{R}^n$, donc c'est un compact, comme f est continue sur K_r, elle attient son seul minimum local x_0 dans K_r, donc $f(x_0) < f(x_1)$. Comme $f(x_1) < f(x)$, pour tout x dans l'extérieur de la boule fermée $\tilde{B}(x_1, r)$, il s'en suit que $f(x_0) < f(x)$ pour tout x dans K (qui est la réunion de $\tilde{B}(x_1, r)$ et son extérieur).

4- La projection orthogonale, méthode du gradient projeté

4.1-La projection orthogonale

Si x est un point du plan π, sa projection orthogonale sur une droite D de ce plan est un élément x^* de D tel que: $|x - x^*| \leq |x - y|$, pour tout $y \in D$. Autrement dit, x^* est solution du problème de minimisation $\inf_{y \in D} f(y)$ où $f(y) = |x - y|$ avec $y \in D$. Cette idée se généralise à l'espace $\mathbb{R}^n$, en remplaçant la valeur absolu dans $\mathbb{R}$ par la norme dans $\mathbb{R}^n$. Avant de donner le théorème de la projection orthogonale aussi appelé le **théorème de la meilleure approximation**. Rappelons que:

Si $\{F_n\}$ est une suite décroissantes de fermés dans $\mathbb{R}^n$ tel que le diamètre de F_n, $\delta(F_n) \longrightarrow 0$, alors $\bigcap_{n \in \mathbb{N}^*} F_n = \{a\}$ (l'intersection est réduite à un seul point).

Théorème 4.1. (Théorème de la projection orthogonale). Si $K \subset \mathbb{R}^n$ est un convexe ferme, pour tout $x \in \mathbb{R}^n$, il existe un unique $x^* \in K$ solution du problème de minimisation

$$\inf_{y\in K} f(y) \text{ où } f(y) = \|x-y\|,\ y \in K.$$

Autrement dit, $\|x-x^*\| \leq \|x-y\|,\ \forall y \in K$.

Preuve.

Si $x \in K$, $x^* = x$, car $\|x-x^*\| = 0 \leq \|x-y\|,\ \forall y \in K$. Supposons que $x \notin K$ et $a = \inf_{y\in K}\|x-y\|$, alors pour tout $n \in \mathbb{N}^*$, il existe une suite $\{y_n\}$ dans K, tel que $\|x-y_n\| < a+\frac{1}{n}$, donc $y_n \in \tilde{B}\left(x, a+\frac{1}{n}\right)$, la suite $\left\{\tilde{B}\left(x, a+\frac{1}{n}\right)\right\}$ est une suite décroissante de boules convexes et fermés dans $\mathbb{R}^n$. Soit $K_n = K \cap \tilde{B}\left(x, a+\frac{1}{n}\right)$, pour tout $n \in \mathbb{N}^*$, il es clair que pour tout $n \in \mathbb{N}^*$, $K_n \neq \emptyset$, $\{K_n\}$ est une suite décroissante de convexes fermés. Démontrons que $\delta(K_n) \longrightarrow 0$. Soient $u, v \in K_n$ d'après l'identité du parallélogramme on a $\|(x-u)+$ *x−v2+x−u−x−v2=2x−u2+x−v2*, donc *4x−u+v22+v−u2=2x−u2+x−v2* ou bien $\|v-u\|^2 = 2[\|x-u\|^2+\|x-v\|^2] - 4\left\|x-\frac{u+v}{2}\right\|^2$ pour tout $u, v \in K_n$, comme $\frac{u+v}{2} \in K_n$ qui est convexe, alors $0 \leq \|v-u\|^2 \leq 4\left(a+\frac{1}{n}\right)^2$ pour tout $u, v \in K_n$, quand $n \longrightarrow \infty$, $\delta(K_n) \longrightarrow 0$. En déduit que $\bigcap_{n\in\mathbb{N}^*} K_n = \{x^*\} \subset K$ et que $\|x-x^*\| \leq a+\frac{1}{n}$, quand $n \longrightarrow \infty$ $\|x-x^*\| \leq a \leq \|x-y\|\ \forall y \in K$. Evidement x^* est unique.

Lemme 4.1 (Caractérisation de la projection)**.** Si $\mathrm{K} \subset \mathbb{R}^{\mathrm{n}}$ est un convexe ferme. Alors $x^* \in K$ est une projection orthogonal d'un élément $x \in \mathbb{R}^n$ ssi $\langle x-x^*, y-x^*\rangle \leq 0,\ \forall y \in K$.

Preuve. Si $x^* \in K$ est une projection orthogonal d'un élément $x \in \mathbb{R}^n$, alors $\|x-x^*\| \leq \|x-y\|\ \forall y \in K$, comme pour tout $t \in]0,1]$, $x^*+t(y-x^*) \in K$, donc $\|x-x^*\|^2 \leq \|x-x^*-t(y-x^*)\|^2 = \langle x-x^*-t(y-x^*), x-x^*-t(y-x^*)\rangle = \|x-x^*\|^2 - 2t\langle x-x^*, y-x^*\rangle + t^2\|y-x^*\|^2$, $\forall y \in K$, donc $0 \leq -2\langle x-x^*, y-x^*\rangle + t\|y-x^*\|^2$, $\forall y \in K$ quand $t \longrightarrow 0$, on a $\langle x-x^*, y-x^*\rangle \leq 0\ \forall y \in K$. Supposons maintenant que $\langle x-x^*, y-x^*\rangle \leq 0\ \forall y \in K$, alors $\|x-x^*\|^2 - \|x-y\|^2 = \langle x-x^*, x-x^*\rangle - \langle x-y, x-y\rangle = \|x\|^2 - 2\langle x, x^*\rangle + \|x^*\|^2 - \|x\|^2 + 2\langle x, y\rangle - \|y\|^2 = 2[\langle x, y\rangle - \langle x, x^*\rangle] + \|x^*\|^2 - \|y\|^2 = 2[\langle x, y-x^*\rangle + \langle -x^*, y-x^*\rangle + \langle x^*, y-x^*\rangle] + \|x^*\|^2 - \|y\|^2 = 2[\langle x-x^*, y-x^*\rangle + \langle x^*, y-x^*\rangle] + \|x^*\|^2 - \|y\|^2 = 2\langle x-x^*, y-x^*\rangle - \|x^*-y\|^2 \leq 0$, $\forall y \in K$, donc $\|x-x^*\| \leq \|x-y\|\ \forall y \in K$.

Corollaire 4.1. Dans le as où K est un sous espace vectoriel fermé. Alors, $x^* \in K$ est une projection orthogonal d'un élément $x \in \mathbb{R}^n$ ssi $\langle x-x^*, y\rangle = 0,\ \forall y \in K$.

Preuve. Puisque K est un sous espace vectoriel fermé, donc c'est un convexe fermé d'après le lemme 4.1, $x^* \in K$ est une projection orthogonal d'un élément $x \in \mathbb{R}^n$ ssi $\langle x-x^*, y-x^*\rangle \leq 0$, $\forall y \in K$, donc pour tout $y \in K$ et pour tout $\lambda \in \mathbb{R}$, $\langle x-x^*, (\lambda y+x^*)-x^*\rangle = \lambda\langle x-x^*, y\rangle \leq 0$ d'où pour $\lambda = \pm 1$, $\langle x-x^*, y\rangle = 0$, pour tout $y \in K$.

