République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Larbi Tébessi - Tébessa
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département : Mathématiques et Informatique

Mémoire de fin d'étude
Pour l'obtention du diplôme de MASTER
Domaine : Mathématiques et Informatique
Filière : Mathématiques
Option : EDP et appliquées

Thème

La stabilité de la convergence globale d'une nouvelle méthode du descente conjuguée

Présenté Par :
Boudiar Soumaia
Abdou Ilhem

Devant le jury :

Mr.HAOUAM Kamel	*Prof*	*Univérsité Larbi Tébessi Tébessa*	*Président*
Mr.BOUALI Tahar	*MCA*	*Univérsité Larbi Tébessi Tébessa*	*Examinateur*
Mr. BOUAZIZ Khelifa	*MCB*	*Univérsité Larbi Tébessi Tébessa*	*Encadreur*

Date de soutenance : 19/06/2019

Résumé

La méthode du gradient conjugué est une méthode utile pour résoudre les problèmes de minimisation à grande échelle. Peuvent survenir des erreurs à cause d'un calcul inexact. Dans ce travail, nous étudions la convergence globale d'une nouvelle méthode du descente conjuguée sous la recherche linéaire de Wolfe forte avec d'existence des erreurs. Des expériences numériques préliminaires sont données pour présenter l'efficacité de la méthode.

Les mots clés : Recherche linéaire inexacte. Algorithme. Règle d'Armijo. Règle de Goldstein. Règle de Wolfe. Méthode de Fletcher-Reeves. Méthode de la descente conjuguée. Convergence globale.

Abstract

The conjugate gradient method is a useful method for solving large-scale minimization problems. Errors may arise because of inexact computation. In this work we study the global convergence of a modified conjugate descent method with strong Wolfe line search in the presence of errors. Preliminary numerical experiments are given to show the efficiency and robustness of the method.

Keywords: Inexact line search, Algorithm, Armijo line search, Goldstein line search, Wolfe line search, Fletcher-Reeves Method, Conjugate descent Method, Global convergence.

ملخص

تعد طريقة التدرج المترافق طريقة مفيدة لحل المشكلات كبيرة الحجم. يمكن أن تحدث هذه الطريقة أخطاء بسبب الحساب غير الدقيق. في هذا العمل ندرس التقارب الكلي لطريقة الانحدار المتقارن الجديدة وفق بحث وو لف الخطي القوي مع وجود أخطاء.

يتم إعطاء تجارب عددية أولية لتوضيح فعالية الطريقة.

الكلمات المفتاحية : البحث الخطي غير الدقيق. خوارزمية. قاعدة ارميجو. قاعدة قولدشتاين. قاعدة وولف. طريقة ريفرس فليتشر. طريقة الانحدار المتقارن .التقارب الكلي.

Remerciments

Nous tenons tout d'abord à remercier **Dieu** le tout puissant, qui nous a donné la force et la patience d'accomplir ce Modeste travail.

Nous tenons à adresser notre remerciements les plus chaleureux et notre profonde gratitude à notre encadreur **Dr.BOUAZIZ Khelifa**, MCB à l'Université Larbi Tébessi-Tébessa pour nos avoir proposé le sujet de ce mémoire. C'est grâce à sa grande disponibilité, ses conseils, ses orientations et ses encouragements que nous avons pu mener à bien ce travail.

Nous sommes conscientes de l'honneur que nous a fait **Pr. HAOUAM Kamel** en étant président du jury et **Dr. BOUALI Tahar** d'avoir accepté d'examiner ce travail.

Nos remerciement s'adresse également à tout nos professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leurs charges académiques et professionnelles.

Nos profonds remerciements vont également à toutes les personnes qui nous ont aidés et soutenue de prés ou de loin.

Enfin, Merci à tous mes amis et tous mes collègues sans exception, nous n'oublions pas de remercier toutes les personnes qui nos ont facilité la tâche.

Liste des figures

Fig2.1 Armijo……..……..11

Fig2.2 Goldstein……..16

Fig2.3 Wolfe……..20

Liste des tableaux

Tableau 1 : liste des résultats numériques.....44

Table des matières

Introduction générale **iii**

1 Notions, définitions et préliminaires **1**
1.1 Notations et définitions 1
1.2 Conditions d'optimalité 4
1.3 Conditions nécessaires d'optimalité 5
1.4 Différentiabilité 6
1.4.1 Gradient, Hessien 6
1.4.2 Dérivée directionnelle 6
1.4.3 Direction de descente 7
1.4.4 Formule de Taylor 7

2 Recherches linéaires inexactes **8**
2.1 Recherche linéaire inexacte d'Armijo 8
2.1.1 Règle d'Armijo 8
2.1.2 Interprétation graphique de la règle d'Armijo 9
2.2 Recherche linéaire inexacte de Goldstein 13
2.2.1 La règle de Goldstein 13
2.2.2 Interprétation graphique 14
2.3 Recherche linéaire inexacte de Wolfe 17
2.3.1 La règle de wolfe 17
2.3.2 Interprétation graphique 17
2.3.3 Règle de Wolfe 20

3 Méthode de descente conjuguée **23**
3.1 Hypothèses C1 et C2 et théorème de Zoutendijk 23
3.1.1 Hypothèses C1 et C2 (de Lipschitz et de bornetude) . . 23

3.1.2 Théorème de Zoutendijk 24
3.2 Les méthodes dans lesquelles le termes $\|g_{k+1}\|^2$ figure dans le numérateur de β_k 25
3.2.1 Méthode de Fletcher-Reeves 25
3.2.2 Méthode de descente conjuguée 31

4 La convergence globale d'une méthode de descente conjuguée avec existence des erreurs 34
4.1 Propriétés de la direction 35
4.2 La convergence globale de la méthode de descente conjuguée . 37
4.3 Résultats numériques 43

4.4 Conclusion 44

Introduction générale

La méthode du Gradient Conjugué est une méthode itérative très puissante et très utilisée pour résoudre des systèmes linéaire et non linéaire de grande taille. Elle repose 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 de directions de descente permettant de minimiser une fonction (P) : $\min\{f(x) : x \in \mathbb{R}^n\}$ de $\mathbb{R}^n$ dans $\mathbb{R}$.

Cette méthode a été découverte en 1952 par Hestenes et Steifel [9], pour la minimisation de 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 [6] puis en 1969 par Polak, Ribière [15] et Ployak [16]. Une autre variante a été étudiée en 1987 par Fletcher [5] (Méthode de la descente conjuguée).

Les méthodes du gradient conjugué engendrent une suite d'itérés $\{x_k\}_{k\in\mathbb{N}}$ comme suit :

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

Où α_k est le pas optimal déterminé par une recherche linéaire inexacte.

Ces méthodes considèrent d_k comme suit :

$$d_k = \begin{cases} -g_1 \text{ si } k = 1 \\ -g_k + \beta_k d_k \text{ si } k \geq 2. \end{cases} \tag{2}$$

Où $g_k = \nabla f(x_k)$ et β_k est le paramètre qui caractérise les différentes versions des méthodes du gradient conjugué.

On obtient les variantes suivantes :

$$\beta_k^{PRP} = \frac{g_k^T (g_{k+1} - g_k)}{\|g_{k-1}\|^2}, \text{ GC variante Polak-Ribière-Polyak.} \tag{3}$$

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \text{ GC variante Fletcher Reeves.} \tag{4}$$

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}, \text{ GC variante descente conjugué.} \tag{5}$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \text{ GC variante Hestenes-Stiefel.} \tag{6}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}, \text{ GC variante de Dai-Yuan.} \tag{7}$$

$$\beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}}, \text{ GC variante de Liu-Storey.} \tag{8}$$

On étudie dans ce mémoire la convergence globale d'une méthode du descente conjuguée avec des erreurs.

Le mémoire comporte quatre chapitres :

Chapitre 1

On introduit dans ce chapitre les notions préléminaires et de base.

Chapitre 2

On expose dans ce chapitre les grandes lignes du directions de descente et les recherches linéaires.

Chapitre 3

On rappele les conditions que la méthode de la descente conjuguée doit être vérifier pour prouver les résultats de convergence.

Chapitre 4

Ce chapitre est la partie la plus importante de ce travail, car elle contient les résultas de descente et de convergence globale d'une nouvelle méthode du descente conjuguée, avec les conditions de Wolfe.

Chapitre 1

Notions, définitions et préliminaires

1.1 Notations et définitions

Ensemble convexe [10]

Définition 1.1.1 *Soit C un sous ensemble non vide de $\mathbb{R}^n$. On dit que C est convexe si*

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

Définition 1.1.2 *Soit $x_1, x_2, ..., x_k \in \mathbb{R}^n$ 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 combinaisan convexe des points x_j ou barycentre.*

Fonction convexe

Définition 1.1.3 *Soit $C \subseteq \mathbb{R}^n$ un ensemble convexe non vide. On dit que la fonction $f : C \to \mathbb{R}$ est convexe si*

$$\forall (x,y) \in C^2, \forall \lambda \in [0,1] : f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Remarque 1.1.1 *On dit que f est une fonction concave si $(-f)$ est convexe.*

Définition 1.1.4 *Soit $C \subseteq \mathbb{R}^n$ un ensemble convexe non vide. On dit que f est strictement convexe dans C si*

$$\forall (x,y) \in C^2, x \neq y, \forall \lambda \in [0,1] : f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$$

Fonction fortement ou uniformément convexe

Définition 1.1.5 *Soit $C \subseteq \mathbb{R}^n$ un ensemble convexe non vide. Une fonction $f : C \to \mathbb{R}$ est fortement ou uniformément convexe de module $\upsilon > 0$ si*

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\upsilon}{2} \lambda (1-\lambda) \|x-y\|^2, \forall x, y \in C^2, \forall \lambda \in [0,1]$$

