



N° d'ordre :

N° de série :

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique

UNIVERSITÉ ECHAHID HAMMA LAKHDAR
EL OUED

FACULTÉ DES SCIENCES ET DE TECHNOLOGIE

Mémoire de fin d'étude

LICENCE ACADEMIQUE

Domaine: Mathématiques et Informatique

Filière: Mathématiques

Spécialité: Modélisation mathématiques & simulation
numérique

Thème

Optimisation Défférentiable et
Convexité

Présenté par:

AMMAR Afifa

LOUETRI Saida

OKBI Zahia

Sous la supervision de :

LETOUFA Yassine

Année universitaire 2014 – 2015

Remerciement

La louange est à Allah, qui nous a facilité l'accomplissement de ce travail de recherche chose ne peut être qu'avec la volonté de Dieu -à lui la toute puissance et la Majesté- et que la louange initiale et finale appartient à Allah, Seigneur des mondes

*Aussi, il nous fait plaisir que nous, au commencement de ce travail, présentons nos grands remerciements, estimations et reconnaissances à notre encadreur puissant " **letoufa yassine**" de nous avoir encouragé moralement la durée de recherche et que ce travail est le fruit de ces encouragements*

Nous présentons nos véritables remerciements à toute personne, du proche ou du loin, qui nous a donné un coup de main, à fin de terminer ce travail de recherche

En fin, nous remercions vivement nos familles pour l'aide matérielle et morale durant la période de préparation

"Que la Grace et la paix soient sur notre prophète Muhammad ainsi que sur sa famille et ses compagnons".

Table des matières

Introduction	1
Notations et coventions	2
1 Quelques rappels de calcul différentiel et analyse convexe	2
1.1 Rappel de calcul différentiel	2
1.1.1 Différentiabilité on ordre un	2
1.1.2 Différentiabilité d'ordre deux	4
1.1.3 Continuité et différentiabilité	4
1.1.4 Différentiabilité et gradient	6
1.1.5 Formules de Taylor	6
1.2 Convexité	9
1.2.1 Fortement Convexe et Strictement Convexe	9
1.2.2 Exemples	10
1.2.3 Continuité des fonctions convexes	11
1.2.4 Caractérisations de la différentiable et convexité	12
2 Optimisation sans contraintes	15
2.1 Définitions et propriétés générales	15
2.1.1 <i>Existence et unicité d'un point de minimum</i>	16
2.1.2 Résultats d'existence et d'unicité	18
2.2 Conditions d'optimalité	19
2.2.1 Conditions nécessaires du premier ordre	19

2.2.2	Conditions du deuxième ordre	20
2.3	Convexité, ellipticité	22
3	Algorithmes d'optimisation sans contrainte	24
3.1	Méthodes de relaxation	24
3.2	Méthodes de descente	25
3.2.1	Algorithme du gradient à pas fixe	26
3.2.2	Algorithme du gradient à pas optimal	28
3.2.3	Algorithme du gradient à pas variable	
3.3	Algorithmes du gradient conjugué	30
3.4	Méthodes de Newton et Quasi-Newton	33
	Bibliographie	33

$C(\mathbb{R}^N, \mathbb{R}) :$	l'ensemble des fonctions continues de \mathbb{R}^N dans \mathbb{R} .
$C^1(E, F) :$	l'ensemble des fonctions continument différentiables d'ordre 1.
$C^2(E, F) :$	l'ensemble des fonctions continument différentiables d'ordre 2.
$\mathcal{L}(E, F) :$	l'ensemble des applications linéaires de E dans F .
$E, F :$	deux espaces vectoriels normés.
$T_x :$	la différentielle de f au point x .
$\nabla f(x) :$	le gradient de f .
$\partial f(x) :$	la dérivée partielle de f .
$H_f(x) :$	la matrice hessienne de f en x .
$J_f(x) :$	la matrice jacobienne de f en x .
$M_{p,N}(\mathbb{R}) :$	l'espace des matrices de p lignes et N colonnes.
$\ \cdot\ :$	la norme.
$ \cdot :$	la valeur absolue.
$\mathbb{k} :$	sous ensemble de \mathbb{R}^N .
$B(\bar{x}, \alpha) :$	la boule ouverte de centre \bar{x} et de rayon α .
$\text{rang}(C) :$	rang de la matrice C .
$\text{card}(I^*(\bar{x})) :$	cardinal d'une partie $I^*(\bar{x})$.
$(\cdot, \cdot)_N :$	produit scalaire euclidien sur \mathbb{R}^N .
$\inf_{x \in K} f(x) :$	la borne inférieure de fonction f sur l'ensemble K .
$\sup_{x \in K} f(x) :$	la borne supérieure de fonction f sur l'ensemble K .
$\min :$	le minimum.
$\max :$	le maximum.

Introduction générale

En général optimiser signifie le fait de chercher une configuration optimale d'un système, c'est à dire, chercher la meilleure configuration parmi tous les configurations possibles du système et ceci, par rapport à un critère donné.

Nous étudions dans ce travail les problème de programmation convexe et différentiel .Un problème de programmation convexe est un problème de minimisation d'une fonctionnelle convexe sur une partie convexe de \mathbb{R}^N .

Révenons à notre travail , cette mémoire est composé de trois chapitres :

D'une façon plus précise dans le premier chapitre . On trouvera: quelques rappels de calcul différentiel, analyse convexe .

Tant qu'au deuxième chapitre nous considérons l'optimisation sans contraintes: définitions et propriétés générale, existence et unicité d'un point de minimum , les résultats d'existence et d'unicité des solutions de ces problèmes, c'est-à-dire des conditions nécessaires et suffisantes d'optimalité.

Dans le dernier chapitre, on obtient la méthode de relaxation qui permet d'un nombre fini des étapes. qui transforme le problème de minimisation d'une fonctionnelle de \mathbb{R} , méthode de descente .la construction effective d'algorithmes permettant d'approcher un solution \bar{x} de problème (p) et nous donnons algorithme du gradient à pas fixe, algorithme du gradient à pas optimal, algorithme du gradient à pas variable et algorithmes du gradient conjugué et méthodes de Newton et Quasi-Newton.

Chapitre 1

Quelques rappels de calcul différentiel et analyse convexe

1.1 Rappel de calcul différentiel

1.1.1 Différentiabilité on ordre un

Soient E et F deux espaces normés, Ω un ouvert de E et une fonction $f : \Omega \rightarrow F$.

Différentiabilité au sens de Gâteaux

Définition 1.1.1 (*Dérivées directionnelle*) On dit que f a une dérivée directionnelle en $x \in \Omega$ (ici Ω n'est pas forcément) dans la direction $h \in E$ si, pour $t > 0$ suffisamment petit $x + th \in \Omega$ et si la limite

$$\lim_{t \rightarrow 0} \frac{1}{t} (f(x + th) - f(x))$$

exist, on la note

$$f'(x, h) = \lim_{t \rightarrow 0} \frac{1}{t} (f(x + th) - f(x)).$$

Remarque 1.1.1 : Pour les fonctions à valeurs dans \mathbb{R} , il arrivera que l'on admette des dérivées directionnelles valant $\pm\infty$, c'est à-dire de prendre la limite dans $\overline{\mathbb{R}}$.

Définition 1.1.2 (*Gâteaux différentiabilité*) On dit qu'une fonction f est GÂTEAUX - différentiable en $x \in \Omega$ si elle admet une dérivée directionnelle en x suivant toutes les directions $h \in E$ et si l'application

$$h \in E \rightarrow f'(x, h) \in F.$$

est linéaire continue. On note $f'(x)$ cet opérateur. on a alors

$$\forall h \in E, f'(x) \cdot h = f'(x, h).$$

Dérivée directionnelle (ou de Gateau) $J : E \rightarrow \mathbb{R}$ admet au point x une dérivée suivant la direction y si et seulement si

$$\lim_{t \rightarrow 0} \frac{J(x + ty) - J(x)}{t}.$$

existe; on la note $J'(x, y)$

Différentiabilité au sens de Fréchet

Définition 1.1.3 [2] (*Fréchet différentiabilité*) On dit que f est fréchet-différentiable en $x \in \Omega$ s'il existe un opérateur continu L de E dans F tel que

$$\lim_{\substack{\|h\|_E \rightarrow 0 \\ \|h\|_E > 0}} \frac{1}{\|h\|_E} (f(x + h) - f(x) - Lh) = 0.$$

L'opérateur L est appelé de dérivée de f en x .

Soit $\langle \cdot, \cdot \rangle$ un produit scalaire sur E et $F = \mathbb{R}$.