Corollaire 4.2. Soit K un convexe fermé dans $\mathbb{R}^n$, l'application projection $P_K: \mathbb{R}^n \longrightarrow K$ est 1-Lipschienne.

Preuve. Soient , $x' \in \mathbb{R}^n$ $(x \neq x')$, $P_K(x) = x^*$ et $P_K(x') = x^{**}$. Alors $\langle x-x^*, x^{**}-x^*\rangle \leq 0$ et $\langle x'-x^{**}, x^*-x^{**}\rangle \leq 0$ donc $\langle x^*-x, x^*-x^{**}\rangle + \langle x'-x^{**}, x^*-x^{**}\rangle = \langle x^*-x^{**}+$ *x′−x, x∗−x∗∗=x∗−x∗∗2+x′−x, x∗−x∗∗≤0*, donc *x∗−x∗∗2≤x−x′, x∗−x∗∗≤x−x′x∗−x∗∗*, alors $\|P_K(x)-P_K(x')\| \leq \|x-x'\|$, pour tout $x, x' \in \mathbb{R}^n$.

Remarque 4.1. Dans le cas où K est un sous espace vectoriel fermé de $\mathbb{R}^n$, l'application P_K est linéaire. En effet, pour tout $\lambda \in \mathbb{R}$ et pour tout $x, x' \in \mathbb{R}^n$, $\langle x-x^*, y\rangle = 0$ et $\langle x'-$ *x∗∗, y=0* pour tout *y∈K*, donc *λx−λx∗, y+x′−x∗∗, y=λx+x′−λx∗+x∗∗, y=0*, pour tout $y \in K$, d'où $\lambda x^*+x^{**} = \lambda P_K(x) + P_K(x') = P_K(\lambda x+x')$.

Lemme 4.2. Soient K un convexe fermé de $\mathbb{R}^n$ et f une fonction différentiable et convexe sur K. S'il existe $\alpha \in \mathbb{R}_+^*$ et $x_0 \in K$ tel que $P_K\big(x_0 - \alpha\nabla f(x_0)\big) = x_0$, alors x_0 est un minimum local de f.

Preuve. D'après le lemme 4.1 on a pour tout $h \in \mathbb{R}^n$ vérifiant $x_0 + h \in K$, $\langle x_0 - \alpha\nabla f(x_0) - PKx0-\alpha\nabla fx0,x0+h-PKx0-\alpha\nabla fx0=x0-\alpha\nabla fx0-x0,x0+h-x0=-\alpha\nabla fx0,h=-\alpha\nabla fx0,h \le 0$. Donc, $\nabla fx0,h\ge 0$ pour tout $h\in \mathbb{R}n$ vérifiant $x0+h\in K$, d'après le lemme 3.1, $x0$ est un minimum local de f.

Proposition 4.1. Si $f \in C^2(K)$ est fortement convexe sur un convexe fermé et borné K de $\mathbb{R}^n$. Alors, l'ensemble $\widehat{K} = \{x \in K, \mu \le f(x) \le \nu\}$ où $\mu > \min_{x\in K} f(x)$ est fermé et borné.

Preuve. Puisque, $\widehat{K} \subset K$ et K est borné alors $\widehat{K}$ est borné. Démontrons que $\widehat{K}$ est fermé. Soit $\{x_k\}$ une suite dans $\widehat{K}$ convergente vers x_0, puisque K est fermé il contient x_0, et puisque $\mu \le f(x_k) \le \nu$ et f est continue sur K, $f(x_k) \longrightarrow f(x_0)$ quand $k \longrightarrow +\infty$, donc $\mu \le f(x_0) \le \nu$, d'où $x_0 \in \widehat{K}$ et $\widehat{K}$ est fermé.

Définition 4.1. Pour une fonction différentiable f sur $\mathbb{R}^n$. Si, au point $x \in \mathbb{R}^n$, la dérivée dans la direction du vecteur non nul $d \in \mathbb{R}^n$, i.e. $\frac{\partial f}{\partial d}(x) = \langle \nabla f(x), d\rangle \ge 0$, la fonction est croissante et si $\langle \nabla f(x), d\rangle < 0$, la fonction est décroissante, dans ce cas on dit que $d \in \mathbb{R}^n$ est une direction de descente. Dans le cas où $\|d\| = \|\nabla f(x)\|$, donc $\langle \nabla f(x), d\rangle \le \|\nabla f(x)\|^2$ (on dit qu'on a une plus forte pente) et dans le cas où $\langle -\nabla f(x), \nabla f(x)\rangle = -\|\nabla f(x)\|^2 \le \langle \nabla f(x), d\rangle$ direction opposée au gradient (on dit qu'on a une plus forte descente de la fonction f).

Proposition 4.2. Si d est une direction de descente au point $\bar{x}$. Alors:

a) il existe $\delta > 0$, tel que $f(\bar{x} + \alpha d) < f(x)$ pout tout $\alpha \in]0, \delta]$.

b) pour tout $0 < \beta < 1$, il existe $\hat{\delta} > 0$ tel que $f(\bar{x} + \alpha d) < f(\bar{x}) + \alpha\beta\langle \nabla f(\bar{x}), d\rangle$, pour tout $\alpha \in \big]0, \hat{\delta}\big]$.

Preuve. a) Puisque, $\frac{f(\bar{x}+\alpha d)-f(\bar{x})}{\alpha} = \langle \nabla f(\bar{x}), d\rangle + \|d\|\rho(\alpha d)$ avec $\rho(\alpha d) \longrightarrow 0$ quand $\alpha \longrightarrow 0$. Alors, pour tout $\varepsilon > 0$, il existe $\delta > 0$, tel que $\forall\, \alpha \in I(0, \delta)$ on a $-\varepsilon < \rho(\alpha d) < \varepsilon$ d'où $\langle \nabla f(\bar{x}), d\rangle - \varepsilon\|d\| < \frac{f(\bar{x}+\alpha d)-f(\bar{x})}{\alpha} < \langle \nabla f(\bar{x}), d\rangle + \varepsilon\|d\|$, quand $\varepsilon \longrightarrow 0$, $\frac{f(\bar{x}+\alpha d)-f(\bar{x})}{\alpha} = \langle \nabla f(\bar{x}), d\rangle < 0$, donc $f(\bar{x} + \alpha d) - f(\bar{x}) < 0$. b) Puisque d'après a), il existe $\delta > 0$, tel que $\forall\, \alpha \in I(0, \delta)$ on a $\frac{f(\bar{x}+\alpha d)-f(\bar{x})}{\alpha} = \langle \nabla f(\bar{x}), d\rangle < 0$, comme $0 < \beta < 1$, alors $\frac{f(\bar{x}+\alpha d)-f(\bar{x})}{\alpha} < \beta\frac{f(\bar{x}+\alpha d)-f(\bar{x})}{\alpha} = \beta\langle \nabla f(\bar{x}), d\rangle$ d'où $f(\bar{x} + \alpha d) < f(\bar{x}) + \alpha\beta\langle \nabla f(\bar{x}), d\rangle$ pour tout $0 < \alpha < \delta$ comme $0 < \alpha\beta < \alpha < \delta$, alors $f(\bar{x} + \alpha d) < f(\bar{x}) + \alpha\beta\langle \nabla f(\bar{x}), d\rangle$ pour tout $\alpha \in \big]0, \hat{\delta}\big]$ où $\hat{\delta} = \frac{\delta}{\beta} > 0$.