Fonction convexe différentiable

Définition 1.1.6 *Soit $C \subseteq \mathbb{R}^n$ un ensemble convexe non vide. $f : \mathbb{R}^n \to \mathbb{R}$ et $\widehat{x} \in int(C)$, f est dite différentiable au point $\widehat{x}$, s'il existe un vecteur $A \in \mathbb{R}^n$ et une fonction $\alpha : \mathbb{R}^n \to \mathbb{R}$ tels que*

$$f(x) = f(\widehat{x}) + A^T (x - \widehat{x}) + \|x - \widehat{x}\| \alpha(\widehat{x}, x - \widehat{x}),$$

où $\alpha(\widehat{x}, x-\widehat{x}) \underset{x \to \widehat{x}}{\to} 0$. On peut note le vecteur A comme suit :

$$A = \nabla f(\widehat{x}) = \left(\frac{\partial f(\widehat{x})}{\partial x_1}, ..., \frac{\partial f(\widehat{x})}{\partial x_n} \right)^T.$$

Fonction convexe deux fois différentiable

Définition 1.1.7 *Soit $C \subseteq \mathbb{R}^n$ un ensemble convexe non vide et $f : \mathbb{R}^n \to \mathbb{R}$, f est dite deux fois différentiable au point $\widehat{x} \in int(C)$ s'il existe un vecteur $\nabla f(\widehat{x})$ et une matrice symétrique $H(\widehat{x})$ d'ordre (n,n) appellée matrice hessienne, et une fonction $\alpha : \mathbb{R}^n \to \mathbb{R}$ tels que*

$$\forall x \in C : f(x) = f(\widehat{x}) + \nabla f(\widehat{x})^T (x - \widehat{x}) + \frac{1}{2} (x - \widehat{x})^T H(\widehat{x}) (x - \widehat{x}) + \|x - \widehat{x}\|^2 \alpha(\widehat{x}, x - \widehat{x}),$$

$$\text{où } \alpha(\widehat{x}, x-\widehat{x}) \underset{x \to \widehat{x}}{\to} 0 \text{ et } H(\widehat{x}) = \begin{bmatrix} \frac{\partial^2 f(\widehat{x})}{\partial x_1 x_1} & \frac{\partial^2 f(\widehat{x})}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(\widehat{x})}{\partial x_1 x_n} \\ \frac{\partial^2 f(\widehat{x})}{\partial x_2 x_1} & \frac{\partial^2 f(\widehat{x})}{\partial x_2 x_2} & \cdots & \frac{\partial^2 f(\widehat{x})}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\widehat{x})}{\partial x_n x_1} & \frac{\partial^2 f(\widehat{x})}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(\widehat{x})}{\partial x_n x_n} \end{bmatrix}$$

Normes dans les espaces vectoriels

Définition 1.1.8 *On appelle norme sur un espace vectoriel E une application de E dans $\mathbb{R}^+$ telle que si on note $\|v\|$ la norme de v*

a) $\|v\| = 0 \Leftrightarrow v = 0$.
b) $\|\lambda v\| = |\lambda| \|v\| \quad \forall v \in E \ \forall \lambda \in \mathbb{R}$ ou $\mathbb{C}$.
c)$\|u + v\| \leq \|u\| + \|v\| \quad \forall u, v \in E$ (inégalité triangulaire).
Quelques propriétés
1) On appelle distance associée à la norme $\|.\|$ la quantité

$$d(u, v) = \|u - v\| .$$

2) $||x| - |y|| \leq |x + y| \leq |x| + |y| .$
3) $|x - y| \leq |x| + |y| .$
4) $\|xy\| \leq \|x\| \|y\|$ (inégalité de **Cauchy Schwarz**).

Définition 1.1.9 *Une application de E dans $\mathbb{R}^+$ est appellée une semi norme si elle satisfait les propriétés b) et c).*

Normes Fondamentals sur $\mathbb{R}^n$

Définition 1.1.10 *Soit $X = (x_1, x_2, ..., x_n)$ un élément de $\mathbb{R}^n$, on obtient*

a) $\|X\|_1 = \sum_{i=1}^{n} |x_i|$
b) $\|X\|_2 = \left(\sum_{i=1}^{n} (x_i)^2 \right)^{\frac{1}{2}}$
c) $\|X\|_\infty = \sup_{1 \leq i \leq n} |x_i| .$

Les normes $\|.\|_2$ et $\|.\|_\infty$ s'appellent norme euclidienne et norme de la convergence uniforme respectivement.

Normes sur $C([a, b], \mathbb{R})$

$C([a, b], \mathbb{R})$ espace d'une fonction continue de $[a, b]$ dans $\mathbb{R}$.
Pour tout f de $C([a, b], \mathbb{R})$ on a :
a) $\|f\|_1 = \int_a^b |f(t)| \, dt$
b) $\|f\|_2 = \left(\int_a^b f^2(t) \, dt \right)^{\frac{1}{2}}$
c) $\|f\|_\infty = \sup_{a \leq t \leq b} |f(t)| .$

Fonction Lipschitzienne

Définition 1.1.11 *Soient E une partie de $\mathbb{R}$, $f : E \to \mathbb{R}$ une application et k un réel positif.*

On dit que f est k lipschitzienne si

$$\forall (x, y) \in E^2, |f(x) - f(y)| \leq k |x - y| .$$

Définition 1.1.12 *Soit $f : E \to F$ une application et k un réel positif, (E, d_E) et (F, d_F) deux espaces métriques.*

On dit que f est k lipschitzienne si

$$\forall (x, y) \in E^2, d_F (f(x), f(y)) \leq k d_E (x, y) .$$

1.2 Conditions d'optimalité

Définition 1.2.1 *[10]Soit $f : \mathbb{R}^n \to \mathbb{R}$ on appelle problème de minimisation sans contraintes le problème (P) suivant*

$$\min \{f(x) : x \in \mathbb{R}^n\} . \quad (P) \tag{1.1}$$

1) $\widehat{x} \in \mathbb{R}^n$ est un minimum global de (P) si et seulement si

$$f(\widehat{x}) \leq f(x) : \forall x \in \mathbb{R}^n.$$

2) $\widehat{x} \in \mathbb{R}^n$ est un minimum local de (P) si et seulement s'il existe un voisinage $V_\varepsilon(\widehat{x})$ tel que

$$f(\widehat{x}) \leq f(x) : \forall x \in V_\varepsilon(\widehat{x}) .$$

3) $\widehat{x} \in \mathbb{R}^n$ est un minimum local strict de (P) si et seulement s'il existe un voisinage $V_\varepsilon(\widehat{x})$ tel que

$$f(\widehat{x}) < f(x) : \forall x \in V_\varepsilon(\widehat{x}) , \ x \neq \widehat{x}.$$

1.3 Conditions nécessaires d'optimalité

Théorème 1.3.1 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ telle que f est différentiable au point $\widehat{x} \in \mathbb{R}^n$. Soit $d \in \mathbb{R}^n$ telle que $\nabla f(\widehat{x})^\intercal d < 0$. Alors il existe $\delta > 0$ tel que $f(\widehat{x} + \alpha d) < f(\widehat{x})$ pour tout $\alpha \in]0, \delta[$.*

Preuve. [2] Comme f est différentiable en $\widehat{x}$ alors

$$f(\widehat{x} + \alpha d) = f(\widehat{x}) + \alpha \nabla f(\widehat{x})^\intercal d + \alpha \|d\| \lambda(\widehat{x}, \alpha d), \tag{1.2}$$

où $\lambda(\widehat{x}\ , \alpha d) \to 0$ pour $\alpha \to 0$. Ceci implique :

$$\frac{f(\widehat{x} + \alpha d) - f(\widehat{x})}{\alpha} = \nabla f(\widehat{x})^\intercal \cdot d + \|d\| \lambda(\widehat{x}, \alpha d),\ \alpha \neq 0,$$

et comme $\nabla f(\widehat{x})^\intercal \cdot d < 0$ et $\lambda(\widehat{x}, \alpha d) \to 0$ pour $\alpha \to 0$, il existe $\delta > 0$ tel que

$$\nabla f(\widehat{x})^\intercal d + \|d\| \lambda(\widehat{x}, \alpha d) < 0 \text{ pour tout } \alpha \in]0, \delta[\,,$$

et par conséquent on obtient :

$$f(\widehat{x} + \alpha d) < f(\widehat{x}) \text{ pour tout } \alpha \in]0, \delta[$$

■

Théorème 1.3.2 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ différentiable en $\widehat{x}$, si $\widehat{x}$ est un minimum local du problème (P) alors $\nabla f(\widehat{x}) = 0$.*

Preuve. Par contradiction, on suppose que $\nabla f(\widehat{x}) \neq 0$. Si on pose $d = -\nabla f(\widehat{x})$, on obtient :

$$\nabla f(\widehat{x})^\intercal d = -\|\nabla f(\widehat{x})\|^2 < 0,$$

et par le théorème précédent, il existe $\delta > 0$ tel que :

$$f(\widehat{x} + \alpha d) < f(\widehat{x})$$

pour tout $\alpha \in]0, \delta[$, mais ceci est contradictoire avec le fait que $\widehat{x}$ est un minimum local, d'où $\nabla f(\widehat{x}) = 0$. ■

Théorème 1.3.3 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux foix différentiable en $\widehat{x}$. Supposons que $\widehat{x}$ soit un minimum local du problème (P). Alors $\nabla f(\widehat{x}) = 0$ et $H(\widehat{x})$ est semi définie positive.*

Preuve. [14] Le théorème ci-dessus montre la première proposition, pour la deuxième proposition on a :

$$f(\widehat{x}+\alpha d) = f(\widehat{x}) + \alpha \nabla f(\widehat{x})^{\intercal}.d + \frac{1}{2}\alpha^2 d^{\intercal} H(\widehat{x}) d + \alpha^2 \|d\|^2 \lambda(\widehat{x}, \alpha d), \quad (1.3)$$

où $\lambda(\widehat{x}, \alpha d) \to 0$ pour $\alpha \to 0$. Ceci implique :

$$\frac{f(\widehat{x}+\alpha d) - f(\widehat{x})}{\alpha^2} = \frac{1}{2} d^{\intercal} H(\widehat{x}) d + \|d\|^2 \lambda(\widehat{x}, \alpha d), \alpha \neq 0.$$

