

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Université 8 Mai 1945 Guelma

Faculté des Mathématiques et de l'Informatique  
et des Sciences de la Matière  
Département de Mathématiques



## Mémoire

Présenté en vue de l'obtention du diplôme de

Master en Mathématiques

Option : EDP Et Analyse numérique

Par :

Karkour Nadja

## Intitulé

**Synthèse sur les résultats de convergence de quelques récentes méthodes du gradient conjugué non linéaires**

Dirigé par : Dr. Mellal Romaïssa

Devant le jury

PRESIDENT	Dr. Boulares Hamid	MCA	Univ-Guelma
RAPPORTEUR	Dr. Mellal Romaïssa	MCB	Univ-Guelma
EXAMINATEUR	Dr. Bellaouar Djamel	MCA	Univ-Guelma

Session Juin 2022

---

# *Remerciements*

Tout d'abord, je remercie ALLAH le tout puissant de m'avoir donné le courage, la santé, la volonté et la patience de mener à terme ce modeste travail.

J'adresse tout particulièrement mes remerciements à Dr « Mellal Romaïssa », qui m'a encadrée, tout au long de ce mémoire. Je vous remercie pour votre précieuse présence assistance, votre disponibilité et votre patience. J'ai pris un grand plaisir à travailler avec vous.

Je remercie les membres de jury ici réunis, d'avoir accepté de participer à cette soutenance, et la lecture attentive qu'ils ont réservée à mon mémoire.

Je dédie ce mémoire spécialement aux êtres au très grand cœur pour leurs confiances en moi, leurs encouragements, leurs patiences, leurs aides et leurs soutiens et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours. Que ce modeste travail soit l'exaucement de vos vœux tant formulés, le fruit de vos innombrables sacrifices, bien que je ne vous en acquitterai jamais assez.. Sans eux je ne serais pas là à partager avec vous tout ce qui me tient à cœur.

Ces êtres sont bien sûr MAMAN et PAPA ainsi qu'à mes frères « Saïf – eddin » et « Badri ». Je remercie sincèrement ma petite famille, mon mari « Abd – ELAZIZE » et ma fille « Meriem soujoud ».

Je voudrais adresser mes derniers remerciements à tous mes collègues qui m'ont accompagné au cours de mon parcours universitaire et à tous mes professeurs avec qui j'ai beaucoup appris d'eux.

## ملخص

نعرض في هذه المذكرة الطريقة العددية للتدرج المترافق الخطية وغير الخطية  
نقدم لكم جميعا لنتائج التقارب لبعض طرق التدرج الغير الخطية الحديثة على سبيل المثال نأخذ  
طريقة LMY-DY, LMY-FR, JMG.

كل هذه المتغيرات لها خاصية نزول الاتجاهات المتولدة التقارب العالمي سيتم تقديم عمليات  
المحاكاة العددية لتوضيح تقارب الطرق المذكورة أعلاه

## الكلمات المفتاحية

طريقة التدرج المترافق اللاخطي, البحث الخطي غير الدقيق, التقارب الشامل, اتجاه النسب.

---

## Abstract

### Abstract :

In this thesis we present a synthesis on the convergence results of some recent nonlinear conjugate gradient method namely methods of JMG, LMY-DY and LMY-FR. All these variants generate descent directions and converge globally.

Numerical simulations will be presented in order to illustrate the convergence of the methods cited above.

Keywords : Descent direction, Non linear Conjugate gradient method, Inexact linear search, Global convergence.

## Résumé

Dans ce mémoire, on présente une synthèse sur les résultats de convergence de quelques récentes méthodes du gradient conjugué non linéaire. Il s'agit des méthodes de **JMG**, **LMY-DY** et **LMY-FR**. Toutes ces variantes possèdent la propriété de la descente des directions générées et la convergence globale.

Des simulations numériques seront présentées afin d'illustrer la convergence des méthodes cités au dessus.

**Mots-clés** : Direction de descente, Méthode du gradient conjugué non linéaire, Recherche linéaire inexacte, convergence globale.

---

# Table des matières

<b>1</b>	<b>Notions Préliminaires</b>	<b>3</b>
1.1	Rappel et définitions . . . . .	3
1.2	Problème d'optimisation sans contraintes . . . . .	4
1.2.1	Condition nécessaires d'optimalité . . . . .	5
1.2.2	Condition suffisantes d'optimalité . . . . .	5
1.3	Direction de descente . . . . .	6
<b>2</b>	<b>Méthode du gradient conjugué linéaire</b>	<b>7</b>
2.1	Méthode du gradient . . . . .	7
2.2	Méthode du gradient conjugué linéaire . . . . .	8
2.2.1	Détermination du pas $\alpha_k$ obtenu par une recherche linéaire exacte . . . . .	9
2.2.2	Détermination du coefficient $\beta_k$ . . . . .	10
2.2.3	Algorithme du gradient conjugué linéaire . . . . .	11
2.2.4	Convergence de méthode du gradient conjugué linéaire . . . . .	11
<b>3</b>	<b>Méthode du gradient conjugué non linéaire</b>	<b>15</b>
3.1	Recherches linéaire inexacte . . . . .	15
3.1.1	La règle d'Armijo . . . . .	16
3.1.2	Existence du pas d'Armijo . . . . .	16
3.1.3	Recherche linéaire inexacte de Wolfe . . . . .	17
3.1.4	Recherche linéaire inexacte de Wolfe forte . . . . .	18
3.1.5	Existence du pas de Wolfe . . . . .	18

---

3.2	Contribution de la recherche linéaire inexacte à la convergence des algorithmes à directions de descente. . . . .	19
3.2.1	Algorithme des recherches linéaire inexacte . . . . .	20
3.3	Différentes versions de la méthode du gradient conjugué non linéaire . . . . .	21
3.3.1	Méthode de Hestenes-Stiefel . . . . .	21
3.3.2	Méthode de Fletcher-Reeves . . . . .	21
3.3.3	Méthode de descente conjuguée . . . . .	22
3.3.4	Méthode de Dai et Yuan . . . . .	22
3.3.5	Algorithmes des méthodes du gradient conjugué non linéaire . . . . .	23
3.4	Convergence des méthodes du gradient conjugué non linéaires . . . . .	24
<b>4</b>	<b>Quelques récentes méthodes du gradient conjugué non linéaire</b>	<b>26</b>
4.1	Méthode de Jianghua-Meixing-Guodong . . . . .	26
4.1.1	Algorithme de la méthode JMG . . . . .	27
4.1.2	Convergence globale . . . . .	27
4.2	Méthode Liu-Ma-Yin (LMY-DY) . . . . .	28
4.2.1	Algorithme de la méthode LMY-DY . . . . .	28
4.2.2	Propriété de descente . . . . .	28
4.2.3	Convergence globale . . . . .	30
4.3	Méthode Liu-Ma-Yin (LMY-FR) . . . . .	33
4.3.1	Algorithme de méthode LMY-FR . . . . .	33
4.3.2	Propriété de descente . . . . .	34
4.3.3	Convergence globale . . . . .	35
<b>5</b>	<b>Simulations Numériques</b>	<b>37</b>
5.1	Performances des méthodes du gradient conjugué non linéaire de JMG LMY-DY et LMY-FR . . . . .	37
5.1.1	La fonction BRYBND (Broyden banded) . . . . .	38
5.1.2	La fonction test TRIDIA (cute) . . . . .	39
5.2	Conclusion . . . . .	40

---

# Introduction

La méthode du gradient conjugué est considérée comme une amélioration de la méthode du gradient. Elle est l'une des méthodes les plus utilisées pour résoudre les problèmes d'optimisation non-linéaire surtout de grande taille.

L'optimisation possède ses racines au 18-ième siècle dans les travaux de :

- Taylor, Newton , Lagrange, qui ont élaboré les bases des développements limités.
- Cauchy fut le premier à mettre en oeuvre une méthode d'optimisation, méthode du pas de descente (méthode du gradient), pour la résolution de problèmes sans contraintes.

Il faut attendre le milieu du vingtième siècle, avec l'émergence des calculateurs et surtout la fin de la seconde guerre mondiale pour voir apparaître des avancées spectaculaires en termes de techniques d'optimisation. A noter, ces avancées ont été essentiellement obtenues en Grande Bretagne.

En 1952, Hestenes et Steifel ([13, 1952]) à proposé une nouvelle méthode itérative qui sert à minimiser des fonctions quadratiques strictement convexes.

Plusieurs mathématiciens ont étendu cette méthode pour le cas non linéaire. Ceci a été réalisé pour la première fois en 1964 par Fletcher et Reeves ([10, 1964]) puis en 1987 par Fletcher (Méthode de la descente conjuguée)[11, 1987] Une autre variante a été proposé par Dai et Yuan en 1999 (méthode de Dai-Yuan)[9, 1999].

En 2011, Jianghua Y., Mexing L. et Guodong M. ont proposé une nouvelle variante du gradient conjugué non linéaire appelée JMG [14, 2011].

Récemment, les même auteurs ont découvert deux nouvelles variantes nommées LMY-DY et LMY-FR [15, 2020].

Dans ce modeste travail on va étudier les méthodes citées au dessus, tout en prouvant la descente de leurs directions, citant leurs algorithmes et démontrant leur convergence globale.

Nous avons mis en oeuvre notre propre simulation numérique (SCILAB) qui montre l'efficacité des méthodes JMG et LMY-FR pour le cas traité.

Ce manuscrit se divise en cinq chapitres:

-Dans le premier chapitre, on va présenter quelques notions préliminaires de l'optimisation sans contraintes, qui seront utilisés par la suite.

-Dans le deuxième chapitre, on va décrire la méthode du gradient conjugué linéaire dans le cadre de la minimisation des fonctions quadratiques strictement convexes, tout en déterminant le pas  $\alpha_k$  et les coefficients  $\beta_{k+1}$ , et en citant le théorème qui assure la convergence.

-Le troisième chapitre présente une synthèse sur quelques variantes de la méthode du gradient conjugué non linéaire. Avant d'étudier ces méthodes (algorithmes et convergence globale). On va d'abord d'écrire les recherches linéaires inexacts, ainsi élaborer leur contribution à la convergence.

-Quant au quatrième chapitre, on va étudier trois nouvelles variantes de la méthode du gradient conjugué non linéaires. Méthode JMG proposée en 2011 [14], les méthodes LMY-DY et LMY-FR, proposée en 2020 [15]. Ces trois variantes ont été introduites par Jianghua Y., Mexing L. et Guodong M.

On va étudier la descente de direction des méthodes LMY-DY et LMY-FR, et on termine par une étude de convergence de trois variantes citées au dessus.

- On termine ce manuscrit par des simulations numériques et une conclusion.

Dans ce chapitre on va rappeler quelques notions de base qui seront utilisées par la suite.

## 1.1 Rappel et définitions

**Définition 1.1.1** [3] Soit  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$  le gradient de cette fonction en  $x_0$  est le vecteur de  $\mathbb{R}^n$  définie par:

$$\overrightarrow{\text{Grad}f} = \nabla f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \frac{\partial f}{\partial x_2}(x_0) \\ \frac{\partial f}{\partial x_n}(x_0) \end{bmatrix} \quad (1.1.1)$$

**Définition 1.1.2** [3] a) On dit qu'une matrice carrée  $A \in M_{n \times n}$  est symétrique si  $A^T = A$ .

b) On dit qu'une matrice carrée  $A \in M_{n \times n}$  est définie positive si et seulement si:

$$\mathbb{R}^n, x^T A x > 0$$

**Définition 1.1.3** [3] Une partie  $\Omega$  de  $\mathbb{R}^n$  est dite convexe si :

$$\forall \lambda \in [0, 1], \forall (x, y) \in \Omega^2, (1 - \lambda)x + \lambda y \in \Omega \quad (1.1.2)$$

▷ Cela signifie que,  $\forall (x, y) \in \Omega^2$ , le segment  $[x, y]$  est tout entier inclus dans  $\Omega$ .

