

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE
SCIENTIFIQUE



Université Ibn Khaldoun- Tiaret
FACULTÉ DES MATHÉMATIQUES ET INFORMATIQUE
DÉPARTEMENT DES MATHÉMATIQUES

MEMOIRE MASTER

Présentée par :

BOUHENNI ASMAA

MIMOUNI BOUCHRA

Spécialité : Mathématiques

Option : Analyse fonctionnelle et Applications

Intitulée

Méthode du gradient conjugué utilisant des recherches linéaires

Soutenue le 20/06/2022.

Devant le jury composé de :

Président: Pr. LARABI

Univ. Ibn Khaldoun Tiaret

Examineurs: Pr. MAZOUZ

Univ. Ibn Khaldoun Tiaret

Encadreur: Pr. REZZOUG

Univ. Ibn Khaldoun Tiaret

2021-2022

Remerciements

Avant toute chose, nous remercions **ALLAH** Le tout-Puissant de nous avoir donné force et
patience.

Nous exprimons notre profonde gratitude et nos remerciements,
à notre encadreur de mémoire **M. N.Rezzoug** pour avoir accepté de nous encadrer, pour son
enseignement, son support, ses encouragements, sa patience qu'il n'a jamais cessé de nous
apporter tout au long de ce travail.

Un remerciement spécial et sincère au **M. A.Larabi**, Doyen de la Faculté de Mathématiques
et d'Informatique qui nous a fait l'honneur d'accepter la présidence du jury de ce mémoire.

Un grand merci à **M. K.Maazouz** qui nous a accordé l'honneur d'accepter l'examen de
notre mémoire.

Nous remercions aussi **M. Mostefaoui** pour son aide et sa sympathie.

Dédicace

Je dédie ce modeste travail :

À mes chers parents.

Je vous remercie pour tout le soutien que vous me portez depuis mon enfance.

Puisse ALLAH, le Très Haut, vous accorde santé, bonheur et longue vie.

À mes frères.

À mes soeurs.

Je vous souhaite un avenir plein de joie, de bonheur, de réussite .

À tous mes professeurs.

À tous mes adorables amis.

Bouhenni Asmaâ

Dédicace



...Je dédie ce travail :

À

Mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

À

Mes chères sœurs.

Mon frère.

Tous Mes Enseignants,

du département de Mathématiques qui ont contribué à mon formation.

Tous mes amis

surtout Tayeb kobci, Tayeb khedir, Sofiane belarbi, Saad boussoura, Asmaâ, Souad, Ahlam, Fatima, Hasna, Nadjat, Nadjat, Fatiha pour leur soutien tout au long de mon parcours universitaire.

Bouchra

Table des matières

Notation	1
INTRODUCTION	1
1 Préliminaires	4
1.1 Dérivée partielle	4
1.2 Dérivée directionnelle	4
1.3 Différentiabilité	5
1.4 Gradient	6
1.5 Matrice Hessienne	6
1.5.1 Matrice semi-définie positive	7
1.5.2 Matrice définie positive	7
1.6 Formules de Taylor	8
1.7 Inégalité d'accroissements finis	8
1.8 Convexité	10
1.8.1 Ensemble convexe	10
1.8.2 Fonction convexe	11
1.8.3 Fonction concave	11
1.9 Classification des problèmes d'optimisation	13
1.9.1 Forme générale d'un problème d'optimisation	13

1.9.2	Problème d'optimisation sans contraintes	14
1.9.3	Problème d'optimisation avec contraintes	14
1.10	Direction de descente	15
1.11	Conditions d'optimalité	16
1.11.1	Conditions nécessaires d'optimalité du premier ordre	16
1.11.2	Conditions nécessaires d'optimalité du second ordre	16
1.11.3	Conditions suffisantes d'optimalité	17
2	Recherches linéaires	19
2.1	Recherches linéaires	19
2.2	Recherche linéaire d'Armijo	22
2.2.1	Algorithme d'Armijo	22
2.3	Recherche linéaire de Goldstein	23
2.3.1	Principe	23
2.3.2	Algorithme de Goldstein	24
2.3.3	Programme en fortran	25
2.4	Recherche linéaire de Wolfe	27
2.4.1	Recherche linéaire de Wolfe forte	28
2.4.2	Algorithme de Wolfe forte	30
3	Méthode du gradient conjugué(G.C)	32
3.1	Méthode des directions conjuguées	32
3.1.1	Définitions et propriétés générales	32
3.1.2	Algorithme des directions conjuguées	33
3.2	Méthode du G.C. Cas quadratique	36
3.2.1	Algorithme	37
3.2.2	Propriétés	38
3.2.3	Convergence dans le cas quadratique	39

3.3	Méthode du G.C. Cas non quadratique	40
3.3.1	Algorithme	42
3.4	Convergence de la méthode du G.C	45
3.5	Théorème de Zoutendijk	46
3.5.1	Hypothèses C1 et C2 (de Lipschitz et de bornitude)	46
3.5.2	Théorème de Zoutendijk	47
3.5.3	Théorème de Zoutendijk et la convergence globale	47
3.6	Méthode de Fletcher-Reeves	49
3.7	Méthode de Dai-Yuan	51
3.8	Méthode de Dai-Yuan généralisée	52

Notation

\mathbb{R}	Ensemble des nombres réels.
$H(x)$	Matrice hessienne.
d_k	Direction de descente.
α_k	Pas de minimisation.
$f(\cdot)$	Fonction objective.
$\ x\ $	Norme d'un vecteur x .
$\langle \cdot, \cdot \rangle$	Produit scalaire.
x^t	Transposé d'un vecteur x .
∇f	Gradient de f .
$M_{m,n}$	Matrice de m lignes et n colonnes.
$C(\mathbb{R}^n, \mathbb{R})$	L'espace des fonctions continues de \mathbb{R}^n dans \mathbb{R}
\hat{x}	Solution optimale.
H^{-1}	Matrice inverse.
$\frac{\partial f}{\partial x}$	Dérivée de f suivant x .
$\frac{\partial^2 f}{\partial x_i \partial x_j}$	Dérivée partielle seconde de f .
(PQSC)	Problème quadratique sans contraintes.
(PNQSC)	Problème non quadratique sans contraintes.
(FR)	Fletcher Reeves.
(PRP)	Polak-Ribière-Polyak.
(DY)	Dai-Yuan.
(HZ)	Hestenes-Stiefel.

INTRODUCTION

L'optimisation est une branche mathématiques cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes qui consistent à minimiser ou maximiser une fonction sur un ensemble.

L'optimisation joue un rôle important en recherche opérationnelle (domaine à la frontière entre l'informatique et les mathématiques), en les mathématiques appliquées (fondamentales pour l'industrie et l'ingénierie), en analyse et en analyse numérique pour la recherche de stratégies dans le cadre de la théorie des jeux ou encore en théorie du contrôle optimal. Beaucoup de systèmes peuvent être décrits par un modèle mathématique sont optimisés.

Plus formellement l'optimisation est l'étude des problèmes qui sont de la forme :

Étant donné : une fonction $f : A \rightarrow \mathbb{R}$ d'un ensemble $A \subset \mathbb{R}^n$ dans l'ensemble \mathbb{R} .

Chercher : un élément x_0 de A tel que $f(x_0) \geq f(x)$ pour tous les x dans A (maximisation) ou tel que $f(x_0) \leq f(x)$ pour tous les x dans A (minimisation).

L'objectif de ce travail est de présenter et étudier **la méthode du gradient conjugué utilisant des recherches linéaires pour résoudre un problème d'optimisation.**

Ce travail est composé de 4 chapitres. Dans le premier, nous rappelons des définitions, propriétés et quelques théorèmes sur la dérivée partielle, la dérivée directionnelle, la différentiabilité, le gradient, la matrice hessienne, la convexité, l'inégalité d'accroissements finis et les conditions d'optimalité, introduisons aussi les formules de Taylor.

Le deuxième chapitre consacré aux recherches linéaires : d'Armijo, de Goldstein et de Wolfe. Dans ce chapitre, nous présentons les définitions, les algorithmes et les propriétés de ces recherches linéaires.

Dans le troisième chapitre nous traitons le part principal dans notre travail. Commençons par la méthode des directions conjuguées, puis passons à les deux cas de la méthode du gradient conjuguées, le cas quadratique et le cas non quadratique.

Dans les deux cas nous introduisons les définitions, les propriétés importantes et nous étudions aussi la convergence de ses algorithmes.

Dans l'étude de la convergence nous utilisons hypothèses C_1 et C_2 , théorème de Zoutendijk.

En fin, nous terminons notre travail par une synthèse des résultats de convergence en présentant la méthode de Fletcher-Reeves et de Dai-Yuan.

Chapitre 1

Préliminaires

Dans ce chapitre, on introduit les outils fonctionnels de base pour les notions utilisés dans ce mémoire.

1.1 Dérivée partielle

Définition 1.1.1 Soient $n \in \mathbb{N}^*$, U un ouvert de \mathbb{R}^n et f une fonction de \mathbb{R}^n vers \mathbb{R} définie sur U .

La fonction f est dite d'avoir une dérivée partielle à $a = (a_1, \dots, a_n)$, par rapport à la i ème variable si la limite

$$\lim_{t \rightarrow a_i} \frac{f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n) - f(a)}{t - a_i},$$

existe.

On note cette limite par

$$\frac{\partial f}{\partial x_i}(a) \quad \text{ou} \quad \partial_i f(a) \quad \text{ou} \quad D_i f(a).$$

1.2 Dérivée directionnelle

Les dérivées directionnelles sont des dérivées calculées suivant une direction .

Définition 1.2.1 Soient $n \in \mathbb{N}^*$, U un ouvert de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$, et $v \in \mathbb{R}^n$. La fonction f

est dite d'avoir une dérivée directionnelle au point $a \in U$ dans la direction v si la limite

$$\lim_{h \rightarrow 0} \frac{1}{h} [f(a + hv) - f(a)]$$

existe.

Proposition 1.2.1 (14) Soient $n \in \mathbb{N}^*$, $U \subset \mathbb{R}^n$ un ouvert de \mathbb{R}^n , $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable sur U . Alors

$$D_v f(a) = \nabla f(a)v.$$

1.3 Différentiabilité

Définition 1.3.1

1) Soit U un ouvert de \mathbb{R}^n et soit la fonction $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$.

On dit que f est différentiable en $a \in U$ s'il existe une application L linéaire continue de \mathbb{R}^n dans \mathbb{R} telle que

$$\lim_{h \rightarrow 0} \frac{f(a + h) - f(a) - L(h)}{\|h\|} = 0, \quad (1.1)$$

où $h = (h_1, \dots, h_n)$.

2) Lorsque l'application L existe, elle s'appelle la différentielle de f en a , on la note $Df(a)$, ainsi (1.1) devient

$$\lim_{h \rightarrow 0} \frac{f(a + h) - f(a) - Df(a)(h)}{\|h\|} = 0.$$

3) Si f est différentiable en tout point de U on dit que f est différentiable sur U .

Proposition 1.3.1 (14) Soient U un ouvert et $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$.

Si $f \in C^1(U)$ alors f est différentiable sur U .

1.4 Gradient

Si les dérivées partielles $\partial f/\partial x_i$ existent pour tout i , le gradient de f est défini de la façon suivante.

Définition 1.4.1 *On note par*

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

le gradient de f au point $x = (x_1, \dots, x_n)$.

Le gradient joue un rôle essentiel dans le développement et l'analyse des algorithmes d'optimisation.