Comme $\widehat{x}$ est un minimum local alors $f(\widehat{x}+\alpha d) > f(\widehat{x})$ pour α suffisament petit, d'où

$$\frac{1}{2} d^{\intercal} H(\widehat{x}) d + \|d\|^2 \lambda(\widehat{x}, \alpha d) \geq 0 \text{ pour } \alpha \text{ petit.}$$

En passant à la limite quand $\alpha \to 0$, on obtient que $d^{\intercal} H(\widehat{x}) d \geq 0$, d'où $H(\widehat{x})$ est semi définie positive. ■

1.4 Différentiabilité

1.4.1 Gradient, Hessien

On note par

$$\nabla^T f(x) = \frac{\partial f}{\partial x}(x) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)(x), \quad (1.4)$$

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

On appelle Hessien de f la matrice symétrique de $M_n(\mathbb{R})$

$$H(x) = \nabla\left(\nabla^T f\right)(x) = \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)(x),\ i = 1, \ldots, n,\ j = 1, \ldots, n$$

Définition 1.4.1 *On dit que x^* est un point stationnaire de f si $\nabla f(x^*) = 0$*

1.4.2 Dérivée directionnelle

Définition 1.4.2 *On appelle dérivée directionnelle de f dans la direction d au point x, notée $\delta f(x, d)$, la limite du rapport :*

$$\frac{f(x+hd) - f(x)}{h} \text{ lorsque } h \text{ tend vers } 0.$$

Autrement dit :

$$\delta f(x,d) = \lim_{h\to 0} \frac{f(x+hd) - f(x)}{h} = \nabla^T f(x)\, d$$

♦ Si $\|d\| = 1$: la dérivée directionnelle est le taux d'accroissement de f dans la direction d au point x.

Remarque 1.4.1 ★ *Le taux d'accroissement est maximal dans la direction du gradient.*

★ *Le gradient indique la direction de la plus grande pente.*

1.4.3 Direction de descente

Définition 1.4.3 *Soit $f : \mathbb{R}^n \to \mathbb{R}$, $\widehat{x} \in \mathbb{R}^n$, d un vecteur non nul de $\mathbb{R}^n$ est dit direction de descente au point $\widehat{x}$ si et seulement s'il existe un réel $\delta > 0$ tel que $\forall \alpha \in]0, \delta[$ on a :*

$$f(\widehat{x} + \alpha d) < f(\widehat{x})$$

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

Théorème 1.4.1 *Soit $f : \mathbb{R}^n \to \mathbb{R}$, différentiable au point $\widehat{x} \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ telle que :*

$$\nabla^T f(\widehat{x})\, d < 0$$

Alors d est une direction de descente en $\widehat{x}$.

1.4.4 Formule de Taylor

Si une fonction f est définie et continue sur $[a, b]$, ainsi que ses n premières dérivées, et si elle admet dans l'intervalle $]a, b[$ une dérivée d'ordre $n+1$, alors il existe une valeur $c \in]a, b[$ pour laquelle

$$f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!}f''(a) + ... + \frac{(b-a)^n}{n!}f^n(a) + \frac{(b-a)^{n+1}}{(n+1)!}f^{(n+1)}(c). \tag{1.5}$$

Cette égalité peut encore s'écrire avec $h = b - a$

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + ... + \frac{h^{n+1}}{(n+1)!}f^{(n+1)}(a+\theta h), \tag{1.6}$$

avec $0 < \theta < 1$.

Chapitre 2

Recherches linéaires inexactes

Pour résourde le problème $\min\{f(x) : x \in \mathbb{R}^n\}$ a chaque itération x_k, il existe deux types des recherches linéaires

i) Recherche linéaire exacte

ii) Recherche linéaire inexacte

dans ce chapitre, on étudions la recherche linéaire inexacte.

2.1 Recherche linéaire inexacte d'Armijo

2.1.1 Règle d'Armijo

Définition 2.1.1 *[17]*

Soient $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable, $x_k \in \mathbb{R}^n$ et $d_k \in \mathbb{R}^n$ une direction de descente, i.e, d_k vérifiant :

$$\nabla f(x_k)^\intercal d_k < 0.$$

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

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta_1 \nabla f(x_k)^\intercal d_k, \qquad 0 < \beta_1 < 1. \qquad \text{(Armijo)}$$

Si on note

$$\varphi_k(\alpha) = f(x_k + \alpha_k d_k). \tag{2.1}$$

Alors on a

$$\varphi_k(0) = f(x_k) \tag{2.2}$$

$$\varphi_k^{'}(\alpha) = \nabla f(x_k + \alpha_k d_k)^\intercal d_k, \tag{2.3}$$

et

$$\varphi_k^{'}(0) = \nabla f(x_k)^\intercal d_k < 0. \tag{2.4}$$

2.1.2 Interprétation graphique de la règle d'Armijo

Calculons l'équation de la tangente T_{M_0} au graphe de la fonction, et qui passe par le point $M_0(0, \varphi_k(0))$

$$T_{M_0} = \{(\alpha, y) : y = \varphi_k(0) + \alpha\varphi_k^{'}(0)\}, \tag{2.5}$$

soit en prenant en compte les relations (2.2) et (2.4)

$$T_{M_0} = \{(\alpha, y) : y = f(x_k) + \alpha\nabla f(x_k)^\intercal d_k\}. \tag{2.6}$$

Calculons maintenant l'équation de la droite notée Δ_{M_0} de coéfficient directeur $\beta_1\varphi_k^{'}(0)$ et qui passe par le point $M_0(0, \varphi_k(0))$. On a

$$\Delta_{M_0} = \{(\alpha, y) : y - \varphi_k(0) = \beta_1\varphi_k^{'}(0)(\alpha - 0)\},$$

ou encore

$$\Delta_{M_0} = \{(\alpha, y) : y = \varphi_k(0) + \alpha\beta_1\varphi_k^{'}(0)\}, \tag{2.7}$$

soit en considérant (2.2) et (2.4)

$$\Delta_{M_0} = \{(\alpha, y) : y = f(x_k) + \alpha\beta_1\nabla f(x_k)^\intercal d_k\}. \tag{2.8}$$

Les droites T_{M_0} et Δ_{M_0} passant par le même point $M_0(0, \varphi_k(0))$. Le coéfficient directeur de Δ_{M_0} est égal à β_1 fois le coéfficient directeur de T_{M_0}. Remarquons que $\varphi_k^{'}(0) = \nabla f(x_k)^\intercal d_k < 0$, alors

$$0 < \beta_1 < 1 \Rightarrow 0 > \varphi_k^{'}(0)\beta_1 > \varphi_k^{'}(0),$$

par conséquent, pour tout $\alpha > 0$, on a

$$\alpha\varphi_k^{'}(0)\beta_1 > \alpha\varphi_k^{'}(0),$$

soit en rajoutant aux deux membres de l'égalité précédente la même quantité $\varphi_k(0)$, on obtient

$$\varphi_k(0) + \alpha\varphi_k^{'}(0)\beta_1 > \varphi_k(0) + \alpha\varphi_k^{'}(0), \tag{2.9}$$

ou encore

$$f(x_k) + \alpha\beta_1\nabla f(x_k)^\intercal d_k > f(x_k) + \alpha\nabla f(x_k)^\intercal d_k. \tag{2.10}$$

Les relations (2.2), (2.4), (2.5) et (2.7) impliquent que pour $\alpha \in]0, +\infty[$, la droite Δ_{M_0} se trouve au dessus de la droite T_{M_0}. Par conséquent, il existe $\tau > 0$ tel que, pour tout $\alpha \in]0, \tau[$, le graphe de $\varphi_k(\alpha)$ se trouve au dessus de la droite Δ_{M_0}. Ecrivons celà de façon analytique

$$\forall\alpha \in]0, \tau[: \varphi_k(\alpha) \leq f(x_k) + \alpha\beta_1\nabla f(x_k)^\intercal d_k. \tag{2.11}$$

Utilisons le fait que $\varphi_k(\alpha) = f(x_k + \alpha_k d_k)$, (2.11) devient

$$\forall\alpha \in]0, \tau[: f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha\beta_1\nabla f(x_k)^\intercal d_k, \tag{2.12}$$

(2.12) est exactement la condition "**Armijo**"

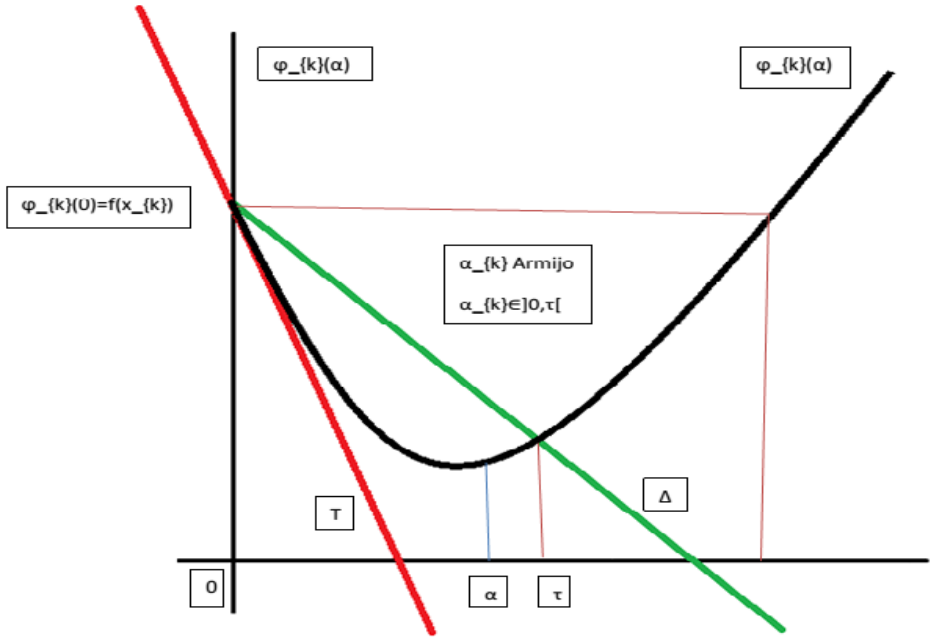


Fig2.1 Armijo

Algorithme 2.1.1
Etape 0 (Initialisation)
Choisir $\widehat{\alpha} \in]0, \infty[$ et $\beta_1 \in]0, 1[$.
Calculez