On définit le gradient de f en x comme l'unique vecteur $\nabla f(x) \in E$:

$$\langle h \nabla f(x), h \rangle = f'(x, h), \forall h \in E.$$

Définition 1.1.4 $J : E \rightarrow \mathbb{R}$ est différentiable (ou dérivée au sens de fréchet) en x si et seulement s'il existe une forme linéaire notée $J'(x)$ telle que pour tout

$$y \in E, \left| J(x + y) - J(x) - J'(x)y \right| = o(\|y\|) \text{ quand } y \rightarrow 0$$

$$\text{ou } \lim_{\|y\| \rightarrow 0} \frac{|J(x + y) - J(x) - J'(x)y|}{\|y\|} = 0.$$

Définition 1.1.5 Si une fonction $J : E \rightarrow \mathbb{R}$ admet des dérivées partielles dans un voisinage de x et que ces dérivées partielles sont continues en x , alors J est différentiable en x (ou dérivable au sens de Fréchet) et l'on a :

$$J'(x)y = \sum \frac{\partial j}{\partial x_i}(x_i) y_i.$$

1.1.2 Différentiabilité d'ordre deux

Définition 1.1.6 Supposons que $f : \Omega \rightarrow F$ soit deux fois différentiable (pour une définition rigoureuse, voir [1]).

Propriétés

$f'(x) \cdot (h, k)$ est la dérivée directionnelle de $x \rightarrow f(x) \cdot h$ dans la direction k

$$f'(x) \cdot (h, k) = \lim_{t \rightarrow 0^+} \frac{1}{t} (f(x + tk) \cdot h - f(x) \cdot h).$$

L'application

$$(h, k) = f''(x) \cdot (h, k)$$

est bilinéaire et symétrique.

Soit $\langle \cdot, \cdot \rangle$ un produit scalaire sur E et $F = \mathbb{R}$. On définit le hessien de f en x comme l'unique opérateur linéaire symétrique $\nabla^2 f(x)$ sur E tel que

$$\langle \nabla^2 f(x) h, k \rangle = f''(x) \cdot (h, k) : \forall (h, k) \in E^2$$

1.1.3 Continuité et différentiabilité

Dans la suite f est une fonction convexe propre définie sur un ouvert convexe C de \mathbb{R}^n

Lemme 1.1.1 (*Monotonie des accroissements*)

Soit $x \in C$ et d une direction de \mathbb{R}^n .

Alors, la fonction $q(t) = \frac{f(x + td) - f(x)}{t}$, définie pour des réels suffisamment petits, est une fonction croissante de t .

Démonstration. Prenons un intervalle $[h, k]$ avec $0 < h \leq k$. On a: $x + h d = (1 - \alpha)x + \alpha(x + k d)$ avec $\alpha = \frac{h}{k} < 1$. La convexité implique $f(x + h d) \leq \left(1 - \frac{h}{k}\right) f(x) + \frac{h}{k} f(x + k d)$ qui s'écrit

$$\frac{f(x + h d) - f(x)}{h} \leq \frac{f(x + k d) - f(x)}{k}.$$

■

- Le même raisonnement s'applique pour des pas négatifs.

Ce résultat implique, en étudiant les limites à gauche et à droite de $q(t)$ quand t tend vers 0, l'existence de dérivées directionnelles en tout point de C , définies par :

$$f'(x, d) = \lim_{t \rightarrow 0} q(t) = \inf_{t > 0} q(t).$$

Observez que l'infimum est borné car, C étant un ouvert, il existe $t < 0$ tel que

$$f'(x, d) \geq q(t).$$

Théorème 1.1.1 Si f est différentiable sur l'ouvert convexe C , alors les 3 affirmations suivantes sont équivalentes

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

Démonstration 1. Montrons qu'il implique que f est convexe. Cette inégalité peut s'écrire

$$\forall y \in C, \sup_{x \in C} \{f(x) + \langle \nabla f(x), y - x \rangle\} \leq f(y).$$

qui est en fait une égalité car on peut échanger les rôles de x et y . Donc, f est convexe tant que supremum d'une famille de fonctions affines..

2. Montrons maintenant que (2) implique (1) par l'absurde. Supposons un segment $[x, y]$ tel que $\langle \nabla f(x), y - x \rangle < f(y) - f(x)$. Grâce au théorème de Rolle, il existe un $z = (1 - \lambda)x + \lambda y$, $\lambda \in [0, 1]$, satisfaisant $f(y) = f(x) + \langle \nabla f(z), y - x \rangle$. On en déduit

$$\langle \nabla f(z) - \nabla f(x), z - x \rangle = \langle \nabla f(z) - \nabla f(x), (y - x) \rangle < 0,$$

ce qui contredit (2) pour le couple (x, z) .

Les autres implications sont immédiates.

Monotonie de l'opérateur Gradient : la relation (2) signifie que l'opérateur Gradient $\nabla f : C \rightarrow \mathbb{R}^n$ est monotone sur C .

Fonctions fortement convexes différentiable [7]

Proposition 1.1.1 *Soit une fonction f de \mathbb{R}^n dans \mathbb{R} fortement convexe de constante sur un ensemble convexe C et différentiable sur C ; alors*

$$\forall x, y \in C : \langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \alpha \|x - y\|^2$$

1.1.4 Différentiabilité et gradient

Soit $J : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle. J est différentiable au point x si et seulement s'il existe $\nabla J(x) \in \mathbb{R}^n$ et $h \mapsto \varepsilon(h, x)$ vérifiant

$$J(x + h) = J(x) + [\nabla J(x)]^T h + \|h\| \varepsilon(h, x), \quad \forall h \in \mathbb{R}^n,$$

avec

$$\varepsilon(h, x) \xrightarrow{h \rightarrow 0} 0.$$

Lorsqu'il existe, le gradient peut être exprimé à l'aide de dérivées partielles

$$\nabla J(X) = \begin{bmatrix} \partial_{x_1} J(x) \\ \partial_{x_2} J(x) \\ \dots \\ \partial_{x_n} J(x) \end{bmatrix}$$

Soit $J : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonctionnelle. J est de classe C^1 .

- J est différentiable au point x ,
- Chacune des composant de ∇J est continue. [2]

1.1.5 Formules de Taylor

Pour une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, les formules de Taylor s'écrivent

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \dots + \frac{h^n}{n!}f^{(n)}(x) + \sigma(h^{n+1}). \quad (1.1.1)$$

- La dérivée $f'(x)$ au point x définit le terme linéaire de ce développement, $h \rightarrow f'(x)h$.
- La dérivée $f''(x)$ définit le terme quadratique de ce développement, $h \rightarrow \frac{1}{2}h^2 f''(x)$.

$$g(h) = \sigma(h^p) \Leftrightarrow \begin{cases} \exists a > 0, \exists M > 0, |h| < a \Rightarrow \left| \frac{g(h)}{h^p} \right| \leq M : \\ \text{"} g(h) \text{ tend vers 0 avec } h \text{ aussi vite que } h^p \text{"} \end{cases}$$

$$g(h) = o(h^p) \Leftrightarrow \begin{cases} \lim_{h \rightarrow 0} \frac{g(h)}{h^p} = 0. \\ \text{"} g(h) \text{ tend vers 0 avec } h \text{ plus vite que } h^p \text{"} \end{cases}$$

pour une fonction $f: \mathbb{R}^p \rightarrow \mathbb{R}$, les formules de Taylor s'écrivent

$$f(x+h) = f(x) + \sum_{i=1}^p h_i \partial_{x_i} f(x) + \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p h_j h_k \partial_{x_j x_k}^2 f(x) + \sigma(\|h^3\|).$$

$$= f(x) + \sum_{j=1}^p h_j \partial_{x_j} f(x) + \frac{1}{2} \sum_{j=1}^p h_j \sum_{k=1}^p h_k \partial_{x_k x_j}^2 f(x) + \sigma(\|h^3\|).$$

• $J_f(x)$ = Matrice Jacobienne = $[\partial_{x_1} f(x), \dots, \partial_{x_p} f(x)] = (\nabla f(x))^T$.

Le gradient de f est le vecteur transposé de sa matrice jacobienne.