**Définition 1.1.4** [3] Soit  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$  une application, on dit que :

a) L'application  $f$  est convexe sur  $\mathbb{R}^n$  si :

$\forall \lambda \in [0, 1]; \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$  alors

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

b) L'application  $f$  est strictement convexe sur  $\mathbb{R}^n$  si :

$\forall \lambda \in [0, 1]; \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$  et  $x \neq y$  alors

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$$

**Définition 1.1.5** [3] Soit  $A \in M_{n \times n}$  une matrice symétrique,  $b$  est un vecteur dans  $\mathbb{R}^n$ ,  $c$  une constante dans  $\mathbb{R}$ . On dit que la fonction  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$  est une forme quadratique si on peut l'écrire sous la forme suivante:

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c \quad (1.1.3)$$

**Remarque 1.1.1** La fonction  $f(x)$  est dite une forme quadratique définie positive (resp semi-définie positive), lorsque  $A$  est une matrice définie positive (resp semi-définie positive).

**Définition 1.1.6** [11] Soit  $\Omega \subset \mathbb{R}^n$ ,  $f : \Omega \rightarrow \mathbb{R}$  une application, On dit que  $f$  est localement lipschitzienne si:  $\forall M > 0, \forall (x, y) \in \Omega^2$  :

$$\|f(x) - f(y)\| \leq M\|x - y\|. \quad (1.1.4)$$

## 1.2 Problème d'optimisation sans contraintes

Soit  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ , on appelle le problème de minimisation sans contraintes le problème suivant :

$$(P) \{ \min f(x), x \in \mathbb{R}^n \}$$

**Définition 1.2.1** [12] Soit  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$  une fonction continûment différentiable.

a) Soit  $x^* \in \mathbb{R}^n$ ,  $x^*$  est dite solution optimale globale (autrement dit minimum globale) du problème (P) si et seulement si:

$$\forall x \in \mathbb{R}^n, f(x^*) \leq f(x) \quad (1.2.1)$$

b) Soit  $x^* \in \mathbb{R}^n$ ,  $x^*$  est dite solution optimale locale (autrement dit minimum locale) du problème (P) si et seulement s'il existe un voisinage  $V_\varepsilon(x^*)$  de  $x^*$  tel que

$$\forall x \in V_\varepsilon(x^*), f(x^*) \leq f(x) \quad (1.2.2)$$

c) Soit  $x^* \in \mathbb{R}^n$ ,  $x^*$  est dite solution optimale locale stricte du problème (P) si et seulement s'il existe un voisinage  $V_\varepsilon(x^*)$  de  $x^*$  tel que

$$\forall x \in V_\varepsilon(x^*) \text{ et } x \neq x^*, f(x^*) < f(x) \quad (1.2.3)$$

### 1.2.1 Condition nécessaires d'optimalité

**Théorème 1.2.1** [5] Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  différentiable au point  $x^* \in \mathbb{R}^n$ .

Si  $x^*$  est un minimum locale de (P) alors  $\nabla f(x^*) = 0$ .

**Théorème 1.2.2** [5] Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  deux fois différentiable au point  $x^* \in \mathbb{R}^n$ .

Si  $x^*$  est un minimum local de (P) alors  $\nabla f(x^*) = 0$  et la matrice hessienne de  $f$  au point  $x^*$  est semi définie positive.

### 1.2.2 Condition suffisantes d'optimalité

**Théorème 1.2.3** [5] Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  deux fois différentiable au point  $x^* \in \mathbb{R}^n$ . Si  $\nabla f(x^*) = 0$  et  $H(x^*)$  est définie positive, alors  $x^*$  est un minimum locale stricte du problème d'optimisation sans contraintes (P).

**Théorème 1.2.4** [5] Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  telle que  $f$  est convexe et différentiable.

Alors  $x^*$  est un minimum globale de  $f$  si et seulement si  $\nabla f(x^*) = 0$ .

## 1.3 Direction de descente

**Définition 1.3.1** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x^* \in \mathbb{R}^n$  est dit direction de descente au point  $x^*$  si et seulement si il existe  $\delta > 0$  tel que:

$$f(x^* + \alpha d) < f(x^*) \forall \alpha \in ]0, \delta[ \quad (1.3.1)$$

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

**Théorème 1.3.1** [12] Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  différentiable au point  $x^* \in \mathbb{R}^n$  et  $d \in \mathbb{R}^n$  une direction vérifiant la condition suivante:

$$f'(x^*, d) = \nabla f(x^*)^T \cdot d < 0$$

Alors,  $d$  est une direction de descente au point  $x^*$ .

**Preuve.**  $f$  est différentiable au point  $x^*$  donc

$$f(x^* + \alpha d) = f(x^*) + \alpha \nabla f(x^*)^T \cdot d + \alpha \|d\| o(\alpha d) \quad \text{quand } \alpha \rightarrow 0$$

Ceci implique que

$$f'(x^*, d) = \lim_{\alpha \rightarrow 0} \frac{f(x^* + \alpha d) - f(x^*)}{\alpha} = \nabla f(x^*)^T \cdot d < 0 \quad (1.3.2)$$

La limite étant strictement négative, alors il existe un voisinage de zéro  $V(0) = ]-\delta, +\delta[$  tel que :

$$\frac{f(x^* + \alpha d) - f(x^*)}{\alpha} < 0, \forall \alpha \in ]-\delta, +\delta[ \quad (1.3.3)$$

La relation (1.3.3) est particulièrement vraie pour tout  $\alpha \in ]0, +\delta[$ . On obtient le résultat cherché en multipliant la relation (1.3.3) par  $\alpha > 0$  c'est à dire:  $f(x^* + \alpha d) < f(x^*)$  pour tout  $\alpha \in ]0, +\delta[$  L'ensemble des directions de descente de  $f$  en  $x$  est donné par

$$\left\{ d \in \mathbb{R}^n : \nabla f(x)^T \cdot d < 0 \right\}$$

Ce qui achève la démonstration. ■

**Remarque 1.3.1** La direction de descente  $d$  fait un angle  $\theta$  avec l'opposé du gradient  $-\nabla f(x)$  qui est strictement inférieur à  $90^\circ$  et on a :

$$\theta = \arccos \left( \frac{-\nabla^T f(x) \cdot d}{\|\nabla^T f(x)\| \|d\|} \right) \in \left[ 0, \frac{\pi}{2} \right]$$

---

# Méthode du gradient conjugué linéaire

La méthode du gradient conjugué est un algorithme développé dans le but de minimiser des fonctions quadratiques strictement convexes.

Il s'agit d'une méthode itérative à direction de descente qui converge en un nombre fini d'itérations, introduite en 1952 par *HESTENES* et *STEIFEL* [13].

Dans ce chapitre on va d'écrire la méthode du gradient conjugué linéaire, tout en donnant son principe, son algorithme et démontrer le théorème qui assure la convergence.

On va d'abord étudier brièvement la méthode du gradient .

## 2.1 Méthode du gradient

Il s'agit d'une méthode fondamentale proposée par *CAUCHY* [6] en 1847 afin de minimiser des problèmes d'optimisation sans contrainte. En effet, cette méthode prend comme direction  $d_k = -\nabla f(x)$ , dans ce cas on peut dire que la méthode du gradient est une méthode à direction de descente.

L'algorithme du gradient est également connu sous le nom d'algorithme de la plus forte pente ou de la plus profonde descente parce que le gradient est la pente de la fonction au point courant et est donc, localement sa plus forte pente.

**Algorithme 2.1 (Algorithme de la méthode du gradient)****Etape 0 (initialisation)**

Soit  $x_0$  le point de départ.

Poser  $d_0 = -\nabla f(x_0)$ ,

Poser  $k = 0$  et aller à l'étape 1.

**Etape 1**

Si  $\|\nabla f(x_k)\| \simeq 0$  :stop ( $x^* = x_k$ ). "Test d'arrêt".

Si non aller à l'étape 2.

**Etape 2**

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

$\alpha_k$  est déterminé par une recherche linéaire

$d_k = -\nabla f(x_k)$

Poser  $k = k + 1$  et aller à l'étape 1.

## 2.2 Méthode du gradient conjugué linéaire

Les méthodes du gradient conjugué ont été introduites en 1952 par HESTENES et STEIFEL [13], dans le but de minimiser des fonctions quadratiques strictement convexes. Elles reposent sur le concept des directions conjuguées parce que les gradients successifs sont orthogonaux entre eux et aux directions précédentes.

L'idée initiale était de trouver une suite à direction de descente qui est mutuellement conjuguée par rapport à la fonction quadratique permettant de résoudre le problème:

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

où  $f$  est une forme quadratique strictement convexe.

Ça veut dire que la minimisation de  $f(x)$  revient à la minimisation de  $q(x)$ , où :

$$q(x) = \frac{1}{2}x^T A x - b^T x + c$$

avec  $x \in \mathbb{R}^n$  et  $A \in M_{n \times n}$  est symétrique définie positive,  $b$  et  $c \in \mathbb{R}$ .

La méthode du gradient conjugué linéaire est une méthode itérative, elle consiste à générer une suite  $(x_k)_{k \in \mathbb{N}}$  de la manière suivante:

On démarre par un vecteur de départ  $x_0$  donné. A l'itération  $k$ , si on a  $x_k \in \mathbb{R}^n$ , le successeur  $x_{k+1}$  de  $x_k$  est donné par la relation suivante:

$$x_{k+1} = x_k + \alpha_k d_k, \quad (2.2.1)$$

où  $\{d_1, d_2, \dots, d_n\}$  est une famille de directions de descente possédant la propriété d'être mutuellement conjuguées par rapport à la fonction quadratique  $q(x)$ .

La direction de descente recherchée est définie par la formule de récurrence suivante:

$$d_k = \begin{cases} -g_1 = \nabla q(x_1) & \text{si } k=1 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 2 \end{cases} \quad (2.2.2)$$

Le gradient de  $f$  au point  $x_k$  est notée par  $g_k$ . On détermine la valeur du pas  $\alpha_k \in \mathbb{R}^+$  par une recherche linéaire c'est ce qu'on appelle une optimisation unidimensionnelle et le coefficient  $\beta_k \in \mathbb{R}$  et déterminé d'une manière où les directions  $d_k$  soient mutuellement conjugué par rapport à la forme quadratique  $q(x)$ .