$$\varphi_k(\widehat{\alpha}) = f(x_k + \widehat{\alpha} d_k), \varphi_k(0) = f(x_k)$$

$$\varphi_k^{'}(0) = \nabla f(x_k)^{\intercal} d_k < 0,$$

et allez à l'étape 1
Etape 1 (Principale)
Si

$$\varphi_k(\widehat{\alpha}) \leq \varphi_k(0) + \widehat{\alpha}\beta_1 \varphi_k^{'}(0),$$

choisir le plus grand entier $n_0 \geq 0$ tel que :

$$\varphi_k(2^{n_0}\widehat{\alpha}) \leq \varphi_k(0) + 2^{n_0}\widehat{\alpha}\beta_1 \varphi_k^{'}(0),$$

et prendre $\widehat{\alpha} = 2^{n_0}\widehat{\alpha}$,
si on choisir le plus petit entier $m_0 \geq 0$ tel que :

$$\varphi_k(\frac{\widehat{\alpha}}{2^{m_0}}) \leq \varphi_k(0) + \frac{\widehat{\alpha}}{2^{m_0}}\beta_1\varphi_k^{'}(0),$$

et $\widehat{\alpha} = \frac{\widehat{\alpha}}{2^{m_0}}$.

Dans le théorème suivant on va assurer l'existence du pas d'Armijo en posant quelques conditions sur la fonction φ_k.

Théorème 2.1.1 *Considérons la fonction $\varphi_k : \mathbb{R}_+ \to \mathbb{R}$, définie par*

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

supposons que φ_k est une direction de descente en x_k $(\nabla f(x_k)^\intercal d_k < 0)$ et si $\beta_1 \in]0, 1[$, alors l'ensemble des pas vérifiant la règle d'Armijo et non vide.

Preuve. [11] On a

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

et

$$\Psi_k(\alpha) = f(x_k) + \beta_1\alpha\nabla^\intercal f(x_k)d_k.$$

Le developpement de Taylor-Young en $\alpha = 0$ de φ est :

$$\varphi(\alpha) = f(x_k + \alpha d_k) = f(x_k) + \beta_1\alpha\nabla^\intercal f(x_k)d_k + \alpha\xi(\alpha),$$

où $\xi(\alpha) \to 0$, $\alpha \to 0$, et comme $\beta_1 \in]0, 1[$ et $\varphi^{'}(0) = \nabla^\intercal f(x_k)d_k < 0$ on déduit :

$$f(x_k) + \alpha\nabla^\intercal f(x_k)d_k \leq f(x_k) + \beta_1\alpha\nabla^\intercal f(x_k)d_k,$$

pour $\alpha > 0$.

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

$$\varphi(\alpha) < \Psi(\alpha).$$

De ce qui précède et du fait que φ est bornée inférieurement, $\Psi(\alpha) \to -\infty$, $\alpha \to +\infty$, on déduit que la fonction $\Psi(\alpha) - \varphi(\alpha)$ a la propriété :

$$\begin{cases} \Psi(\alpha) - \varphi(\alpha) > 0, \text{ pour } \alpha \text{ assez petit,} \\ \Psi(\alpha) - \varphi(\alpha) < 0, \text{ pour } \alpha \text{ assez grand.} \end{cases}$$

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

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

$$\varphi(\overline{\alpha}) = \Psi(\overline{\alpha}),$$

et

$$\varphi(\alpha) < \Psi(\alpha),$$

pour $0 < \alpha < \overline{\alpha}$, ce qui achève la preuve. ■

2.2 Recherche linéaire inexacte de Goldstein

2.2.1 La règle de Goldstein

Soient $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable, $x_k \in \mathbb{R}^n$, le vecteur de descente $d_k \in \mathbb{R}^n (\nabla f(x)^\intercal d_k < 0)$.

On dit que le pas $\alpha_k > 0$ vérifié les deux conditions de Goldstein, si on a :

$$f(x_k + \alpha d_k) \leq f'(x_k) + \rho\alpha_k \nabla f(x)^\intercal d_k, \rho \in \left]0, \frac{1}{2}\right[\qquad (Goldstein1)$$

et

$$f(x_k + \alpha d_k) \geq f'(x_k) + (1-\rho)\alpha_k \nabla f(x)^\intercal d_k.\ \rho \in \left]0, \frac{1}{2}\right[\qquad (Goldstein2)$$

Si on note

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

alors (*Goldstein1*) et (*Goldstein2*) deviennent

$$\varphi_k(\alpha) \leq \varphi_k(0) + \rho\alpha_k \varphi_k^{'}(0)$$

$$\varphi_k(\alpha) \geq \varphi_k(0) + (1-\rho)\alpha_k \varphi_k^{'}(0)$$

2.2.2 Interprétation graphique

a) $0 < \rho < \frac{1}{2} \Rightarrow -\frac{1}{2} < -\rho < 0 \Rightarrow \frac{1}{2} < 1 - \rho < 1$. Donc

$$0 < \rho < \frac{1}{2} \Rightarrow (1 - \rho) > \rho.$$

D'autre part

$$\varphi_k^{'}(0) < 0 \Rightarrow (1 - \rho)\varphi_k^{'}(0) < \rho\varphi_k^{'}(0).$$

Par conséquent

$$\forall \alpha \in]0, +\infty[: \alpha(1 - \rho)\varphi_k^{'}(0) < \alpha\rho\varphi_k^{'}(0).$$

Ce qui emplique

$$\forall \alpha \in]0, +\infty[: \varphi_k(0) + \alpha(1 - \rho)\varphi_k^{'}(0) < \varphi_k(0) + \alpha\rho\varphi_k^{'}(0). \tag{2.13}$$

D'autre part, puisque $\varphi_k^{'}(0) < 0$, on a pour tout $\alpha \in \mathbb{R}_+^*$

$$0 < \rho < 1 \Rightarrow \rho\varphi_k^{'}(0) > \varphi_k^{'}(0) \Rightarrow \varphi_k(0) + \alpha\rho\varphi_k^{'}(0) > \varphi_k(0) + \alpha\varphi_k^{'}(0), \tag{2.14}$$

et

$$0 < 1-\rho < 1 \Rightarrow (1-\rho)\varphi_k^{'}(0) > \varphi_k^{'}(0) \Rightarrow \varphi_k(0)+\alpha(1-\rho)\varphi_k^{'}(0) > \varphi_k(0)+\alpha\varphi_k^{'}(0). \tag{2.15}$$

Notons par :

Δ_1 : la droite de coéfficient directeur $\rho\varphi_k^{'}(0)$ et passant par le point $(0, \varphi_k(0))$

Δ_2 : la droite de coéfficéent directeur $(1 - \rho)\varphi_k^{'}(0)$ et passant par le point $(0, \varphi_k(0))$

T : la tangente au graphe de $\varphi_k(\alpha)$, passant par le point $(0, \varphi_k(0))$.

On a

$$\Delta_1 = \{(\alpha, y) : y = \varphi_k(0) + \alpha\rho\varphi_k^{'}(0)\} \tag{2.16}$$

$$\Delta_2 = \{(\alpha, y) : y = \varphi_k(0) + \alpha(1 - \rho)\varphi_k^{'}(0)\} \tag{2.17}$$

$$T = \{(\alpha, y) : y = \varphi_k(0) + \alpha\varphi_k^{'}(0)\}. \tag{2.18}$$

Les relations (2.14), (2.15), (2.16), (2.17) et (2.18) impliquent que :

a) La droite Δ_1 est au dessus de la droite Δ_2 pour tout $\alpha \in \mathbb{R}_+^*$.

b) Les deux droites $\Delta_1$et Δ_2 sont au dessus de la tangente T, pour tout $\alpha \in \mathbb{R}_+^*$.

Notons par β_1 l'abscisse du premier point d'intersection de Δ_2 avec le graphe de $\varphi_k(\alpha)$.

Notons par β_2 l'abscisse du premier point d'intersection de Δ_1 avec le graphe de $\varphi_k(\alpha)$.

On a donc [2]

c) $]0, \beta_2[= \{\alpha : \alpha$ vérifie $(\textit{Goldstein1})\}$. Le graphe de φ_k est au dessous de la droite Δ_1 de coéffécient directeur $\rho\varphi^{'}(0)$

d)$]\beta_1, +\infty[= \{\alpha : \alpha$ vérifie $(\textit{Goldstein2})\}$. Le graphe de φ_k est au dessus de la droite Δ_2 l'équation $(1-\rho)\varphi^{'}(0)$

e) $]\beta_1, \beta_2[= \{\alpha : \alpha$ vérifie $(\textit{Goldstein1})$ et $(\textit{Goldstein2})\}$. Le graphe de φ_k est situé entre les deux droites $\Delta_1$et Δ_2 des coéffécients directeurs $\rho\varphi^{'}(0)$ et $(1-\rho)\varphi^{'}(0)$.