Lemme 4.3. Si, K est un convexe fermé de $\mathbb{R}^n$ et $f \in C^1(K)$. Alors, pour tout $x \in K$ et pour tout $\alpha \in \mathbb{R}_+^*$, on a $\langle \nabla f(x), h\rangle \le -\frac{1}{\alpha}\|h\|^2$ où $h = P_K\big(x - \alpha\nabla f(x)\big) - x$. Si, de plus K est borné et $x \in \widehat{K}$, il existe $\gamma \in \mathbb{R}_+^*$ tel que $\|h\| \ge \gamma$.

Preuve. Démontrons d'abord que $\langle \nabla f(x), h\rangle \le -\frac{1}{\alpha}\|h\|^2$ pour tout $\alpha \in \mathbb{R}_+^*$. Appliquant le lemme 4.1 pour $x \in K$ et $x - \alpha\nabla f(x)$, on a $\big\langle x - \alpha\nabla f(x) - P_K\big(x - \alpha\nabla f(x)\big), x - P_K\big(x - \alpha\nabla f(x)\big)\big\rangle = \langle -(\alpha\nabla f(x) + h), -h\rangle = \langle (\alpha\nabla f(x) + h), h\rangle = \langle \alpha\nabla f(x), h\rangle + \langle h, h\rangle = \alpha\langle \nabla f(x), h\rangle + \|h\|^2 \le 0$ pour tout $\alpha \in \mathbb{R}_+^*$, d'où $\langle \nabla f(x), h\rangle \le -\frac{1}{\alpha}\|h\|^2$ pour tout $\alpha \in \mathbb{R}_+^*$. Supposons de plus que K est borné et $x \in \widehat{K}$ et démontrons qu'il existe $\gamma \in \mathbb{R}_+^*$ tel que $\|h\| \ge \gamma$. Puisque f est fortement convexe sur K, $\nabla f(x)$ est continue sur K, et d'après le corollaire 4.2, l'application P_K est continue sur K, alors la fonction composée $g(x) = \big\|P_K\big(x - \alpha\nabla f(x)\big) - x\big\|$ est continue sur K, par conséquent elle est continue sur le compact

$\widehat{K}$, il existe alors $\gamma \in \mathbb{R}_+^*$ tel que $\gamma = \min_{x\in\widehat{K}} g(x) \leq \|h\|$. Car si $\gamma = 0$, il existe $x_0 \in \widehat{K}$ tel que $P_K\big(x_0 - \alpha\nabla f(x_0)\big) = x_0$ ce qui implique d'après le lemme 4.2, x_0 est un minimum local unique de f dans K et donc $x_0 \notin \widehat{K}$, contradiction.

Lemme 4.4. Si, K est un convexe fermé de $\mathbb{R}^n$ et $f \in C^2(K)$ fortement convexe de paramètre $m \in \mathbb{R}_+^*$. En passant de n'importe quel point $x \in K$ au point $x^* = P_K\big(x - \alpha\nabla f(x)\big)$ où α satisfaisant $0 < \alpha < \frac{2}{k_2}$ et k_2 est donnée par le corollaire 3.1, on a $f(x) - f(x^*) \geq \left(\frac{1}{\alpha} - \frac{k_2}{2}\right)\|h\|^2$ où $h = P_K\big(x - \alpha\nabla f(x)\big) - x$. Si, de plus K est borné et $x \in \widehat{K}$, il existe $\gamma \in \mathbb{R}_+^*$ tel que $f(x) - f(x^*) \geq \left(\frac{1}{\alpha} - \frac{k_2}{2}\right)\gamma^2$

Preuve. Il suffit, de démontrer la première inégalité, la deuxième inégalité se déduit de la première par le lemme 4.3. On démontre, la première inégalité d'abord pour x dans l'intérieur de K. En appliquant (TL) à la fonction f au point $x^* = x + h$, on a $f(x^*) = f(x) + \langle\nabla f(x), h\rangle + \frac{1}{2}d^2 f(x + \theta h)$ avec $0 < \theta < 1$. Comme d'après le lemme 4.3, $\langle\nabla f(x), h\rangle \leq -\frac{1}{\alpha}\|h\|^2$ et d'après le corollaire 3.1, il existe $k_2 > 0$, tel que $d^2 f(x + \theta h) \leq k_2\|h\|^2$, alors $f(x^*) - f(x) \leq \left(\frac{k_2}{2} - \frac{1}{\alpha}\right)\|h\|^2$, d'où $f(x) - f(x^*) \geq \left(\frac{1}{\alpha} - \frac{k_2}{2}\right)\|h\|^2$. Si, maintenant x est dans la frontière de K qui est un fermé, il existe une suite $\{x_n\}$ dans l'intérieur de K qui converge vers x, et $f(x_n) - f(x^*) \geq \left(\frac{1}{\alpha} - \frac{k_2}{2}\right)\|h_n\|^2$ où $h_n = P_K\big(x_n - \alpha\nabla f(x_n)\big) - x_n$. Comme f, ∇f, P_K et $\|.\|$ sont des fonctions continues sur K, quand $n \longrightarrow +\infty$, $f(x) - f(x^*) \geq \left(\frac{1}{\alpha} - \frac{k_2}{2}\right)\|h\|^2$.

4.2-Méthode du gradient projeté

Dans cette section on va démontrer que, la suite récurrente définie si dessous converge vers le minimum local de la fonction f. En partant d'une première approximation x_1 dans K et d'un nombre arbitraire α satisfaisant $0 < \alpha < \frac{2}{k_2}$ où k_2 est donnée par le corollaire 3.1. On définie une suite $\{x_k\}$ à l' aide de la relation de récurrence dite **méthode** (ou **algorithme) du gradient projeté à pas constant** α dans **la direction de descente** $d_k = -\nabla f(x_k)$.

$$x_{k+1} = P_K[x_k - \alpha\nabla f(x_k)],\ k \in \mathbb{N}^*$$

Comme, d_k est dirigé vers la plus grande descente de la fonction f, en passant d' un point x_k au point x_{k+1}, le taux $f(x_{k+1}) - f(x_k)$ diminue, par conséquent, la suite $\{f(x_k)\}$ décroit vers son minimum réalisé par la limite x_0 de la suite $\{x_k\}$ sur K. Le théorème fondamental suivant assure l'existence et l'unicité de la solution du **problème avec contrainte**

$$(\mathcal{P})\ \inf_{x\in K} f(x)$$

Théorème 4.2 (Théorème fondamental). Si, K est un convexe fermé de $\mathbb{R}^n$ et si $f \in C^2(K)$ et elle est fortement convexe de paramètre $m \in \mathbb{R}_+^*$. Alors, pour un point initial arbitraire x_1 dans K, la suite récurrente $\{x_k\}$ converge vers le minimum local x_0 de f sur K.