### 2.2.1 Détermination du pas $\alpha_k$ obtenu par une recherche linéaire exacte

Une recherche linéaire exacte est une méthode qui sert à déterminer un pas  $\alpha_k > 0$ , qui minimise suffisamment  $f(x_k + \alpha d_k)$ , sachant que le vecteur  $x_k$  et la direction  $d_k$  sont donnés à l'itération  $k$ . Autrement dit, résoudre le problème d'optimisation unidimensionnel suivant:

$$\alpha_k = \min_{\alpha > 0} f(x_k + \alpha d_k).$$

On note

$$g_k = \nabla f(x_k) = \nabla q(x_k) = Ax_k - b$$

Comme  $\alpha_k$  minimise  $q$  dans la direction  $d_k$  alors le pas  $\alpha_k$  vérifie la condition nécessaire d'optimalité et on a

$$\begin{aligned}\forall k : \frac{dq}{d\alpha}(\alpha_k) &= q'(\alpha_k) = d_k^T \nabla q(x_{k+1}) = 0 \\ d_k^T \nabla q(x_{k+1}) &= d_k^T (Ax_{k+1} - b) = 0\end{aligned}$$

D'où

$$\begin{aligned}d_k^T A(x_k + \alpha_k d_k) - d_k^T b &= 0 \\ \Rightarrow d_k^T Ax_k + d_k^T A\alpha_k d_k - d_k^T b &= 0\end{aligned}$$

Alors

$$\begin{aligned}d_k^T A\alpha_k d_k &= -d_k^T Ax_k + d_k^T b \\ \Rightarrow d_k^T A\alpha_k d_k &= -d_k^T (Ax_k - b)\end{aligned}$$

D'où l'on tire

$$\alpha_k = \frac{-d_k^T (Ax_k - b)}{d_k^T Ad_k}$$

Donc

$$\alpha_k = \frac{-d_k^T g_k}{d_k^T Ad_k} \tag{2.2.3}$$

Le pas  $\alpha_k$  obtenu ainsi s'appelle le pas optimal.

### 2.2.2 Détermination du coefficient $\beta_k$

Les coefficients  $\beta_k$  sont choisis de telle sorte que  $d_k$  soit conjuguée avec toutes les directions précédentes. En d'autres termes on doit avoir:

En d'autres termes on doit avoir  $d_{k+1}^T Ad_k = 0$

On en déduit :

$$\begin{aligned}d_{k+1}^T Ad_k = 0 &\Rightarrow (-\nabla q(x_{k+1}) + \beta_{k+1} d_k)^T Ad_k = 0 \\ \Rightarrow -\nabla^T q(x_{k+1}) Ad_k + \beta_{k+1} d_k^T Ad_k &= 0 \\ \Rightarrow \beta_{k+1} &= \frac{\nabla^T q(x_{k+1}) Ad_k}{d_k^T Ad_k} = \frac{g_{k+1}^T Ad_k}{d_k^T Ad_k}\end{aligned}$$

### 2.2.3 Algorithme du gradient conjugué linéaire

Notons  $g_k = \nabla q(x_k)$  le gradient de la fonction  $q$  en  $x_k$ , l'algorithme du gradient conjugué est comme suit :

**Algorithme 2.2 (Algorithme du gradient conjugué linéaire)**

#### Etape 0 (initialisation)

Soit  $x_0$  le point de départ,  $g_0 = \nabla q(x_0) = Ax_0 + b$ , poser  $d_0 = -g_0$

Poser  $k = 0$  et aller à l'étape 1.

#### Etape 1

Si  $\|g_k\| = 0$  : STOP ( $x^* = x_k$ ). "Test d'arrêt"

Si non aller à l'étape 2.

#### Etape 2

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

$$\begin{aligned}\alpha_k &= \frac{-d_k^T g_k}{d_k^T A d_k} \\ d_{k+1} &= -g_{k+1} + \beta_{k+1} d_k \\ \beta_{k+1} &= \frac{g_{k+1}^T A d_k}{d_k^T A d_k}\end{aligned}\tag{2.2.4}$$

Poser  $k = k + 1$  et aller à l'étape 1.

### 2.2.4 Convergence de méthode du gradient conjugué linéaire

La méthode du gradient conjugué linéaire converge vers la solution optimale en  $n$  itération.

Pour démontrer sa validité, nous devons tout d'abord énoncer les lemmes suivants:

**Lemme 2.2.1** [16] *Notons par  $g_k = \nabla q(x_k) = Ax_k - b$  où  $x_k \in \mathbb{R}^n$  les itérées générées*

*par l'algorithme du gradient conjugué linéaire, pour tout  $k = 0, \dots, n - 1$ . Alors on a :*

$$g_{k+1}^T g_i = 0 ; \text{ pour tout } k = 0, \dots, n - 1 \text{ et } i = 0, \dots, k.$$

**Lemme 2.2.2** [5] *Les direction  $(d_1, d_2, \dots, d_{n-1})$  engendrées par l'algorithme du gradient conjugué linéaire sont  $A$ -conjuguées.*

Le théorème suivant assure la convergence de la méthode du gradient conjugué linéaire en  $n$  itération.

**Lemme 2.2.3 Théorème 2.2.1** [16] *A une itération  $k$  quelconque de l'algorithme où l'optimum de  $q(x)$  n'est pas encore atteint (c'est-à-dire  $g_i \neq 0, i = 0, 1, \dots, k$ ) on a :*

**Théorème 2.2.2** [16] a)

$$\alpha_k = \frac{g_k^T g_k}{d_k^T A d_k} \quad (2.2.5)$$

b)

$$\beta_{k+1}^T = \frac{g_{k+1}^T [g_{k+1} - g_k]}{g_k^T g_k} \quad (2.2.6)$$

$$= \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \quad (2.2.7)$$

**Preuve.** On démontre par récurrence. Supposons que  $\{d_1, d_2, \dots, d_k\}$  sont mutuellement conjuguées.

a) Montrons d'abord l'équivalence de (2.2.7) et de (2.2.4)

On a:  $d_k = -g_k + \beta_k d_{k-1}$

Donc (2.2.4) s'écrit :

$$\begin{aligned} \alpha_k &= \frac{-d_k^T g_k}{d_k^T A d_k} \\ &= \frac{-[-g_k + \beta_k d_{k-1}]^T g_k}{d_k^T A d_k} \\ &= \frac{g_k^T g_k}{d_k^T A d_k} = -\beta_k \frac{d_{k-1}^T g_k}{d_k^T A d_k} \end{aligned}$$

comme  $(d_1, d_2, \dots, d_{k-1})$  sont mutuellement conjuguées,  $x_k$  est l'optimum de  $q(x)$  sur la variété  $v_k$  passant par  $x_0$  et engendrée par  $(d_1, d_2, \dots, d_k)$  donc  $d_k^T g_{k+1} = 0$ . D'où l'on déduit(2.2.7).

b) Pour démontrer (2.2.8) remarquons que :

$$\begin{aligned} g_{k+1} - g_k &= A(x_{k+1} - x_k) = \alpha_k A d_k \\ \Rightarrow A d_k &= \frac{1}{\alpha_k} [g_{k+1} - g_k] \end{aligned}$$

On a alors:

$$g_{k+1}Ad_k = \frac{1}{\alpha_k} g_{k+1}^T [g_{k+1} - g_k]$$

Et en utilisant (2.2.7)

$$\alpha_k = \frac{g_k^T g_k}{d_k^T Ad_k}$$

Il vient :

$$\begin{aligned} g_{k+1}^T Ad_k &= \frac{d_k^T Ad_k}{g_k^T g_k} g_{k+1}^T [g_{k+1} - g_k] \\ &\Rightarrow \frac{g_{k+1}^T Ad_k}{d_k^T Ad_k} = \frac{g_{k+1}^T [g_{k+1} - g_k]}{g_k^T g_k} \end{aligned}$$

Or

$$\beta_{k+1} = \frac{g_{k+1}^T Ad_k}{d_k^T Ad_k} = \frac{g_{k+1}^T [g_{k+1} - g_k]}{g_k^T g_k}$$

Ce qui démontre (2.2.8)

Alors du fait que :

$$g_{k+1}^T g_k = 0$$

Car

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

Appartient au sous-espace engendré par  $(d_1, d_2, \dots, d_k)$  et que  $g_{k+1}$  est orthogonal à ce sous-espace, de cela nous obtenons la formule(2.2.9) .

c) Montrons enfin que  $d_{k+1}$  est conjugué par rapport à  $(d_1, d_2, \dots, d_k)$ , on a :

$$\begin{aligned} d_{k+1}^T Ad_k &= (-g_{k+1} + \beta_{k+1} d_k)^T Ad_k \\ &= -g_{k+1}^T Ad_k + \beta_{k+1} d_k^T Ad_k \\ &= -g_{k+1}^T Ad_k + \frac{g_{k+1}^T Ad_k}{d_k^T Ad_k} d_k^T Ad_k = 0 \end{aligned}$$

Vérifions maintenant que  $d_{k+1}^T Ad_i = 0$  pour  $i = 0, 1, \dots, k - 1$

On a :

$$d_{k+1}^T Ad_i = -g_{k+1}^T Ad_i + \beta_{k+1} d_k^T Ad_i$$

Le seconde terme est nul par l'hypothèse de récurrence ( $(d_1, d_2, \dots, d_k)$  sont mutuellement conjuguées).

Montrons qu'il en est de même du premier terme.

Puisque  $x_{i+1} = x_i + \alpha_i d_i$  et que  $\alpha_i \neq 0$  on a :

$$\begin{aligned} Ad_i &= \frac{1}{\alpha_i} (Ax_{i+1} - Ax_i) \\ &= \frac{1}{\alpha_i} (g_{i+1} - g_i) \end{aligned}$$

En écrivant :

$$\begin{aligned} g_{i+1} &= d_{i+1} - \beta_i d_i; \\ g_i &= d_i - \beta_{i-1} d_{i-1}; \end{aligned}$$

On voit que  $Ad_i$  est une combinaison linéaire de  $d_{i+1}$  et de  $d_{i-1}$  seulement.

Mais, puisque  $(d_1, d_2, \dots, d_k)$  sont mutuellement conjugué, on sait que le point  $x_{k+1}$  est l'optimum de  $q(x)$  sur la variété  $v_{k+1}$ , engendrée par  $(d_1, d_1, \dots, d_k)$ .

Donc  $g_{k+1}$  est orthogonal au sous-espace engendré par  $(d_1, d_2, \dots, d_k)$  et comme  $Ad_i$  appartient a ce sous-espace pour  $i = 0, 1, \dots, k - 1$ , on en déduit  $g_{k+1}Ad_i = 0$  ce qui achève la démonstration. ■

---

# Méthode du gradient conjugué non linéaire

La méthode du gradient conjugué linéaire sert à minimiser des fonctions quadratiques strictement convexes. A chaque itération le pas  $\alpha_k$  est déterminé par une recherche linéaire exacte. Dans le cas non linéaire aucune condition n'est imposé sur la fonction et donc le pas ne peut être déterminé par une recherche linéaire exacte.

La première extension de la méthode du gradient conjugué linéaire pour le cas non linéaire a été introduite par Fletcher-Reeves en 1964 [10], où le pas  $\alpha_k$  est déterminé par une recherche linéaire inexacte.

Dans ce chapitre on va présenter une synthèse sur quelques variantes de la méthode du gradient conjugué non linéaire. Avant d'étudier ces méthodes (algorithme et convergence globale).

On va d'abord d'écrire les recherches linéaires inexactes d'Armijo [4] et de Wolfe [18] et illustrer leur contribution à la convergence.

## 3.1 Recherches linéaire inexacte

La recherche linéaire inexacte est une brik essentiel dans la minimisation des fonctions non linéaires par la méthode du gradient conjugué non linéaire.