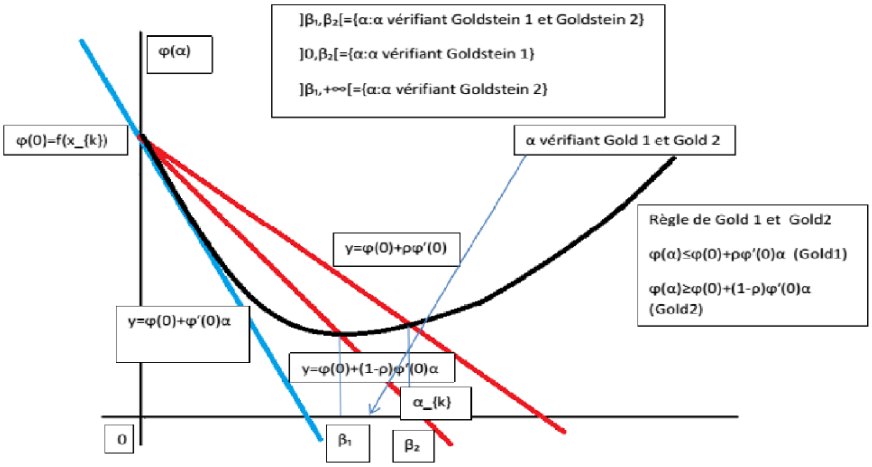


Fig2.2 Goldstein

Algorithme 2.2.1
Etape 1(Initialisation)
Choisir $\alpha_0 > 0$ et $\rho \in]0, 1[$. Poser $a_0 = 0$, $b_0 = 0$,
poser $k = 0$ et aller à l'étape 2
Etape 2(Testé *Goldstein1***)**
Itération k, on a $[a_k, b_k]$ et α_k, calculez $\varphi_k(\alpha_k)$,
si

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + (1 - \rho)\alpha_k\varphi_k^{'}(0),$$

allez à l'étape 3
si non
poser $b_{k+1} = \alpha_k$, $\alpha_{k+1} = \alpha_k$, et allez à l'étape 4

Etape 3(Testé *Goldstein2*)
Si $\varphi_k(\alpha_k) \geq \varphi_k(0) + (1-\rho)\alpha_k\varphi_k^{'}(0)$ stop. $\alpha^* = \alpha_k$,
sinon poser $\alpha_{k+1} = \alpha_k$, $b_{k+1} = b_k$ et allez à l'étape 4
Etape 4
Poser $\alpha_{k+1} = \frac{\alpha_{k+1}+b_{k+1}}{2}$,
poser $k = k+1$ et allez à l'étape 2.

Théorème 2.2.1 *Si $\varphi_k : \mathbb{R}^+ \to \mathbb{R}$; définie par $\varphi_k(\alpha) = f(x_k + \alpha d_k)$ est continue et bornée inférieurement, si d_k est une direction de descente en x_k et si $\rho \in \left]0, \frac{1}{2}\right[$, alors l'ensemble des pas vérifiant la règle de Goldstein1 et Goldstein2 est non vide.*

2.3 Recherche linéaire inexacte de Wolfe

2.3.1 La règle de wolfe

La règle de wolfe fait appel au calcul de $\varphi_k^{'}(\alpha)$, elle est donc en théorie plus coûteuse que la règle de Goldstein. Cependant dans le neumbreuses applications, le calcul du gradient $\nabla f(x)$ représente un faible coût additionnel en comparaison du coût d'évaluation de $f(x)$, c'est pourquoi cette règle est trés utilisée.

α_k est acceptable par la recherche linéaire inexacte de wolfe si α_k vérifie les deux conditions (Wolfe1) et (Wolfe 2) suivantes

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

$$\varphi_k^{'}(\alpha_k) \geq \sigma\varphi_k^{'}(0),\ 0 < \sigma < 1,\ \sigma > \rho \qquad \text{(Wolfe2)}$$

2.3.2 Interprétation graphique

a) On a

$$\left\{ \begin{array}{c} 0 < \rho < \sigma < 1 \\ \varphi_k^{'}(0) < 0 \\ \alpha \in \mathbb{R}_+^* \end{array} \right\} \Rightarrow \left\{ \varphi_k(0) + \alpha\rho\varphi_k^{'}(0) > \varphi_k(0) + \alpha\sigma\varphi_k^{'}(0) > \varphi_k(0) + \alpha\varphi_k^{'}(0) \right\}. \tag{2.19}$$

Notons par :

Δ_3 : la droite de coéfficient directeur $\rho\varphi_k^{'}(0)$ et passant par le point $(0, \varphi_k(0))$

Δ_4 : la droite de coéfficient directeur $\sigma\varphi_k^{'}(0)$ et passant par le point $(0, \varphi_k(0))$

T_1 : la tangente au graphe de $\varphi_k(\alpha)$, passant par le point $(0, \varphi_k(0))$.

T_2 : la tangente au graphe de $\varphi_k(\alpha)$, dont le coéfficient directeur est égal à $\sigma\varphi_k^{'}(0)$. Bien sur $T_2 // \Delta_4$

μ_1 : l'abscisse du premier point d'intersection de T_2 et du graphe de $\varphi_k(0)$

μ_2 : l'abscisse du premier point d'intersection de Δ_3 avec le graphe de $\varphi_k(0)$

T_3 : la tangente au graphe de $\varphi_k(\alpha)$, dont le coéfficient directeur est égal à $\varphi_k^{'}(\alpha_g)$, α_g vérifiant $0 < \alpha_g < \mu_1$

T_4 : la tangente au graphe de $\varphi_k(\alpha)$, dont le coéfficient directeur est égal à $\varphi_k^{'}(\alpha_d)$, α_g vérifiant $\alpha_d > \mu_1$

Δ_5 : la droite de coéfficient directeur $\varphi_k^{'}(\alpha_g)$ et passant par le point $(0, \varphi_k(0))$. Bien sur $T_3 // \Delta_5$

Δ_6 : la droite de coéfficient directeur $\varphi_k^{'}(\alpha_d)$ et passant par le point $(0, \varphi_k(0))$. Bien sur $T_4 // \Delta_6$

On a

$$\Delta_3 = \left\{(\alpha, y) : y = \varphi_k(0) + \alpha\rho\varphi_k^{'}(0)\right\} \tag{2.20}$$

$$\Delta_4 = \left\{(\alpha, y) : y = \varphi_k(0) + \alpha\sigma\varphi_k^{'}(0)\right\} \tag{2.21}$$

$$\Delta_5 = \left\{(\alpha, y) : y = \varphi_k(0) + \alpha\varphi_k^{'}(\alpha_g)\right\} \tag{2.22}$$

$$\Delta_6 = \left\{(\alpha, y) : y = \varphi_k(0) + \alpha\varphi_k^{'}(\alpha_d)\right\} \tag{2.23}$$

$$T_1 = \left\{(\alpha, y) : y = \varphi_k(0) + \alpha\varphi_k^{'}(0)\right\} \tag{2.24}$$

$$\begin{aligned} T_2 &= \left\{(\alpha, y) : y - \varphi_k(\mu_1) = \varphi_k^{'}(\mu_1)(\alpha - \mu_1)\right\} \\ &= \left\{(\alpha, y) : y = \varphi_k(\mu_1) + \alpha\varphi_k^{'}(\mu_1) - \mu_1\varphi_k^{'}(\mu_1)\right\} \end{aligned}$$

$$\begin{aligned} T_3 &= \left\{(\alpha, y) : y - \varphi_k(\alpha_g) = \varphi_k^{'}(\alpha_g)(\alpha - \alpha_g)\right\} \\ &= \left\{(\alpha, y) : y = \varphi_k(\alpha_g) + \alpha\varphi_k^{'}(\alpha_g) - \alpha_g\varphi_k^{'}(\alpha_g)\right\} \end{aligned}$$

$$\begin{aligned} T_4 &= \left\{(\alpha, y) : y - \varphi_k(\alpha_d) = \varphi_k^{'}(\alpha_d)(\alpha - \alpha_d)\right\} \\ &= \left\{(\alpha, y) : y = \varphi_k(\alpha_d) + \alpha\varphi_k^{'}(\alpha_d) - \alpha_d\varphi_k^{'}(\alpha_d)\right\}. \end{aligned}$$

On a aussi d'après la relation (2.19) :

$$\Delta_3 \text{ est au dessus de } \Delta_4$$

$$\Delta_3 \text{ et } \Delta_4 \text{ sont au dessus de } T_1.$$

Remarquons maintenant que

$$\Delta_5 \text{ est au dessous de } \Delta_4 \tag{2.25}$$

$$\Delta_6 \text{ est au dessus de } \Delta_4. \tag{2.26}$$

Analysons analytiquement (2.25) et (2.26). La relation (2.25) implique en prenant en considération (2.21) et (2.22)

$$\varphi_k(0) + \alpha\varphi_k^{'}(\alpha_g) < \varphi_k(0) + \alpha\sigma\varphi_k^{'}(0) \Rightarrow \varphi_k^{'}(\alpha_g) < \sigma\varphi_k^{'}(0). \tag{2.27}$$

La relation (2.26) implique en prenant en considération (2.21) et (2.23)

$$\varphi_k(0) + \alpha\varphi_k^{'}(\alpha_d) > \varphi_k(0) + \alpha\sigma\varphi_k^{'}(0) \Rightarrow \varphi_k^{'}(\alpha_d) > \sigma\varphi_k^{'}(0) \tag{2.28}$$

(2.27) et (2.28) impliquent que

$$\left\{ \begin{array}{c} \alpha_g \text{ ne vérifie pas Wolfe2} \\ \alpha_d \text{ vérifie Wolfe2.} \end{array} \right\} \tag{2.29}$$

Par conséquent μ_1 vérifie :

$$\left\{ \begin{array}{c} \varphi_k^{'}(\mu_1) = \sigma\varphi_k^{'}(0) \\ \alpha < \mu_1 \Rightarrow \varphi_k^{'}(\alpha) < \sigma\varphi_k^{'}(0) \\ \alpha > \mu_1 \Rightarrow \varphi_k^{'}(\alpha) > \sigma\varphi_k^{'}(0), \end{array} \right\} \tag{2.30}$$

et on a par la suite :
$]0, \mu_2[= \{\alpha : \alpha \text{ vérifie (Wolfe1)}\}$.
$]\mu_1, +\infty[= \{\alpha : \alpha \text{ vérifie (Wolfe2)}\}$.
$]\mu_1, \mu_2[= \{\alpha : \alpha \text{ vérifie (Wolfe1) et (Wolfe2)}\}$.