Preuve. Nous allons démontrer le théorème en supposant de plus que K est borné. Soient x_1 un point arbitraire dans K et $x_{k+1} = P_K[x_k - \alpha\nabla f(x_k)]$, $k \in \mathbb{N}^*$ où, α satisfaisant $0 < \alpha < \frac{2}{k_2}$ et k_2 est donnée par le corollaire 3.1. Il est clair que d'après le lemme 4.4 que $f(x_k) - f(x_{k+1}) \geq \left(\frac{1}{\alpha} - \frac{k_2}{2}\right)\|x_{k+1} - x_k\|^2 \geq 0$. La suite $\{f(x_k)\}$ est décroissante dans K, comme f et continue sur le compact K, il existe un $x_0 \in K$ tel que$f(x_0) = \min_{x\in K} f(x)$ et donc la suite $\{f(x_k)\}$ est convergente et $\lim_{k\to+\infty} f(x_k) = f(x_0)$. Démontrons que, la suite $\{x_k\}$ converge vers x_0. Soient $\varepsilon > 0$ et $\widehat{K}_\varepsilon = K \setminus B(x_0, \varepsilon)$, il est clair que $\widehat{K}_\varepsilon$ est un fermé et borné dans K, donc il existe un $x_\varepsilon \in \widehat{K}_\varepsilon$ tel que $f(x_\varepsilon) = \min_{x\in\widehat{K}_\varepsilon} f(x)$. Le minimum de f sur K étant unique, donc $f(x_\varepsilon) > f(x_0)$. $\widehat{K}_\varepsilon$ contient seulement un nombre finie d'élément de la suite $\{x_k\}$, si non il existe $N \in \mathbb{N}^*$ tel que, pour tout $k > N$, $f(x_k) \geq f(x_\varepsilon)$ donc $\lim_{k\to+\infty} f(x_k) = f(x_0) \geq f(x_\varepsilon) > f(x_0)$, contradiction. Donc, pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}^*$ tel que, pour tout $k > N$, $x_k \in B(x_0, \varepsilon)$ i.e. $\lim_{k\to+\infty} x_k = x_0$. A présent, nous avons démontrer le théorème dans le cas où K est un convexe fermé et borné. Dans le cas où K est un convexe fermé et non borné. Soient x_1 un point arbitraire dans K et $x_{k+1} = P_K[x_k - \alpha\nabla f(x_k)]$, $k \in \mathbb{N}^*$ où, α satisfaisant $0 < \alpha < \frac{2}{k_2}$ et k_2 est donnée par le corollaire 3.1. Comme dans la démonstration du théorème 3.2, le point minimum de la fonction f se trouve dans l'ensemble convexe fermé et borné $K_r = \mathrm{K} \cap \tilde{B}(x_1, r)$ où le rayon $r > 0$ de la boule fermée $\tilde{B}(x_1, r)$ satisfait $\frac{k_1}{2}\mathrm{r} - \|\nabla f(x_1)\| > 0$, et $f(x_1) < f(x)$, pour tout x dans l'extérieur de la boule fermée $\tilde{B}(x_1, r)$ et donc dans l'extérieure de K_r. Comme d'après le lemme 4.4 la suite $\{f(x_k)\}$ est décroissante dans K, il vient que la suite récurrente $\{x_k\}$ se trouve à l'extérieure de K_r, par conséquent,

$$x_{k+1} = P_K[x_k - \alpha\nabla f(x_k)] = P_{K_r}[x_k - \alpha\nabla f(x_k)], \text{ pour tout } k \in \mathbb{N}^*$$

et tous les arguments qui suivront se réduisent à l'ensemble convexe fermé borné K_r, c'est-à-dire au cas où K est un convexe fermé et borné que nous avons considéré avant.

Remarque 4.2.

a) Dans le cas où $K = \mathbb{R}^n$ et parce que $x = P_K(x)$ pour tout $x \in K$, la suite itérative $\{x_k\}$ s'écrit

$$x_{k+1} = x_k - \alpha\nabla f(x_k).$$

b) La méthode exposée précédemment nous permet de chercher la solutions x_0 du système à n équations suivant:

$$f_i(x) = 0,\ 1 \leq i \leq n.$$

Il suffit de remarquer que, la solution de ce système coïncide avec le minimum de la fonction $f^2(x) = f_1^2(x) + \cdots + f_n^2(x)$.

c) La théorie que nous avons exposés, pour rechercher un minimum d'une fonction fortement convexe, s'étende sans complication à la recherche du maximum d'une fonction f fortement concave i.e. la fonction $(-f)$ est fortement convexe, en tenant compte de la relation $\inf_{x\in K}(-f) = -\sup_{x\in K} f$.

5-Méthode du gradient à pas optimal et à pas variable

5.1-Description

La méthode ou algorithme du gradient fait partie d'une vaste classe de méthodes numériques appelées méthode de descente, dont le schéma général est:

$$x_{k+1} = x_k + \alpha_k d_k$$

où α_k est un élément de $\mathbb{R}_+^*$, et d_k est un vecteur non nul de $\mathbb{R}^n$. α_k et d_k sont choisis de tel sorte que la suite $\{f(x_{k+1})\}$ est décroissante vers la valeur minimal de la fonction f.

Dans le cas où $d_k = -\nabla f(x_k)$ et $\alpha \in \mathbb{R}_+^*$, la suite

$$x_{k+1} = x_k - \alpha\nabla f(x_k)$$

décrit **l'algorithme du gradient à pas constant**.

Dans le cas où $d_k = -\nabla f(x_k)$ et α_k est la solution quand elle existe du problème de minimisation de la fonction $g_k(\alpha) = f\big(x_k + \alpha\nabla f(x_k)\big)$ sur $\mathbb{R}_+^*$, la suite

$$x_{k+1} = x_k - \alpha_k\nabla f(x_k)$$

décrit **la méthode du gradient à pas optimal**.

Dans le cas où $d_k = -\nabla f(x_k)$ et α_k est une variable dans $\mathbb{R}_+^*$, la suite

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$
décrit **la méthode du gradient à pas variable**.

5.2-Résultats de convergence pour les fonctions elliptiques

La démonstration des résultats de convergences pour les fonctions elliptiques utilise les propriétés suivantes: les fermés et bornés dans $\mathbb{R}^n$ sont des compactes. Les fonctions continues sur un compact sont uniformément continues. Une fonction f est uniformément continue ssi pour toutes suites $\{x_n\}$ et $\{y_n\}$ vérifiant $\|x_n - y_n\| \longrightarrow 0$, on a $\|f(x_n) -$ *fyn⟶0.*

Théorème 5.1. Si, f est une fonction elliptique de paramètre $m \in \mathbb{R}_+^*$, la méthode du gradient à pas optimal converge vers le minimum de f.

Preuve. S'il existe $N \in \mathbb{N}^*$ tel que $\nabla f(x_k) = 0$, pour tout $k \geq N$, la suite $\{x_k\}$ est évidement convergente pour un nombre fini d'itérations. Supposons que $\nabla f(x_k) \neq 0$ pour tout $k \in \mathbb{N}^*$ et démontrons que:

$a)$ $f(x_k) - f(x_{k+1}) \geq \frac{m}{2}\|x_k - x_{k+1}\|^2$. Puisque, pour tout $k \in \mathbb{N}^*$ et pour tout $\alpha \in \mathbb{R}_+^*$, la fonction $\varphi_k(\alpha) = f\big(x_k - \alpha \nabla f(x_k)\big)$ est elliptique car $[\varphi_k(\alpha) - \varphi_k(\beta)](\alpha - \beta) = \langle f\big(x_k - \beta \nabla f(x_k)\big) - f\big(x_k - \alpha \nabla f(x_k)\big), (\alpha - \beta)\nabla f(x_k)\rangle \geq m\|(\alpha - \beta)\nabla f(x_k)\|^2 = \bar{\alpha}\|\nabla f(x_k)\|^2|\alpha - \beta|^2 = \bar{m}|\alpha - \beta|^2$ où $\bar{m} = m\|\nabla f(x_k)\|^2 > 0$ est la constante d'ellipticité de φ_k, donc φ_k est strictement convexe et coercive sur $\mathbb{R}$, elle admet un unique élément $\alpha_k \in \mathbb{R}_+^*$, satisfaisant $\varphi'_k(\alpha_k) = -\langle \nabla f\big(x_k - \alpha_k \nabla f(x_k)\big), \nabla f(x_k)\rangle = 0$, comme $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$, on déduit que $\nabla(x_{k+1}) \perp \nabla f(x_k)$ (les directions de descentes successives sont perpendiculaires) et $\forall k \in \mathbb{N}^*$ $\langle \nabla(x_{k+1}), -\alpha_k \nabla f(x_k)\rangle = \langle \nabla(x_{k+1}), x_{k+1} - x_k\rangle = 0$. Comme f est elliptique de paramètre $m \in \mathbb{R}_+^*$, d'après $(*)$, $f(x_k) - f(x_{k+1}) \geq \frac{m}{2}\|x_k - x_{k+1}\|^2$, $\forall k \in \mathbb{N}^*$.