En effet, à chaque itération  $k$  on cherche à minimiser d'une façon importante la valeur de l'objectif par un pas  $\alpha_k > 0$  tel que

$$\alpha_k = \min_{\alpha > 0} f(x_k + \alpha d_k)$$

Le choix positif garde la descente de direction et la détermination du pas  $\alpha_k$  nous permet d'introduire la nouvelle itération

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

### 3.1.1 La règle d'Armijo

On bute à réduire “de façon importante” la valeur de l'objectif  $f$  par un pas  $\alpha_k > 0$  dans la direction  $d_k$ , telle que  $f(x_k + \alpha_k d_k) < f(x_k)$ .

Or cette condition de décroissance stricte n'est pas suffisante pour minimiser  $f$  au moins localement.

Une condition naturelle est de demander que  $f$  décroisse autant qu'une portion  $\rho \in ]0, 1[$  parfois appelée condition d'Armijo ou condition de décroissance linéaire

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla f(x_k)^T d_k \quad (3.1.1)$$

Le choix du pas positif assure que la descente de la direction et  $f(x_k + \alpha_k d_k) < f(x_k)$  car  $\rho$  devra être choisi dans  $]0, 1[$ .

### 3.1.2 Existence du pas d'Armijo

**Théorème 3.1.1** [4] *Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x_k \mapsto f(x_k + \alpha_k d_k)$  une fonction continue et bornée inférieurement,  $d_k$  est une direction de descente en  $x_k$ ,  $\alpha_k > 0$ , et  $\rho \in ]0, 1[$ , alors l'ensemble des pas vérifiant la règle d'Armijo est non vide.*

**Preuve.** Posons

$$\psi_\rho(\alpha) = f(x_k) + \rho \alpha \nabla^T f(x_k) d_k \quad (3.1.2)$$

Le développement de Taylor-Yong en  $\alpha = 0$  de  $f(x_k + \alpha d_k)$  est:

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

et comme  $\rho \in ]0,1[$  et  $\frac{d}{d\alpha} f(x_k + \alpha d_k) = \nabla^T f(x_k) d_k < 0$  on déduit :

$$f(x_k) + \alpha \nabla^T f(x_k) d_k < f(x_k) + \rho \alpha \nabla^T f(x_k) d_k \text{ pour } \alpha > 0$$

On voit que bien pour  $\alpha > 0$  assez petit on a :

$$f(x_k + \alpha d_k) < \psi_\rho(\alpha)$$

De ce qui précède et du fait que  $f(x_k + \alpha d_k)$  est bornée inférieurement,

$$\text{et } \psi_\rho(\alpha) \rightarrow -\infty; \alpha \rightarrow +\infty,$$

On déduit que la fonction  $\psi_\rho(\alpha) - f(x_k + \alpha d_k)$  possède la propriété :

$$\psi_\rho(\alpha) - f(x_k + \alpha d_k) > 0 \text{ pour } \alpha \text{ assez petit.}$$

$$\psi_\rho(\alpha) - f(x_k + \alpha d_k) < 0 \text{ pour } \alpha \text{ assez grand}$$

Donc s'annule au moins une fois pour  $\alpha > 0$ .

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

$$f(x_k + \bar{\alpha} d_k) = \psi_\rho(\bar{\alpha}) \text{ et } f(x_k + \alpha d_k) < \psi_\rho(\alpha) \text{ pour } 0 < \alpha < \bar{\alpha}$$

Ce qui achève la démonstration. ■

### 3.1.3 Recherche linéaire inexacte de Wolfe

On dit que  $\alpha_k$  est un pas de Wolfe [18] si  $\alpha_k$  vérifie les deux condition suivantes :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla^T f(x_k) d_k \quad 0 < \rho < \frac{1}{2} \quad (3.1.3)$$

$$\frac{d}{d\alpha} f(x_k + \alpha_k d_k) \Big|_{\alpha=\alpha_k} \geq \sigma \frac{d}{d\alpha} f(x_k) \Big|_{\alpha=0} \quad \rho < \sigma < 1. \quad (3.1.4)$$

Ainsi, on aura

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla^T f(x_k) d_k, \quad 0 < \rho < \frac{1}{2} \quad (3.1.5)$$

$$\nabla^T f(x_k + \alpha_k d_k) d_k \geq \sigma \nabla^T f(x_k) d_k, \quad \rho < \sigma < 1. \quad (3.1.6)$$

### 3.1.4 Recherche linéaire inexacte de Wolfe forte

Comme leur nom l'indique, les conditions de Wolfe forte sont plus fortes que les conditions de Wolfe.

En effet les conditions de Wolfe forte impliquent les condition de Wolfe.

**Définition 3.1.1** [18] *On dit que  $\alpha_k$  vérifie les condition de Wolfe forte si les deux conditions suivantes sont vérifiées :*

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla f(x_k) \cdot d_k, \quad 0 < \rho < \frac{1}{2} \quad (3.1.7)$$

$$|\nabla f(x_k + \alpha_k d_k) \cdot d_k| \leq -\sigma \nabla f(x_k) \cdot d_k, \quad \rho < \sigma < 1. \quad (3.1.8)$$

### 3.1.5 Existence du pas de Wolfe

Le théorème suivant assure l'existence du pas de Wolfe

**Théorème 3.1.2** [18] *Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x_k \mapsto f(x_k + \alpha_k d_k)$  une fonction continue et bornée inférieurement,  $d_k$  est une direction de descente en  $x_k$ ,  $\alpha_k > 0$ ,  $0 < \rho < \frac{1}{2}$  et  $\rho < \sigma < 1$ , alors l'ensemble des pas vérifiant la règle de Wolfe (3.1.5) et (3.1.6) est non vide .*

**Preuve.** Supposons

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

Le développement de Taylor-Yong en  $\alpha = 0$  de  $f$  est

$$f(x_k + \alpha d_k) = f(x_k) + \rho \alpha \nabla^T f(x_k) d_k + \alpha \xi(\alpha) \quad \text{ou } \xi(\alpha) \rightarrow 0, \alpha \rightarrow 0$$

Et comme  $\rho \in ]0,1[$  et  $\nabla^T f(x_k) d_k < 0$  on déduit :

$$f(x_k) + \alpha \nabla^T f(x_k) d_k < f(x_k) + \rho \alpha \nabla^T f(x_k) d_k \quad \text{pour } \alpha > 0$$

On voit que pour  $\alpha > 0$  assez petit on a:

$$f(x_k + \alpha d_k) < \psi_\rho(\alpha)$$

3.2. Contribution de la recherche linéaire inexacte à la convergence des algorithmes à directions de descente.

---

De ce qui précède et du fait que  $f$  est bornée inférieurement, et  $\psi_\rho(\alpha) \rightarrow -\infty$ , quand  $\alpha \rightarrow +\infty$ , on déduit que la fonction  $\psi_\rho(\alpha) - f(x_k + \alpha d_k)$  à la propriété

$$\begin{aligned}\psi_\rho(\alpha) - f(x_k + \alpha d_k) &> 0 \text{ pour } \alpha \text{ assez petit} \\ \psi_\rho(\alpha) - f(x_k + \alpha d_k) &< 0 \text{ pour } \alpha \text{ assez grand}\end{aligned}$$

Donc s'annule au moins une fois pour  $\alpha > 0$

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

$$f(x_k + \bar{\alpha}d_k) = \psi_\rho(\bar{\alpha}) \text{ et } f(x_k + \alpha d_k) < \psi_\rho(\alpha) \text{ pour } 0 < \alpha < \bar{\alpha} \quad (3.1.9)$$

La formule d'accroissements finis fournit alors un nombre  $\hat{\alpha}$ ,  $0 < \hat{\alpha} < \bar{\alpha}$  tel que

$$\begin{aligned}f(x_k + \bar{\alpha}d_k) - f(x_k) &= \bar{\alpha} \nabla^T f(x_k + \hat{\alpha}d_k) d_k \\ \Rightarrow \rho \bar{\alpha} \nabla^T f(x_k) d_k &= \bar{\alpha} \nabla^T f(x_k + \hat{\alpha}d_k) d_k \\ \Rightarrow \nabla^T f(x_k + \hat{\alpha}d_k) d_k &= \rho \nabla^T f(x_k) d_k \geq \sigma \nabla^T f(x_k) d_k\end{aligned}$$

Car  $0 < \rho < \sigma < 1$  et  $\nabla^T f(x_k) d_k < 0$

Donc  $\hat{\alpha}$  satisfait (3.1.3)

D'autre part  $\alpha = \hat{\alpha}$  satisfait (3.1.5)

En effet,  $\hat{\alpha}$  satisfait (3.1.8)  $f(x_k + \hat{\alpha}d_k) < \psi_\rho(\hat{\alpha})$  n'est autre que :

$$f(x_k + \hat{\alpha}d_k) < f(x_k) + \rho \hat{\alpha} \nabla^T f(x_k) d_k$$

Ce qu'il fallait démontrer. ■

## 3.2 Contribution de la recherche linéaire inexacte à la convergence des algorithmes à directions de descente.

On va étudier la contribution de la recherche linéaire inexacte à la convergence des algorithmes à directions de descente.

### 3.2. Contribution de la recherche linéaire inexacte à la convergence des algorithmes à directions de descente.

---

Ce n'est qu'une contribution, parce que la recherche linéaire ne peut à elle seule assurer la convergence des itérés, autrement dit, que le choix de la direction de descente joue aussi un rôle.

Cela se traduit par une condition, dite de Zoutendijk [19], dont on peut tirer quelques informations qualitatives intéressantes.

**Définition 3.2.1** [19] (*Condition de Zoutendijk*)

On dit qu'une règle de recherche linéaire à satisfait la condition de zoutendijk s'il existe une constante  $c > 0$  telle que pour tout  $k \geq 1$  on ait

$$f(x_{k+1}) \leq f(x_k) - c \|\nabla f(x_k)\|^2 \cos^2(\theta_k),$$

où  $\theta_k$  est l'angle que fait  $d_k$  avec  $-\nabla f(x_k)$ , défini par

$$\cos \theta_k = \frac{-\nabla f(x_k) d_k}{\|d_k\| \|d_k\|}.$$

Voici comment on se sert de la condition de Zoutendijk

**Proposition 3.2.1** [19] *Supposons que la suite  $(x_k)_{k \in \mathbb{N}}$  est générée par un algorithme d'optimisation vérifie la condition de zoutendijk et que la suite  $(f(x_k))_{k \in \mathbb{N}}$  est minorée, alors*

$$\sum_{k \geq 1} \|\nabla f(x_k)\|^2 \cos^2 \theta_k < \infty \quad (3.2.1)$$

#### 3.2.1 Algorithme des recherches linéaire inexacte

Présentons maintenant l'algorithme des recherches linéaires inexactes d'Armijo et de Wolfe.

**Algorithme 3.1 (Règle d'Armijo, et Wolfe)**

**Etape 0 (initialisation)**

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

**Etape 1**

Si  $\alpha_k$  vérifie les critères d'Armijo ou de Wolfe: STOP ( $\alpha^* = \alpha_k$ ).

Sinon

Poser  $\alpha_{k+1} = \rho \alpha_k$

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

### 3.3 Différentes versions de la méthode du gradient conjugué non linéaire

Dans cette section on va citer quatre versions différentes de la méthode du gradient conjugué non linéaire.