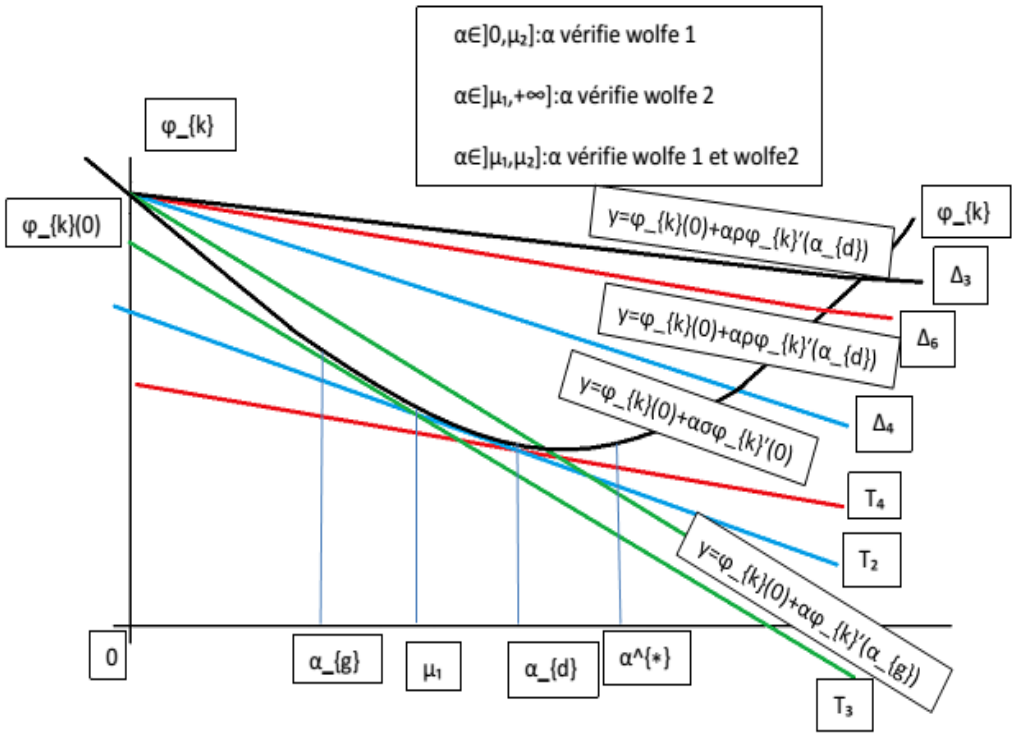


Fig2.3 Wolfe

2.3.3 Règle de Wolfe

Règle de Wolfe faible

Nous allons présenter les conditions de Wolfe faible sur $\alpha > 0$:

$$f\left(x_k + \alpha_k d_k\right) \leq f\left(x_k\right) + \rho\alpha_k \nabla f\left(x_k\right)^\top d_k,\ 0 < \rho < \frac{1}{2} \qquad \text{(W1)}$$

$$\nabla f\left(x_k + \alpha_k d_k\right)^\top d_k \geq \sigma \nabla f\left(x_k\right)^\top d_k,\ 0 < \sigma < 1,\ \sigma > \rho \qquad \text{(W2)}$$

Règle de Wolfe forte

Elles sont trés utilisées en théorie et les grands résultats de convergence ont été prouvés en utilisant des recherches linéaires inéxactes de Wolfe-forte.

Définition 2.3.1 *α_k vérifie les conditions de Wolfe-forte si les deux conditions suivantes sont vérifié :*

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho\alpha_k\varphi_k'(0),\ 0 < \rho < \frac{1}{2} \tag{W3}$$

$$\left|\varphi_k'(\alpha_k)\right| \leq -\sigma\varphi_k'(0),\ 0 < \sigma < 1,\ \sigma > \rho \tag{W4}$$

Algorithme 2.3.1
Etape 1(Initialisation)
Choisir $\alpha_0 > 0$, calculez $\varphi_k(0)$, $\varphi_k'(0)$. Prendre $\rho \in]0,1[$, $\sigma \in]\rho,1[$.
Poser $a_0 = b_0 = 0$, $k = 0$ et allez à l'étape 2.
Etape 2 (testé de **Wolfe1** ou **W3**)
Calculez $\varphi_k(\alpha_k)$. Si $\varphi_k(\alpha_k) \leq \varphi_k(0) + \rho\alpha_k\varphi_k'(0)$, aller à l'étape 3.
Sinon poser $a_{k+1} = a_k$, $b_{k+1} = \alpha_k$ et allez à l'étape 4.
Etape 3 (testé de **Wolfe2** ou **W4**))
Calculez $\varphi_k'(\alpha_k)$. Si $\varphi_k'(\alpha_k) \geq \nabla\varphi_k'(0)(\left|\varphi_k'(\alpha_k)\right| \leq -\sigma\varphi_k'(0))$. STOP
Prendre $\overline{\alpha} = \alpha_k$.
Sinon
Poser $a_{k+1} = \alpha_k$, $b_{k+1} = b_k$ et allez à l'étape 4
Etape 4 (calculé de α_{k+1})
$\alpha_{k+1} = \frac{a_{k+1}+b_{k+1}}{2}$.
Poser $k = k+1$ et allez à l'étape 2.

Remarque 2.3.1 *1)On voit bien que les conditions de Wolfe fortes impliquent les conditions de Wolfe faibles, en effet :*

$$\begin{aligned}\left|\nabla^\top f(x_k + \alpha_k d_k)\, d_k\right| &\leq -\sigma\nabla^\top f(x_k)\, d_k \\ &\Leftrightarrow \sigma\nabla^\top f(x_k)\, d_k \leq \nabla^\top f(x_k + \alpha_k d_k)\, d_k \leq -\sigma\nabla^\top f(x_k)\, d_k \\ &\Rightarrow \sigma\nabla^\top f(x_k)\, d_k \leq \nabla^\top f(x_k + \alpha_k d_k)\, d_k.\end{aligned}$$

2)L'existence de valeurs vérifiants les conditions de Wolfe faibles et fortes est donnèe par la Proposition suivante.

Proposition 2.3.1 *Soit $f \in C^1$ et d une direction de descente de f en x, on suppose que f est minorée. Alors, si $0 < \sigma < \sigma_1 < 1$, il existe des intervalles dans $\mathbb{R}_+$ qui vérifient les conditions de Wolfe faibles et fortes.*

La règle de Wolfe relaxée

Proposée par Dai et Yuan, cette règle consiste à choisir le pas α_k satisfaisant aux conditions :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho\alpha_k \nabla^\top f(x_k) d_k, \tag{W5}$$

$$\sigma_1 \nabla^\top f(x_k) d_k \leq \nabla f(x_k + \alpha_k d_k)^\top d_k \leq -\sigma_2 \nabla^\top f(x_k) d_k, \tag{W6}$$

où $0 < \rho < \sigma_1 < 1$ et $\sigma_2 > 0$.

Remarque 2.3.2 *Les conditions de Wolfe relaxée (W5) et (W6) impliquent les conditions de (Wolfe forte) tandis que pour les cas particulier $\sigma = \sigma_1 = \sigma_2$. En effet :*

$$\sigma_1 \nabla^\top f(x_k) d_k \leq \nabla f(x_k + \alpha_k d_k)^\top d_k \leq -\sigma_2 \nabla^\top f(x_k) d_k,$$

$$\Rightarrow \sigma \nabla^\top f(x_k) d_k \leq \nabla f(x_k + \alpha_k d_k)^\top d_k \leq -\sigma \nabla^\top f(x_k) d_k,$$

$$\Rightarrow \left| \nabla f(x_k + \alpha_k d_k)^\top d_k \right| \leq -\sigma \nabla^\top f(x_k) d_k. \tag{W4}$$

Théorème 2.3.1 *Si $\varphi_k : \mathbb{R}_+ \to \mathbb{R}$; définie par $\varphi_k(\alpha) = f(x_k + \alpha d_k)$ est derivable et bornée inférieurement et si d_k est une direction de descente en x_k et si $\rho \in]0, 1[$ et $\sigma \in]\rho, 1[$, alors l'ensemble des pas vérifiant la règle de Wolfe (Wolfe1 et Wolfe2) est non vide.*

Chapitre 3

Méthode de descente conjuguée

Rappelons dans ce qui suit les deux conditions suivantes que la méthode de descente conjuguée doit être vérifier pour prouver les résultats de convergence.

3.1 Hypothèses C1 et C2 et théorème de Zoutendijk

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

Condition $C1$

Soit $f : \mathbb{R}^n \to \mathbb{R}$. On dit que vérifie la condition $C1$ si f est continuement 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 $\nabla f(x)$ vérifie la condition de Lipschitz dans $V(\Gamma)$, c'est à dire, il existe une constante $L > 0$ telle que

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \text{ : pour tout } x, y \in V(\Gamma) \tag{3.1}$$

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

$$\|x\| \leq B : \forall x \in \Gamma \tag{3.2}$$

3.1.2 Théorème de Zoutendijk

Théorème 3.1.1 *[7] Considérons une méthode itérative quelconque générant une suite $\{x_k\}$ de la forme :*

$$x_{k+1} = x_k + \alpha_k d_k, \tag{3.3}$$

d_k étant une direction de descente et α_k est une recherche linéaire inéxacte de Wolfe. Supposons aussi que f vérifie les Conditions (3.1) et (3.2). Alors on a

$$\sum_{k=0}^{\infty} \frac{\left(g_k^\top d_k\right)^2}{\|d_k\|^2} < \infty. \tag{3.4}$$

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

$$f(x_{k+1}) \leq f(x_k) - C \|\nabla f(x_k)\|^2 \cos^2 \theta_k, \tag{3.5}$$

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

$$\cos \theta_k = \frac{-\nabla^\top f(x_k)\, d_k}{\|\nabla f(x_k)\| \, \|d_k\|}. \tag{3.6}$$