• $H_f(x)$ = Matrice Hessienne = $[\partial_{x_i x_j}^2 f(x)]_{1 \leq i, j \leq p} \in M_p(\mathbb{R})$. elle est symétrique

$$f(x+h) = f(x) + J_f(x)h + \frac{1}{2}h^T H_f(x)h + \sigma(\|h^3\|). \quad (1.1.2)$$

• La notation AB désigne le produit matriciel. Dans le cas du produit d'une matrice ligne par une matrice colonne, on peut une notation de produit scalaire de deux matrices (vecteurs) colonnes

$$(1.1.2) \Leftrightarrow f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle H_f(x)h, h \rangle + \sigma(\|h^3\|). \quad (1.1.3)$$

Pour une fonction $f: \mathbb{R}^p \rightarrow \mathbb{R}^n$, les formules de Taylor s'écrivent

$$f(x+h) = f(x) + L_f(h) + \frac{1}{2}Q_f(h) + \sigma(\|h^3\|). \quad (1.1.4)$$

- $x, h \in \mathbb{R}^p$ et $f(x) \in \mathbb{R}^n$: (1.1.4) est une égalité dans \mathbb{R}^n . $L_f(h)$ désigne un terme (vecteur de \mathbb{R}^n) linéaire en h , $Q_f(h)$ désigne un terme quadratique en h . Pour les expliciter, on peut écrire une formule de type (1.1.2) pour chaque composante f_i de f ; pour $1 \leq i \leq n$,

$$f_i(x+h) = f_i(x) + J_{f_i}(x)h + \frac{1}{2}h^T H_{f_i}(x)h + \sigma(\|h^3\|).$$

· En rassemblant toutes les lignes et suivant matrice-vecteur, on obtient $L_f(h) = J_f(x)h$ avec

$$J_f(x) = \text{matrice Jacobienne} = \begin{bmatrix} J_{f_1}(x) \\ \dots \\ J_{f_n}(x) \end{bmatrix} \in M_{n,p}(\mathbb{R}). \quad (1.1.5)$$

· Pour chaque i , le terme quadratique du f_i est donné par $\frac{1}{2}h^T H_{f_i}(x)h$.
En rassemblant toutes les lignes et en factorisant h à droite, on obtient:

$$\begin{bmatrix} h^T H_{f_1}(x)h \\ \dots \\ h^T H_{f_n}(x)h \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^p h_k \partial_{x_k x_1}^2 f_1(x) & \dots & \sum_{k=1}^p h_k \partial_{x_k x_p}^2 f_1(x) \\ \dots & \dots & \dots \\ \sum_{k=1}^p h_k \partial_{x_k x_1}^2 f_n(x) & \dots & \sum_{k=1}^p h_k \partial_{x_k x_p}^2 f_n(x) \end{bmatrix} \begin{bmatrix} h_1 \\ \dots \\ h_p \end{bmatrix} \quad (1.1.6)$$

· On a obtenu une expression de $Q_f(h)$ qui permet d'écrire (1.1.5) sous la forme

$$f(x+h) = f(x) + \begin{bmatrix} J_{f_1}(x) \\ \dots \\ J_{f_n}(x) \end{bmatrix} h + \frac{1}{2} \begin{bmatrix} h^T H_{f_1}(x)h \\ \dots \\ h^T H_{f_n}(x)h \end{bmatrix} h + \sigma(\|h^3\|).$$

[2]

1.2 Convexité

1.2.1 Fortement Convexe et Strictement Convexe

Définition 1.2.1 . Un ensemble $U \subset \mathbb{R}^n$ est dit convexe si $\forall x, y \in U$ on a $[x, y] \subset U$ (quelque soient deux points dans U , tout le segment qui les unit est dans U).

Définition 1.2.2 . Soit $U \in \mathbb{R}^n$ un ensemble convexe et $f: U \rightarrow \mathbb{R}$ une fonction

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

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

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

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

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

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

4. On dit que f est **concave** (respectivement **strictement concave**, respectivement **fortement concave**). si f est convexe (respectivement strictement convexe, respectivement fortement convexe). [2]

Remarque 1.2.1 : Il est facile de voir qu'on a: fortement convexe \implies strictement convexe

\implies convexe. Les réciproques ne sont pas vraies en général; par exemple une application affine $f(x) = Ax + b$ est convexe (et aussi concave) mais elle n'est pas strictement convexe

(ni strictement concave) donc elle n'est pas fortement convexe (ni fortement concave).

Proposition 1.2.1 : Soient $U \subset \mathbb{R}^n$ un ensemble convexe, $p \in \mathbb{N}^*$, $f_1, f_2, \dots, f_p: U \rightarrow \mathbb{R}$ des fonctions convexes et $\gamma_1, \gamma_2, \dots, \gamma_n$ des constantes strictement positives.

$$\text{Posons } f = \gamma_1 f_1 + \gamma_2 f_2 + \dots + \gamma_p f_p.$$

Alors, on a :

1. La fonction f est convexe (donc toute combinaison linéaire avec des coefficients strictement positifs de fonctions convexes est convexe).
2. Si au moins l'une des fonctions f_1, \dots, f_p est strictement convexe alors f est strictement convexe.
3. Si au moins l'une des fonctions f_1, \dots, f_p est fortement convexe alors f est fortement convexe.

1.2.2 Exemples

Exemple 1.2.1 Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ donnée par $f(x) = x^2$. Comme $\nabla f(x) = f'(x) = 2x$ on a $\forall x, y \in \mathbb{R}$:

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

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

C'est encore plus facile si on utilise f :

$$f''(x) = 2 \equiv \beta > 0.$$

donc f est fortement convexe.

Exemple 1.2.2 Plus généralement, si $n = 1$ et $\Omega \subset \mathbb{R}$ un intervalle ouvert :

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

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

est une fonction fortement convexe.

Si $\Omega = \mathbb{R}$ alors la fonction f est elliptique:

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

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

Exemple 1.2.3 Le cas général ($n \in \mathbb{N}^*$)

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

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

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

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

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

1.2.3 Continuité des fonctions convexes

Donnons maintenant quelques propriétés (topologiques) importantes des fonctions convexes.

Dans tout ce qui suit J est une fonction de H vers $\mathbb{R} \cup \{+\infty\}$. On suppose que le domaine de J est non vide.

Soit $J: H \rightarrow \mathbb{R} \cup \{+\infty\}$ convexe. Il est équivalent de dire que:

- i) Il existe un ouvert non vide sur lequel J est majorée par une constante a réelle et ne vaut pas constamment $-\infty$
- ii) J est propre, $\text{int}(\text{dom}(J))$ est d'intérieur non vide et J est continue sur l'intérieur de son domaine.

Pour démontrer le résultat réciproque nous aurons besoin du lemme suivant:

Lemme 1.2.1 *Si, au voisinage d'un point $u_0 \in H$, une fonction convexe J est majorée par une constante finie, alors J est continue en u_0 .*

Nous démontrerons ce lemme ensuite.

Supposons donc (i) vérifié: Ω est donc inclus dans l'intérieur du domaine de J qui est en particulier non vide. Donc J est propre. Soit $u \in \Omega$

tel que $J(u) > -\infty$. D'après le lemme, J sera continue en u , donc finie sur un voisinage de u . Pour tout $v \in \text{int}(\text{dom}(J))$, il existe $\rho > 1$ tel que $w = u + \rho(v - u)$ appartienne encore à $\text{int}(\text{dom } J)$ car l'intérieur d'un convexe est convexe (ce que nous admettrons). L'homothétie h de centre w et de rapport $1 - \frac{1}{\rho}$ transforme u en v et Ω en un ouvert $h(\Omega)$ contenant v . Pour tout v' de $h(\Omega)$, on a par convexité

$$j(v') \leq \frac{\rho-1}{\rho} j(h^{-1}(v')) + \frac{1}{\rho} j(w) \leq \frac{\rho-1}{\rho} a + \frac{1}{\rho} j(w).$$

Par conséquent : tout point v de $\text{int}(\text{dom } J)$ possède un voisinage $h(\Omega)$ sur lequel J est majorée par une constante finie. D'après le lemme, J est continue en v .

Démonstration du lemme: On se ramène par translation au cas où $u_0 = 0$ et $J(0) = 0$. Soit V un voisinage de l'origine tel que $J(u) \leq a < +\infty$, pour tout u de V . Posons $W = V \cap -V$ et donnons nous $\varepsilon \in]0, 1[$. Si:

$u \in \varepsilon W$, on par convexité

$$\frac{u}{\varepsilon} \in w, \text{ donc } j(u) \leq (1-\varepsilon) j(0) + \varepsilon j\left(\frac{u}{\varepsilon}\right) \leq \varepsilon a.$$

$$-\frac{u}{\varepsilon} \in w, \text{ donc } j(u) \geq (1+\varepsilon) j(0) - \varepsilon j\left(-\frac{u}{\varepsilon}\right) \geq \varepsilon a.$$

Finalement

$$\forall u \in \varepsilon W \quad |J(u)| \leq \varepsilon a,$$

d'où la continuité. [4]

1.2.4 Caractérisations de la différentiable et convexité

Théorème 1.2.1 Soit Ω un ouvert de \mathbb{R}^n , C convexe, et $f : \Omega \rightarrow \mathbb{R}$, différentiable. On a les équivalences suivantes:

a. f est convexe sur C si et seulement si

$$\forall x, x_0 \in C; f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle.$$

b. f est strictement convexe sur C si et seulement si

$$\forall x, x_0 \in C; x \neq x_0; f(x) > f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle.$$

c. f est α -convexe sur C si et seulement si

$$\forall x; x_0 \in C, f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{\alpha}{2} \|x - x_0\|^2.$$

Théorème 1.2.2 *Le membre de droite des inégalités a et b donne l'équation de l'hyperplan tangent au graphe en x_0 , tandis que celui de l'inégalité c donne l'équation d'un paraboloïde tangent au graphe en x_0 .*

Démonstration 1. Montrons d'abord la première équivalence. Supposons donc f convexe.