#### 3.3.1 Méthode de Hestenes-Stiefel

Cette méthode fut découverte par Hestenes et Stiefel [13, 1952] dont la variante est :

$$\beta_{k+1}^{HS} = \frac{g_{k+1}^T [g_{k+1} - g_k]}{d_k^T [g_{k+1} - g_k]} \quad (3.3.1)$$

On voit bien que  $\beta_{k+1}^{HS}$  est égale à la variante de la méthode du gradient conjugué linéaire, en effet

$$\begin{aligned} \beta_{k+1} &= \frac{g_{k+1}^T A d_k}{d_k^T A d_k} = \frac{g_{k+1}^T \frac{1}{\alpha_k} [g_{k+1} - g_k]}{d_k^T \frac{1}{\alpha_k} [g_{k+1} - g_k]} \\ &= \frac{g_{k+1}^T [g_{k+1} - g_k]}{d_k^T [g_{k+1} - g_k]} = \beta_{k+1}^{HS} \end{aligned}$$

#### 3.3.2 Méthode de Fletcher-Reeves

Cette méthode a été découverte en 1964 par Fletcher et Reeves [10, 1964] dont la variante est

$$\beta_{k+1}^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \quad (3.3.2)$$

### 3.3. Différentes versions de la méthode du gradient conjugué non linéaire

On voit bien que  $\beta_{k+1}^{FR}$  est égale à la variante de la méthode du gradient conjugué linéaire, en effet

$$\begin{aligned}\beta_{k+1} &= \frac{g_{k+1}^T Ad_k}{d_k^T Ad_k} = \frac{g_{k+1}^T [g_{k+1} - g_k]}{d_k^T [g_{k+1} - g_k]} \\ &= \frac{g_{k+1}^T [g_{k+1} - g_k]}{\|g_k\|^2} = \frac{\|g_{k+1}\|^2 - g_{k+1}^T g_k}{\|g_k\|^2}\end{aligned}$$

D'autre part, on a :

$$\begin{aligned}g_{k+1}^T g_k &= g_{k+1}^T (-d_k + \beta_k d_{k-1}) = -g_{k+1}^T d_k + \beta_k g_{k+1}^T d_{k-1} = \beta_k g_{k+1}^T d_{k-1} \\ &= \beta_k (g_k + \alpha_k Ad_k)^T d_{k-1} = \beta_k g_k^T d_{k-1} + \beta_k d_k^T Ad_{k-1} = 0\end{aligned}$$

On aura donc :

$$\beta_{k+1} = \frac{\|g_{k+1}\|^2 - g_{k+1}^T g_k}{\|g_k\|^2} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \beta_{k+1}^{FR}$$

#### 3.3.3 Méthode de descente conjuguée

Cette méthode a été proposée en 1987 par Fletcher [11, 1987] dont la variante est :

$$\beta_{k+1}^{CD} = \frac{\|g_{k+1}\|^2}{-d_k^T g_k} \quad (3.3.3)$$

On voit bien que  $\beta_{k+1}^{CD}$  est égale à la variante de la méthode du gradient conjugué linéaire, en effet

$$\begin{aligned}\beta_{k+1} &= \frac{g_{k+1}^T [g_{k+1} - g_k]}{d_k^T g_{k+1} - d_k^T g_k} \\ &= \frac{g_{k+1}^T [g_{k+1} - g_k]}{-d_k^T g_k} = \frac{\|g_{k+1}\|^2 - g_{k+1}^T g_k}{-d_k^T g_k} \\ &= \frac{\|g_{k+1}\|^2}{-d_k^T g_k} = \beta_{k+1}^{CD}\end{aligned}$$

#### 3.3.4 Méthode de Dai et Yuan

Cette méthode fut découverte par Dai et Yuan [9, 1999], dont la variante est :

$$\beta_{k+1}^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T [g_{k+1} - g_k]} \quad (3.3.4)$$

### 3.3. Différentes versions de la méthode du gradient conjugué non linéaire

On voit bien que  $\beta_{k+1}^{CD}$  est égale à la variante de la méthode du gradient conjugué linéaire, en effet :

$$\begin{aligned}\beta_{k+1} &= \frac{g_{k+1}^T [g_{k+1} - g_k]}{d_k^T [g_{k+1} - g_k]} = \frac{g_{k+1}^T g_{k+1} - g_{k+1}^T g_k}{d_k^T [g_{k+1} - g_k]} \\ &= \frac{\|g_{k+1}\|^2 - g_{k+1}^T g_k}{d_k^T [g_{k+1} - g_k]} = \frac{\|g_{k+1}\|^2}{d_k^T [g_{k+1} - g_k]} = \beta_{k+1}^{DY}\end{aligned}$$

**Remarque 3.3.1** Dans le cas quadratique on a vu que:

$$\beta^{HS} = \beta^{FR} = \beta^{CD} = \beta^{DY}$$

Dans le cas non quadratique, ces quantités ont en générale des valeurs différentes.

### 3.3.5 Algorithmes des méthodes du gradient conjugué non linéaire

#### **Algorithm 3.2 Méthodes de HS, FR, CD, et DY**

##### **Etape 0 (initialisation)**

Soit  $x_0$  le point de départ,  $g_0 = \nabla f(x_0)$ , poser  $d_0 = -g_0$

Poser  $k = 0$  et aller à l'étape 1.

##### **Etape 1**

Si  $g_k = 0$  : STOP (  $x^* = x_k$  ). "Test d'arrêt"

Sinon aller à l'étape 2.

##### **Etape 2**

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

$$\begin{aligned}\alpha_k &= \min_{\alpha > 0} f(x_k + \alpha d_k) \\ d_{k+1} &= -g_{k+1} + \beta_{k+1} d_k\end{aligned}\tag{3.3.5}$$

Où  $\beta_{k+1}$  est définie selon la méthode.

Poser  $k = k + 1$  et aller à l'étape 1.

## 3.4 Convergence des méthodes du gradient conjugué non linéaires

Dans cette section on va étudier la convergence globale des méthodes du gradient conjugué non linéaire cités au dessus.

Il y a eu de nombreuses études et résultats sur la convergence globale de la méthode du gradient conjugué avec une recherche linéaire exacte, mais le premier et le plus important résultat de convergence globale de cette méthode avec une recherche linéaire inexacte a été donné par Al-Baali[1, 1985].

Dai-Yuan ont démontré la convergence des méthodes de la descente conjugué [7, 1996] et la méthode de Dai-Yuan [8, 1998] où le pas  $\alpha_k$  est déterminé par la règle de wolfe forte.

Le résultat suivant prouve la convergence de toutes les variantes de la méthode du gradient conjugué non linéaire cités dans la section précédente

**Théorème 3.4.1** [3] *Supposons que la fonction objectif  $f$  est minorée sur l'ensemble  $\Lambda$  tel que :  $\Lambda = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ , où  $x_1 \in \mathbb{R}^n$  est un point initiale quelconque. Supposons que  $f$  vérifie la condition de Lipschitz, autrement dit il existe une constante  $L > 0$  tel que :  $\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Lambda$ . Considérons une méthode du gradient conjugué non linéaire du type (2.2.1) et (2.2.2) où  $\alpha_k$  est déterminée par la recherche linéaire inexacte de Wolfe forte (3.1.8) et (3.1.7). Alors*

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0 \quad (3.4.1)$$

$$\text{ou} \sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty$$

**Preuve.** De (3.3.5) pour tout  $k \geq 1$

$$d_{k+1} + g_{k+1} = \beta_{k+1} d_k \quad (3.4.2)$$

En appliquant le carré sur les deux côtés de (3.4.2) résulte :

$$\|d_{k+1}\|^2 = -\|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1} + \beta_{k+1}^2 \|d_k\|^2 \quad (3.4.3)$$

Puisque  $d_k$  satisfait à la condition de descente  $g_k d_k < 0$  d'après (3.4.4), il en déduit que

$$\|d_{k+1}\|^2 \geq -\|g_{k+1}\|^2 + \beta_{k+1}^2 \|d_k\|^2 \quad (3.4.4)$$

De (3.3.5)

$$g_{k+1}^T d_{k+1} - \beta_{k+1} g_k^T d_k = -\|g_{k+1}\|^2 \quad (3.4.5)$$

Comme  $\alpha_k$  vérifie les conditions de recherche linéaire de Wolfe forte (3.1.7) on aura

$$|g_{k+1}^T d_k| + \sigma |\beta_{k+1}| |g_k^T d_k| \geq \|g_{k+1}\|^2 \quad (3.4.6)$$

En appliquant l'inégalité  $(a + \sigma b)^2 \leq (1 + \sigma^2)(a^2 + b^2)$  pour tout  $a, b$  et  $\sigma \geq 0$  avec  $a = |g_{k+1} d_{k+1}|$  et  $b = |\beta_{k+1} g_k^T d_k|$ , de (3.4.6) on peut écrire

$$(g_{k+1}^T d_{k+1})^2 + \beta_{k+1}^2 (g_k^T d_k)^2 \geq c \|g_{k+1}\|^4 \quad (3.4.7)$$

où  $c = (1 + \sigma^2)^{-1}$ . Par conséquent de (3.4.5) et (3.4.7), il en résulte que

$$\begin{aligned} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_k^T d_k)^2}{\|d_k\|^2} &= \frac{1}{\|d_{k+1}\|^2} \left[ (g_{k+1}^T d_{k+1})^2 + \frac{\|d_{k+1}\|^2}{\|d_k\|^2} (g_k^T d_k)^2 \right] \\ &\geq \frac{1}{\|d_{k+1}\|^2} \left[ (g_{k+1}^T d_{k+1})^2 + \beta_{k+1}^2 (g_k^T d_k)^2 - \frac{(g_k^T d_k)^2}{\|d_k\|^2} \|g_{k+1}\|^2 \right] \\ &\geq \frac{1}{\|d_{k+1}\|^2} \left[ c \|g_{k+1}\|^4 - \frac{(g_k^T d_k)^2}{\|d_k\|^2} \|g_{k+1}\|^2 \right] \end{aligned} \quad (3.4.8)$$

Or, si la relation (3.4.1) n'est pas vraie, alors (3.4.8) et la condition de Zoutendijk (3.2.1) impliquent que l'inégalité suivante

$$\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{c}{2} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \quad (3.4.9)$$

est vérifiée pour  $k$  suffisamment grand. Ce qui achève la démonstration. ■

---

# Quelques récentes méthodes du gradient conjugué non linéaire

Dans ce chapitre on va étudier trois nouvelles variantes différentes de la méthode du gradient conjugué non linéaire. Ces méthodes ont été proposées par Jianghua Yin, Mexing Liu et Guodong Ma. La première en 2011 nommée méthode de JMG [14], la deuxième et troisième variantes en 2020 sont LMY-DY, LMY-FR [15, 2020] respectivement.

Cette partie est consacrée à une étude de ces trois variantes.

On va donner leurs algorithmes et prouver les théorèmes qui assurent leurs convergence.