Proposition 3.1.1 *Si la suite $\{x_k\}$ générée par un algorithme vérifie la condition de Zoutendijk et la suite $\{f(x_k)\}$ est minorée, alors*

$$\sum_{k\geq 1} \|\nabla f(x_k)\|^2 \cos^2 \theta_k < \infty. \tag{3.7}$$

Preuve. En sommant les inégalités $\left(f(x_{k+1}) \leq f(x_k) - C \|\nabla f(x_k)\|^2 \cos^2 \theta_k\right)$, on a

$$\sum_{k\geq 1}^{l} \|\nabla f(x_k)\|^2 \cos^2 \theta_k \leq \frac{1}{C} \left(f(x_1) - f(x_{l+1})\right).$$

En effet :

$$\begin{aligned}
k &= 1 : f(x_2) \leq f(x_1) - C \|\nabla f(x_1)\|^2 \cos^2 \theta_1. \\
k &= 2 : f(x_3) \leq f(x_2) - C \|\nabla f(x_2)\|^2 \cos^2 \theta_2. \\
k &= 3 : f(x_4) \leq f(x_3) - C \|\nabla f(x_3)\|^2 \cos^2 \theta_3. \\
&\vdots \\
k &= l : f(x_{l+1}) \leq f(x_l) - C \|\nabla f(x_l)\|^2 \cos^2 \theta_l.
\end{aligned}$$

En sommant terme à terme, on a

$$f(x_2)+f(x_3)+\cdots+f(x_{l+1}) \leq f(x_1)+f(x_2)+\cdots+f(x_l)-C\sum_{k\geq 1}^{l}\|\nabla f(x_k)\|^2\cos^2\theta_k.$$

Donc

$$f(x_{l+1}) \leq f(x_1) - C\sum_{k\geq 1}^{l}\|\nabla f(x_k)\|^2\cos^2\theta_k.$$

Alors

$$\sum_{k\geq 1}^{l}\|\nabla f(x_k)\|^2\cos^2\theta_k \leq \frac{1}{C}\left(f(x_1) - f(x_{l+1})\right).$$

La série est donc convergente puisqu'il existe une constant C' telle que pour tout k, $f(x_k) \geq C'$. ■

3.2 Les méthodes dans lesquelles le termes $\|g_{k+1}\|^2$ figure dans le numérateur de β_k

Les méthodes **FR** et **CD** ont toutes comme numérateur commun le terme $\|g_{k+1}\|^2$.

3.2.1 Méthode de Fletcher-Reeves

Cette méthode été découverte en 1964 par Fletcher et Reeves, β_k est égale à :

$$\beta_{k+1}^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}. \tag{3.8}$$

Les avantages de la méthode de Fletcher-Reeves

Cette méthode est très intéressante, parce qu'elle nécessite le stockage de très peu d'informations (essentiellement trois vecteurs de dimension n).

Résultats d'El Baali

Comme on l'a remarqué auparavant le premier et le plus important résultat de convergence de la méthodes Fletcher Reeves (FR), avec une recherche linéaire inexacte est celui d'El Baali [1]. Nous exposons dans ce qui suit en détail ses résultats.

Théorème 3.2.1 *[1] Supposons que L'hypothèse (3.1) soit satisfaite. Considérons une méthode du type (3.3) avec* $\beta_k = \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$ *et le pas* α_k *satisfaisant à la règle de Wolfe forte (W3) et (W4) où* $\sigma \in \left]0, \frac{1}{2}\right[$. *Alors on a*

$$\frac{-1}{1-\sigma} \leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\sigma - 1}{1-\sigma} \; ; \; k = 1, \ldots \tag{3.9}$$

Preuve. Pour tout $k\,;\, g_k \neq 0$, la démonstration se fait par récurrence

1) Pour $k = 1$:

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

D'autre part :

$$0 < \sigma < \frac{1}{2} \Rightarrow \left\{ \begin{array}{l} \frac{-1}{1-\sigma} \leq -1 \\ \frac{2\sigma-1}{1-\sigma} \geq -1 \end{array} \right.$$

2) Supposons que (3.9) est satisfaite pour $k > 1$ et démontrons qu'elle pour $k+1$:

Supposons que :

$$\frac{-1}{1-\sigma} \leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\sigma - 1}{1-\sigma} \; ; \; k = 1, \ldots \tag{3.10}$$

On a :

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

Alors

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} = -1 + \frac{d_k^T g_{k+1}}{\|g_k\|^2}. \tag{3.11}$$

En utilisant la condition

$$\left|d_k^T g_{k+1}\right| \leq -\sigma d_k^T g_k \tag{3.12}$$

$$\Rightarrow \sigma d_k^T g_k \leq d_k^T g_{k+1} \leq -\sigma d_k^T g_k.$$

En divisant par $\|g_k\|^2$, on a

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

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

Alors

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

De (3.9), on a

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

$$\Rightarrow \frac{-1}{1-\sigma} = -1 - \frac{\sigma}{1-\sigma} \leq -1 + \sigma \frac{d_k^T g_k}{\|g_k\|^2}, \tag{3.13}$$

et

$$\frac{\sigma}{1-\sigma} \geq -\sigma \frac{d_k^T g_k}{\|g_k\|^2} \Rightarrow \frac{2\sigma - 1}{1-\sigma} = -1 + \frac{\sigma}{1-\sigma} \geq -1 - \sigma \frac{d_k^T g_k}{\|g_k\|^2}. \tag{3.14}$$

Par conséquent on aura :

$$\frac{-1}{1-\sigma} \leq \frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} \leq \frac{2\sigma - 1}{1-\sigma}.$$

Ce qui achéve la démonstration. ■

Corollaire 3.2.1 *Sous les mèmes hypothèses et conditions du Théorème précédent, on a pour tout $k \geq 1$, d_k est une direction de descente suffisante i.e., il existe une constante $C > 0$ telle que*

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

Preuve. D'aprés (3.9), on a

$$\begin{aligned} \frac{-1}{1-\sigma} &\leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\sigma - 1}{1-\sigma} \\ &\Rightarrow \quad d_k^T g_k \leq \frac{2\sigma-1}{1-\sigma} \|g_k\|^2 = -\frac{1-2\sigma}{1-\sigma} \|g_k\|^2 = -C \|g_k\|^2 , \end{aligned}$$

avec $C = \frac{1-2\sigma}{1-\sigma}$ ■

Convergence de la méthode de Fletcher-Reeves

Théorème 3.2.2 *[1] Supposons que L'hypothèse (3.1) soit satisfaite. Considérons une méthode du type (3.3) avec $\beta_k = \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$ et le pas α_k satisfaisant à la règle de Wolfe forte (W4) et (W3) où $\sigma \in \left]0, \frac{1}{2}\right[$. Alors cette méthode est globalement convergente, dans le sens suivant :*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{3.15}$$

Preuve. Puisque les conditions du théorème précédent sont satisfaites alors on a :

$$\frac{-1}{1-\sigma} \leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\sigma-1}{1-\sigma} \Rightarrow -\sigma d_{k-1}^T g_{k-1} \leq \frac{\sigma}{1-\sigma} \|g_{k-1}\|^2 .$$

D'autre part de condition

$$\left|d_k^T g_{k+1}\right| \leq -\sigma d_k^T g_k \Rightarrow \left|d_{k-1}^T g_k\right| \leq -\sigma d_{k-1}^T g_{k-1},$$

d'où

$$\left|d_{k-1}^T g_k\right| \leq -\sigma d_{k-1}^T g_{k-1} \leq \frac{\sigma}{1-\sigma} \|g_{k-1}\|^2 . \tag{3.16}$$

De (3.9), (3.16) et :

$$\begin{aligned}
\|d_k\|^2 &= \left|\|g_k\|^2 - 2\beta_k^{FR} d_{k-1}^T g_k + (\beta_k^{FR})^2 \|d_{k-1}\|^2\right| \\
&\leq \|g_k\|^2 + 2\beta_k^{FR}\left|d_{k-1}^T g_k\right| + (\beta_k^{FR})^2 \|d_{k-1}\|^2 \\
&\leq \|g_k\|^2 + 2\frac{\|g_k\|^2}{\|g_{k-1}\|^2}\cdot\frac{\sigma}{1-\sigma}\|g_{k-1}\|^2 + (\beta_k^{FR})^2 \|d_{k-1}\|^2 \\
&\leq \left(\frac{1+\sigma}{1-\sigma}\right)\|g_k\|^2 + (\beta_k^{FR})^2 \|d_{k-1}\|^2 .
\end{aligned}$$

Posons $\widehat{\sigma} = \frac{1+\sigma}{1-\sigma} \geq 1$, on aura :

$$\begin{aligned}
\|d_k\|^2 &\leq \widehat{\sigma}\|g_k\|^2 + \left(\beta_k^{FR}\right)^2 \|d_{k-1}\|^2 \\
&\leq \widehat{\sigma}\|g_k\|^2 + \left(\beta_k^{FR}\right)^2 \left[\widehat{\sigma}\|g_{k-1}\|^2 + \left(\beta_{k-1}^{FR}\right)^2 \|d_{k-2}\|^2\right] \\
&\leq \widehat{\sigma}\|g_k\|^4 \sum_{j=2}^{k}\frac{1}{\|g_j\|^2} + \widehat{\sigma}\|g_k\|^4 \|g_1\|^{-2} .
\end{aligned}$$

Alors

$$\|d_k\|^2 = \widehat{\sigma}\|g_k\|^4 \sum_{j=1}^{j=k}\frac{1}{\|g_j\|^2}. \tag{3.17}$$

Supposons que g_k est borné en dehors du zéro $\left(\lim_{k\to\infty}\inf\|g_k\| \neq 0\right)$, c'est-à-dire :

$$\|g_k\| \geq \omega > 0\,;\, \forall k \Rightarrow \frac{1}{\|g_k\|^2} \leq \frac{1}{\omega^2}.$$

De (3.17) on a :

$$\begin{aligned}
\|d_k\|^2 &\leq \widehat{\sigma}\|g_k\|^4 \sum_{j=1}^{k}\frac{1}{\|g_j\|^2} \leq \widehat{\sigma}\frac{\lambda^4}{\omega^2}\sum_{j=1}^{k}1 \\
&\Rightarrow \|d_k\|^2 \leq \widehat{\sigma}\frac{\lambda^4}{\omega^2}k,
\end{aligned}$$

d'où

$$\sum_{k\geq 1}\frac{1}{\|d_k\|^2} \geq \frac{\omega^2}{\widehat{\sigma}\lambda^4}\sum_{k\geq 1}\frac{1}{k} > \infty. \tag{3.18}$$

Ce qui veut dire que $\sum_{k\geq 1}\frac{1}{\|d_k\|^2}$ est divergente.