Soient $x, x_0 \in C$, on a donc, pour tout $\theta \in]0, 1[$.

$$f(\theta x + (1 - \theta)x_0) \leq \theta f(x) + (1 - \theta)f(x_0).$$

soit

$$\frac{f(\theta x + (1 - \theta)x_0) - f(x_0)}{\theta} \leq f(x) - f(x_0).$$

En passant à la limite quand θ tend vers 0, on obtient le résultat voulu.

Réciproquement, supposons les inégalités vérifiées et montrons que f est convexe.

Soient donc $x, y \in C, \theta \in]0, 1[$. Posons $x_0 = \theta x + (1 - \theta)y$. On a donc:

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle.$$

$$f(y) \geq f(x_0) + \langle \nabla f(x_0), y - x_0 \rangle.$$

En multipliant la première inégalité par θ , la seconde par $1 - \theta$, et en les ajoutant, on obtient exactement :

$$\theta f(x) + (1 - \theta)f(y) \geq f(x_0) = f(\theta x + (1 - \theta)y).$$

Pour le point 2, la réciproque se montre comme ci-dessus en remplaçant les inégalités larges par des inégalités strictes. Par contre, on ne peut pas calquer la preuve du sens direct, car les inégalités strictes deviendraient larges à la limite $\theta \rightarrow 0$. On peut cependant appliquer le résultat du point 1. Soit donc f supposée strictement convexe, et deux points $x, x_0 \in C, x \neq x_0$. On a alors, pour tout $\theta \in]0, 1[$.

$$f(x_\theta) < f(x) + (1 - \theta)f(x_0), \text{ ou } x_\theta = \theta x + (1 - \theta)x_0.$$

Mais on a aussi, d'après le résultat 1. qu'on vient de montrer,

$$f(x_\theta) \geq f(x_0) + \langle \nabla f(x_0), x_\theta - x_0 \rangle.$$

En comparant ces deux inégalités, et en simplifiant $f(x_0)$, on obtient:

$$\theta \langle \nabla f(x_0), x - x_0 \rangle < \theta(f(x) - f(x_0)).$$

Il reste à diviser par θ pour obtenir le résultat recherché.

Pour le troisième point, on utilise la remarque

f α -convexe $\Leftrightarrow f - \frac{\alpha}{2} \|\cdot\|^2$ convexe $\Leftrightarrow \forall x, x_0 \in C$,

$$f(x) - \frac{\alpha}{2} \|x\|^2 \geq f(x_0) - \frac{\alpha}{2} \|x_0\|^2 + \langle \nabla f(x_0) - \alpha x_0, x - x_0 \rangle.$$

Après simplification, on arrive à l'équivalence voulue. [3]

Chapitre 2

Optimisation sans contraintes

2.1 Définitions et propriétés générales

Introduction: Dans ce chapitre nous allons étudier les problèmes d'optimisation évoqués dans le chapitre précédent dans le cas où $H = \mathbb{R}^n$ muni du produit scalaire usuel et lorsqu'il n'y a pas de contraintes : On effectue la minimisation de la fonction J sur tout l'espace. Nous considérons donc le problème formulé de la façon suivante

$$(P) \quad \left\{ \begin{array}{l} \min J(x) \\ x \in \mathbb{R}^n \end{array} \right\}$$

où J est une fonction de \mathbb{R}^n vers $\mathbb{R} \cup \{+\infty\}$.

C'est un problème de minimisation sans contraintes.

Trouver $u^* \in \mathbb{R}^n$ tel que $J(u^*) \leq J(u), \forall u \in \mathbb{R}^n$.

On supposera que u^* existe (éventuellement qu'il est unique) et on se propose de trouver une approximation numérique de u^* , en construisant une suite $\{u^{(k)}\} k \in \mathbb{N} \subset \mathbb{R}^n$ telle que $u^{(k)} \rightarrow u^*$ pour $k \rightarrow +\infty$.

Définition 2.1.1 Soit $U \subset \mathbb{R}^n, u^* \in U$ et $f : U \rightarrow \mathbb{R}$.

1. On dit que u^* est un point de minimum absolu (ou global) de f sur U si

$$f(u) \geq f(u^*); \forall u \in U.$$

2. On dit que u est un point de minimum relatif (ou local) de f sur U si $\exists V$ voisinage de u^* dans \mathbb{R}^n , tel que $f(u) \geq f(u^*)$; $\forall u \in U \cap V$:

3. On dit que u^* est un point de maximum absolu (respectivement relatif) de f sur U si u est un point de minimum absolu (respectivement relatif) de $-f$ sur U .

4. On dit que u^* est un point d'extremum absolu (respectivement relatif) de f sur U si u^* est : soit un point de minimum absolu (respectivement relatif) de f sur U , soit un point de maximum absolu (respectivement relatif) de f sur U .

Dans toute la suite du cours on parlera uniquement de la minimisation d'une fonction f , pour la maximisation, faire la minimisation de la fonction $-f$. [2]

2.1.1 Existence et unicité d'un point de minimum

Théorème 2.1.1 (Existence)

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

1. U est borné, ou
2. U est non bornée et f est une fonction coercive.

Alors il existe au moins un point de minimum de f sur U (c'est à dire, $\exists u^* \in U$ tel que $f(u^*) \leq f(u), \forall u \in U$)

Démonstration. On distingue deux cas :

Cas1) L'ensemble U est borné.

Alors comme U est aussi fermé, U est compact. Comme f est continue, le Théorème de Weierstrass nous assure que f est bornée sur U et elle atteint ses bornes. Donc il existe au moins un point de minimum absolu de f sur U .

Cas 2) L'ensemble U est non borné.

Soit $a \in U$ et considérons l'ensemble

$$E = \{x \in U, f(x) \leq f(a)\}. [5]$$

Remarque: $a \in E$.

Il est facile de montrer :

1. E est fermé

(car $E = f^{-1}([-\infty, f(a)])$), donc E est l'image inverse d'un intervalle fermé par une fonction continue)

2. E est borné

(supposons le contraire : alors il existe une suite $x_k \in E$ avec $\|x_k\| \rightarrow +\infty$ pour $k \rightarrow +\infty$. Comme f est coercive, ceci entraîne $f(x_k) \rightarrow +\infty$ ce qui est absurde, car

$$f(x_k) \leq f(a), \forall k \in \mathbb{N}.)$$

On déduit alors que E est un ensemble compact dans \mathbb{R}^n . Du Théorème de Weierstrass, $\exists u^* \in E$ tel que

$$f(u^*) \leq f(u), \forall u \in E.$$

Mais d'autre part, on a

$$f(u^*) < f(u), \forall u \in U - E$$

(car $f(u^*) \leq f(a) < f(u), \forall u \in U - E$).

Ceci prouve que u^* est un point de minimum absolu de f sur U , ce qui finit la preuve.

Théorème 2.1.2 . (Unicité)

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

Démonstration. On va raisonner par absurd. Soient $u_1, u_2 \in U$ avec $u_1 \neq u_2$ deux points de minimum de f sur U . Nous avons donc :

$$f(u_1) = f(u_2) \leq f(u), \forall u \in U. \tag{2.1.1}$$

Comme f est strictement convexe, on a

$$f\left(\frac{1}{2}u_1 + \frac{1}{2}u_2\right) < \frac{1}{2}f(u_1) + \frac{1}{2}f(u_2) = f(u_1)$$

(car $f(u_1) = f(u_2)$) et ceci contredit (2.1.1) . [5]

Corollaire 2.1.1 . (Existence et unicité)

Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $U \subset \Omega$ un ensemble fermé et convexe et $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 sur Ω et fortement convexe sur U . Alors il existe un unique point de minimum de f sur U .

Démonstration. Existence : Il y a deux possibilités :

1. Soit U est borné et alors l'existence est immédiate du Théorème (2.1.1) .
2. Soit U est non borné et alors comme conséquence de la Proposition f est coercive sur U et donc l'existence résulte encore du Théorème (2.1.1) .

Unicité : Comme f est strictement convexe sur U , l'unicité est une conséquence immédiate du Théorème (2.1.2) .

Remarque : On utilisera très souvent ce corollaire avec $\Omega = \mathbb{R}^n$ et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de classe C^1 et fortement convexe sur \mathbb{R}^n (c'est à dire f est une fonction elliptique).

Alors pour tout ensemble $U \subset \mathbb{R}^n$ convexe et fermé on a l'existence et l'unicité d'un point de minimum de f sur U .

2.1.2 Résultats d'existence et d'unicité

Avant d'étudier les propriétés de la solution (ou des solutions) de (P) il faut s'assurer de leur existence. Nous donnerons ensuite des résultats d'unicité.