Proposition 1.4.1 *(Gradient de la composée)*

Supposons qu'on a deux ouverts $\Omega \subset \mathbb{R}^n$ et $U \subset \mathbb{R}$ et deux fonctions $f : \Omega \rightarrow \mathbb{R}$ et $g : U \rightarrow \mathbb{R}$ avec en plus $f(\Omega) \subset U$ (on peut alors définir $g \circ f : \Omega \rightarrow \mathbb{R}$).

Supposons que f et g sont de classe C^1 .

Alors $g \circ f$ est aussi de classe C^1 avec

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

1.5 Matrice Hessienne

Si les dérivées partielles $(\partial^2 f/\partial x_i \partial x_j)$ existent pour tout $i, j \in N^*$, la matrice hessienne définie comme suit.

Définition 1.5.1 *Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$ une fonction dont toutes les dérivées partielles d'ordre deux.*

Alors la matrice

$$H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

est appelée matrice hessienne

Théorème 1.5.1 Soit $\Omega \subset \mathbb{R}^n$ un ouvert. Si $f \in C^2(\Omega)$ alors $\nabla^2 f(x)$ est une matrice symétrique $\forall x \in \Omega$.

Exemple 1.5.1 Soit $f(x_1, x_2) = (x_1)^2(x_1 + x_2)$.

On a

$$\nabla f(x) = \begin{pmatrix} 3x_1^2 + 2x_1x_2 \\ 2x_1 \end{pmatrix},$$

alors

$$H(x) = \begin{pmatrix} 6x_1 + 2x_2 & 2x_1 \\ 2x_1 & 0 \end{pmatrix}.$$

1.5.1 Matrice semi-définie positive

Définition 1.5.2 Soit M une matrice symétrique réelle d'ordre n .

Elle est dite **semi-définie positive** si elle vérifie l'une des 2 propriétés équivalentes suivantes :

1. Pour toute matrice colonne X à n éléments réels,

nous avons

$$x^T M x \geq 0.$$

2. Toutes les valeurs propres de M sont positives ou nulles.

1.5.2 Matrice définie positive

Définition 1.5.3 Soit M Une matrice symétrique dont les éléments sont des nombres réels.

M est dite définie positive si pour tout vecteur $x \in \mathbb{R}^n$ non nul on a $x^T M x > 0$.

1.6 Formules de Taylor

Formule de Taylor avec reste intégrale

Théorème 1.6.1 Soit U un ouvert de \mathbb{R}^n et $a, h \in \mathbb{R}^n$ tel que $[a, a + h] \subset U$.

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de $C^2(\mathbb{R}^n)$.

Alors on a

$$f(a + h) = f(a) + \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(a) + \sum_{i,j=1}^n h_i h_j \int_0^1 (1-t) \frac{\partial^2 f}{\partial x_i \partial x_j}(a + th) dt.$$

Formule de Taylor-Lagrange

Théorème 1.6.2 Soit U un ouvert de \mathbb{R}^n et $a, h \in \mathbb{R}^n$ tels que $[a, a + h] \subset U$.

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction deux fois différentiable sur U . La formule de Taylor-Lagrange d'ordre 2 est

$$f(a + h) = f(a) + \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(a) + \frac{1}{2} \sum_{i,j=1}^n h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a + \theta h) \quad \text{avec } \theta \in]0, 1[.$$

Formule de Taylor-young

Théorème 1.6.3 Soit U un ouvert de \mathbb{R}^n et $a \in U$. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de classe C^2 , la formule de Taylor-young d'ordre 2 s'écrit

$$f(a + h) = f(a) + \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i}(a) + \frac{1}{2} \sum_{i,j=1}^n h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + \varepsilon(h) \|h\|^2.$$

1.7 Inégalité d'accroissements finis

Théorème 1.7.1 Soit f une application continue sur un segment $[a, b]$ et dérivable sur $]a, b[$.

S'il existe $M \in \mathbb{R}$ tel que $f'(x) \leq M$ pour tout $x \in]a, b[$ alors

$$f(b) - f(a) \leq M(b - a).$$

Preuve

Soit $\epsilon > 0$. Considérons la fonction h définie sur $[a, b]$ par

$$h(x) = f(x) - f(a) - M(x - a) - \epsilon(x - a).$$

Si l'on montre que $h(b) \leq \epsilon$ alors la proposition est démontrée. En effet

$$h(b) \leq \epsilon \Leftrightarrow \frac{f(b) - f(a) - M(b - a)}{b - a + 1} \leq \epsilon$$

d'où, l'inégalité étant vraie pour tout $\epsilon > 0$, $\frac{f(b) - f(a) - M(b - a)}{b - a + 1} \leq 0$,

où $b - a + 1 > 0$ et donc $f(b) - f(a) \leq M(b - a)$.

Montrons donc que $h(b) \leq \epsilon$.

Soit $A = \{x \in [a, b] \mid h(x) \leq \epsilon\}$.

Comme $h(a) = 0$, $a \in A$ et la fonction h étant continue en a , il existe $d \in]a, b]$ tel que $h(\bar{d}) \leq \epsilon/2$, donc $A \neq \{a\}$.

Soit $c = \sup A \in]a, b]$.

Il existe une suite (x_n) d'éléments de A qui converge vers c .

Pour tout n , $h(x_n) \leq \epsilon$ et donc la fonction h étant continue en c , $h(c) \leq \epsilon$ d'où $c \in A$.

Supposons que $c < b$.

Il existe $t \in]c, b]$ telque $\frac{f(t) - f(c)}{t - c} \leq f'(c) + \epsilon$, car

$$\lim_{x \rightarrow c, x > c} \frac{f(x) - f(c)}{x - c} = f'(c).$$

On a

$$f(t) - f(c) \leq (t - c)f'(c) + (t - c)\epsilon \leq M(t - c) + (t - c)\epsilon.$$

Comme $c \in A$,

$$f(c) - f(a) \leq M(c - a) + \epsilon(c - a) + \epsilon,$$

d'où par addition,

$$f(t) - f(a) \leq M(t - a) + \epsilon(t - a) + \epsilon,$$

ce qui signifie $t \in A$. C'est absurde.

Donc $c = b$ et $h(b) \leq \epsilon$. ■

1.8 Convexité

1.8.1 Ensemble convexe

Définition 1.8.1 Soit S un sous ensemble de \mathbb{R}^n .

S est dit convexe si

$$\forall x, y \in U, \quad \forall t \in [0, 1] \text{ on a } tx + (1 - t)y \in U.$$

Autrement dit U est convexe si pour tous points x, y de U le segment $[x, y]$ est inclu dans U (c'est à dire, $\forall x, y \in U$ on a $[x, y] \subset U$).

Définition 1.8.2 Soit $x_1, x_2, \dots, x_n \in \mathbb{R}^n$ et $\lambda_j \geq 0$ tels que $\sum_{j=1}^k \lambda_j = 1$. Toute expression de la forme $\sum_{j=1}^k \lambda_j x_j$ s'appelle combinaison convexe des points x_j .

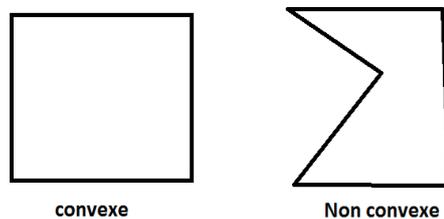


FIGURE 1.1 –

Proposition 1.8.1 *i) L'intersection d'une famille quelconque de convexes est convexe.*

ii) Le produit cartésien de deux ensembles convexes est convexe.

iii) L'image directe d'un convexe par une application linéaire est convexe.

1.8.2 Fonction convexe

Définition 1.8.3 *Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \rightarrow \mathbb{R}$ une fonction. On dit que f est convexe sur U si $\forall x, y \in U, \quad \forall t \in [0, 1]$*

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

Autrement dit, pour tout $\alpha, \beta \in \mathbb{R}_+$ tel que $\alpha + \beta = 1$, on a

$$f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y).$$

Remarque 1.8.1 *La somme de deux fonctions convexes est convexe.*

Le produit d'une fonction convexe par un scalaire n'est pas toujours convexe.

1.8.3 Fonction concave

Définition 1.8.4 *Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite concave si $(-f)$ est une fonction convexe, c'est à dire*

$$\forall x, y \in \mathbb{R}^n, \quad \forall \lambda \in [0, 1], \quad f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y).$$

Définition 1.8.5 *Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \rightarrow \mathbb{R}$ une fonction.*

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

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

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

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

3. On dit que f est concave sur U (respectivement strictement concave, respectivement fortement concave) si $-f$ est convexe (respectivement strictement convexe, respectivement fortement convexe).

Remarque 1.8.2 Il est facile de voir que

fortement convexe \Rightarrow strictement convexe \Rightarrow convexe. Mais la réciproque n'est pas vraie en général, par exemple

une application affine

$f(x) = ax + b$, $a > 0$ est convexe mais elle n'est pas strictement convexe donc elle n'est pas fortement convexe.

Proposition 1.8.2 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. Posons

$$f = \gamma_1 f_1 + \gamma_2 f_2 + \dots + \gamma_n f_n.$$

Alors

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

2. Si au moins l'une des fonctions f_1, f_2, \dots, f_p est strictement convexe alors f est strictement convexe.

3. Si au moins l'une des fonctions f_1, f_2, \dots, f_p est fortement convexe alors f est fortement convexe.

Proposition 1.8.3 Soit U un ouvert de \mathbb{R}^n , $h \in \mathbb{R}^n$. Alors

$\forall x \in U$ nous avons

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

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

Proposition 1.8.4 Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $U \subset \Omega$ avec U convexe et $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 alors

a) Les trois assertions suivantes sont équivalentes

- 1) f est convexe sur U .
- 2) $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in U$.
- 3) ∇f est monotone sur U c'est à dire

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

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

$$\langle \nabla^2 f(x)(y - x), y - x \rangle \geq 0 \quad \forall x, y \in U. \quad (1.2)$$

1.9 Classification des problèmes d'optimisation

1.9.1 Forme générale d'un problème d'optimisation

Etant donné un ensemble $X \subset \mathbb{R}^n$ et une fonction $f : X \rightarrow \mathbb{R}$, la forme générale d'un problème d'optimisation est la suivante

$$\min\{f(x); x \in X\}, \quad (1.3)$$

où $X \subset \mathbb{R}^n$ est appelée ensemble des solutions admissibles ou réalisables, $x \in \mathbb{R}^n$ s'appelle variable de décision, $f(x)$ s'appelle fonction économique ou fonction objective.

1.9.2 Problème d'optimisation sans contraintes

Si $X = \mathbb{R}^n$, le problème d'optimisation (1.4) s'appelle problème d'optimisation sans contraintes et aura donc la forme suivante

$$\min \{f(x); x \in \mathbb{R}^n\}, \quad (1.4)$$

Problème d'optimisation sans contraintes non linéaire

Le problème (1.4) s'appelle le problème d'optimisation sans contraintes non linéaire si f n'est pas affine.

Problème d'optimisation sans contraintes quadratique

Le problème (1.4) s'appelle le problème d'optimisation sans contraintes non linéaire quadratique si

$$f(x) = x^T A x - b x,$$

où A est une matrice symétrique d'ordre n , x^T est le transposé du vecteur x .

1.9.3 Problème d'optimisation avec contraintes

Soit \mathcal{U} un ouvert non vide de \mathbb{R}^n (le plus souvent $\mathcal{U} = \mathbb{R}^n$) et soit $f : \mathcal{U} \rightarrow \mathbb{R}$ une application différentiable sur \mathcal{U} , et $\varphi_1, \varphi_2, \dots, \varphi_p : \mathcal{U} \rightarrow \mathbb{R}$ des applications de classe C^1 sur \mathcal{U} .