$b)$ $\lim_{k \to +\infty}\|x_k - x_{k+1}\| = 0$. Puisque d'après $a)$ la suite $\{f(x_k)\}$ est décroissante et d'après le théorème 3.1 $d)$ il existe un unique $x^* \in \mathbb{R}^n$ tel que $f(x^*) \leq f(x)$ pour tout $x^* \in \mathbb{R}^n$, donc $\lim_{k \to +\infty} f(x_k) = f(x^*) = \lim_{k \to +\infty} f(x_{k+1})$, en utilisant $a)$ on a $0 \geq \lim_{k \to +\infty}\|x_k - x_{k+1}\|^2 \geq 0$, d'où $\lim_{k \to +\infty}\|x_k - x_{k+1}\| = 0$

$c)$ $\|\nabla f(x_k)\| \leq \|\nabla f(x_k) - \nabla f(x_{k+1})\|$. Puis que $\nabla(x_{k+1}) \perp \nabla f(x_k)$, alors $\|\nabla f(x_k)\|^2 = \|\langle \nabla f(x_k) - \nabla f(x_{k+1}), \nabla f(x_k)\rangle\| \leq \|\nabla f(x_k)\|\|\nabla f(x_k) - \nabla f(x_{k+1})\|$, donc $\|\nabla f(x_k)\| \leq \|\nabla f(x_k) - \nabla f(x_{k+1})\|$.

$d)$ $\lim_{k \to +\infty}\|\nabla f(x_k)\| = 0$. Puisque la suite $\{f(x_k)\}$ est décroissante et f est coercive, la suite $\{x_k\}$ est bornée, donc elle est contenue dans une boule fermée et bornée dans $\mathbb{R}^n$, donc dans un compact. Comme $f \in C^1(\mathbb{R}^n)$ par hypothèse, f' est continue sur ce compact, elle est uniformément continue sur ce compact. En utilisant $b)$, $c)$ et le lien entre les suites convergentes et l'uniforme continuité, on déduit que $\lim_{k \to +\infty}\|\nabla f(x_k) - \nabla f(x_{k+1})\| = 0$ et donc $\lim_{k \to +\infty}\|\nabla f(x_k)\| = 0$.

$e)$ La convergence. Puisque $\nabla f(x^*) = 0$ et f est elliptique de paramètre $m \in \mathbb{R}_+^*$, d'après le théorème 3.1 $a)$ $m\|x_k - x^*\| \leq \|\nabla f(x_k) - \nabla f(x^*)\| = \|\nabla f(x_k)\|$, ou bien $0 \leq \|x_k - x^*\| \leq \frac{1}{m}\|\nabla f(x_k)\|$, en utilisant $d)$ on a $x_k \longrightarrow x^*$ quand $k \longrightarrow +\infty$.

Remarque 5.1. Les constantes a et b non pas intervenues, dans la démonstration précédente. Mais elles interviennent dans le procédé de recherche de α_k, sans passer par le problème de minimisation de la fonction φ_k, comme dans **la règle de recherche linéaire de pas optimal** du à Wolf

Un autre critère de convergence, pour les fonctions non nécessairement elliptiques, est donné par le théorème suivant:

Théorème 5.2. Si, une fonction $f \in C^1(\mathbb{R}^n)$ est strictement convexe et coercive et si, sa dérivée f' est M-Lipschitzienne sur $\mathbb{R}^n$, i.e. il existe $M > 0$ tel que $\|f'(x) - f'(y)\| \leq M\|x - y\|$, pour tout $x, y \in \mathbb{R}^n$. Alors, pour tout $\alpha_k \in [a, b] \subset \left]0, \frac{2}{M}\right[$, la méthode du gradient à pas optimal converge vers le minimum de f.

Preuve. Démontrons que:

a) $\lim_{k\to\infty}\|x_{k+1} - x_k\| = 0$.

Appliquant (TI) à la fonction f à l'ordre 1 au point x_k, alors $f(x_{k+1}) = f(x_k) + \int_0^1 \langle f'[x_k + t(x_{k+1} - x_k)] - f'(x_k), x_{k+1} - x_k\rangle + \langle f'(x_k), x_{k+1} - x_k\rangle\, dt = f(x_k) + \langle f'(x_k), x_{k+1} - x_k\rangle + \int_0^1 \langle f'[x_k + t(x_{k+1} - x_k)] - f'(x_k), x_{k+1} - x_k\rangle\, dt$. En utilisant le notation ∇f au lieu de f' et le fait que $-\frac{1}{\alpha_k}(x_{k+1} - x_k) = \nabla f(x_k)$, alors $f(x_{k+1}) = f(x_k) - \frac{1}{\alpha_k}\|x_{k+1} - x_k\|^2 + \int_0^1 \langle \nabla f[x_k + t(x_{k+1} - x_k)] - \nabla f(x_k), x_{k+1} - x_k\rangle\, dt$. D'après l'inégalité de Cauchy Schawarz $f(x_{k+1}) - f(x_k) \leq -\frac{1}{\alpha_k}\|x_{k+1} - x_k\|^2 + \int_0^1 \|\nabla f[x_k + t(x_{k+1} - x_k)] - \nabla fxkxk+1-xkdt$, comme ∇f est M-Lipschitzienne sur $\mathbb{R}n$, alors $f(x_{k+1}) - f(x_k) \leq -\frac{1}{\alpha_k}\|x_{k+1} - x_k\|^2 + \int_0^1 M\|x_k + t(x_{k+1} - x_k) - x_k\|\|x_{k+1} - x_k\|\, dt$ ou bien $f(x_{k+1}) - f(x_k) \leq -\frac{1}{\alpha_k}\|x_{k+1} - x_k\|^2 + \frac{M}{2}\|x_{k+1} - x_k\|^2$, donc $f(x_{k+1}) - f(x_k) \leq \left(\frac{M}{2} - \frac{1}{\alpha_k}\right)\|x_{k+1} - x_k\|^2 < 0$ car $0 < \alpha_k < \frac{2}{M}$ par hypothèse. La suite $\{f(x_k)\}$ est décroissante et minorée par sa valeur minimal $f(x^*)$, donc $\nabla f(x^*) = 0$ et $f(x_k) \longrightarrow f(x^*)$, comme f est continue $\lim_{k\to\infty}[f(x_{k+1}) - f(x_k)] = 0$, d'où $0 \leq \left(\frac{M}{2} - \frac{1}{\alpha_k}\right)\lim_{k\to\infty}\|x_{k+1} - x_k\|^2 \leq 0$ donc, $\lim_{k\to\infty}\|x_{k+1} - x_k\| = 0$.