Définition 2.1.2 On dit que $J : H \rightarrow \mathbb{R}$ est coercive si

$$\lim_{\|x\| \rightarrow +\infty} J(x) = +\infty$$

Ici $\|\cdot\|$ désigne la norme de l'espace de Hilbert H . Dans le cas où $H = \mathbb{R}^n$ les normes sont toutes équivalentes et $\|\cdot\|$ désigne une norme quelconque de \mathbb{R}^n . On notera $\|\cdot\|_p$ ($p \in \mathbb{N}$) la norme ℓ_p

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n \quad \|x\|_p = \left[\sum_{i=1}^n |x_i|^p \right]^{\frac{1}{p}}$$

La norme infinie de \mathbb{R}^n est

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Rappelons que parmi les normes ci-dessus, seule la norme ℓ_2 munit \mathbb{R}^n d'une structure d'espace de Hilbert. Nous noterons (\cdot, \cdot) le produit scalaire associé. [3]

2.2 Conditions d'optimalité

2.2.1 Conditions nécessaires du premier ordre

Les conditions que nous allons donner sont des conditions différentielles qui portent sur la dérivée de la fonction à minimiser. On va donc se restreindre au cas des fonctions Gâteaux-différentiables.

Théorème 2.2.1 (*Condition nécessaire d'optimalité du premier ordre*)

Soit H un espace de Hilbert réel et $J : H \rightarrow \mathbb{R}$ une fonctionnelle Gâteaux-différentiable sur H . Si

x^* réalise un minimum (global ou local) de J sur H alors

$$\nabla J(x^*) = 0. \quad (2.2.1)$$

Démonstration - Si x^* réalise un minimum de J sur H alors

$$\forall x \in B(x^*, \rho) \quad J(x^*) \leq J(x),$$

où $B(x^*, \rho)$ est une boule de rayon $\rho > 0$ centrée en x^* .

Soit $h \in H, h \neq 0$; on peut trouver $t_h = \frac{\rho}{\|h\|} > 0$ tel que

$$\forall t \in]0, t_h[\quad x^* + th \in B(x^*, \rho)$$

et donc

$$\forall t \in]0, t_h[\quad J(x^*) \leq J(x^* + th),$$

Or J est Gâteaux-différentiable en x^* , donc

$$\lim_{t \rightarrow +0} \frac{J(x^* + th) - J(x^*)}{t} = (\nabla J(x^*), h).$$

Donc

$$\forall h \in H \quad (\nabla J(x^*), h) \geq 0, \quad (2.2.2)$$

c'est-à-dire $\nabla J(x^*) = 0$.

Définition 2.2.1 *Un point x^* de H vérifiant $\nabla J(x^*) = 0$ est appelé point critique ou point stationnaire .*

La relation $\nabla J(x^*) = 0$ est aussi appelée équation d'Euler .

Ce théorème n'a pas de sens si la fonction J n'est pas différentiable

Théorème 2.2.2 *(CNS du premier ordre dans le cas convexe)*

Soit $J : H \rightarrow \mathbb{R}$ Gâteaux- différentiable et convexe sur H . Un point x réalise un minimum global de J sur H si et seulement si $\nabla J(x^*) = 0$.

Démonstration - On a vu que la condition est toujours nécessaire. Montrons qu'elle est suffisante.

Soit $x \in H$ tel que $\nabla J(x) = 0$. Comme J est convexe on peut utiliser le théorème (1.2.1) du chapitre 1 et on obtient :

$$\forall x \in H \quad J(x) \geq J(x^*) + (\nabla J(x^*), x - x^*) = J(x^*).$$

On a donc immédiatement le fait que x réalise un minimum de J sur H . [4]

2.2.2 Conditions du deuxième ordre

Nous commençons par une condition nécessaire permettant de préciser encore les éventuels minima.

Théorème 2.2.3 *(Condition nécessaire du second ordre)*

On suppose que x^* est un minimum (local) de J et que J est deux fois dérivable sur H . Alors

- i. $\nabla J(x^*) = 0$ et
- ii. $\forall x \in H \quad (D^2 J(x^*)x, x) \geq 0$

Démonstration - (i) a déjà été vue: montrons (ii).

Soit $x \in H$. Appliquons la formule de Taylor à la fonction $\varphi : t \rightarrow \varphi(t) = J(x^* + tx)$.

Comme

$\nabla J(x^*) = 0$ on obtient

$$0 \leq J(x^* + tx) - J(x^*) \frac{t^2}{2} (D^2 J(x^*) x, x) + o(t^2).$$

Après division par t^2 , on fait tendre t vers 0 et on a le résultat voulu.

Dans le cas où $H = \mathbb{R}^n$, (ii) est équivalent à dire que la matrice Hessienne de J en x^* : $D^2 J(x^*)$ est semi-définie positive.

Rappelons qu'un critère pour que $D^2 J(x^*)$ (qui est une matrice symétrique) soit semi-définie positive est que toutes ses valeurs propres soient positives ou nulles.

La réciproque du théorème précédent est fautive (il suffit de penser à la fonction $t \rightarrow t^3$ pour s'en convaincre). Nous pouvons toutefois donner une réciproque sous forme de condition suffisante du second ordre plus forte (pour un résultat plus faible) de ce qui précède. [3]

Théorème 2.2.4 (*Condition suffisante du second ordre*)

Soit J deux fois dérivable sur H vérifiant $\nabla J(x^*) = 0$ et

$$\exists \alpha > 0, \forall x \in H (D^2 J(x^*) x, x) \geq \alpha \|x\|^2 \quad (2.2.3)$$

Alors la fonction J admet un minimum local strict en x^* .

Démonstration - Soit x dans H . On utilise de nouveau la formule de Taylor appliquée à la fonction

$\varphi : t \rightarrow \varphi(t) = J(x^* + tx)$. Nous avons

$$J(x^* + tx) - J(x^*) = \frac{t^2}{2} (D^2 J(x^*) x, x) + o(t^2) \geq \frac{t^2}{2} \alpha \|x\|^2 + o(t^2).$$

Ceci montre que x^* réalise un minimum local strict de J la condition 2 est une condition d'ellipticité locale [3]

Exemple 2.2.1 Soit $J : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $J(x_1, x_2) = 3x_1^4 - 4x_1^2 x_2 + x_2^2$.

Il est facile de voir que le point $(0, 0)$ est un point critique mais on ne peut pas conclure immédiatement

car la matrice Hessienne $D^2J(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$, est seulement semi-définie positive puisque ses valeurs propres sont 0 et 2. On peut toutefois vérifier que ce point critique n'est ni un minimum ni un maximum en remarquant que $J(0,1) = 1 > J(0,0) = 0 > J(0.5,0.5) = -0.0625$.

On peut aussi remarquer que J n'est pas coercive puisque $J(n, n^2) = 0$.

2.3 Convexité, ellipticité

La convexité (voir le cours d'analyse pour les propriétés) est un cadre naturel pour la minimisation de fonctions de plusieurs variables; nous utiliserons une classe plus réduite de fonctions pour lesquelles des démonstrations sont un peu moins techniques: fonctions elliptiques (notions différente de celle des fonctions elliptiques de l'analyse classique) ou -convexes.

Proposition 2.3.1 . *La première propriété définit les fonctions elliptiques ; cette première condition est impliquée par les autres ; sous hypothèse de différentiabilité, la première condition entraîne les 2 suivantes ; enfin pour les fonctions 2 fois*

différentiables, toutes les conditions sont équivalentes. l'une des conditions équivalentes suivantes avec $\alpha > 0$:

1. $\forall u \in \mathbb{R}^n, \forall v \in \mathbb{R}^n, \forall \delta \in [0,1]$:

$$J((1-\delta)u + \delta v) \leq (1-\delta)J(u) + \delta J(v) - \frac{\alpha}{2} \delta(1-\delta) \|uv\|^2 \mathbb{R}^n$$

2. si J est différentiable, $\forall u \in \mathbb{R}^n, \forall v \in \mathbb{R}^n$,

$$J(v) \geq J(u) + J'(u)(v-u) + \frac{\alpha}{2} \|u-v\|^2 \mathbb{R}^n.$$

3. si J est différentiable, $\forall u \in \mathbb{R}^n, \forall v \in \mathbb{R}^n$,

$$(J'(v) - J'(u), v-u) \geq \alpha \|u-v\|^2 \mathbb{R}^n.$$

4. si J est 2 fois différentiable, $\forall u \in \mathbb{R}^n, \forall w \in \mathbb{R}^n$,

$$J''(u)(w,w) \geq \alpha \|w\|^2 \mathbb{R}^n.$$

Montrons que la deuxième condition implique la troisième:

$$J(v) \geq J(u) + J'(u)(v - u) + \frac{\alpha}{2} \|u - v\|^2 \mathbb{R}^n \quad J(u) \geq J(v) + J'(v)(u - v) + \frac{\alpha}{2} \|u - v\|^2 \mathbb{R}^n.$$

d'où la troisième condition en additionnant.