Soit le domaine \mathcal{D}

$$\mathcal{D} = \{\mathbf{x} \in \mathcal{U} \mid \varphi_i(\mathbf{x}) = 0, \forall i = 1, 2, \dots, p\}.$$

Le problème

$$\min_{x \in \mathcal{D}} f(\mathbf{x})$$

est un problème de minimisation sous contraintes.

1.10 Direction de descente

Définition 1.10.1 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $\hat{x} \in \mathbb{R}^n$. $d \in \mathbb{R}^n$ est dite direction de descente au point \hat{x} si et seulement si il existe $\delta > 0$ tel que

$$f(\hat{x} + \lambda d) < f(\hat{x}) : \quad \forall \lambda \in]0, \delta[.$$

Donnons maintenant une condition suffisante pour que d soit une direction de descente.

Théorème 1.10.1 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $\hat{x} \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ une direction vérifiant la condition suivante

$$\nabla f(\hat{x})^t \cdot d < 0.$$

Alors d est une direction de descente au point \hat{x} .

Preuve

f est différentiable au point \hat{x} . Donc

$$f(\hat{x} + \lambda d) = f(\hat{x}) + \lambda \nabla f(\hat{x})^t \cdot d + \lambda \|d\| \alpha(\hat{x}, \lambda d),$$

avec

$$\alpha(\hat{x}, \lambda d) \xrightarrow{\lambda \rightarrow 0} 0,$$

ceci implique que

$$f'(\hat{x}, d) = \lim_{\lambda \rightarrow 0} \frac{f(\hat{x} + \lambda d) - f(\hat{x})}{\lambda} = \nabla f(\hat{x})^t \cdot d < 0.$$

La limite étant strictement négative, alors il existe un voisinage de zéro ,

$V(0) =] - \delta, +\delta[$ tel que

$$\frac{f(\hat{x} + \lambda d) - f(\hat{x})}{\lambda} < 0, \quad \forall \lambda \in] - \delta, +\delta[, \quad (1.5)$$

la relation (1.5) est particulièrement vraie pour tout $\lambda \in]0, +\delta[$. On obtient le résultat cherché en multipliant la relation (1.5) par $\lambda > 0$ ■

1.11 Conditions d'optimalité

1.11.1 Conditions nécessaires d'optimalité du premier ordre

Théorème 1.11.1 *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $\hat{x} \in \mathbb{R}^n$. Si \hat{x} est un minimum local de (P) alors $\nabla f(\hat{x}) = 0$.*

1.11.2 Conditions nécessaires d'optimalité du second ordre

Théorème 1.11.2 *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable au point $\hat{x} \in \mathbb{R}^n$.*

Si \hat{x} est un minimum local de (P) alors $\nabla f(\hat{x}) = 0$ et la matrice hessienne de f au point \hat{x} .

Qu'on note $H(\hat{x})$, est semi définie positive.

Preuve

Soit $x \in \mathbb{R}^n$ quelconque, f est une fonction deux fois différentiable au point \hat{x} , donc on a pour tout $\lambda \neq 0$

$$f(\hat{x} + \lambda x) = f(\hat{x}) + \frac{1}{2} \lambda^2 x^t H(\hat{x}) x + \lambda^2 \|x\|^2 \alpha(\hat{x}, \lambda x), \quad \alpha(\hat{x}, \lambda x) \xrightarrow{\lambda \rightarrow 0} 0.$$

Ceci implique

$$\frac{f(\hat{x} + \lambda x) - f(\hat{x})}{\lambda^2} = \frac{1}{2} x^t H(\hat{x}) x + \|x\|^2 \alpha(\hat{x}, \lambda x), \quad (1.6)$$

\hat{x} est un minimum local, il existe alors $\delta > 0$ tel que

$$\frac{f(\hat{x} + \lambda x) - f(\hat{x})}{\lambda^2} \geq 0, \quad \forall \lambda \in]-\delta, +\delta[.$$

Si on prend en considération (1.5) et on passe à la limite quand $\lambda \rightarrow 0, \lambda \neq 0$, on obtient

$$x^t H(\hat{x})x \geq 0, \quad \forall x \in \mathbb{R}^n \blacksquare$$

1.11.3 Conditions suffisantes d'optimalité

Théorème 1.11.3 *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable au point $\hat{x} \in \mathbb{R}^n$.*

Si $\nabla f(\hat{x}) = 0$ et $H(\hat{x})$ est définie positive alors \hat{x} est un minimum local strict de (1.4).

Preuve

f est deux fois différentiable au point \hat{x} , pour tout $x \in \mathbb{R}^n$ on aura

$$f(x) = f(\hat{x}) + \frac{1}{2}(x - \hat{x})^t H(\hat{x})(x - \hat{x}) + \|(x - \hat{x})\|^2 \alpha(\hat{x}, (x - \hat{x})), \quad \alpha(\hat{x}, (x - \hat{x})) \xrightarrow{x \rightarrow \hat{x}} 0 \quad (1.7)$$

Supposons que \hat{x} n'est pas un minimum local strict.

Alors il existe une suite $\{x_k\}_{k \in \mathbb{N}}$ telle que

$$x_k \neq \hat{x}, \forall k \in \mathbb{N}, \quad x_k \xrightarrow{k \rightarrow \infty} \hat{x} \text{ et } f(x_k) \leq f(\hat{x}). \quad (1.8)$$

Dans (1.7) prenons $x = x_k$, divisons le tout par $\|(x_k - \hat{x})\|^2$ et notons

$d_k = \frac{(x_k - \hat{x})}{\|(x_k - \hat{x})\|}$, on obtient

$$\frac{f(x_k) - f(\hat{x})}{\|(x_k - \hat{x})\|^2} = \frac{1}{2} d_k^t H(\hat{x}) d_k + \alpha(\hat{x}, (x_k - \hat{x})), \quad \alpha(\hat{x}, (x_k - \hat{x})) \xrightarrow{k \rightarrow \infty} 0, \quad (1.9)$$

(1.7) et (1.8) impliquent

$$\frac{1}{2} d_k^t H(\hat{x}) d_k + \alpha(\hat{x}, (x_k - \hat{x})) \leq 0, \quad \forall k \in \mathbb{N}.$$

D'autre part la suite $\{d_k\}_{k \in \mathbb{N}}$ est bornée ($\|d_k\| = 1, \forall k \in \mathbb{N}$), donc il existe une sous suite $\{d_k\}_{k \in N_1 \subset \mathbb{N}}$ telle que

$$d_k \xrightarrow[k \rightarrow \infty, k \in N_1]{} \tilde{d}.$$

Finalement lorsque $k \rightarrow \infty, k \in N_1$, on obtient

$$\frac{1}{2} \tilde{d} H(\hat{x}) \tilde{d} \leq 0.$$

La dernière relation et le fait que $\tilde{d} \neq 0$ ($\|\tilde{d}\| = 1$) impliquent que la matrice hessienne $H(\hat{x})$ n'est pas définie positive. Ceci est contradiction avec l'hypothèse.

Recherches linéaires

2.1 Recherches linéaires

Considérons le problème d'optimisation sans contraintes

$$\text{minimiser } \{f(x) : x \in \mathbb{R}^n\} \quad (P),$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Les algorithmes qu'on étudie par la suite suivent les schémas généraux de la forme

$$x_{k+1} = x_k + \lambda_k d_k,$$

où λ_k est une solution optimale du problème d'optimisation suivant

$$\min_{\lambda > 0} f(x_k + \lambda d_k),$$

c'est à dire que λ_k vérifie

$$f(x_k + \lambda_k d_k) \leq f(x_k + \lambda d_k), \forall \lambda > 0.$$

x_k, d_k sont fixes et la fonction à minimiser est une fonction d'une variable réelle définie comme suit

$$\lambda \mapsto h_k(\lambda) = f(x_k + \lambda d_k).$$

Il faut noter que dans le problème d'optimisation sans contraintes, on a besoin de résoudre à chaque itération x_k , un problème d'optimisation dans \mathbb{R} .

Objectifs de la recherche linéaire

Il s'agit de réaliser deux objectifs

Le premier objectif

Consiste à faire décroître f suffisamment.

Cela se traduit le plus souvent par la réalisation d'une inégalité de la forme

$$f(x_k + \lambda d_k) \leq f(x_k) + \text{un terme négatif} \quad (2.1)$$

Le terme négatif joue un rôle-clé dans la convergence de l'algorithme utilisant cette recherche linéaire.

L'argument est le suivant.

Si $f(x_k)$ est minorée (il existe une constante C telle que $f(x_k) \geq C$ pour tout k), alors ce terme négatif tend nécessairement vers zéro : $\nu_k \rightarrow 0$.

C'est souvent à partir de la convergence vers zéro de cette suite que l'on parvient à montrer que le gradient lui-même doit tendre vers zéro.

Le terme négatif devra prendre une forme bien particulière si on veut pouvoir en tirer de l'information. En particulier, il ne suffit pas d'imposer $f(x_k + \lambda d_k) \leq f(x_k)$.

Le second objectif

Consiste d'empêcher le pas $\lambda_k > 0$ d'être trop petit, trop proche de zéro.

Le premier objectif n'est en effet pas suffisant car l'inégalité (2.1) est en général satisfaite par du pas $\lambda_k > 0$ arbitrairement petit.

Or ceci peut entraîner une "fausse convergence", c'est-à-dire la convergence des itérés vers un

point non stationnaire, comme le montre l'observation suivante.

Si on prend

$$0 < \lambda_k \leq \frac{\varepsilon}{2^k \|d_k\|},$$

la suite $\{x_k\}$ est de Cauchy, puisque pour $1 < l < k$ on a

$$\|x_k - x_l\| = \left\| \sum_{i=1}^{i=k-1} \lambda_i d_i \right\| \leq \sum_{i=1}^{i=k-1} \frac{\varepsilon}{2^i} \rightarrow 0, \quad l \rightarrow \infty$$

Donc $\{x_k\}$ converge, disons vers un point x^* .

En prenant $l = 1$ et $k \rightarrow \infty$ dans l'estimation ci-dessus, on voit que $x^* \in \bar{B}(x_1, \varepsilon)$ et x^* ne soit pas une solution s'il n'y a pas de solution dans $\bar{B}(x_1, \varepsilon)$.

On a donc arbitrairement forcer la convergence de $\{x_k\}$ en prenant des pas très petits.

Pour simplifier les notations, on définit $h_k : \mathbb{R}_+ \rightarrow \mathbb{R}$ comme suit

$$\lambda \mapsto h_k(\lambda) = f(x_k + \lambda d_k).$$

But de la recherche linéaire

Dans le cas non quadratique les méthodes de descente, nécessitent la recherche d'une valeur de $\lambda_k > 0$ optimale ou non, vérifiant

$$f(x_k + \lambda_k d_k) \leq f(x_k).$$

Rappelons que si f est différentiable, le pas optimal λ^* peut être caractérisé par

$$h^*(\lambda^*) = 0,$$

$$h(\lambda^*) \leq h(\lambda), \text{ pour } 0 < \lambda \leq \lambda^*.$$

autrement dit, λ^* est un minimum local de h qui assure de plus la décroissance de f .

En fait, dans la plupart des algorithmes d'optimisation modernes, on ne fait jamais de recherche

linéaire exacte, car trouver λ^* signifie qu'il va falloir calculer un grand nombre de fois la fonction f et cela peut être dissuasif du point de vue du temps de calcul.