## 4.1 Méthode de Jianghua-Meixing-Guodong

La méthode de JMG a été proposée en 2011 par Jianghua Y., Mexing L. et Guodong M., dont  $\beta_k$  est donnée par la formule suivante:

$$\beta_{k+1}^{JMG} = \begin{cases} \frac{g_{k+1}^T (g_{k+1} - (\|g_{k+1}\|/\|d_k\|)d_k)}{d_k (g_{k+1} - g_k)} & \text{si } g_{k+1}^T d_k > 0 \\ \frac{g_{k+1}^T (g_{k+1} + (\|g_{k+1}\|/\|d_k\|)d_k)}{d_k (g_{k+1} - g_k)} & \text{sinon} \end{cases}$$

Ceci peut être réécrit sous la forme

$$\beta_{k+1}^{JMG} = \frac{\|g_{k+1}\|^2 - (\|g_{k+1}\|/\|d_k\|)(g_{k+1}^T d_k)}{d_k^T (g_{k+1} - g_k)} \quad (4.1.1)$$

En effet,

$$\begin{aligned}\beta_{k+1}^{JMG} &= \frac{g_{k+1}^T (g_{k+1} - (\|g_{k+1}\|/\|d_k\|) d_k)}{d_k (g_{k+1} - g_k)} = \frac{g_{k+1}^T g_{k+1} - g_{k+1}^T (\|g_{k+1}\|/\|d_k\|) d_k}{d_k (g_{k+1} - g_k)} \\ &= \frac{\|g_{k+1}\|^2 - (\|g_{k+1}\|/\|d_k\|) (g_{k+1}^T d_k)}{d_{k-1}^T (g_k - g_{k-1})}\end{aligned}$$

### 4.1.1 Algorithme de la méthode JMG

#### **Algorithme 4.1** Méthode de JMG

**Initialisation** Étant donne les constantes  $\varepsilon > 0$  et  $0 < \delta < \sigma < 1$ .

Soit  $x_1 \in \mathbb{R}^n$  un point initiale quelconque. Poser  $d_1 = g_1$  et  $k = 1$ .

**Étape 1** Si  $\|g_k\| < \varepsilon$ , alors stop. Sinon passez à l'étape 2.

**Étape 2** Déterminez une longueur de pas  $\alpha_k$  par la recherche linéaire de Wolfe.

**Étape 3** Poser  $x_{k+1} = x_k + \alpha_k d_k$  calculer  $g_{k+1} = \nabla f(x_{k+1})$  et  $\beta_{k+1} = \beta_{k+1}^{JMG}$ .

**Étape 4** Calculer  $d_{k+1} = -g_{k+1} + \beta_{k+1} d_k$

Poser  $k = k + 1$  et allaer à l'étape 1.

### 4.1.2 Convergence globale

**Théorème 4.1.1** [14] *Supposons que la fonction objectif  $f$  est minorée sur l'ensemble  $\Lambda$  tel que :  $\Lambda = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ , où  $x_1 \in \mathbb{R}^n$  est un point initiale quelconque. Supposons que  $f$  vérifie la condition de Lipschitz, autrement dit il existe une constante  $L > 0$  tel que :  $\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Lambda$ . Considérons une méthode du gradient conjugué non linéaire du type (2.2.1) et (2.2.2) où  $\alpha_k$  est déterminée par la recherche linéaire inexacte de Wolfe forte (3.1.8) et (3.1.7) et  $\beta_k$  donnée par la formule (4.1.2). Alors*

$$\begin{aligned}\liminf_{k \rightarrow \infty} \|g_k\| &= 0 \\ \text{ou } \sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} &< \infty\end{aligned}\tag{4.1.2}$$

**Preuve.** En suivant les même étapes de la démonstration du Théorème (3.4.1). ■

## 4.2 Méthode Liu-Ma-Yin (LMY-DY)

En 2020, Liu M., Ma G., et Yin J. [15], ont proposé une nouvelle variante de  $\beta_k$  donnée par la formule suivante:

$$\beta_{k+1}^{LMY-DY} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)}$$

Cette méthode possède la propriété de descente de la direction à chaque itération en plus elle converge globalement et ceci même pour des fonctions continuellement différentiable non nécessairement convexes où le pas est déterminé avec la recherche linéaire inexacte de Wolfe.

### 4.2.1 Algorithme de la méthode LMY-DY

#### Algorithme 4.2 Méthode de LMY-DY

**Initialisation** Étant donné les constantes  $\varepsilon > 0$  et  $0 < \delta < \sigma < 1$

Soit  $x_1 \in \mathbb{R}^n$  un point initial quelconque. Soit  $d_1 = g_1$  et  $k = 1$ .

**Étape 1** Si  $\|g_k\| < \varepsilon$ , alors stop. Sinon passez à l'étape 2.

**Étape 2** Déterminer une longueur du pas  $\alpha_k$  par la recherche linéaire de Wolfe forte.

**Étape 3** Soit  $x_{k+1} = x_k + \alpha_k d_k$  calculer  $g_{k+1} = \nabla f(x_{k+1})$  et  $\beta_{k+1} = \beta_{k+1}^{LMY-DY}$ .

**Étape 4** Calculer  $d_{k+1} = -g_{k+1} + \beta_{k+1} d_k$ ,

Poser  $k = k + 1$  et aller à l'étape 1.

### 4.2.2 Propriété de descente

**Lemme 4.2.1** [15] Soit  $d_{k+1} \in \mathbb{R}^n$  la direction de la méthode LMY-DY, alors:

$$g_{k+1}^T d_{k+1} \leq -\frac{1}{1+\sigma} \|g_{k+1}\|^2 \text{ pour tout } 0 < \sigma < 1 \quad (4.2.1)$$

*Preuve.* La démonstration se fait par récurrence.

1) Pour  $k = 0$ , comme  $0 < \sigma < 1$ , on a:

$$g_1^T d_1 = -\|g_1\|^2 < -\frac{1}{1+\sigma} \|g_1\|^2$$

2) Supposons que l'inégalité (4.2.2) est satisfaite pour  $k$  c'est-à-dire

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

Et démontrons qu'elle le restera pour  $k+1$ .

Comme

$$\beta_{k+1}^{CD} = \frac{\|g_{k+1}\|^2}{-d_k^T g_k} > 0$$

(parce que  $d_k$  est une direction de descente).

De plus, la satisfaction de la règle de Wolfe forte nous donne :

$$0 < -(1-\sigma) g_k^T d_k \leq d_k^T (g_{k+1} - g_k) \leq -(1+\sigma) g_k^T d_k \quad (4.2.2)$$

et

$$0 < \beta_{k+1}^{CD} |g_{k+1}^T d_k| \leq \beta_{k+1}^{CD} (-\sigma d_k^T g_k) = \sigma \|g_{k+1}\|^2.$$

Et comme

$$\beta_{k+1}^{LMY-DY} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)}$$

On obtient :

$$0 < \frac{(1-\sigma)\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} \leq \beta_{k+1}^{LMY-DY} \leq \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} \quad (4.2.3)$$

Ainsi on distingue trois cas différents :

(i) Si  $g_{k+1}^T d_k = 0$ , alors

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1}^{LMY-DY} g_{k+1}^T d_k = -\|g_{k+1}\|^2 < -\frac{1}{\sigma+1} \|g_{k+1}\|^2 \quad (4.2.4)$$

(ii) Si  $g_{k+1}^T d_k > 0$  alors

$$\begin{aligned} g_{k+1}^T d_{k+1} &= \\ &= -\|g_{k+1}\|^2 + \beta_{k+1}^{LMY-DY} g_{k+1}^T d_k \end{aligned} \quad (4.2.5)$$

$$\begin{aligned} &= -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 - \beta_{k+1}^{CD} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)} g_{k+1}^T d_k \\ &= \frac{\|g_{k+1}\|^2 g_k^T d_k - \beta_{k+1}^{CD} (g_{k+1}^T d_k)^2}{d_k^T (g_{k+1} - g_k)} \\ &< \frac{\|g_{k+1}\|^2 g_k^T d_k}{d_k^T (g_{k+1} - g_k)} \leq \frac{\|g_{k+1}\|^2 g_k^T d_k}{-(1+\sigma) g_k^T d_k} = \frac{1}{(1+\sigma)} \|g_{k+1}\|^2 \end{aligned} \quad (4.2.6)$$

(iii) Si  $g_{k+1}^T d_k < 0$ , alors

$$g_{k+1}^T d_{k+1} = \|g_{k+1}\|^2 + \beta_{k+1}^{LMY-DY} g_{k+1}^T d_k < -\|g_{k+1}\|^2 < -\frac{1}{(1+\sigma)} \|g_{k+1}\|^2 \quad (4.2.7)$$

On conclut que la formule de récurrence est vraie pour tout  $k+1$ .

Ce qui achève la démonstration. ■

### Cas particulier

Il est évident que la formule susmentionnée se réduit à la formule de la variante de DY dans le cas où le pas est déterminé par une recherche linéaire exacte.

En effet, dans le cas où  $\alpha$  est déterminé par une règle exacte, alors il vérifie la condition nécessaire d'optimalité, autrement dit  $g_{k+1}^T d_k = 0$ .

Ainsi on aura :

$$\begin{aligned} \beta_{k+1}^{LMY-DY} &= \frac{\|g_{k+1}\|^2 - \beta_{k+1}^{CD} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)} \\ &= \frac{\|g_{k+1}\|^2}{d_k^T [g_{k+1} - g_k]} = \beta_{k+1}^{DY} \end{aligned}$$

### 4.2.3 Convergence globale

Avant d'étudier la convergence de la méthode LMY-DY, on va d'abord lancer le lemme suivant

**Lemme 4.2.2** [15] *La variante de la méthode LMY-DY vérifie l'inégalité suivante :*

$$0 < \beta_{k+1}^{LMY-DY} \leq \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}, \quad \forall k \geq 0.$$

**Preuve.** On voit bien que  $\beta_{k+1}^{LMY-DY} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)} > 0$ .

On spécifie deux cas :

a) Si  $g_{k+1}^T d_k = 0$ , alors en utilisant la relation (4.2.3), on aura

$$\beta_{k+1}^{LMY-DY} = \frac{\|g_{k+1}\|^2}{(-g_k^T d_k)} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}$$

b)  $g_{k+1}^T d_k \neq 0$  on distingue le deux cas suivant

(i) Si  $g_{k+1}^T d_k > 0$  d'après la relation (4.2.6), on a

$$g_{k+1}^T d_{k+1} < \frac{\|g_{k+1}\|^2 g_k^T d_k}{d_k^T (g_{k+1} - g_k)}$$

Alors, en divisant cette inégalité par le terme négatif  $g_k^T d_k$ , il s'ensuit que

$$\frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} < \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}$$

Puis en combinant ceci avec (4.2.6) on aura:

$$\beta_{k+1}^{LMY-DY} < \left( \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \right)$$

(ii) Si  $g_{k+1}^T d_k < 0$ , on a:

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_{k+1}^{LMY-DY} g_{k+1}^T d_k & (4.2.8) \\ &= \frac{\|g_{k+1}\|^2 g_k^T d_k - \beta_{k+1}^{CD} |g_{k+1}^T d_k| g_{k+1}^T d_k}{d_k^T (g_{k+1} - g_k)} \\ &\leq \frac{\|g_{k+1}\|^2 g_k^T d_k - \sigma \beta_{k+1}^{CD} |g_{k+1}^T d_k| g_k^T d_k}{d_k^T (g_{k+1} - g_k)} \\ &= \frac{(\|g_{k+1}\|^2 - \sigma \beta_{k+1}^{CD} |g_{k+1}^T d_k|) g_k^T d_k}{d_k^T (g_{k+1} - g_k)} \end{aligned}$$

et donc

$$\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \geq \frac{(\|g_{k+1}\|^2 - \sigma \beta_{k+1}^{CD} |g_{k+1}^T d_k|)}{d_k^T (g_{k+1} - g_k)}$$

Puisque  $g_k^T d_k < 0$ . depuis  $0 < \sigma < 1$ , on a

$$\beta_{k+1}^{LMY-DY} = \frac{\|g_{k+1}\|^2 - \beta_{k+1}^{CD} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)} \leq \frac{\|g_{k+1}\|^2 - \sigma \beta_{k+1}^{CD} |g_{k+1}^T d_k|}{d_k^T (g_{k+1} - g_k)} \quad (4.2.9)$$

Cela implique que

$$\beta_{k+1}^{LMY-DY} \leq \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}$$

et la preuve est terminée. ■

Notons que les lemmes précédents présents un brik essentiel dans la démonstration du théorème suivant, qui assure la convergence globale de la méthode LMY-DY.