Réciproquement, introduisons $j(t) = J(u + t(v - u))$ on a : $J'(t) = J'(u + t(v - u))(v - u)$, donc la troisième condition donne

$$J'(t) - J'(0) \geq \alpha t \|u - v\|^2 \mathbb{R}^n \quad \text{d'où la deuxième condition en intégrant de 0 à 1. [1]}$$

Chapitre 3

Algorithmes d'optimisation sans contrainte

Soit $E = \mathbb{R}^n$ et $f \in C(E, \mathbb{R})$. On suppose qu'il existe $\bar{x} \in E$ tel que $f(\bar{x}) = \inf_E f$.

On cherche à calculer \bar{x}

(si f est de classe C^1 , on a nécessairement $\nabla f(\bar{x}) = 0$). On va donc maintenant développer des algorithmes (ou méthodes de calcul) du point \bar{x} qui réalise le minimum de f .

Définition 3.0.1 (*algorithme*)

Un algorithme est défini par une application A de \mathbb{R}^n dans \mathbb{R}^n permettant la génération d'une suite d'éléments de \mathbb{R}^n par la formule: $x_0 \in \mathbb{R}^n$ donné, $k = 0$ Etape d'initialisation $x_{k+1} = A(x_k)$, $k = k + 1$ Itération k .

Ecrire un algorithme n'est ni plus ni moins que se donner une suite $(x_k)_k \in \mathbb{N}$ de \mathbb{R}^n ; étudier la convergence de l'algorithme, c'est étudier la convergence de la suite $(x_k)_k \in \mathbb{N}$.

3.1 Méthodes de relaxation

On suppose ici

$$k = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_N, b_N] = \prod_{i=1}^N [a_i, b_i]. \quad (3.1.1)$$

L'algorithme dans ce cas rassemble à l'algorithme de relaxation pour minimisation sans contraintes.

Algorithme:

On se donne $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_N^{(k)})^T$ et on va calculer $x^{(k+1)} = (x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_N^{(k+1)})^T$ en N pas successifs. On a: $f(x_1^{(k+1)}, x_2^{(k)}, \dots, x_N^{(k)}) = \min_{y \in [a_1, b_1]} f(y, x_2^{(k)}, \dots, x_N^{(k)})$,

$$f(x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k)}, \dots, x_N^{(k)}) = \min_{y \in [a_2, b_2]} f(x_1^{(k+1)}, y, x_3^{(k)}, \dots, x_N^{(k)}),$$

$$f(x_1^{(k+1)}, \dots, x_N^{(k+1)}) = \min_{y \in [a_N, b_N]} f(x_1^{(k+1)}, \dots, x_{N-1}^{(k+1)}, y).$$

Comme dans le cas sans contraintes, on a le théorème de convergence suivant :

. Si $f : \mathbb{R}^N \rightarrow \mathbb{R}$ est une fonction elliptique et l'ensemble K est donné par $(C, 0)$, alors la méthode de relaxation pour la minimisation de f sur K est bien définie et converge [2]

3.2 Méthodes de descente

Définition 3.2.1 Soient $f \in C(E, \mathbb{R})$ et $E = \mathbb{R}^N$.

1. Soit $x \in E$, on dit que $\omega \in E \setminus \{0\}$ est une direction de descente en x s'il existe $\rho_0 > 0$ tel que

$$f(x + \rho \omega) \leq f(x) : \forall \rho \in [0, \rho_0].$$

2. Soit $x \in E$, on dit que $\omega \in E \setminus \{0\}$ est une direction de descente stricte en x s'il existe $\rho_0 > 0$ tel que

$$f(x + \rho \omega) < f(x) : \forall \rho \in]0, \rho_0].$$

3. Une "méthode de descente" pour la recherche de \bar{x} tel que $f(\bar{x}) = \inf_E f$ consiste à construire une suite $(x_n)_n$ de la manière suivante :

(a) Initialisation $x_0 \in E$;

(b) Itération n : on suppose $x_0 \dots x_n$ connus ($n \geq 0$);

i. On cherche ω_n direction de descente stricte de x_n .

ii. On prend $x_{n+1} = x_n + \rho_n \omega_n$ avec $\rho_n > 0$ "bien choisi". [6]

Proposition 3.2.1 Soient $E = \mathbb{R}^N$, $f \in C^1(E, \mathbb{R})$, $x \in E$ et $\omega \in E \setminus \{0\}$; alors

1. si ω direction de descente en x alors $\omega \cdot \nabla f(x) \leq 0$
2. si $\nabla f(x) \neq 0$ alors $\omega = -\nabla f(x)$ est une direction de descente stricte en x .

Démonstration 1. Soit $\omega \in E \setminus \{0\}$ une direction de descente en x alors par définition,

$$\exists \rho_0 > 0 \text{ tel que } f(x + \rho \omega) \leq f(x), \forall \rho \in [0, \rho_0].$$

Soit φ la fonction de \mathbb{R} dans \mathbb{R} définie par: $\varphi(\rho) = f(x + \rho \omega)$. On a $\varphi \in C^1(\mathbb{R}, \mathbb{R})$ et $\varphi'(\rho) = \nabla f(x + \rho \omega) \cdot \omega$. Comme ω est une direction de descente, on peut écrire: $\varphi(\rho) \leq \varphi(0), \forall \rho \in [0, \rho_0]$, et donc

$$\forall \rho \in]0, \rho_0[, \frac{\varphi(\rho) - \varphi(0)}{\rho} \leq 0.$$

en passant à la limite lorsque ρ tend vers 0, on déduit que $\varphi'(0) \leq 0$, c.à.d. $\nabla f(x) \cdot \omega \leq 0$.

2. Soit $\omega = -\nabla f(x) \neq 0$. On veut montrer qu'il existe $\rho_0 > 0$ tel que si $\rho \in]0, \rho_0]$ alors $f(x + \rho \omega) < f(x)$

ou encore que $\varphi(\rho) < \varphi(0)$ où φ est la fonction définie en 1 ci-dessus. On a:

$\varphi'(0) = \nabla f(x) \cdot \omega = -|\nabla f(x)|^2 < 0$. Comme φ' est continue, il existe $\rho_0 > 0$ tel que si $\rho \in]0, \rho_0]$ alors $\varphi'(\rho) < 0$. Si $\rho \in]0, \rho_0]$

alors $\varphi(\rho) - \varphi(0) = \int_0^\rho \varphi'(t) dt < 0$, et on a donc bien $\varphi(\rho) < \varphi(0)$ pour tout $\rho \in]0, \rho_0]$, ce qui prouve que

ω est une direction de descente stricte en x .

3.2.1 Algorithme du gradient à pas fixe

Soient $f \in C^1(E, \mathbb{R})$ et $E = \mathbb{R}^N$. On se donne $\rho > 0$.

Initialisation: $x_0 \in E$,

Itération n : x_n connu, ($n \geq 0$) (C, 1)

$$\omega_n = -\nabla f(x_n),$$

$$x_{n+1} = x_n + \rho \omega_n.$$

Théorème 3.2.1 (Convergence du gradient à pas fixe) Soient $E = \mathbb{R}^N$ et $f \in C^1(E, \mathbb{R})$ On suppose que :

1. $\exists \alpha > 0$ tel que $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2, \forall (x, y) \in E^2,$
2. $\exists M > 0$ tel que $\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|, \forall (x, y) \in E^2,$

alors:

1. f est strictement convexe,
2. $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty,$
3. il existe un et un seul $\bar{x} \in E$ tel que $f(\bar{x}) = \inf_E f$ (conséquence de 1. et 2.),
4. si $0 < \rho < \frac{2\alpha}{M^2}$ alors la suite $(x_n)_{n \in \mathbb{N}}$ construite par (C, 1) converge vers \bar{x} lorsque

$n \rightarrow +\infty.$

Démonstration.