En pratique, on recherche plutôt une valeur de λ qui assure une décroissance suffisante de f .

Cela conduit à la notion d'intervalle de sécurité.

Dans ce qui suit nous introduisons trois types de recherches linéaires.

2.2 Recherche linéaire d'Armijo

Principe

Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable et $x_k \in \mathbb{R}^n$ une direction (de descente) et $d_k \in \mathbb{R}^n$ telle que $\nabla f(x_k)^T d_k < 0$.

On dit que le pas $\alpha_k > 0$ vérifie la condition d'Armijo, si on a

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta \nabla f(x_k)^T d_k, \quad 0 < \beta < 1. \quad (\text{Armijo})$$

2.2.1 Algorithme d'Armijo

Etape 0 (initialisation)

$\alpha_{g,1} = \alpha_{d,1} = 0$, choisir $\alpha_1 > 0, \rho \in]0, 1[$ poser $k = 1$ et aller à l'étape 1.

Etape 1

si $\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho \varphi_k'(0) \alpha_k$: STOP ($\alpha^* = \alpha_k$)

si $\varphi_k(\alpha_k) > \varphi_k(0) + \rho \varphi_k'(0) \alpha_k$,

alors $\alpha_{d,k+1} = \alpha_d, \alpha_{g,k+1} = \alpha_k$ et aller à l'étape 2.

Etape 2

si $\alpha_{d,k+1} = 0$ déterminer $\alpha_{k+1} \in]\alpha_{g,k+1}, +\infty[$

si $\alpha_{d,k+1} \neq 0$ déterminer $\alpha_{k+1} \in]\alpha_{g,k+1}, \alpha_{d,k+1} |$

remplacer k par $k + 1$ et aller à l'étape 1 .

2.3 Recherche linéaire de Goldstein

2.3.1 Principe

Supposons que :

À l'iteration k , on ait le point $x_k \in \mathbb{R}^n$ et le vecteur de descente $d_k \in \mathbb{R}^n$ ($\nabla f(x_k)^T d_k < 0$).

On cherche le pas $\alpha_k > 0$ de Goldstein vérifiant les deux conditions suivantes :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla f(x_k)^T d_k, \quad \rho \in \left] 0, \frac{1}{2} \right[\quad (\text{Goldstein 1})$$

et

$$f(x_k + \alpha_k d_k) \geq f(x_k) + (1 - \rho) \alpha_k \nabla f(x_k)^T d_k, \quad \rho \in \left] 0, \frac{1}{2} \right[\quad (\text{Goldstein 2}).$$

Si on note

$$\varphi_k(\alpha) = f(x_k + \alpha d_k),$$

alors (Goldstein 1) et (Goldstein 2) deviennent

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho \alpha_k \varphi_k'(0)$$

et

$$\varphi_k(\alpha_k) \geq \varphi_k(0) + (1 - \rho) \alpha_k \varphi_k'(0).$$

(Goldstein1) assure une décroissance suffisante de f .

Les pas α_k trop petits sont nocifs pour la convergence.

La suite $\{x_h\}$ pourrait converger prématurément vers un point \tilde{x} qui n'a rien à voir avec le problème d'optimisation.

La condition (Goldstein 2) élimine justement les pas α_k trop petits.

2.3.2 Algorithme de Goldstein

L'Algorithme essaye de trouver $\alpha_k \in]\beta_1, \beta_2[$. On démarre avec un intervalle $[a_0, b_0]$ assez grand.

On prend $\alpha_0 \in [a_0, b_0]$

Si α_0 vérifié Goldstein1 et Goldstein2

alors $\alpha_0 \in]\beta_1, \beta_2[$ et on s'arrête.

Si $\alpha_0 > \beta_2$

alors α_0 ne vérifié pas Goldstein1, alors on prend $b_1 = \alpha_0$ et $a_1 = b_0$ et $\alpha_1 = \frac{a_1 + b_1}{2}$

et on recommence avec α_1

Si $\alpha_0 < \beta_1$

alors α_0 ne vérifié pas Goldstein2, on prend $a_1 = \alpha_0, b_1 = b_0$ et $\alpha_1 = \frac{a_1 + b_1}{2}$ et on teste de nouveau α_1

A l'itération k

Supposons qu'on ait $[a_k, b_k]$ et $\alpha_k = \frac{a_k + b_k}{2}$

Si α_k vérifie Goldstein1 et Goldstein2, $\alpha_k \in]\beta_1, \beta_2[$. Stop.

Si α_k ne vérifié pas Goldstein1 alors $\alpha_k > \beta_2$.

On prend $b_{k+1} = \alpha_k, a_{k+1} = a_k, \alpha_{k+1} = \frac{a_{k+1} + b_{k+1}}{2}$.

Si α_k ne vérifie pas Goldstein2 alors $\alpha_k < \beta_1$.

On prend $a_{k+1} = \alpha_k, b_{k+1} = b_k, \alpha_{k+1} = \frac{a_{k+1} + b_{k+1}}{2}$

On obtient ainsi l'algorithme suivant :

Algorithme de Goldstein

Etape 1 (Initialisation)

Choisir $\alpha_0 \in [0, 10^{100}]$ et $\rho \in]0, 1[$.

Poser $a_0 = 0, b_0 = 10^{100}$ Poser $k = 0$ et aller à Etape2

Etape 2 (Test Goldstein1)

Iteration k on a $[a_k, b_k]$ et α_k , calculez $\varphi_k(\alpha_k)$

Si $\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho \alpha_k \varphi_k'(0)$, allez à Etape 3

Sinon

Poser $b_{k+1} = \alpha_k$, $a_{k+1} = a_k$, et allez à Etape 4

Etape 3 (Test Goldstein 02)

Si $\varphi_k(\alpha_k) \geq \varphi_k(0) + (1 - \rho)\alpha_k\varphi'_k(0)$ stop. $\alpha^* = \alpha_k$

Sinon Poser $a_{k+1} = \alpha_k$, $b_{k+1} = b_k$ et allez à Etape 4

Etape 4

Poser $\alpha_{k+1} = \frac{a_{k+1} + b_{k+1}}{2}$

Poser $k = k + 1$ et allez à Etape 2 .

2.3.3 Programme en fortran

Programme en fortran

```

real x(2), xp(2), g(2), gp(2), f0, alpha, f, p, a, a, b, pf, t, y, phopt
integer i, k, test, n
ro = 0.1; delt = 0.3; b = 10000000; alpha = 10; test = 0; k = 0
print, ', 'le dimention n = ', read, n
do i = 1, n print *, xo(, i') = ', read *, xp(i) end do
10 if (test=0.and.k.le.1000) then call fonc (f0, gp,xp)
do i = 1, n x(i) = xp(i) - alpha*gp(i) end do
call fonc (f, g, x); p = 0
do i = 1, n p = p + gp(i)* qp(i) end do
s = f(-r0*alpha*p; y = f0- delt *alpha*p
if (f.gt.s) then b = alpha else if (f.lt.y) then a = alpha else
phopt = alpha; test = 1 endif
endif
print*, alpha(', k') = , , alpha; alpha = (a + b)/2; k = k + 1
80 to 10
end if
print*, 'alpha optimal=', phopt

```

end

subroutine fonc (4, 8, x)

real f, 8(2), x(2)

$f(x(1))^{**2} + (x(2))^{**4}; 8(1) = 2^*x(1); 8(2) = 4^*(x(2))^{**3}$ end

Dans le théorème suivant on va assurer l'existence du pas de Goldstein et Price en posant quelques conditions sur la fonction h_k .

Théorème 2.3.1 *Si $\varphi_k : \mathbb{R}_+ \rightarrow \mathbb{R}$ définie par $\varphi_k(a) = f(x_k + ad_k)$ est continue et bornée inférieurement, si d_n est une direction de descente en w_n et at $p \in]0, 1[, \delta \in]\rho, 1[$, alors l'ensemble des pas vérifiant la règle de Goldstein et Price est non vide.*

Preuve

On a

$$\varphi_k(a) = f(a_k + ad_k),$$

$$\psi_\rho(a) = f(x_k) + \rho a \nabla^T f(x_k) d_k,$$

$$\psi_\delta(a) = f(x_k) + \delta a \nabla^T f(x_k) d_k.$$

Le développement de Taylor-Yong en $a = 0$ de φ_k est ;

$$\varphi_k(a) = f(x_k + ad_k) = f(x_k) + \rho a \nabla^T f(x_k) d_k + a\xi(a) \text{ où } \xi(a) \rightarrow 0, a \rightarrow 0.$$

et comme $\rho \in]0, 1[$ et $\varphi'_k(0) = \nabla^T f(x_k) d_k < 0$ on déduit ;

$$f(x_k) + a \nabla^T f(x_k) \alpha_k < f(x_k) + \delta a \nabla^T f(x_k) d_k < f(x_k) + \rho a \nabla^T f(x_k) d_k \text{ pour } a > 0.$$

On voit que pour $a > 0$ assez petit on a

$$\varphi_k(a) < \psi_\delta(a) < \psi_\rho(a)$$

De ce qui précède et du fait que φ_k est bornée inférieurement, et $\psi_\rho(a) \rightarrow -\infty; a \rightarrow +\infty$, on déduit que la fonction $\psi_\rho(a) - \varphi_k(a)$ a la propriété :

$$\begin{cases} \psi_\rho(a) - \varphi_k(a) > 0 & \text{pour } a \text{ assez petit} \\ \psi_\rho(a) - \varphi_k(a) < 0 & \text{pour } a \text{ assez grand,} \end{cases}$$

donc s'annule au moins une fois pour $a > 0$.

En choisissant le plus petit de ces zéros on voit qu'il existe $\bar{a} > 0$ tel que

$$\varphi_k(\bar{a}) = \psi_\rho(\bar{a}) \text{ et } \varphi_k(a) < \psi_\rho(a) \text{ pour } 0 < a < \bar{a}.$$

De la même manière, il existe $\bar{a} > 0$ tel que

$$\varphi_k(\bar{a}) = \psi_\delta(\bar{a}) \text{ et } \varphi_k(a) < \psi_\delta(a) \text{ pour } 0 < a < \bar{a}$$

et comme $\psi_\sigma(a) < \psi_\rho(a)$ pour $a > 0$, forcément $\bar{a} < \bar{a}$ et $a = \bar{a}$ est satisfié.