**Théorème 4.2.1** [15] *Supposons que la fonction objectif  $f$  est minorée sur l'ensemble  $\Lambda$  tel que :  $\Lambda = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ , où  $x_1 \in \mathbb{R}^n$  est un point initiale quelconque.*

Supposons que  $f$  vérifie la condition de Lipschitz, autrement dit, il existe une constante  $L > 0$  tel que :  $\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Lambda$ . Considérons une méthode du gradient conjugué non linéaire du type (2.2.1) et (2.2.2), où  $\alpha_k$  est déterminée par la recherche linéaire inexacte de Wolfe forte (3.1.8) et (3.1.7) et  $\beta_k$  donnée par la formule (4.2.1). Alors

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (4.2.10)$$

**Preuve.** La démonstration se fait par contradiction.

Supposons que

$$\lim_{k \rightarrow \infty} \|g_k\| \neq 0$$

alors il existe une constante  $\gamma > 0$  telle que

$$\|g_{k+1}\|^2 \geq \gamma, \forall k \geq 0.$$

Puisque  $d_{k+1} + g_{k+1} = \beta_{k+1}^{LMY-DY} d_k$ , il en résulte du lemme (4.2.3) que:

$$\begin{aligned} \|d_{k+1}\|^2 &= (\beta_{k+1}^{LMY-DY})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \\ &\leq \left( \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \right)^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \end{aligned} \quad (4.2.11)$$

En divisant par  $(g_{k+1}^T d_{k+1})^2$ , c'est facile d'avoir ça

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &\leq \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \frac{2}{g_{k+1}^T d_{k+1}} - \frac{\|g_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \\ &= \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \left( \frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \\ &\leq \frac{\|d_k\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_{k+1}\|^2} \end{aligned} \quad (4.2.12)$$

Comme  $\frac{\|d_1\|^2}{(g_1^T d_1)^2} = \frac{1}{\|g_1\|^2}$ , il en résulte que

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &\leq \frac{\|d_k\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_{k+1}\|^2} \\ &\leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2} + \frac{1}{\|g_{k+1}\|^2} \\ &\leq \dots \leq \sum_{i=1}^k \frac{1}{\|g_i\|^2} \leq \frac{k}{\gamma} \end{aligned} \quad (4.2.13)$$

Ainsi,

$$\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} \geq \frac{\gamma}{k+1}$$

Ce qui implique

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty$$

Donc la condition de Zoutendijk n'est pas vérifiée, ceci n'est qu'une contradiction avec le théorème (3.4.1)

Et donc

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

Ce qui achève la démonstration. ■

### 4.3 Méthode Liu-Ma-Yin (LMY-FR)

En 2020, Liu M., Ma G., et Yin J. [15] ont proposé une nouvelle variante de  $\beta_k$  :

$$\beta_{k+1}^{LMY-FR} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{\|g_k\|^2} \quad (4.3.1)$$

Cette méthode possède la propriété de descente de la direction à chaque itération en plus elle converge globalement et ceci même pour des fonctions différentiable non nécessairement convexes où le pas est déterminé par la règle de Wolfe.

#### 4.3.1 Algorithme de méthode LMY-FR

##### **Algorithme 4.3 Méthode de LMY-FR**

**Initialisation** Étant donné les constantes  $\varepsilon > 0$  et  $0 < \delta < \sigma < 1$ .

Soit  $x_1 \in \mathbb{R}^n$  un point initial quelconque. Poser  $d_1 = g_1$  et  $k = 1$ .

**Étape 1** Si  $\|g_k\| < \varepsilon$ , alors stop. Sinon passer à l'étape 2.

**Étape 2** Déterminer une longueur de pas  $\alpha_k$  par la recherche linéaire de Wolfe.

**Étape 3** Poser  $x_{k+1} = x_k + \alpha_k d_k$  calculer  $g_{k+1} = \nabla f(x_{k+1})$  et  $\beta_{k+1} = \beta_{k+1}^{LMY-FR}$ .

**Étape 4** Calculer  $d_{k+1} = -g_{k+1} + \beta_{k+1} d_k$

Poser  $k = k + 1$  et aller à l'étape 1.

### 4.3.2 Propriété de descente

**Lemme 4.3.1** [15] *La direction de descente  $d_{k+1} \in \mathbb{R}^n$  de la méthode LMY-FR, vérifie :*

$$-\frac{1}{1-\sigma} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -\frac{1-2\sigma}{1-\sigma} \text{ pour tout } 0 < \sigma < \frac{1}{2} \quad (4.3.2)$$

De plus

$$0 < \beta_{k+1}^{LMY-FR} \leq \beta_{k+1}^{FR}$$

**Preuve.** *La démonstration se fait par récurrence.*

1) pour  $k = 0$ , on a

$$\frac{g_1^T d_1}{\|g_1\|^2} = -1$$

Donc (4.3.2) pour tout  $0 < \sigma < \frac{1}{2}$  est vérifiée.

2) Supposons que la relation (4.3.2) soit satisfaite pour  $k$ .

Comme le pas  $\alpha_k$  est déterminé par la recherche linéaire de Wolfe et  $d_k$  est une direction de descente, on aura

$$0 \leq |g_{k+1}^T d_k| \leq -\sigma g_k^T d_k$$

Autrement dit

$$0 < 1 - \sigma \leq 1 - \frac{|g_{k+1}^T d_k|}{-g_k^T d_k} \leq 1 \quad (4.3.3)$$

On a  $\beta_{k+1}^{LMY-FR} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{\|g_k\|^2}$  ce qui donne

$$\begin{aligned} 0 &< \frac{(1-\sigma) \|g_{k+1}\|^2}{\|g_k\|^2} \leq \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{\|g_k\|^2} \\ &= \frac{\|g_{k+1}\|^2 (1 - |g_{k+1}^T d_k| / (-g_k^T d_k))}{\|g_{k+1}\|^2} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \beta_{k+1}^{FR} \end{aligned} \quad (4.3.4)$$

De plus,

$$\begin{aligned} \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} &= -1 + \frac{\|g_{k+1}\|^2 - (\|g_{k+1}\|^2 / -d_k^T g_k) |g_{k+1}^T d_k|}{\|g_k\|^2} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} \\ &= -1 + \left(1 - \frac{|g_{k+1}^T d_k|}{-d_k^T g_k}\right) \frac{g_{k+1}^T d_k}{\|g_k\|^2} \end{aligned} \quad (4.3.5)$$

En combinant ceci avec la forme (4.3.3) on aura

$$-1 - \frac{|g_{k+1}^T d_k|}{\|g_k\|^2} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{|g_{k+1}^T d_k|}{\|g_{k+1}\|^2} \quad (4.3.6)$$

D'après l'hypothèse, on aura

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \geq -1 + \sigma \frac{|g_k^T d_k|}{\|g_k\|^2} \geq -1 - \frac{\sigma}{1-\sigma} = -\frac{1}{1-\sigma} \quad (4.3.7)$$

D'après la forme (4.3.6), on obtient

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{|g_k^T d_k|}{\|g_k\|^2} \leq -1 + \frac{\sigma}{1-\sigma} - \frac{(1-2\sigma)}{1-\sigma} \quad (4.3.8)$$

Ce qui achève la démonstration. ■

### Cas particulier

Il est évident que la formule susmentionnée se réduit à la formule de la variante de FR dans le cas où le pas est déterminé par une recherche linéaire exacte.

En effet, dans le cas où  $\alpha$  est déterminé par une recherche linéaire exacte, alors il vérifie la condition nécessaire d'optimalité d'où  $g_{k+1} d_k = 0$ .

En effet,

$$\begin{aligned} \beta_{k+1}^{LMY-FR} &= \frac{\|g_{k+1}\|^2 - \beta_{k+1}^{CD} |g_{k+1}^T d_k|}{\|g_k\|^2} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2}{-d_k^T g_k} |g_{k+1}^T d_k|}{\|g_k\|^2} \\ &= \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \beta_{k+1}^{FR} \end{aligned}$$

### 4.3.3 Convergence globale

En se basant sur le lemme (4.3.1) et la condition de zoutendijk (3.2.1), nous obtenons également la convergence globale de la méthode de LMY-FR pour des fonctions continuellement différentiable.

**Théorème 4.3.1** [15] *Supposons que la fonction objectif  $f$  est minorée sur l'ensemble  $\Lambda$  tel que :  $\Lambda = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ , où  $x_1 \in \mathbb{R}^n$  est un point initial quelconque. Supposons que  $f$  vérifie la condition de Lipschitz, autrement dit, il existe une constante  $L > 0$  tel que :  $\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Lambda$ . Considérons une méthode du gradient conjugué non linéaire du type (2.2.1) et (2.2.2) où  $\alpha_k$  est déterminée par la recherche linéaire inexacte de Wolfe forte (3.1.8) et (3.1.7) et  $\beta_k$  est donnée par la formule (4.3.1).*

Alors

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0 \quad (4.3.9)$$

**Preuve.** La démonstration se fait par contradiction. En effet supposons que

$$\liminf_{k \rightarrow \infty} \|g_k\| \neq 0$$

alors il existe une constante  $\gamma > 0$  telle que

$$\|g_{k+1}\| \geq \gamma, \forall k \geq 0.$$

d'où

$$\|d_{k+1}\|^2 = \|g_{k+1}\|^2 - 2\beta_{k+1}^{LMY-FR} g_{k+1}^T d_k + (\beta_{k+1}^{LMY-FR})^2 \|d_k\|^2$$

D'après Lemme (4.3.1) et la satisfaction du pas de Wolfe forte, on a

$$\begin{aligned} -2\beta_{k+1}^{LMY-FR} g_{k+1}^T d_k &\leq -2\beta_{k+1}^{LMY-FR} |g_{k+1}^T d_k| \leq 2\beta_{k+1}^{FR} |g_{k+1}^T d_k| \\ &\leq 2 \frac{\|g_{k+1}\|^2}{\|g_k\|^2} (\sigma |g_k^T d_k|) \leq \frac{2\sigma \|g_{k+1}\|^2}{1 - \sigma} \end{aligned} \quad (4.3.10)$$

Donc, on obtient

$$\|d_{k+1}\|^2 \leq \left( \frac{1 + \sigma}{1 - \sigma} \right) \|g_{k+1}\|^2 + \left( \frac{\|g_{k+1}\|^4}{\|g_k\|^4} \right) \|d_k\|^2$$

Encore une fois, en divisant cette inégalité par  $\|g_{k+1}\|^4$  nous avons

$$\left( \frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \right) \leq \left( \frac{1 + \sigma}{1 - \sigma} \right) \left( \frac{1}{\|g_{k+1}\|^2} \right) + \frac{\|d_k\|^2}{\|g_k\|^4}$$

Prendre  $\|d_1\|^2 = \|g_1\|^2$  et  $\|g_{k+1}\| \geq \gamma$  en utilisant la formule ci-dessus on a

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \left( \frac{1 + \sigma}{1 - \sigma} \right) \sum_{i=1}^k \frac{1}{\|g_i\|^2} \leq \frac{1 + \sigma}{1 - \sigma} \frac{k}{\gamma^2}$$

Cela implique que

$$\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \geq \left( \frac{(1 - \sigma) \gamma^2}{(1 + \sigma)} \right) \left( \frac{1}{k + 1} \right)$$

Ainsi,

$$\sum_{k=0}^{\infty} \frac{\|g_{k+1}\|^4}{\|g_{k+1}\|^2} = \infty$$

Ce qui contredit la condition de Zoutendijk (3.2.1).