1. $f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2} \|y - x\|^2$

■

$$\varphi(t) = f(x + t(y - x))$$

$$\varphi(0) = f(x); \varphi(1) = f(y) : \varphi'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle$$

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle dt$$

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle dt - \int_0^1 \langle \nabla f(x), y - x \rangle dt$$

$$= \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle dt$$

$$\int_0^1 \frac{1}{t} \langle \nabla f(x + t(y - x)) - \nabla f(x), t(y - x) \rangle dt \geq \int_0^1 \frac{1}{t} \alpha \|t(y - x)\|^2 dt = \alpha \|y - x\|^2 \int_0^1 t dt = \frac{\alpha}{2} \|y - x\|^2 \text{ donc:}$$

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2} \|x - y\|^2$$

2. $f(y) - f(0) \geq \langle \nabla f(x), y \rangle + \frac{\alpha}{2} \|y\|^2$

$$f(y) \geq f(0) - \|\nabla f(0)\| \|y\| + \frac{\alpha}{2} \|y\|^2$$

$$\lim_{\|y\| \rightarrow +\infty} f(y) = +\infty. f \text{ coercive}$$

$$f(y) - f(x) > \langle \nabla f(x), y - x \rangle : \forall x, y \in \mathbb{R}^N / x \neq y. \text{ donc } f \text{ strictement convexe}$$

existe un est un seul

$$\bar{x} \in \mathbb{R}^N \text{ tel que } f(\bar{x}) = \min_{x \in \mathbb{R}^N} f(x). [6]$$

Exemple 3.2.1 On considère la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par:

$$f(x, y) = x^2 + 2y^2 - xy - 3y - x + 4$$

calculer le premier itéré $((x^{(1)}, y^{(1)}))$ donnée par l'algorithme du gradient à pas fixe (GPF)

L'algorithme du gradient à pas fixe (GPF)

Initialisation $x^{(0)} = (0, 0) \in \mathbb{R}^2$

Itération n ; $x^{(n)}$ connu

$$w^{(0)} = (3, 1) \text{ on en déduit } x^{(1)} = (3\rho, -\rho) = \left(\frac{3}{2}, \frac{1}{2}\right) \text{ et } f(x^{(1)}) = \frac{5}{2} = 2.5$$

$$\nabla f(0, 0) = (-3, -1)$$

3.2.2 Algorithme du gradient à pas optimal

L'idée de l'algorithme du gradient à pas optimal est d'essayer de calculer à chaque itération le paramètre qui minimise la fonction dans la direction de descente donnée par le gradient.

Soient

$f \in C^1(E, \mathbb{R})$ et $E = \mathbb{R}^N$, cet algorithme s'écrit :

Initialisation: $x_0 \in \mathbb{R}^N$.

Itération n : x_n connu.

On calcule $\omega_n = -\nabla f(x_n)$.

On choisit $\rho_n \geq 0$ tel que (C, 2)

$$f(x_n + \rho_n \omega_n) \leq f(x_n + \rho \omega_n) \quad \forall \rho \geq 0.$$

On pose $x_{n+1} = x_n + \rho_n \omega_n$.

Les questions auxquelles on doit répondre pour s'assurer du bien fondé de ce nouvel algorithme sont les suivantes :

1. Existe-t-il ρ_n tel que $f(x_n + \rho_n \omega_n) \leq f(x_n + \rho \omega_n), \forall \rho \geq 0$?
2. Comment calcule-t-on ρ_n ?
3. La suite $(x_n)_{n \in \mathbb{N}}$ construite par l'algorithme converge-t-elle ?

La réponse aux questions 1. et 3. est apportée par le théorème suivant :

Théorème 3.2.2 (Convergence du gradient à pas optimal)

Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$ telle que $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$. Alors:

1. La suite $(x_n)_{n \in \mathbb{N}}$ est bien définie par (C, 2). On choisit $\rho_n > 0$ tel que

$$f(x_n + \rho_n \omega_n) \leq f(x_n + \rho \omega_n): \quad \forall \rho \geq 0.$$

(ρ_n existe mais n'est pas nécessairement unique).

2. La suite $(x_n)_{n \in \mathbb{N}}$ est bornée et si $(x_{nk})_{nk \in \mathbb{N}}$ est une sous suite convergente, i.e. $x_{nk} \rightarrow x$ lorsque $k \rightarrow +\infty$,

on a nécessairement $\nabla f(x) = 0$. De plus si f est convexe on a $f(x) = \inf_{\mathbb{R}^N} f$

3. Si f est strictement convexe on a alors $x_n \rightarrow \bar{x}$ quand $n \rightarrow +\infty$, avec $f(\bar{x}) = \inf_{\mathbb{R}^N} f$
[6]

Exemple 3.2.2 *l'algorithme du gradient à pas optimale s'écrit:*

Initialisation $x^{(0)} = (0, 0) \in \mathbb{R}^2$

Itération n: $x^{(n)}$ connu

on calcule $\omega^{(0)} = (3, 1)$ le ρ_0 optimale minimum la fonction $\varphi(\rho) = f(x^{(0)} + \rho\omega^{(0)}) = f(3\rho, \rho)$

on doit donc avoir $\varphi'(\rho_0) = 0$ calculons

$$\varphi'(\rho) = \nabla f(x^{(0)} + \rho\omega^{(0)}) \cdot \omega^{(0)} = \begin{bmatrix} 11\rho & -3 \\ -\rho & -1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 3(11\rho - 3) + (-\rho - 1) = 32\rho - 10$$

on en déduit que $\rho_0 = \frac{5}{16} \geq 0$, On obtient alors :

$$x^{(1)} = x^{(0)} + \rho_0\omega^{(0)} = \left(\frac{15}{16}, \frac{5}{16}\right), \text{ et } f(x^{(1)}) = 2,4375$$

3.2.3 Algorithme du gradient à pas variable

Dans ce nouvel algorithme, on ne prend pas forcément le paramètre optimal pour ρ , mais on lui permet d'être variable d'une itération à l'autre. L'algorithme s'écrit:

Initialisation: $x_0 \in \mathbb{R}^N$.

Itération: On suppose x_n connu; soit $\omega_n = -\nabla f(x_n)$ où: $\omega_n \neq 0$

(si $\omega_n = 0$ l'algorithme s'arrête).

On prend $\rho_n > 0$ tel que $f(x_n + \rho_n \omega_n) < f(x_n)$. (C, 3)

On pose $x_{n+1} = x_n + \rho_n \omega_n$.

3.3 Algorithmes du gradient conjugué

La méthode du gradient conjugué a été posée pour la minimisation de fonctionnelles quadratiques, c'est-à-dire de fonctionnelles de la forme

$$f(x) = \frac{1}{2}Ax \cdot x - b \cdot x,$$

où $A \in M_n(\mathbb{R})$ est une matrice symétrique définie positive et $b \in \mathbb{R}^n$.

Définition 3.3.1 (*Vecteurs conjugués*) Soit $A \in M_N(\mathbb{R})$ une matrice symétrique définie positive,

1. Deux vecteurs v et w de $\mathbb{R}^N \setminus \{0\}$ sont dits A -conjugués si $Av \cdot w = w \cdot Av = 0$.
2. Une famille $(w^{(1)}, \dots, w^{(p)})$ de $\mathbb{R}^N \setminus \{0\}$ est dite A -conjuguée si $w^{(i)} \cdot Aw^{(j)} = 0$ pour

tout couple

$$(i, j) \in \{1, \dots, p\}^2 \text{ tel que } i \neq j.$$

Proposition 3.3.1 Soit $A \in M_N(\mathbb{R})$ une matrice symétrique définie positive, $(w^{(1)}, \dots, w^{(p)})$ une famille de \mathbb{R}^N , alors :

1. si la famille $(w^{(1)}, \dots, w^{(p)})$ est A -conjuguée alors elle est libre;
2. dans le cas où $p = N$, si la famille $(w^{(1)}, \dots, w^{(p)})$ est A -conjuguée alors c'est une base de \mathbb{R}^N .

Démonstration. : Le point 2. est immédiat dès qu'on a démontré le point 1. Supposons donc que $(w^{(1)}, \dots, w^{(p)})$ ■

est une famille A -conjuguée, i.e. $w^{(i)} \neq 0, \forall i$ et $w^{(i)} \cdot Aw^{(j)} = 0$ si $i \neq j$; soit $(\alpha_i)_{i=1, \dots, p} \subset \mathbb{R}$, supposons que

$$\sum_{i=1}^p \alpha_i w^{(i)} = 0 \text{ on a donc } \sum_{i=1}^p \alpha_i w^{(i)} \cdot Aw^{(j)} = 0 \text{ et donc } \alpha_j w^{(j)} \cdot Aw^{(j)} = 0. \text{ or } w^{(j)} \cdot Aw^{(j)} \neq 0$$

car $w^{(j)} \neq 0$ et A est symétrique définie positive. On en déduit que $\alpha_j = 0$ pour $j = 1, \dots, p$. La famille

$$(w^{(1)}, \dots, w^{(p)}) \text{ est donc libre.}$$

Définition 3.3.2 (*Méthode du gradient conjugué*) Soit $A \in M_N(\mathbb{R})$ une matrice symétrique définie positive,

$$b \in \mathbb{R}^N \text{ et } f(x) = \frac{1}{2}Ax \cdot x - b \cdot x.$$

Initialisation

Soit $x^{(0)} \in \mathbb{R}^N$, et soit $r^{(0)} = b - Ax^{(0)} = -\nabla f(x^{(0)})$.

- 1) Si $r^{(0)} = 0$, alors $Ax^{(0)} = b$ et donc $x^{(0)} = \bar{x}$, auquel cas l'algorithme s'arrête.
- 2) Si $r^{(0)} \neq 0$, alors on pose $w^{(0)} = r^{(0)}$, et on choisit ρ_0 optimal dans la direction $w^{(0)}$.

On pose alors $x^{(1)} = x^{(0)} + \rho_0 w^{(0)}$.