2.4 Recherche linéaire de Wolfe

α_k est acceptable par la recherche linéaire inexacte de Wolfe si α_k vérifie les deux conditions (Wolfe1) et (Wolfe2) suivantes

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho\alpha_k\varphi'_k(0), \quad 0 < \rho < \frac{1}{2} \quad (\text{Wolfe1})$$

$$\varphi'_k(\alpha_k) \geq \sigma\varphi'_k(0), \quad 0 < \sigma < 1, \quad \sigma > \rho \quad (\text{Wolfe2})$$

ou en remplaçant $\varphi_k(\alpha_k), \varphi_k(0), \varphi'_k(0), \varphi'_k(\alpha_k)$ par leurs expressions, on trouve

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho\alpha_k \nabla f(x_k)^T d_k, \quad 0 < \rho < \frac{1}{2} \quad (\text{Wolfe3}) \quad (2.2)$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k, \quad 0 < \sigma < 1, \quad \sigma > \rho \quad (\text{Wolfe4}) \quad (2.3)$$

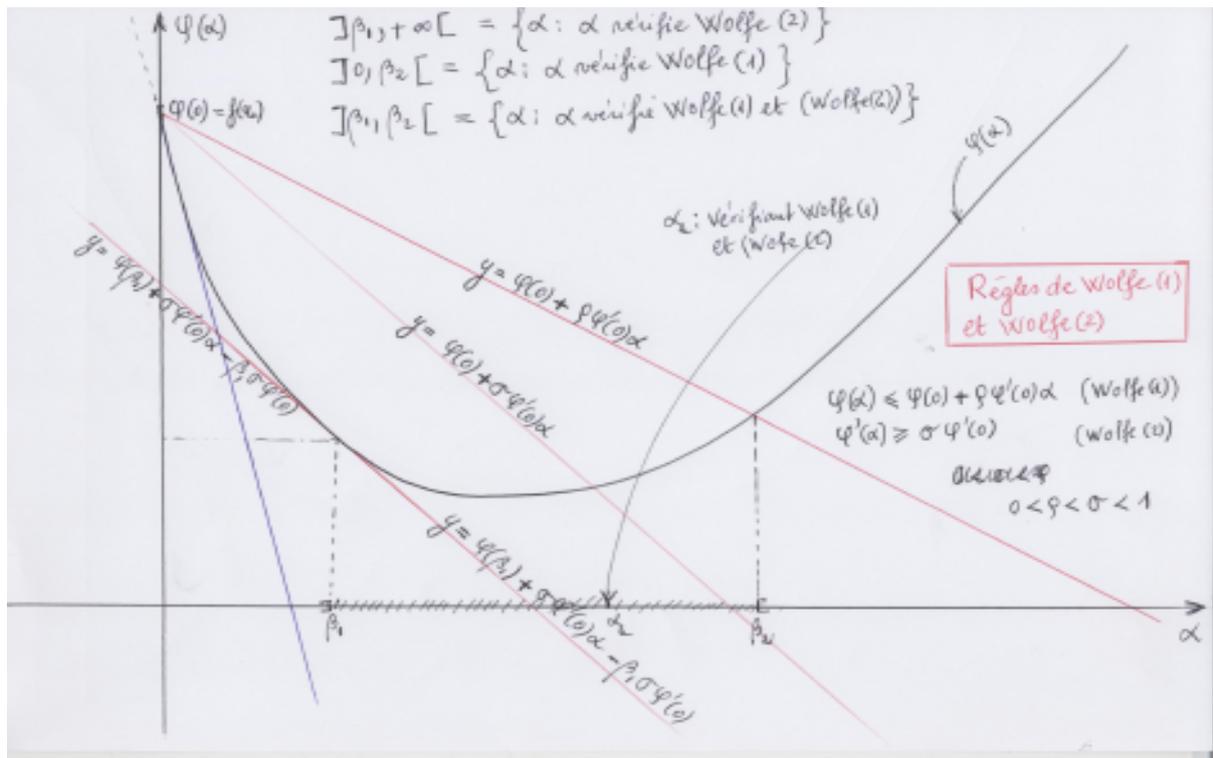


FIGURE 2.1 –

Commentaires sur la figure 2.1

a) $\alpha \in]0, \beta'_1[$: alors α ne vérifie pas (wolfe2) car β'_1 vérifié : $\varphi'(\beta'_1) = \sigma \varphi'(0)$.

Les droites Δ_2 et Δ_3 sont parallèles ($\Delta_2 // \Delta_3$)

et si $\alpha < \beta'_1$ alors $\varphi'(\alpha) < \varphi'(\beta'_1)$. En fait β'_1 est le premier point tel que $\varphi'(\beta'_1) = \sigma \varphi'(0)$.

b) $\alpha \in]\beta'_2, +\infty[$, α ne vérifié pas (wolfe2), le graphe de φ est au dessus de la droite Δ_1 de coefficient directeur $\rho \varphi'(0)$

c) $\alpha \in]\beta'_1, \beta'_2[$, α vérifie (Wolfe 1) et (Wolfe2).

2.4.1 Recherche linéaire de Wolfe forte

α_k vérifie les conditions de Wolfe forte si les deux conditions suivantes sont vérifiées

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho \alpha_k \varphi'_k(0), \quad 0 < \rho < \frac{1}{2} \quad (\text{Wolfe forte1})$$

$$|\varphi'_k(\alpha_k)| \leq -\sigma \varphi'_k(0), \quad 0 < \sigma < 1, \quad \sigma > \rho. \quad (\text{Wolfe forte 2})$$

Théorème 2.4.1 *Si α_k vérifie (wolfe forte 1) et (wolfe forte 2), alors α_k vérifie (wolfe 1) et (wolfe 2).*

Preuve

(wolfe forte 1) et (wolfe 1) sont les mêmes propositions. Reste à montrer que

$$\{(wolfe\ forte2) \implies (wolfe2)\}$$

En effet. Supposons que α_k vérifie (Wolfe forte 2).

Alors on a

$$\sigma \varphi'_k(0) \leq \varphi'_k(\alpha_k) \leq -\sigma \varphi'_k(0),$$

par conséquent on a

$$\varphi'_k(\alpha_k) \geq \sigma \varphi'_k(0) \quad .$$

Ceci implique que α_k vérifie Wolfe 2 ■

Remarque 2.4.1 *Supposons qu'on a une suite de paramètres $\{\sigma_j\}_{j \in \mathbb{N}}$ telle que*

$$\sigma_j \rightarrow 0, j \rightarrow \infty$$

Fixons $k \in \mathbb{N}$. Supposons aussi que pour k fixé, (Wolfe forte 1) et (Wolfe forte 2). On obtient donc une suite $\{\alpha_{k,j}\}_{k \text{ fixé}, j \in \mathbb{N}}$

$$\forall j \in \mathbb{N} : \sigma_j \varphi'_k(0) \leq \varphi'_k(\alpha_{k,j}) \leq -\sigma_j \varphi'_k(0).$$

Ceci implique

$$\varphi'_k(\alpha_{kj}) \xrightarrow{j \rightarrow \infty, k \text{ fixe}} 0.$$

La relation précédente implique que

$$\varphi'_k(\alpha_k) \xrightarrow{k \rightarrow \infty} 0.$$

Si φ'_k est continue et si $\alpha_j \rightarrow \bar{\alpha}$ alors $\bar{\alpha} = \alpha^*$.

Remarque 2.4.2 Dans la recherche linéaire inexacte de Wolfe forte, si on choisit σ très petit, on a beaucoup de chances de tomber sur α_k proche de α^* .

L'inconvénient est que ce choix (σ très petit) entraîne beaucoup d'itérations pour parvenir à obtenir α_k vérifiant (Wolfe forte 1) et (Wolfe forte 2).

On pourrait choisir une suite $\{\sigma_j\}$ telle que $\sigma_j \rightarrow 0$ lentement. Ces choix influent sur la convergence et la vitesse de convergence.

2.4.2 Algorithme de Wolfe forte

Etape 1 (Initialisation)

Prendre $\alpha_0 \in [0, 10^{99}]$, calculez $\varphi(0), \varphi'(0)$.

Prendre $\rho = 0.1$ (ou $\rho = 0.1$ ou $\rho = 0.001$ ou $\rho = 10^{-4}$) $\sigma = 0.9$ (ou σ plus petit encore)

Poser $a_0 = 0, b_0 = 10^{99}, k = 0$ et allez à Etape 2.

Etape 2 (test de (Wolfe 1))

Calculez $\varphi(\alpha_k)$.

Si $\varphi(\alpha_k) \leq \varphi(0) + \rho\alpha_k\varphi'(0)$, aller à Etape 3.

Sinon

Poser $a_{k+1} = a_k, b_{k+1} = \alpha_k$ et allez à Etape 4.

Etape 3 (test (Wolfe 2) ou (Wolfe forte 2))

Calculez $\varphi'(\alpha_k)$.

Si $\varphi'(\alpha_k) \geq \sigma\varphi'(0)$ ($|\varphi'_k(\alpha_k)| \leq -\sigma\varphi'_k(0)$). STOP

Prendre $\bar{\alpha} = \alpha_k$.

Sinon Poser $a_{k+1} = \alpha_k, b_{k+1} = b_k$ et allez à Etape 4.

Etape 4 (calcul de α_{k+1})

$\alpha_{k+1} = \frac{a_{k+1} + b_{k+1}}{2}$ Poser $k = k + 1$ et allez à Etape 2 .

Méthode du gradient conjugué(G.C)

3.1 Méthode des directions conjuguées

3.1.1 Définitions et propriétés générales

Définition 3.1.1 Soit Q une matrice (n, n) symétrique. Les directions d_0, d_1, \dots, d_k sont dites Q conjuguées si on a

$$d_i^T Q d_j = 0, \quad 0 \leq i, j \leq k. \quad (3.1)$$

Théorème 3.1.1 Soit Q une matrice (n, n) symétrique et définie positive. Si les directions d_0, d_1, \dots, d_k , avec $k \leq n - 1$, sont non nuls et Q conjuguées, alors ils sont linéairement indépendants.

Preuve

Supposons que :

$$\alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_k d_k = 0. \quad (3.2)$$

Multiplions l'égalité (3.2) par $d_j^T Q$, $0 \leq j \leq k$, on obtient

$$0 = \sum_{i=1}^k \alpha_i d_j^T Q d_i = \alpha_j d_j^T Q d_j. \quad (3.3)$$

Or

$$\alpha_j d_j^T Q d_j = 0 \implies \alpha_j = 0,$$

car Q est définie positive et par conséquent $d_j^T Q d_j > 0$.

Donc le système $\{d_0, d_1, \dots, d_k\}$ est libre ■

3.1.2 Algorithme des directions conjuguées

Soit Q est une matrice (n, n) symétrique et définie positive et $b \in \mathbb{R}^n$.

Considérons le problème de minimisation quadratique sans contraintes, (PQSC), suivant

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} x^T Q x - b^T x \right\} \quad (PQSC).$$

Algorithme des directions conjuguées

a) Initialisation

On se donne $x_0 \in \mathbb{R}^n$ quelconque et $d_0, d_1, \dots, d_{n-1} Q$ conjugués.

Poser $k = 0$ et allez à b) Etape principale.

b) Etape principale

Pour $k \geq 0$

Calculez

$$g_k = \nabla f(x_k) = Qx_k - b.$$

Si $g_k = 0$. Stop.

Sinon calculez

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k}$$

et

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

Poser $k = k + 1$ et allez à b) Etape principale.

Théorème 3.1.2 Partant d'un point initial $x_0 \in \mathbb{R}^n$ quelconque, l'algorithme des directions

conjuguées précédent converge vers la solution optimale unique \hat{x} du problème (PQSC) dans n iterations, c'est à dire qu'on a

$$x_n = \hat{x} \text{ et } Qx_n = Q\hat{x} = b. \quad (3.4)$$

Preuve

Considérons le vecteur $\hat{x} - x_0 \in \mathbb{R}^n$.

D'autre part $\{d_0, \dots, d_{n-1}\}$ est un système libre et contient n vecteurs. Donc $\{d_0, \dots, d_{n-1}\}$ est une base de \mathbb{R}^n .