b) $\lim_{k\to\infty}\nabla f(x_k) = 0$. Comme dans la démonstration du théorème 3.2, a) on a $\nabla f(x_{k+1}) \perp \nabla f(x_k)$ et
$\|\nabla f(x_k)\|^2 = \|\langle \nabla f(x_k) - \nabla f(x_{k+1}), \nabla f(x_k)\rangle\| \leq \|\nabla f(x_k)\|\|\nabla f(x_k) - \nabla f(x_{k+1})\|$, donc
$0 \leq \|\nabla f(x_k)\| \leq \|\nabla f(x_k) - \nabla f(x_{k+1})\| \leq M\|x_{k+1} - x_k\|$, d'où $\lim_{k\to\infty}\|\nabla f(x_k)\| = 0$ donc, $\lim_{k\to\infty}\nabla f(x_k) = 0$.

c) La convergence. Puisque f est coércive, la suite $\{x_k\}$ est bornée, elle admet une sous suite $\{x_{\sigma(k)}\}$ convergente vers x, comme par hypothèse ∇f est continue sur $\mathbb{R}^n$, alors $\lim_{k\to\infty}\nabla f(x_{\sigma(k)}) = \nabla f(\lim_{k\to\infty} x_{\sigma(k)}) = \nabla f(x) = 0 = \nabla f(x^*)$, comme f est strictement convexe le minimum est unique et $x^* = x$. Comme, toute sous suite de la suite $\{x_k\}$, converge vers x^*. Alors la suite $\{x_k\}$ converge vers x^*.

Du deux théorèmes 5.1 et 5.2 on a immédiatement:

Corollaire 5.1. La méthode du gradient à pas optimal, pour une fonction quadratique est convergente

Corollaire 5.2. Si, une fonction f est elliptique et sa dérivée f' est M-Lipschitzienne sur $\mathbb{R}^n$ Alors, pour tout $\alpha_k \in \left]0, \frac{2}{M}\right[$, la méthode du gradient à pas optimal converge vers le minimum de f.

Remarque 5.2.

a) La dimension finie est nécessaire dans les résultats de convergences des algorithmes de projection et du gradient.

b) La résolution du problème de minimisation à une seule variable dans les démonstrations des méthodes du gradient à pas optimal conduit à l'orthogonalité des directions de descentes successives. On peut donc, prendre dans l'itération k, le pas optimal $\alpha_k = \frac{\|y_k\|^2}{\langle Ay_k, y_k\rangle}$. Parce que: $\nabla f(x) = Ax - b$ et $\nabla f(x_{k+1}) \perp \nabla f(x_k)$. Alors, $\langle A(x_k - \alpha_k(Ax_k - b)) - b, Ax_k - b\rangle = 0$, en posant $r_k = Ax_k - b$ on a $\langle r_k - \alpha_k A r_k, r_k\rangle = 0$ d'où $\alpha_k\langle Ar_k, r_k\rangle = \langle r_k, r_k\rangle = \|r_k\|^2$ donc, $\alpha_k = \frac{\|r_k\|^2}{\langle Ar_k, r_k\rangle}$. Cette méthode est utilisée pour résoudre le système $r(x) = Ax - b = 0$ où A est une matrice carré symétrique, définie positive, i.e. $\langle Ah, h\rangle \geq 0$, pour tout $h \in \mathbb{R}^n$.

Théorème 5.3. Si, une fonction $f \in C^1(\mathbb{R}^n)$ est elliptique de paramètre $m \in \mathbb{R}_+^*$ et si, sa dérivée f' est M-Lipschitzienne sur $\mathbb{R}^n$. Alors, pour tout $\alpha_k \in [a, b]$, où $0 < a \leq b < 2\frac{m}{M}$, il existe $\beta \in]0,1[$ dépendant de a, b, m et M, tel que $\|x_k - x^*\| \leq \beta^{k-1}\|x_1 - x^*\|$. Par conséquent, la méthode du gradient à pas variable converge géométriquement vers le minimum x^* de f.

Preuve. Puisque $\nabla f(x^*) = 0$, alors $x_{k+1} - x^* = (x_k - x^*) - \alpha_k\big(\nabla f(x_k) - \nabla f(x^*)\big)$, donc $\|x_{k+1} - x^*\|^2 = \|x_k - x^*\|^2 - 2\langle \nabla f(x_k) - \nabla f(x^*), x_k - x^*\rangle + \alpha_k^2\|\nabla f(x_k) - \nabla f(x^*)\|^2$. Par hypothèse et l'inégalité de Cauchy Schwarz, on a: $\|x_{k+1} - x^*\|^2 \leq (1 - 2m\alpha_k +$ *αk2M2xk−x∗2.* Soit la fonction *pαk=1−2mαk+αk2M2*, alors *p′αk=2αkM2−m=0* ssi $\alpha_k^0 = \frac{m}{M^2}$, p est décroissante sur $\left[0, \frac{m}{M^2}\right]$ et croissante sur $\left[\frac{m}{M^2}, \frac{2m}{M^2}\right]$, d'où pour tout $\alpha_k \in [a, b] \not\subseteq \left]0, \frac{2m}{M^2}\right[$, $0 \leq p(\alpha_k) \leq p(a) < p(0) = 1$ si $\alpha_k \in \left[a, \frac{m}{M^2}\right]$ et $p(\alpha_k) \leq p(b) < p\left(\frac{2m}{M^2}\right) = 1$ si $\alpha_k \in \left[\frac{m}{M^2}, b\right]$. Donc, pour tout $\alpha_k \in [a, b] \not\subseteq \left]0, \frac{2m}{M^2}\right[$, $0 < p(\alpha_k) \leq \max\big(p(a), p(b)\big) < 1$ d'où $\|x_{k+1} - x^*\| \leq \beta\|x_k - x^*\|$ où $0 < \beta = \sqrt{\max\big(p(a), p(b)\big)} < 1$ pour tout $k \geq 1$, donc $\|x_{k+1} - x^*\| \leq \beta\|x_k - x^*\| \leq \beta^2\|x_{k-1} - x^*\| \leq \cdots \leq \beta^{k-2}\|x_1 - x^*\|$ et donc $0 \leq \beta\|x_k - x^*\| \leq \beta^{k-1}\|x_1 - x^*\|$ quand $k \longrightarrow +\infty$, $\|x_k - x^*\| \longrightarrow 0$ ou bien $x_k \longrightarrow x^*$.

Remarque 5.3.

a) La dimension finie n'intervienne pas dans la démonstration du théorème 5.3 , ce qui permet d'obtenir le même résultat dans un espace de Hilbert.

b) Dans le cas d'une fonctions f deux fois dérivable, f' est M-Lipschitzienne sur $\mathbb{R}^n$ ssi $\sup_{h\in\mathbb{R}^n}\|f''(h)\| \leq M$.

c) Dans le cas de fonctions quadratiques, $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ où A est une matrice carrée et symétrique d'ordre n. La méthode du gradient à pas variable s'écrit:

$$x_{k+1} = x_k - \alpha_k(Ax_k - b)$$

elle est convergente pour tout $\alpha_k \in [a, b] \not\subseteq \left]0, \frac{2\lambda_1}{\lambda_2}\right[$ où $\lambda_1 = m$ est la plus petite (respectivement $\lambda_2 = M$ la plus grande) valeur propre de la matrice A.

d) L'ellipticité est importante, pour la convergence de la plupart des algorithmes, toutefois les conditions données dans les théorèmes sont suffisantes pour la convergence, mais il se peut qu'un algorithme converge même elles ne le sont pas.