Itération $1 \leq n \leq N - 1$:

On suppose $x^{(0)}, \dots, x^{(n)}$ et $w^{(0)}, \dots, w^{(n-1)}$ connus et on pose $r^{(n)} = b - Ax^{(n)}$.

- 1) Si $r^{(n)} = 0$ on a $Ax^{(n)} = b$ donc $x^{(n)} = \bar{x}$ auquel cas l'algorithme s'arrête.
- 2) Si $r^{(n)} \neq 0$, alors on pose $w^{(n)} = r^{(n)} + \lambda_{n-1} w^{(n-1)}$ avec λ_{n-1} tel que $w^{(n)} \cdot Aw^{(n-1)} = 0$, et on choisit ρ_n optimal dans la direction $w^{(n)}$;
On pose alors $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$.

Théorème 3.3.1 Soit A une symétrique définie positive, $A \in M_N(\mathbb{R})$, $b \in \mathbb{R}^N$ et $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$ si $(x^{(n)})_{n=0, \dots, p}$ avec : $p \leq N$ définie par algorithme du gradient conjugué alors $x^{(n)} = \bar{x}$ avec $A\bar{x} = b$. [6]

Lemme 3.3.1 Sous les hypothèses et notations précédents on a :

$$\rho_n = \frac{r^{(n)} \cdot r^{(n)}}{\omega^{(n)} \cdot A\omega^{(n)}}. \quad (1)$$

$$r^{(n)} = r^{(n-1)} + \rho_{n-1} A\omega^{(n-1)}. \quad (2)$$

$$r^{(n)} \cdot r^{(n-1)} = 0. \quad (3)$$

$$\lambda_{n-1} = \frac{r^{(n)} \cdot r^{(n)}}{r^{(n-1)} \cdot r^{(n-1)}}. \quad (4)$$

1. Comme ρ_n est le paramètre optimal on sait

$$\rho_n = \frac{r^{(n)} \cdot w^{(n)}}{Aw^{(n)} \cdot w^{(n)}}.$$

Or par définition, $w^{(n)} = r^{(n)} + \lambda_{n-1}w^{(n-1)}$, et donc $w^{(n)} \cdot r^{(n)} = r^{(n)} \cdot r^{(n)} + \lambda_{n-1}w^{(n-1)} \cdot r^{(n)}$.

Il ne reste

plus à remarquer que $w^{(n-1)} \cdot r^{(n)} = 0$ (d'après le choix de ρ_{n-1} optimal)

2. Par définition, $x^{(n)} = x^{(n-1)} + \rho_{n-1}w^{(n-1)}$, donc $Ax^{(n)} = Ax^{(n-1)} + \rho_{n-1}Aw^{(n-1)}$, ce

qui entraîne

$$r^{(n)} = r^{(n-1)} + \rho_{n-1}Aw^{(n-1)}.$$

3. Par définition, et grâce à 2:

$$r^{(n)} \cdot r^{(n-1)} = r^{(n-1)} \cdot r^{(n-1)} + \rho_{n-1}Aw^{(n-1)} \cdot r^{(n-1)}.$$

Or $w^{(n-1)} = r^{(n-1)} + \lambda_{n-1}w^{(n-2)}$, et donc $r^{(n-1)} = w^{(n-1)} - \lambda_{n-1}w^{(n-2)}$. On en déduit que

$$r^{(n)} \cdot r^{(n-1)} = r^{(n-1)} \cdot r^{(n-1)} - \rho_{n-1}Aw^{(n-1)} \cdot w^{(n-1)} - \rho_{n-1}\lambda_{n-1}Aw^{(n-1)} \cdot w^{(n-2)}.$$

Or $Aw^{(n-1)} \cdot w^{(n-2)} = 0$ (ρ_{n-1} est optimal)

$$r^{(n-1)} \cdot r^{(n-1)} - \rho_{n-1}Aw^{(n-1)} \cdot w^{(n-1)} = 0$$

4. Par définition,

$$\lambda_{n-1} = -\frac{r^{(k)} \cdot Aw^{(k-1)}}{w^{(k-1)} \cdot Aw^{(k-1)}}.$$

Or par (2)

$$Aw^{(k-1)} = \frac{1}{n-1}(r^{(k-1)} - r^{(k)}).$$

On conclut grâce à (3) et (1).

Proposition 3.3.2 *Sous les hypothèses les définitions précédents et si, $n \in \mathbb{N}$ tel que $1 \leq n \leq N$, si $r^{(q)} \neq 0$*

pour $0 \leq q \leq n$, les propriétés suivantes sont vérifiées:

1. $r^{(k)} \cdot w^{(q)} = 0, \forall q = 0, \dots, n-1$,
2. $\text{Vect}(r^{(0)}, \dots, r^{(k)}) = \text{Vect}(r^{(0)}, \dots, A^n r^{(0)})$,
3. $\text{Vect}(w^{(0)}, \dots, w^{(k)}) = \text{Vect}(r^{(0)}, \dots, A^n r^{(0)})$,
4. $w^{(k)} \cdot Aw^{(q)} = 0, \forall q = 0, \dots, n-1$,
5. $r^{(k)} \cdot r^{(q)} = 0, \forall q = 0, \dots, n-1$,

où $Vect(w^{(0)}, \dots, w^{(k)})$ désigne l'espace vectoriel engendré par les vecteurs $w^{(0)}, \dots, w^{(k)}$. En particulier, la

famille $(w^{(0)}, \dots, w^{(n-1)})$ est A -conjuguée.

L'espace $Vect(r^{(0)}, \dots, A^n r^{(0)})$ est appelé espace de Krylov.

3.4 Méthodes de Newton et Quasi-Newton

Soit $f \in C^2(\mathbb{R}^N, \mathbb{R})$ et $g = \nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)$. On a dans ce cas

$$f(x) = \inf_{\mathbb{R}^N} f \Rightarrow g(x) = 0.$$

Si de plus f est convexe alors on a $g(x) = 0 \Rightarrow f(x) = \inf_{\mathbb{R}^N} f$. Dans ce cas d'équivalence, on peut employer la

méthode de Newton pour minimiser f en appliquant l'algorithme de Newton pour chercher un zéro de $g = \nabla f$.

On a $D(\nabla f) = H_f$ où $H_f(x)$ est la matrice hessienne de f en x . La méthode de Newton s'écrit dans ce cas:

Initialisation $x^{(0)} \in \mathbb{R}^N,$

Itération n $H_f(x^{(n)})(x^{(n-1)} - x^{(n)}) = -\nabla f(x^{(n)}).$

Remarque 3.4.1 (Sur l'implantation numérique) La convergence de la méthode de Newton est très rapide, mais

nécessite en revanche le calcul de $H_f(x)$, qui peut s'avérer impossible ou trop coûteux.

On va maintenant donner des variantes de la méthode de Newton qui évitent le calcul de la matrice hessienne.

Proposition 3.4.1 Soient $f \in C^1(\mathbb{R}^N, \mathbb{R})$, $x \in \mathbb{R}^N$ tel que $\nabla f(x) \neq 0$, et soit $B \in M_N(\mathbb{R})$ une matrice

symétrique définie positive; alors $w = -B \nabla f(x)$ est une direction de descente stricte en x .

Démonstration. On a: $w \cdot \nabla f(x) = -B \nabla f(x) \cdot \nabla f(x) < 0$ car B est symétrique définie positive et $\nabla f(x) \neq 0$ donc ■

w est une direction de descente stricte en x . En effet, soit φ la fonction de \mathbb{R} dans \mathbb{R} définie par $\varphi(\rho) = f(x + \rho w)$.

Il est clair que $\varphi \in C^1(\mathbb{R}, \mathbb{R})$, $\varphi'(\rho) = \nabla f(x + \rho w) \cdot w$ et $\varphi'(0) = \nabla f(x) \cdot w < 0$. Donc $\exists \rho_0 > 0$ tel que

$\varphi'(\rho) < 0$ si $\rho \in]0, \rho_0[$. Par le théorème des accroissements finis, $\varphi(\rho) < \varphi(0) \forall \rho \in]0, \rho_0[$ donc w est une

direction de descente stricte. [6]

Bibliographie

- [1] *B. Rousselet. Aide mémoire d'optimisation élémentaire. février 2006.*
- [2] *.ciuperca ionel sorin. cours d'optimisation. cours à l'ISFA, en , paris: 1984*
- [3] *H. Cartan, Cours de calcul différentiel , Hermann, Paris, 1977.*
- [4] *I.Ekeland - R.Temam, Analyse Convexe et Problèmes Variationnels, Dunod, Gauthier-Villars, Paris, 1974.*
- [5] *.P. G. ciarlet, B. Miara, j. M. Thomas. Exercices l'analyse numérique matricielle et d'optimisation. Masson, Paris. 1987.*
- [6] *.Raphaële Herbin. Cours d'Analyse numérique. Université Aix Marseille Licence de mathématiques. 18 novembre 2013.*
- [7] *R.T Rockafellar, Convex Analysis, Princeton U. , 1970.*