Par conséquent il existe $\beta_0, \dots, \beta_{n-1}$ scalaires réels tels que

$$\hat{x} - x_0 = \beta_0 d_0 + \dots + \beta_{n-1} d_{n-1}. \quad (3.5)$$

Multiplions les 2 côtés de (3.5) par $d_k^T Q$, $0 \leq k \leq n-1$, on obtient

$$d_k^T Q (\hat{x} - x_0) = \beta_k d_k^T Q d_k. \quad (3.6)$$

Donc

$$\beta_k = \frac{d_k^T Q (\hat{x} - x_0)}{d_k^T Q d_k}. \quad (3.7)$$

Montrons maintenant que le numérateur de (3.7) est égal à $-d_h^T g_h$. En effet, d'après l'algorithme des directions conjuguées

$$\begin{aligned} x_k &= x_{k-1} + \alpha_{k-1} d_{k-1} \\ &= x_{k-2} + \alpha_{k-2} d_{k-2} + \alpha_{k-1} d_{k-1} \\ &= \dots \\ &= x_0 + \alpha_0 d_0 + \dots + \alpha_{k-1} d_{k-1}. \end{aligned}$$

Par conséquent

$$x_k - x_0 = \alpha_0 d_0 + \dots + \alpha_{k-1} d_{k-1}, \quad k = 1, 2, \dots, n-1. \quad (3.8)$$

Ecrivons maintenant : $\hat{x} - x_0$ sous la forme

$$\hat{x} - x_0 = (\hat{x} - x_k) + (x_k - x_0). \quad (3.9)$$

Multiplions (3.9) par $d_k^T Q$, on obtient

$$d_k^T Q (\hat{x} - x_0) = d_k^T Q (\hat{x} - x_k), \quad (3.10)$$

car

$$d_k^T Q (x_k - x_0) = \alpha_0 d_k^T Q d_0 + \dots + \alpha_{k-1} d_k^T Q d_{k-1} = 0.$$

Calculons maintenant $d_k^T Q (\hat{x} - x_k)$,

on a

$$d_k^T Q (\hat{x} - x_k) = d_k^T [Q\hat{x} - Qx_k] = d_k^T [b - Qx_k] = -d_k^T (Qx_k - b) = -d_k^T g_k, \quad (3.11)$$

(3.11) et (3.7) impliquent donc

$$\beta_k = -\frac{d_k^T g_k}{d_k^T Q d_k} = \alpha_k. \quad (3.12)$$

En conclusion on a

$$\hat{x} - x_0 = \beta_0 d_0 + \dots + \beta_{n-1} d_{n-1}$$

et

$$x_n - x_0 = \alpha_0 d_0 + \dots + \alpha_{n-1} d_{n-1},$$

or

$$\alpha_k = \beta_k \quad k = 0, \dots, n-1.$$

Donc

$$\widehat{x} - x_0 = x_n - x_0.$$

Ce qui implique

$$\widehat{x} = x_n \quad \blacksquare$$

Remarque 3.1.1 *Si on démarre du point x_1 , alors la solution optimale est atteinte au point x_{n+1} , c'est à dire qu'on aura*

$$\widehat{x} = x_{n+1}.$$

3.2 Méthode du G.C. Cas quadratique

Soit Q une matrice (n, n) , symétrique et définie positive. On définit le problème quadratique sans contraintes par

$$\min \{f(x) : x \in \mathbb{R}^n\} = \min \left\{ \frac{1}{2}x^T Qx - b^T x : x \in \mathbb{R}^n \right\} (PQSC).$$

Introduction

Dans la méthode des directions conjugués, les directions d_0, \dots, d_{n-1} sont données à l'avance.

Dans la méthode du gradient conjugué, on démarre d'un point $x_0 \in \mathbb{R}^n$,

$$d_0 = -g_0 = \nabla f(x_0) = Qx_0 - b.$$

Les directions $d_k, k = 1, \dots, n-1$ sont calculés à chaque itération.

A l'iteration k

$$d_k = -g_k + \beta_{k-1}d_{k-1},$$

β_{k-1} est obtenu de sorte que d_k soit Q conjugué avec les autres vecteurs d_i ,
 $i = 0, \dots, k-1$. En d'autres termes on doit avoir

$$d_k^T Q d_i = 0 \quad i = 0, \dots, k-1.$$

Dans l'appellation gradient conjugué, on trouve les deux mots : gradient et conjugué.

a) Le mot gradient est utilisé car d_k est calculé à partir du gradient au point x_k .

b) Le mot conjugué est aussi justifié, car et comme on le verra plus loin, les directions $\{d_k\}_{k=0}^{n-1}$ sont Q conjugués.

3.2.1 Algorithme

Principe de l'Algorithme

On démarre d'un point quelconque $x_0 \in \mathbb{R}^n$.

Calculez $d_0 = -g_0 = b - Qx_0$, $\alpha_0 = -\frac{g_0^T d_0}{d_0^T Q d_0}$.

Supposons que l'itération k on ait : x_k et d_k .

Ceci nous permettra de calculer

$$g_k = Qx_k - b, \quad \alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k}, \quad x_{k+1} = x_k + \alpha_k d_k, \quad g_{k+1} = Qx_{k+1} - b, \quad d_{k+1} = -g_{k+1} + \beta_k d_k \quad (3.13)$$

β_k est choisi de sorte que

$$d_{k+1}^T Q d_k = 0. \quad (3.14)$$

Puisque $d_{k+1} = -g_{k+1} + \beta_k d_k$, alors (4.2) donne

$$(-g_{k+1} + \beta_k d_k)^T Q d_k = 0,$$

ou encore

$$\beta_k d_k^T Q d_k = g_{k+1}^T Q d_k, \quad (3.15)$$

et finalement

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}. \quad (3.16)$$

Algorithme

1. choisir $x_0 \in \mathbb{R}^n$.
2. Calculez $g_0 = Qx_0 - b$. Si $g_0 = 0$ stop. Sinon poser $d_0 = -g_0$. Poser $k = 0$
3. Calculez $\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k}$.
4. Calculez $x_{k+1} = x_k + \alpha_k d_k$.
5. Calculez $g_{k+1} = Qx_{k+1} - b$. Si $g_{k+1} = 0$ stop.
6. Calculez $\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$.
7. Calculez $d_{k+1} = -g_{k+1} + \beta_k d_k$.
8. Poser $k = k + 1$ et allez à 3 .

3.2.2 Propriétés

La propriété fondamentale du gradient conjugué cas quadratique est que les directions $\{d_k\}_{k=0}^{n-1}$ sont Q coconjugués. Ces directions vérifient comme on l'a vu dans l'algorithme

$$d_{k+1} = -g_{k+1} + \beta_k d_k,$$

avec

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}.$$

D'après le théorème 3.1.2, l'algorithme du gradient conjugué, version quadratique converge vers la solution optimale en n itérations.

Théorème 3.2.1 *Les directions $\{d_0, d_1, \dots, d_{n-1}\}$ engendrées par l'algorithme du gradient conju-*

gué quadratique sont Q conjuguées.

3.2.3 Convergence dans le cas quadratique

Théorème 3.2.2 Soit Q une matrice (n, n) symétrique et définie positive et soit (PQSC) le problème de minimisation sans contraintes quadratique suivant

$$\min \{f(x) : x \in \mathbb{R}^n\} = \min \left\{ \frac{1}{2}x^T Qx - b^T x : x \in \mathbb{R}^n \right\} \text{ (PQSC)}.$$

Démarrant d'un point $x_0 \in \mathbb{R}^n$ quelconque, considérons la suite $\{x_k\}$ générée par l'algorithme du gradient conjugué quadratique définie par

$$g_k = \nabla f(x_k) = Qx_k - b, k = 0, 1, \dots \quad (3.17)$$

$$\beta_k = \frac{g_{k+1}^T Qd_k}{d_k^T Qd_k}, \quad k = 0, 1, \dots \quad (3.18)$$

$$d_k = \begin{cases} -g_0 & \text{si } k = 0 \\ -g_k + \beta_{k-1}d_{k-1} & \text{si } k \geq 1 \end{cases} \quad (3.19)$$

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Qd_k} \quad (3.20)$$

et

$$x_{k+1} = x_k + \alpha_k d_k. \quad (3.21)$$

Alors la suite $\{x_n\}$ converge en n itérations vers la solution optimale \hat{x} du problème (PQSC), c'est à dire que x_n vérifie $x_n = \hat{x}$ et

$$Q\hat{x} = Qx_n = b. \quad (3.22)$$

3.3 Méthode du G.C. Cas non quadratique

Introduction

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ non quadratique. On cherche à résoudre le problème non quadratique sans contraintes (PNQSC) suivant

$$\min \{f(x) : x \in \mathbb{R}^n\} (PNQSC).$$

Parmi les plus anciennes méthodes utilisées pour résoudre les problèmes du type (PNQSC), on peut citer la méthode du Gradient conjugué. Cette méthode est surtout utilisée pour les problèmes de grande taille.

Cette méthode a été découverte en 1952 par Hestenes et Steifel ([1]), pour la minimisation des fonctions quadratiques strictement convexes.

Plusieurs mathématiciens ont étendu cette méthode pour le cas non quadratique. Ceci a été réalisé pour la première fois, en 1964 par Fletcher et Reeves ([2]) (méthode de Fletcher-Reeves) puis en 1969 par Polak, Ribière ([3]) et Ployak ([3bis]) (méthode de Polak-Ribière-Ployak). D'autres variantes ont été étudiées plus tard ([4], [5], [6], [7], [8], [9], [10]) Une autre variante a été étudiée en 1987 par Fletcher ([4]) (Méthode de la descente conjuguée).

Toutes ces méthodes génèrent une suite $\{x_k\}_{k \in \mathbb{N}}$ de la façon suivante

$$x_{k+1} = x_k + \alpha_k d_k. \tag{3.23}$$

Le pas $\alpha_k \in \mathbb{R}$ est déterminé par une optimisation ou recherche linéaire exacte ou inexacte du type Armijo, Goldstein ou Wolfe.

Les directions d_k sont calculées de façon récurrente par les formules suivantes

$$d_k = \begin{cases} -g_0 & \text{si } k = 0 \\ -g_k + \beta_{k-1} d_{k-1} & \text{si } k \geq 1, \end{cases} \tag{3.24}$$

3.3 Méthode du G.C. Cas non quadratique *Méthode du gradient conjugué(G.C)*

avec $g_k = \nabla f(x_k)$ et $\beta_k \in \mathbb{R}$.

Les différentes valeurs attribuées à β_k définissent les différentes formes du gradient conjugué.

Si on note

$$y_{k-1} = g_k - g_{k-1}, \quad s_k = x_{k+1} - x_k, \quad (3.25)$$

on obtient les variantes suivantes

1964 ([2])-Gradient conjugué variante Fletcher Reeves (FR)

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}. \quad (3.26)$$

1969([3])- Gradient conjugué variante Polak-Ribière-Polyak (PRP)

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \quad (3.27)$$

1987 ([4])- Gradient conjugué variante descente conjuguée - Fletcher (CD)

$$\beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}} \quad (3.28)$$

1999([6]) - Gradient conjugué variante de Dai-Yuan (DY)

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \quad (3.29)$$

2005([7]) - Gradient conjugué variante Hager-Zhang(HZ)

$$\beta_k^{HZ} = \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k} \right)^T \frac{g_{k+1}}{d_k^T y_k} \quad (3.30)$$

2012([8]) - Gradient conjugué variante Rivaie-Mustafa-Ismail-Leong(RMIL) [60]

$$\beta_{k-1}^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2}. \quad (3.31)$$

Remarque 3.3.1 Dans le cas où f n'est pas quadratique on a

$$\beta_k^{\bar{F}R} \neq \beta_k^{PRP} \neq \beta_k^{CD} \neq \beta_k^{DY} \neq \beta_k^{HZ} \neq \beta_k^{RMIL}. \quad (3.32)$$

Par conséquent, en appliquant l'algorithme du gradient conjugué non quadratique, en utilisant les coefficients β_k figurant dans (3.32), on obtient des suites $\{x_k\}_{k \in \mathbb{N}}$ différentes.