Par conséquent, nous pouvons conclure que la méthode LMY-FR est globalement convergente. ■

Dans ce chapitre, on va étudier les performances de trois variantes différentes de la méthode du gradient conjugué non linéaire étudiées dans ce mémoire. Il s'agit des méthodes JMG, LMY-FR et LMY-DY. Le pas est déterminé par la recherche linéaire inexacte de Wolfe forte. Les fonctions test choisies sont TRIDIA, BALF et BRYBND.

L'algorithme de la recherche linéaire inexacte de Wolfe forte assure que la direction est de descente. Les paramètres sont donnés  $\rho = 0.1$ , et  $\delta = 0.01$ .

Les méthodes CG sont utilisées avec redémarrage, c'est-à-dire dans le cas où le dénominateur de la variante du gradient conjugué  $\beta_k$  est nulle on redémarre en prenant la direction de la plus profonde pente.

Dans nos simulations numériques on a limité le nombre des itérations nécessaires dans l'algorithme de la recherche linéaire inexacte de Wolfe forte, et le pas initial est égal à un.

Les programmes sont écrits sous scilab 6.1.0 et exécutés dans un ordinateur Lenovo i7 de vitesse 3.10 GHz

Notre critère d'arrêt est  $\|\nabla f(x_k)\| \leq \varepsilon$ , où  $\varepsilon = 10^{-2}$ .

## 5.1 Performances des méthodes du gradient conjugué non linéaire de JMG LMY-DY et LMY-FR

Dans cette section on va présenter nos résultats numériques.

### 5.1.1 La fonction BRYBND (Broyden banded )

Cette fonction test [3, 2020] est donnée par :

$$f_i(x) = x_i(2 + 5x_i^2) + 1 - \sum_{j \in J_i} x_i(1 + x_j), \quad 1 \leq i < n,$$

$$\text{où } J_i = \{j : j \neq i, \max(1, i - m_l) \leq j \leq \min(n, m_u)\},$$

$$m_l = 5, \quad m_u = 1, \quad \text{et } x_0 = (-1, \dots, -1).$$

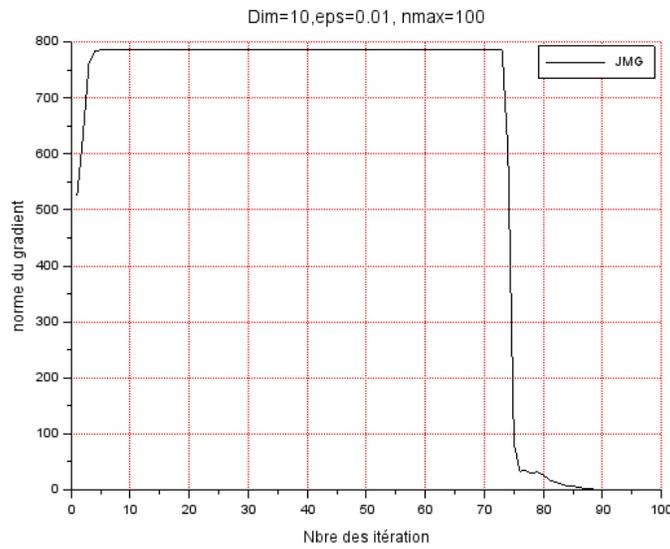


Fig 5.1 JMG pour la fonction BRYBND

D'après la figure 5.1, on voit bien qu'avant 90 itérations, la méthode JMG atteint son minimum, pour la minimisation de la fonction test BRYBND dans le cas où la dimension  $\dim = 10$  et la précision  $\epsilon = 0.01$ .

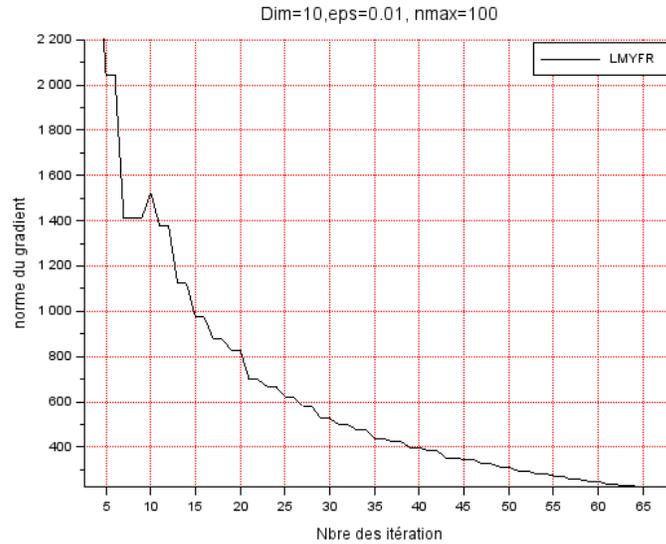


Fig 5.3 LMY-FR pour la fonction BRYBND

Les figures ci dessus illustrent la convergence des méthodes JMG et LMY-FR pour la fonction test BRYND après 90 et 65 itérations succesivement, sachant que la dimension  $n=10$  et la précision  $\epsilon = 0.01$ .

La méthode LMY-DY diverge après 100 itérations.

### 5.1.2 La fonction test TRIDIA (cute )

Cette fonction test [3, 2020] est donnée par :

$$f(x) = \gamma(\delta x_1 - 1)^2 + \sum_{i=2}^n (\alpha x_i - \beta x_{i-1})^2, \text{ où } \alpha = 2, \beta = 1, \gamma = 1, \delta = 1 \text{ et } x_0 = (1, \dots, 1).$$

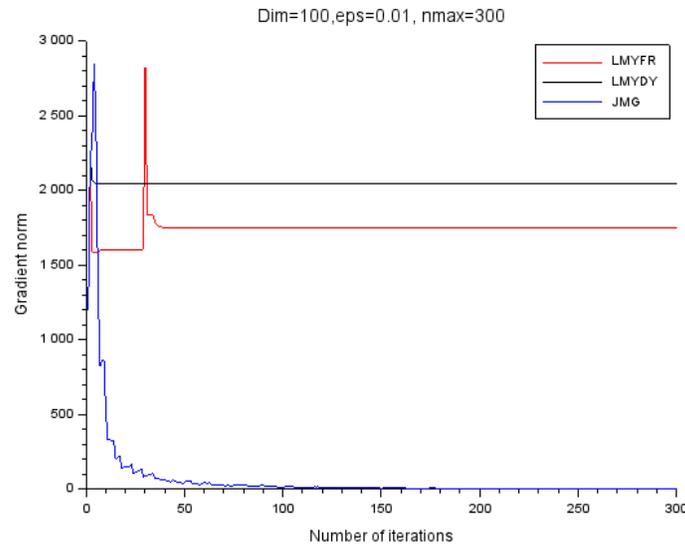


Fig 5.4 Comparaison pour la fonction TRIDIA

D'après la figure 5.4, on voit bien que toutes les méthodes précédentes divergent pour la fonction test TRIDIA, dans le cas où  $n=100$  et ceci après 300 itérations.

## 5.2 Conclusion

Ce memoire est consacré à une étude numérique comparative des performances des méthodes du gradient conjugué non linéaire de JMG, LMY-DY et LMY-FR avec la recherche linéaire inexacte de Wolfe forte. Les variantes de ces méthodes possèdent la propriété de descente ainsi que la convergence globale.

Les méthodes citées dessus sont testées sur deux fonctions testes choisies [3, 2020].

Notons que le redémarrage de la méthode CG (si la variante  $\beta_k$  est proche du zéro) accélère la convergence de la méthode CG et évite l'échec du programme.

Nous avons mis en oeuvre nos propres simulations numériques (SCILAB) qui montrent l'efficacité des méthodes JMG et LMY-FR pour le cas traité.

---

# Bibliographie

- [1] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA J. Numerical Analysis, volume 5, pages 121-124 (1985).
- [2] N. ANDREI. Another accelerated conjugate gradient algorithm with guaranteed descent and conjugacy conditions for large-scale unconstrained optimization, Journal of Optimization Theory and Applications, volume 159, pages 159–182 (2013).
- [3] N. ANDREI. Nonlinear Conjugate Gradient Methods for Unconstrained Optimization. Springer, 2020.
- [4] L. Armijo, Minimization of function having Lipschitz continuous first partial derivatives, Pacific Journal of Mathematics, Volume 16, pages 1-3 (1966).
- [5] M. S. Bazaraa, H. D. Sherali, et C. M. Shetty, Nonlinear Programming, Theory and Algorithms, Wiley-Interscience, 1993.
- [6] A. Cauchy, Analyse mathématique, Méthode générale pour la résolution des systèmes d'équations simultanées, Comptes Rendus de l'Académie des Sciences de Paris, volume 25, pages 536-538 (1847).
- [7] Y. H. Dai and Y. Yuan, Convergence properties of the conjugate descent method, Advanced Mathematics (China), volume 26, pages 552-562 (1996).
- [8] Y. H. Dai and Y. Yuan. Some properties of a new conjugate gradient method. Advances in Nonlinear Programming Kluwer, Boston, pages 251-262 (1998).

- [9] Y.H. Dai and Y. Yuan. A nonlinear conjugate gradient with a strong global convergence property. *SIAM J. Optimization*, Volume 10, pages 177-182 (1999).
- [10] R. Fletcher and C.Reeves. Function minimization by conjugate gradients, *Computational Journal*, volume 7, page 149-154 (1964).
- [11] R. Fletcher, *Practical Methods of Optimization volume 1 : Unconstrained Optimization*, John Wiley Sons, New York, 1987.
- [12] J.C.Gilbert, *Eléments d'Optimisation Différentiable : Théorie et Algorithmes*, Notes de cours, École Nationale Supérieure de Techniques Avancées, Paris 2007.
- [13] M. R. Hestenes and E. L. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards Sect. volume 5*, pages 409-436 (1952).
- [14] X. Z. Jiang,G. D. Ma, and J. B. jian. A new global convergent conjugate gradient methode with Wolfe line search, *chinese journal of Engineering Mathematics*, volume 28,page 779-786 (2011).
- [15] M.Liu, G. Ma and J. Yin. Two new conjugate gradient methodes for unconstrained optimiztion *Wiley-Hindawi Volume 13* (2020).
- [16] M. Minoux, *Programmation Mathématique, Théorie et Algorithmes*, tome 1 , Dunod 1983.
- [17] J. NOCEDAL et Stephen WRIGHT. *Numerical optimization*. Springer Science & Business Media, 2006.
- [18] P.Wolfe. Convergence conditions for ascent method, *SIAM Revue*, 1969.
- [19] G ZOUTENDIJK, *Nonlinear programming, computational methods*, In : *Integer and nonlinear programming*, pages 37-86, 1970.