6-Méthode du gradient conjugué quadratique

6.1-Introduction

La méthode du gradient conjugué pour les fonctions quadratiques appelé **méthode du gradient conjugué quadratique** est l'une des techniques les plus utiles pour résoudre le système linéaire $r(x) = Ax - b = 0$ ($r(x)$ est le résidu du système linéaire $Ax - b = 0$) où A est une matrice carrée symétrique définie positive, $x \in \mathbb{R}^n$ et b est une constante dans $\mathbb{R}^n$. La résolution du système $r(x)$ est équivalent à la résolution du problème de minimisation de la fonction quadratique, $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ où $x \in \mathbb{R}^n$ et A est une matrice carrée symétrique. Posons, $r_k = \nabla f(x_k) = Ax_k - b$, ($-r_k$ est la direction de la plus grande pente), alors $r_{k+1} = \nabla f(x_{k+1}) = A(x_{k+1}) - b = A(x_k + \alpha_k d_k) - b = (Ax_k - b) + \alpha_k Ad_k = r_k + \alpha_k Ad_k$. La méthode du gradient conjugué est basée sur le principe suivant. pout tout $k \in \{1, \dots, n\}$ les directions d_k sont ***A*-conjugués**, i.e. $\langle d_k, Ad_i \rangle = 0$, si $k \neq i$ et $\{d_1, \dots, d_k\}$ est une famille libre et génératrice donc une base d'un sous espace vectoriel de $\mathbb{R}^n$. Ce principe est basé sur une écriture linéaire simple, de la formule donnant la direction d_{k+1}. L'idée principale de cette écriture vient par exemple du procédé d'orthogonalisation de Gram-Schmit, appliqué aux vecteurs libres $\{r_1, \dots, r_k\}$. Ce procédé défini une famille $\{d_1, \dots, d_k\}$ comme suit: on prend $d_1 = -r_1$, et on définit par récurrence $d_2 = -r_2 - \langle r_2, \frac{r_1}{\|r_1\|^2} \rangle r_1$, $d_3 = -r_3 - \langle r_3, \frac{r_1}{\|r_1\|^2} \rangle r_1 - \langle r_3, \frac{r_2}{\|r_2\|^2} \rangle r_2, \dots, d_k = -r_k - \sum_{i=1}^{k-1} \langle r_k, \frac{r_i}{\|r_i\|^2} \rangle r_i$. Ainsi, $\langle d_k, d_i \rangle = 0$ pour tout $k \neq i$, $\|d_i\|^2 = \langle d_i, d_i \rangle > 0$ et $\{d_1, \dots, d_k\}$ est une base du sous espace vectoriel E_k engendré par $\{r_1, \dots, r_k\}$, donc $E_k = [r_1, \dots, r_k] = [d_1, \dots, d_k]$.

6.2-Description da la méthode

Etant donné une valeur initiale arbitraire x_1, on construit la suite récurrente
$$x_{k+1} = x_k + \alpha_k d_k,$$
où la direction d_k est choisie A-conjuguées aux autres directions et le pas α_k est choisi optimal dans la direction d_k. La direction d'ordre $k+1$, s'écrit sous une forme simple, comme **combinaison linéaire** de r_{k+1} et d_k
$$d_{k+1} = -r_{k+1} + \beta_{k+1} d_k$$
où les nombres réels β_{k+1} sont déterminés par la relation de conjugaison
$$\langle d_{k+1}, Ad_k \rangle = \langle -r_{k+1} + \beta_{k+1} d_k^t, Ad_k \rangle = 0,$$
donc, $\beta_{k+1} \langle d_k, Ad_k \rangle = \langle r_{k+1}, Ad_k \rangle$ et quand $d_k \neq 0$, on a:
$$\beta_{k+1} = \frac{\langle r_{k+1}, Ad_k \rangle}{\langle d_k, Ad_k \rangle}.$$
En posant $\varphi_k(\alpha) = f(x_k + \alpha d_k)$, $\alpha \in \mathbb{R}_+^*$, le minimum α_k de φ_k est tel que:
$$\varphi_k'(\alpha_k) = \langle \nabla f(x_k + \alpha_k d_k), d_k \rangle = 0,$$
ou bien $\langle A(x_k + \alpha_k d_k) - b, d_k \rangle = 0$, ce qui implique $\langle d_k, A(x_k + \alpha_k d_k) - b \rangle = 0$. Donc, $\alpha_k \langle d_k, Ad_k \rangle = -[\langle d_k, Ax_k \rangle + \langle d_k, -b \rangle] = -\langle d_k, Ax_k - b \rangle$ d'où $\alpha_k = -\frac{\langle d_k, Ax_k - b \rangle}{\langle d_k, Ad_k \rangle}$, finalement
$$\alpha_k = -\frac{\langle d_k, r_k \rangle}{\langle d_k, Ad_k \rangle}.$$

6.3-Résultat de convergence et propriétés

Théorème 6.1. Pout toute valeur initiale x_1, la méthode du gradient conjugué quadratique, converge vers l'optimum x^*de la fonction quadratique f, en au maximum n pas.

Preuve. Démontrons que la méthode s'arrête au maximum à l'étape n qui est la dimension de $\mathbb{R}^n$. Comme x^* est l'optimum de f, on a $\nabla f(x^*) = Ax^* - b = 0$ ou $Ax^* = b$. Puisque $x^* - x_1 \in \mathbb{R}^n$ et à l'étape n, $\{d_1, \dots, d_k, \dots, d_n\}$ est une base de $\mathbb{R}^n$, il existe $\mu_1, \dots, \mu_k, \dots, \mu_n \in \mathbb{R}$ tel que $x^* - x_1 = \mu_1 d_1 + \dots + \mu_k d_k + \dots + \mu_n d_n$. Alors, $\langle d_k, A(x^* - x_1) \rangle = \langle d_k, \mu_1 Ad_1 + \dots + \mu_k Ad_k + \dots + \mu_n Ad_n \rangle = \mu_1 \langle d_1, Ad_1 \rangle + \dots + \mu_k \langle d_k, Ad_k \rangle + \dots + \mu_n \langle d_n, Ad_n \rangle$. Comme, $\langle d_k, Ad_i \rangle = 0$, pour tout $i \neq k$, alors $\langle d_k, A(x^* - x_1) \rangle - \mu_k \langle d_k, Ad_k \rangle = 0$, d'où $\mu_k = \frac{\langle d_k, A(x^* - x_1) \rangle}{\langle d_k, Ad_k \rangle}$ pour tout $k \in \{1, \dots, n\}$. Comme, par construction, $x_k - x_1 = \alpha_1 d_1 + \alpha_2 d_2 + \dots + \alpha_{k-1} d_{k-1}$, donc $\langle d_k, A(x_k - x_1) \rangle = \langle d_k, \alpha_1 Ad_1 + \alpha_2 Ad_2 + \dots + \alpha_{k-1} Ad_{k-1} \rangle = \alpha_1 \langle d_k, Ad_1 \rangle + \dots + \alpha_{k-1} \langle d_k, Ad_{k-1} \rangle$, or $\langle d_k, Ad_i \rangle = 0$ pour tout $i \in \{1, \dots, k-1\}$ donc $\langle d_k, A(x_k - x_1) \rangle = 0$ d'où $\langle d_k, A(x^* - x_1) \rangle - \mu_k \langle d_k, Ad_k \rangle = \langle d_k, A(x_k - x_1) \rangle$, comme, $\langle d_k, A[(x^* - x_1) + (x_1 - x_k)] \rangle = \langle d_k, A(x^* -$