Que se passe-t'il si f est quadratique strictement convexe et si α_k est obtenue par une recherche linéaire exacte. La réponse à cette question se trouve dans le théorème suivant.

Théorème 3.3.1 Si $f(x) = \frac{1}{2}x^T Qx - b^T x$, avec Q symétrique définie positive, $x \in \mathbb{R}^n, b \in \mathbb{R}^n$ et si α_k est obtenue par une recherche linéaire exacte. Notons

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k},$$

alors on a

$$\beta_k = \beta_k^{FR} = \beta_k^{PRP} = \beta_k^{CD} = \beta_k^{DY} = \beta_k^{HZ} = \beta_k^{RMIL}, \quad (3.33)$$

et l'algorithme du gradient conjugué quadratique génère la même suite $\{x_k\}_{k \in \mathbb{N}}$.

3.3.1 Algorithme

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ non quadratique et (PNQSC) le problème de minimisation non quadratique sans contraintes suivant

$$\min \{f(x) : x \in \mathbb{R}^n\} \text{ (PNQSC)}.$$

Pour construire l'algorithme du gradient conjugué cas non quadratique, on peut s'inspirer de l'algorithme du gradient conjugué quadratique établi au chapitre précédent.

Contrairement au cas quadratique, on n'a pas de matrice Q .

Par conséquent on n'a pas de directions Q conjuguées. Comme dans le cas quadratique, l'algorithme du gradient conjugué cas non quadratique génère une suite $\{x_k\}_{k \in \mathbb{N}}$ de la manière

3.3 Méthode du G.C. Cas non quadratique *Méthode du gradient conjugué(G.C)*

suivante

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

L'algorithme démarre d'un point $x_0 \in \mathbb{R}^n$ quelconque.

A l'itération k

Supposons qu'on ait le vecteur $x_k \in \mathbb{R}^n$ et la direction d_{k-1} . Ceci nous permet de calculer $\nabla f(x_k)$ au lieu $g_k = Qx_k - b$ dans le cas quadratique. Pour avoir x_{k+1} , on a besoin de calculer α_k et d_k .

Calcul de d_k :

$$d_k = -\nabla f(x_k) + \beta_{k-1} d_{k-1} \quad (3.34)$$

On a six façons pour calculer β_{k-1}

$$\beta_{k-1}^{PRB} = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|\nabla f(x_{k-1})\|^2}$$

$$\beta_{k-1}^{FR} = \frac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_{k-1})\|^2}$$

$$\beta_{k-1}^{DY} = \frac{\|\nabla f(x_k)\|^2}{d_{k-1}^T (\nabla f(x_k) - \nabla f(x_{k-1}))}$$

$$\beta_{k-1}^{CD} = -\frac{\|\nabla f(x_k)\|^2}{d_{k-1}^T \nabla f(x_{k-1})}$$

$$\beta_{k-1}^{HZ} = \left((\nabla f(x_k) - \nabla f(x_{k-1})) - 2d_{k-1} \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|^2}{d_{k-1}^T (\nabla f(x_k) - \nabla f(x_{k-1}))} \right)^T \frac{\nabla f(x_k)}{(\nabla f(x_k) - \nabla f(x_{k-1}))^T y_{k-1}}$$

$$\beta_{k-1}^{RMIL} = \frac{(\nabla f(x_k))^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|d_{k-1}\|^2}.$$

Calcul de α_k

Ayant obtenu d_k , rappelons que α_h vérifie

$$f(x_k + \alpha_k d_k) = \min \{ f(x_k + a d_k) : a \in]0, +\infty[\}. \quad (3.35)$$

3.3 Méthode du G.C. Cas non quadratique *Méthode du gradient conjugué(G.C)*

Dans le cas où $f(x) = \frac{1}{2}x^T Qx - b$, Q symétrique définie positive, α_k solution de (3.35), est donnée par la relation suivante

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k}. \quad (3.36)$$

Dans le cas où f n'est pas quadratique, α_k ne peut pas être calculée par la formule (3.36). On calcule dans ce cas α_k par d'autres méthodes. On utilise par exemple la méthode du nombre d'or ou la méthode de dichotomie. Comme on le verra plus loin, α_k peut être calculée par une recherche linéaire inexacte d'Armijo ou Goldstein ou Wolfe.

Algorithme

Etape 1

Choisir $x_0 \in \mathbb{R}^n$ Quelconque et $\epsilon > 0$.

Etape 2

Poser $k = 0$

Calculez $g_0 = \nabla f(x_0)$. Posez $d_0 = -g_0$.

Etape 3

Calculez α_k en utilisant une recherche linéaire exacte ou inexacte d'Armijo ou de Goldstein ou de Wolfe ou de Wolfe Forte

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

Etape 4

Si $\|\nabla f(x_{k+1})\| < \epsilon_1$ Stop, $x^* = x_{k+1}$ Sinon allez à Etape 5.

Etape 5

Calculez $g_{k+1} = \nabla f(x_{k+1})$

Calculez β_k par l'une des manières suivantes

$$\beta_k = \beta_k^{FR} \text{ ou } \beta_k = \beta_k^{PRP} \text{ ou } \beta_k = \beta_k^{CD}$$

$$\text{ou } \beta_k = \beta_k^{DY} \text{ ou } \beta_k = \beta_k^{HZ} \text{ ou } \beta_k = \beta_k^{RMIL}.$$

Calculez

$$d_{k+1} = -g_{k+1} + \beta_k d_k.$$

Posez $k = k + 1$ et allez a Etape 3.

3.4 Convergence de la méthode du G.C

Avant tout rappelons que les différentes variantes du gradient conjugué algorithmes génèrent une suite $\{x_k\}_{k \in \mathbb{N}}$ dans \mathbb{R}^n définie par $x_0 \in \mathbb{R}^n$ quelconque et les itérations

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \dots \quad (GC1)$$

avec α_k est le résultat d'une recherche linéaire inexacte du type Armijo ou Goldstein ou Wolfe ou Wolfe forte. Les directions d_k sont calculées itérativement par la formule

$$d_k = \left\{ \begin{array}{l} -g_0 \text{ si } k = 0 \\ = -g_k + \beta_{k-1} d_{k-1}, k \geq 1 \end{array} \right\} \quad (GC2)$$

avec $g_k = \nabla f(x_k)$ et

$$\beta_k = \beta_k^{FR} \text{ ou } \beta_k = \beta_k^{PRP} \text{ ou } \beta_k = \beta_k^{CD} \quad (GC3)$$

$$\text{ou } \beta_k = \beta_k^{DY} \text{ ou } \beta_k = \beta_k^{HZ} \text{ ou } \beta_k = \beta_k^{RMIL}.$$

Dans ce chapitre on va essayer de présenter une synthèse sur les différents résultats de convergence des méthodes du gradient conjugué pour la minimisation des fonctions sans contraintes. Ces méthodes seront utilisées avec une recherche linéaire inexacte.

L'analyse couvre quatre classes de méthodes qui sont globalement convergentes pour des fonctions régulières non nécessairement convexes.

Dans la première famille, ce sont certaines propriétés de la méthode de Fletcher-Reeves qui jouent un rôle crucial, tandis que la seconde famille c'est la méthode de Polak-Ribière-Polyak qui a une propriété importante.

La troisième concerne la méthode de la descente conjuguée et la dans dernière famille on va présenter quelques propriétés de la nouvelle méthode du gradient conjugué non linéaire dite de Dai-Yuan.

On terminera ce chapitre par Quelques développements récents de la méthode du Gradient conjugué.

3.5 Théorème de Zoutendijk

3.5.1 Hypothèses C1 et C2 (de Lipschitz et de bornetude)

Rappelons dans ce qui suit les deux conditions suivantes que toutes les méthodes du gradient conjugué doivent vérifier (ou au moins l'une d'entre elles) pour prouver les résultats de convergence.

Condition C1

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$. On dit que f vérifie la condition C1 si f est contiuement différentiable dans un voisinage $V(\Gamma)$ de $\Gamma = \{x \in \mathbb{R}^n : f(x) \leq f(x_1) : x_1 \in \mathbb{R}^n \text{ point initial} \}$ et si $\nabla f(\alpha)$ vérifie la condition de Lipschitz dans $V(\Gamma)$, c'est à dire, il existe une conatante L telle que

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \text{ pour tout } x, y \in V(\Gamma).$$

Condition C2

L'ensemble $\Gamma = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ est borné, i.e., il existe une constante $B < \infty$ telle que

$$\|x\| \leq B, \forall x \in \Gamma.$$

3.5.2 Théorème de Zoutendijk

L'outil de base utilisé par les différentes variantes du gradient conjugué avec une recherche linéaire inexacte est le théorème suivant dû à Zoutendijk.

Théorème de Zoutendijk

Considérons une méthode itérative quelconque générant une suite $\{x_k\}$ de la forme :

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

d_k étant une direction de descente et α_k est une recherche linéaire inexacte de Wolfe. Supposons aussi que f vérifie la condition C1. Alors on a

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \quad (3.37)$$

3.5.3 Théorème de Zoutendijk et la convergence globale

La condition $\sum_{k=0}^{\infty} \frac{(g_k^T \alpha_k)^2}{\|d_k\|^2} < \infty$ est équivalente à $\sum_{k=1}^{\infty} \cos^2(\theta_k) \|g_k\|^2 < \infty$ où θ_n est l'angle que fait d_k avec $-g_k$. Il est évident de voir que si

$$\cos(\theta_k) \geq \delta > 0$$

pour tout k , alors on a

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.38)$$

Dans les différentes méthodes du gradient conjugué on n'arrive pas à démontrer le résultat précédent i.e., $\lim_{n \rightarrow \infty} \|g_k\| = 0$, mais seulement

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.39)$$

La condition (3.39) implique qu'il existe une sous suite de $\{\|g_k\|\}$ qui tend vers zéro.

Conditions suffisantes associés au théorème de Zoutendijk pour démontrer la relation(3.39)

Pour démontrer (3.39) on associe au théorème de Zoutendijk les 2 hypothèses suivantes :

Hypothèse1

La descente suffisante est assuré i.e., il existe une constante c indépendante de k telle que

$$g_k^T d_k \leq -c \|g_k\|^2 .$$

Hypothèse 2

Il existe une constante β telle que

$$\|d_k\|^2 \leq \beta_k .$$

On a donc

$$\left\{ \begin{array}{l} \text{relation de Zoutendijk } \left(\sum_{k=0}^{\infty} \frac{(g_k^T \alpha_k)^2}{\|d_k\|^2} < \infty \right) \\ \text{Hypothèse1} \\ \text{Hypothèse2} \end{array} \right\} \Rightarrow \left\{ \liminf_{k \rightarrow \infty} \|g_k\| = 0 \right\} .$$

Les méthodes dans lesquelles $\|g_{k+1}\|^2$ figure dans β_k Les méthodes des **FR, DY et CD** ont toutes comme numérateur commun le terme $\|g_{k+1}\|^2$. Une différence fondamentale qui caractérise ces méthodes par rapport aux autres méthodes du gradient conjugué où β_k est calculé différemment.

3.6 Méthode de Fletcher-Reeves

Cette méthode été découverte en 1964 par Fletcher et Reeves ([35]), β_k est égale à :

$$\beta_{k+1}^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \quad (FR).$$

Synthèse des principaux résultats de convergence pour la méthode de Fletcher-Reeves

Le premier résultat de convergence globale de lo méthode de **FR** a été donné par Zoutendijk [77] en 1970 .

Il a prouvé que la méthode FletcherReeves converge globalement quand α_k est une recherche linéaire exacte, en d'autres mots.