$x_k))\rangle = \langle d_k, Ax^* - Ax_k\rangle = \langle d_k, b - Ax_k\rangle = -\langle d_k, r_k\rangle$, donc $-\langle d_k, r_k\rangle = \mu_k\langle d_k, Ad_k\rangle$, d'où $\mu_k = -\frac{\langle d_k,r_k\rangle}{\langle d_k,Ad_k\rangle} = \alpha_k$ et $x^* = x_1 + \alpha_1 d_1 + \cdots + \alpha_k d_k + \cdots + \alpha_n d_n$ et donc x^* est atteint à n pas aux maximum.

Proposition 6.1. Pout toute valeur initiale x_1, la suite $\{x_k\}$ dans l'algorithme du gradient conjugué quadratique est tel que:

$a)$ $\langle r_{k+1}, d_i\rangle = 0, \forall i \in \{1, \dots, k\}$

$b)$ x_{k+1} est un minimum de $f(x)$ sur l'ensemble convexe $C_k = x_1 + F_k$, où $F_k = [d_1, \dots, d_k]$ est le sous espace vectoriel de $\mathbb{R}^n$, engendré par $d_1, \dots, d_k$.

Preuve. Soit pour tout $\mu \in C_k$, la fonction $h(\mu) = f(x_1 + \mu_1 d_1 + \cdots + \mu_k d_k)$, où $\mu = (\mu_1, \dots, \mu_k)^T$, h est quadratique, donc elle admet un minimum unique sur le convexe C_k, et $\mu^* \in C_k$ est un minimum de h ssi $\nabla h(\mu^*) = 0$ ssi $\forall i \in \{1, \dots, k\}$, $\langle \nabla f(x_1 + \mu_1^* d_1 + \cdots +$ *μk∗dk,di=0* ssi *rμ∗,di=0, ∀i∈1, …,k.* Donc, pour démontrer que *xk+1* est minimum de *f* sur E_k, il suffit de démontrer que $\langle r_{k+1}, d_i\rangle = 0, \forall i \in \{1, \dots, k\}$ par récurrence.

Pour $k = 1$, on a $\langle r_2, d_1\rangle = \langle r_1 + \alpha_1 Ad_1, d_1\rangle = -\langle d_1, d_1\rangle + \frac{\langle d_1,d_1\rangle}{\langle d_1,Ad_1\rangle}\langle d_1, Ad_1\rangle = 0$, où $\alpha_1 = -\frac{\langle d_1,d_1\rangle}{\langle d_1,Ad_1\rangle}$ Supposons que, $\langle r_k, d_i\rangle = 0, \forall i \in \{1, \dots, k\}$ et montrons que $\langle r_{k+1}, d_i\rangle = 0$, $\forall i \in \{1, \dots, k\}$. Puisque, $r_{k+1} = r_k + \alpha_k Ad_k$, alors $\langle r_{k+1}, d_i\rangle = \langle r_k + \alpha_k Ad_k, d_i\rangle = \langle r_k, d_i\rangle + \alpha_k\langle d_i, Ad_k\rangle$, comme par hypothèse de récurrence $\langle r_k, d_i\rangle = 0, \forall i \in \{1, \dots, k\}$ et $\langle d_i, Ad_k\rangle = 0$ par conjugaison $\forall i \in \{1, \dots, k-1\}$, alors $\forall i \in \{1, \dots, k-1\}$, $\langle r_{k+1}, d_i\rangle = 0$ et $\langle r_{k+1}, d_k\rangle = \alpha_k\langle d_k, Ad_k\rangle = -\frac{\langle d_k,r_k\rangle}{\langle d_k,Ad_k\rangle}\langle d_k, Ad_k\rangle = -\langle d_k^t, r_k\rangle = -\langle r_k^t, d_k\rangle = 0$ où $\alpha_k = -\frac{\langle d_k,r_k\rangle}{\langle d_k,Ad_k\rangle}$.

Corolaire 6.1. Si, $\forall i \in \{1, \dots, k\}$ où $r_i \neq 0$ (i.e. l'algorithme du gradient conjugué quadratique n'est pas atteint à l'itération k) on a:

$a)$ $\alpha_k = \frac{\langle r_k,r_k\rangle}{\langle d_k,Ad_k\rangle} > 0$.

$b)$ $\beta_{k+1} = \frac{\langle r_{k+1},r_{k+1}\rangle}{\langle r_k,r_k\rangle}$.

$c)$ Les directions $d_1, \dots, d_{k+1}$ sont mutuellement conjugués.

Preuve. $a)$ Puisque, $d_k = -r_k + \beta_k d_{k-1}$ et $\alpha_k = -\frac{\langle d_k,r_k\rangle}{\langle d_k,Ad_k\rangle}$, donc $\alpha_k = -\frac{\langle -r_k+\beta_k d_{k-1},r_k\rangle}{\langle d_k,Ad_k\rangle} = \frac{\langle r_k,r_k\rangle}{\langle d_k,Ad_k\rangle} - \beta_k\frac{\langle d_{k-1},r_k\rangle}{\langle d_k,Ad_k\rangle}$, comme d'après la proposition 6.1, $\langle d_{k-1}, r_k\rangle = \langle r_k, d_{k-1}\rangle = 0$, on a le résultat. $b)$ Puisque, $r_{k+1} = r_k + \alpha_k Ad_k$ et $\beta_{k+1} = \frac{\langle r_{k+1},Ad_k\rangle}{\langle d_k,Ad_k\rangle}$, alors $\beta_{k+1} = \frac{1}{\langle d_k,Ad_k\rangle}\langle r_{k+1}, \left(\frac{r_{k+1}-r_k}{\alpha_k}\right)\rangle$ où $\alpha_k = \frac{\langle r_k,r_k\rangle}{\langle d_k,Ad_k\rangle} > 0$, donc $\beta_{k+1} = \frac{\langle r_{k+1},r_{k+1}-r_k\rangle}{\langle r_k,r_k\rangle} = \frac{\langle r_{k+1},r_{k+1}\rangle}{\langle r_k,r_k\rangle} - \frac{\langle r_{k+1},r_k\rangle}{\langle r_k,r_k\rangle}$. Comme, les gradients successives sont perpendiculaire alors $\langle r_{k+1}, r_k\rangle = 0$ ou parecque $\langle r_{k+1}, r_k\rangle = \langle r_{k+1}, d_k - \beta_k d_{k-1}\rangle = \langle r_{k+1}, d_k\rangle - \beta_k\langle r_{k+1}, d_{k-1}\rangle = 0$ d'après la proposition 4.1. Alors, $\beta_{k+1} = \frac{\langle r_{k+1},r_{k+1}\rangle}{\langle r_k,r_k\rangle}$ et on a $b)$. $c)$ Montrons que d_{k+1} est A-conjugué avec $\{d_1, \dots, d_k\}$. **(La démonstration serra donnée par la suite dans la deuxième version)**