En 1977 , Powell [58] a fait remarquer que la méthode de Fletcher-Reeves, avec une recherche linéaire exacte, était sensible numériquement.

Autrement dit, l'algorithme pourrait prendre de nombreuses mesures courtes sans faire d'importants progres au minimum.

La mauvaise performance de la méthode de **FR** dans les applications a été souvent attribuée à ce phénomène de brouillage.

Le premier résultat de convergence globale de la méthode de FR pour une recherche linéaire inexacte a été donné par Al- Baali [1] en 1985 .

En utilisant les conditions de Wolfe forte avec $\sigma < \frac{1}{2}$, il a prouvé que la méthode **FR** génère des directions de descente suffisantes. Plus précisément, il a prouvé que :

$$\frac{1 - 2\sigma + \sigma^{k+1}}{1 - \sigma} \leq \frac{-g_k^T d_k}{\|g_k\|^2} \leq \frac{1 - \sigma^{k+1}}{1 - \sigma},$$

pour tout $k \geq 0$.

Comme conséquence, la convergence globale a été démontrée en utilisant la condition Zoutendijk.

Pour $\sigma = 1/2$, d_k est une direction de descente, toutefois, l'analyse n'a pas établi la descente

suffisante.

Dans Liu et al. [50], la preuve de la convergence d'Al-Baali est étendue au $\cos \sigma = 1/2$.

Dai et Yuan [21] ont montré que dans les versions **FR** consécutives, au moins une itération satisfait la propriété de descente suffisante. En d'autres termes,

$$\max \left\{ \frac{-g_k^T d_k}{\|g_k\|^2}, \frac{-g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \right\} \geq \frac{1}{2}.$$

Le théorème de Zoutendijk peut également être utilisé pour obtenir un résultat de convergence globale pour les méthodes FR mis en oeuvre avec une recherche inexacte de Wolfe forte et $\sigma \leq 1/2$, puisque les directions de recherche sont toujours des directions de descente.

Dans [28], Dai et Yuan montrent qu'avec la méthode de **FR**, les conditions Wolfe fortes n'engendrent pas en général des directions de descente quand $\sigma > 1/2$, même pour la fonction $f(x) = \lambda \|x\|^2$, où $\lambda > 0$ est une constante.

Par conséquent, la contrainte $\sigma \leq 1/2$ doit être imposée pour assurer la descente.

Dans les implémentations typiques des conditions Wolfe, il est souvent plus efficace de choisir σ proche de 1 .

Par conséquent, la contrainte $\sigma \leq 1/2$, nécessaire pour assurer la descente, représente une restriction importante dans le choix des paramètres de la recherche linéaire.

D'autre part, Dai et Yuan montrent dans [25] que, lorsque $\sigma > 1/2$ et $g_k^T d_k > 0$, $-d_k$ peut être utilisé pour une direction de recherche, et si $g_k^T d_k = 0$, alors la recherche linéaire peut être ignorée par le choix $x_{k+1} = x_k$.

S'il existe une constante γ telle que $\|g_k\| \leq \gamma$, sous la condition de Lipschitz, la méthode de **FR**, avec une recherche linéaire de Wolfe et ces ajustements spéciaux quand $g_k^T d_k \geq 0$, est globalement convergente.

Dans [21], la recherche linéaire de Wolfe forte est étendue à une recherche linéaire de Wolfe. La convergence globale est obtenue lorsque $\sigma_1 + \sigma_2 \leq 1$.

Pour une recherche linéaire de Wolfe forte, $\sigma_1 = \sigma_2 = \sigma$, dans ce cas, la contrainte $\sigma_1 + \sigma_2 \leq 1$ implique que $\sigma \leq 1/2$.

Par conséquent, la condition $\sigma_1 + \sigma_2 \leq 1$ est plus faible que la contrainte de Wolfe forte $\sigma \leq 1/2$.

Et il est possible de prendre σ_1 proche de 1 , en prenant σ_2 près de 0 .

3.7 Méthode de Dai-Yuan

Introduction

Cette méthode été découverte par Y. H. Dai et Y. Yuan ([24]), β_k est égale à :

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}; \quad y_k = g_{k+1} - g_k(DY).$$

Cette méthode est fondamentalement différente de la méthode Fletcher Reeves et aussi de la méthode CD.

Avec une recherche linéaire de Wolfe, la méthode de **DY** génère toujours des directions de descente.

En plus, elle est globalement convergente avec des hypothèses tres générales sur la fonction objectif f .

On exige seulement que f soit continueuement différentiable et que le gradient soit Lipachitzien, c'est à dire que f vérifie l'hypothèse condition $C1$.

Où

2012([8]) - Gradient conjugué variante Rivaie-Mustafa-Ismail-Leong(RMIL) [60]

$$\beta_{k-1}^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2}. \quad (3.40)$$

3.8 Méthode de Dai-Yuan généralisée

Dans le procédé de l'analyse de la méthode de **DY**, Dai et Yuan ont généralisé leurs résultats pour toutes les méthodes du gradient conjugué où β_k peut se mettre sous la forme suivante

$$\beta_k = \beta_k^{DYG} = \frac{\Phi_{k+1}}{\Phi_k} \quad (DYG).$$

Le procédé de FR correspond au choix $\Phi_k = \|g_k\|^2$. β_k^{DY} peut être réécrit sous la forme,

$$\beta_k^{DY} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}.$$

Par conséquent, le procédé de DY a la forme (DYG) avec $\Phi_k = g_k^T d_k$.

Le résultat suivant a été établi par Dai et Yuan dans [29,31].

Considérons une méthode du type (GC1), (GC2), (GC3) avec $\beta_k = \beta_k^{DYG} = \frac{\phi_{k+1}}{\phi_n}$,

d_k est une direction de descente pour tout k .

Supposons aussi que l'hypothèse condition C1 soit satisfaite.

Si la relation de Zoutendijk $\left(\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \right)$ est vérifiée et si l'une des trois relations suivantes

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\Phi_k^2} = \infty \text{ ou } \sum_{k=0}^{\infty} \frac{\|g_k\|^2}{\Phi_k^2} = \infty \text{ ou } \sum_{k=1}^{\infty} \prod_{i=1}^k \beta_i^{-2} = \infty$$

a lieu, alors la suite générée est globalement convergente, i.e.,

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0.$$

Applications du théorème à la méthode de Dai Yuan

Comme corollaire de ce résultat, la méthode de DY est globalement convergente, si elle est mise en oeuvre avec une recherche linéaire inexacte de Wolfe.

En effet, on a $\Phi_k = g_k^T d_k$.

Par conséquent

$$\sum_{k=0}^N \frac{(g_k^T d_k)^2}{\Phi_k^2} = N + 1 \implies \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\Phi_k^2} = \lim_{n \rightarrow \infty} \sum_{k=0}^N \frac{(g_k^T d_k)^2}{\Phi_k^2} = \infty$$

Applications du théorème à la méthode de Fletcher Reeves

La méthode de FR est globalement convergente, si elle est mise en oeuvre avec une recherche linéaire inexacte de Wolfe forte avec $\sigma \leq 1/2$.

En effet, on a $\Phi_k = \|g_k\|^2$.

Par conséquent

$$\sum_{k=0}^N \frac{\|g_k\|^2}{\Phi_k^2} = N + 1 \implies \sum_{k=0}^{\infty} \frac{\|g_k\|^2}{\Phi_k^2} = \lim_{n \rightarrow \infty} \sum_{k=0}^N \frac{\|g_k\|^2}{\Phi_k^2} = \infty$$

Notez qu'une méthode générale de gradient conjugué peut être exprimée sous la forme (DYG)

en prenant $\Psi_0 = 1$, et

$$\Phi_k = \prod_{j=1}^n \beta_j, \text{ pour } k > 0.$$

Conclusion

Dans ce travail, nous avons présenté la méthode du gradient conjugué en utilisant les recherches linéaires inexactes pour résoudre un problème d'optimisation.

Nous avons discuté les recherches linéaires inexactes d'Armijo, Goldstein et Wolfe et nous avons étendu la recherche de Wolfe à recherche de wolfe forte.

Nous avons aussi étudié la convergence de la méthode du gradient conjugué.

Bibliographie

- [1] A.A. Goldstein and J.F. Price, *An effective algorithm for minimisation*, Num.Math., 10 (1969), pp. 184-189.
- [2] A. A. Goldstein, *On steepest descent*, SIAM J. on Control A, Vol. 3, No. 1 (1965), pp. 147-151.
- [3] A. Buckley, Conjugate gradient methods, in Nonlinear Optimization 1981, M. J. D. Powell, ed., Academic Press, London, (1982), pp. 1722.
- [4] B. Rachid "Optimisation Convexe, Optimisation Sans Contraintes", (2007).
- [5] E. M. L. Beale, A derivative of conjugate gradients, in Numerical Methods for Nonlinear Optimization, F. A. Lootsma, ed., Academic Press, London, (1972), pp. 39-43.
- [6] H. P. Crowder and P. Wolfe, Linear convergence of the conjugate gradient method, IBM J. Res. Dev., 16(1969), pp. 431 – 433.
- [7] L. Armijo, Minimization of function having lipschitz continuous first partial derivatives, Pacific Journal of Mathematics, Vol. 16(1966), pp. 1 – 3.
- [8] L. Andrieu, Optimisation sous contrainte en probabilité, THESE présentée pour l'obtention du titre de Docteur de l'Ecole Nationale des Ponts et Chaussées, décembre (2004)
- [9] M. Belloufi, R. Benzine, and Y. Laskri, Modification of the Armijo line search to satisfy the convergence properties of HS method, An International Journal of Optimization and Control : Theories & Applications (IJOCTA), 141(2013), 145 – 152.

- [10] M. Belloufi and R. Benzine, Global Convergence Properties of the HS Conjugate Gradient Method. Applied Mathematical Sciences, Vol. 7 , 142(2013), 7077 – 7091
- [11] R. Fletcher and C. Reeves, *Function minimisation by conjugate gradients*, Comput. J., 7(1964), pp. 149–154.
- [12] R. Fletcher and M. J. D. Powell, *A rapidly convergent descent méthode for minimisation*, Computer. J., 6 (1963), pp. 163-168.
- [13] R. Fletcher, *Practical Méthodes of Optimisation vol. 1 : Unconstrained Optimisation*, John Wiley Sons, New York, 1987.
- [14] .S.Chekroune, Y.Abdelali,A.Zaich "La résolution d'un problème d'optimisation Les méthodes de Quasi-Newton" Mémoire se Master(2020-2021).
- [15] W. W. Hager and H. Zhang, *A new conjugate gradient méthode with guaranteed descent and an efficient line search*, Novembre 17, (2003) (to appear in SIAM J. Optima.).
- [16] X.D. Chen and J. Sun, Global convergence of a two-parameter family of conjugate gradient methods without line search, Journal of Computational and Applied Mathematics 146(2002)37 – 45.
- [17] Y. H. Dai and Y. Yuan, *An efficient hybrid conjugate gradient méthode for unconstrained optimisation*, Ann. Oper. Res., 103 (2001), pp. 33–47.
- [18] Y. H. Dai and Y. Yuan, *A class of globally convergent conjugate gradient méthodes*, Sci. China Ser. A, 46 (2003), pp. 251–261.
- [19] Y. H. Dai and Y. Yuan, *A class of globally convergent conjugate gradient méthodes*, Research report ICM-98-030, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998.
- [20] Y. H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves méthode*, IMA J. Numer. Anal., 16 (1996), pp. 155–164.