D'autre part, puisque les conditions du deux théorèmes précédent sont satisfaites, alors on a :

$$\sum_{k\geq 1}\cos^2\theta_k\,\|g_k\|^2<\infty,$$

et

$$c_1\frac{\|g_k\|}{\|d_k\|}\leq\cos\theta_k\leq c_2\frac{\|g_k\|}{\|d_k\|},$$

d'où

$$\begin{aligned}\sum_{k\geq 1}c_1^2\frac{\|g_k\|^2}{\|d_k\|^2}\|g_k\|^2 &\leq \sum_{k\geq 1}\cos^2\theta_k\,\|g_k\|^2<\infty\\ &\Rightarrow \sum_{k\geq 1}\frac{\|g_k\|^4}{\|d_k\|^2}<\infty\\ &\Rightarrow \sum_{k\geq 1}\frac{\omega^4}{\|d_k\|^2}<\infty\\ &\Rightarrow \sum_{k\geq 1}\frac{1}{\|d_k\|^2}<\infty.\end{aligned}$$

Ce qui contredit (3.18), d'où le résultat. Donc

$$\lim_{k\to\infty}\inf\|g_k\|=0.$$

■

Remarque 3.2.1 *La méthode de Fletcher-Reeves avec une recherche linéaire exacte génère des directions de descente.*

En effet, à chaque itération $k\geq 1$, on a :

$$\begin{aligned}d_{k+1}^T g_{k+1} &= \left(-g_{k+1}+\beta_{k+1}^{FR}d_k\right)^T g_{k+1}\\ &= -g_{k+1}^T g_{k+1}+\beta_{k+1}^{FR}d_k^T g_{k+1}\\ &= -\|g_{k+1}\|^2.\end{aligned}$$

Puisque

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

Donc α_k vérifie la condition nécessaire d'optimalité :

$$h_k'\left(\alpha_k\right)=d_k^T\nabla f\left(x_k+\alpha_k d_k\right)=d_k^T g_{k+1}=0,\ \forall k\geq 1$$

Inconvinients de la méthode de Fletcher-Reeves

La méthode de Fletcher-Reeves possède de bonnes propriété théoriques, mais en pratique, elle converge parfois lentement, et même prématurément. En effet, si les pas sont trop petits, il se peut que ce comportement s'élargit pour un grand nombre d'itérations, et c'est ce qu'il nous oblige à réinitialiser en posant $\beta_k^{FR} = 0$.

3.2.2 Méthode de descente conjuguée

Cette méthode été découverte en 1987 par Fletcher [5].

Rappelons que pour la méthode de la descente conjugué la variante β_k est égale à :

$$\beta_{k+1}^{CD} = \frac{\|g_{k+1}\|^2}{-d_k^T g_k}. \tag{3.19}$$

La propriété de descente de la méthode de la descente conjuguée

Fletcher [5] a démontré que la méthode de la descente conjuguée est une méthode de descente si le pas α_k est déterminé par la règle forte de Wolfe avec $\sigma \leq \frac{1}{2}$.

Dai et Yuan ont démontré le théorème suivant :

Théorème 3.2.3 *[3] Supposons que l'hypothèse (3.1) soit satisfaite. Considérons une méthode du type (3.3) avec* $\beta_k = \beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}$ *et le pas* α_k *satisfait aux conditions de wolfe relaxée (W5) et (W6) :*

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \rho d_k^T g_k \\ \text{et } \sigma_1 d_k^T g_k &\leq d_k^T g_{k+1} \leq -\sigma_2 d_k^T g_k, \end{aligned}$$

avec $0 < \rho < \sigma_1 < 1$ et $0 \leq \sigma_2 \leq 1$. Alors la méthode génère des directions de descente suffisante à chaque itération $k \geq 1$.

Preuve. On a

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

En divisant par $\|g_k\|^2$

$$\frac{-d_k^T g_k}{\|g_k\|^2} = 1 + \frac{d_{k-1}^T g_k}{d_{k-1}^T g_{k-1}}.$$

D'autre part de (W6), on a

$$\begin{aligned} \sigma_1 d_k^T g_k &\leq d_k^T g_{k+1} \leq -\sigma_2 d_k^T g_k \\ &\Rightarrow \sigma_1 d_{k-1}^T g_{k-1} \leq d_{k-1}^T g_k \leq -\sigma_2 d_{k-1}^T g_{k-1} \\ &\Rightarrow -\sigma_2 \leq \frac{d_{k-1}^T g_k}{d_{k-1}^T g_{k-1}} \leq \sigma_1 \\ &\Rightarrow 1 - \sigma_2 \leq 1 + \frac{d_{k-1}^T g_k}{d_{k-1}^T g_{k-1}} \leq 1 + \sigma_1, \end{aligned}$$

d'où

$$1 - \sigma_2 \leq \frac{-d_k^T g_k}{\|g_k\|^2} \leq 1 + \sigma_1,$$

Donc si $\|g_k\| \neq 0$, on a :

$$d_k^T g_k \leq -C \|g_k\|^2,$$

où $C = 1 - \sigma_2 > 0$ et donc d_k est une direction de descente suffisante. ■

Convergence de la méthode de la descente conjuguée

Théorème 3.2.4 *[3] Supposons que l'hypothèse (3.1) soit satisfaite. Considérons une méthode du type (3.3) avec* $\beta_k = \beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}$ *et le pas* α_k *satisfait aux conditions de wolfe relaxée (W5) et (W6) :*

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \rho d_k^T g_k, \\ \text{et } \sigma_1 d_k^T g_k &\leq d_k^T g_{k+1} \leq -\sigma_2 d_k^T g_k, \end{aligned}$$

avec $0 < \rho < \sigma_1 < 1$ *et* $\sigma_2 = 0$, *est globalement convergente, dans le sens suivant :*

$$\lim_{k\to\infty} \inf \|g_k\| = 0.$$

Preuve. De théorème précédent on a :

$$\begin{aligned}
1 - \sigma_2 \quad &\leq \quad \frac{-d_k^T g_k}{\|g_k\|^2} \leq 1 + \sigma_1 \\
&\Rightarrow \quad 1 \leq \frac{-d_k^T g_k}{\|g_k\|^2} \leq 1 + \sigma_1 \\
&\Rightarrow \quad \frac{1}{1+\sigma_1} \leq \frac{\|g_k\|^2}{-d_k^T g_k} \leq 1 \\
&\Rightarrow \quad \frac{1}{1+\sigma_1} \leq \frac{\|g_{k+1}\|^2 \|g_k\|^2}{-d_k^T g_k \|g_{k+1}\|^2} \leq 1 \\
&\Rightarrow \quad \frac{1}{1+\sigma_1} \leq \frac{\beta_{k+1}^{CD}}{\beta_{k+1}^{FR}} \leq 1 \\
&\Rightarrow \quad \beta_{k+1}^{CD} \leq \beta_{k+1}^{FR}.
\end{aligned}$$

D'après le théorème 3.3.2 on a :

$$\lim_{k\to\infty} \inf \|g_k\| = 0.$$

■

Chapitre 4

La convergence globale d'une méthode de descente conjuguée avec existence des erreurs

Considerons le problème d'optimisation sans contraintes suivant :

$$\min f(x),\ x \in \mathbb{R}^n, \tag{4.1}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est continuement differentiable. Notons par $g_k = \bigtriangledown f(x_k)$ est le gradient de f à x_k. Pour résoudre le problème (4.1) on utilise des méthodes itératives qui génèrent des suites $\{x_k\}$ définies par :

$$x_{k+1} = x_k + \alpha_k d_k \tag{4.2}$$

où $x_k \in \mathbb{R}^n$ est la kième approximation de la solution, α_k est le pas obtenu en effectuant une recherche linéaire exacte ou inexacte et d_k est la direction de recherche définie par

$$d_k = \begin{cases} -g_1 \ , & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geq 2, \end{cases}$$

où $g_k = g(x_k)$, et β_k est un scalaire. Les formes de β_k, sont définies comme suit :

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \qquad \text{Fletcher -Reeves [6],}$$

$$\beta_k^{PRP} = \frac{g_k^\intercal (g_k - g_{k-1})}{d_{k-1}^\intercal (g_k - g_{k-1})} \qquad \text{Polak-Ribiere-Polyak [15]},$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}(g_k - g_{k-1})} \qquad \text{Dai-Yuan [2]},$$

$$\beta_k^{CD} = -\frac{\|g_k\|^2}{g_{k-1}^\intercal d_{k-1}} \qquad \text{Conjugate-Descent [5]}.$$

Récemment Tang et Shi [19] à proposé une méthode modifier de Conjugate-Descent (CD) :

$$\beta_k = \frac{-\rho \|g_k\|^2}{\sigma g_{k-1}^\intercal d_{k-1}}, \tag{4.3}$$

où $\|.\|$ signifie la norme euclidienne, $\sigma \in (0, 1/2)$ et $\rho \in (1, \sigma/(1+\sigma))$.
La direction principale S_k est satisfaite :

$$g_k^\intercal s_k \leq -c_1 \|g_k\|^2, \ c_1 > 0 \tag{4.4}$$

Min [13] étudié la convergence globale de la méthode (4.3) avec le terme des erreurs w_k tel que

$$\|w_k\| \leq \gamma_k (q + p\|g_k\|), \ p, q > 0$$

où

$$\gamma_k = O(\frac{1}{k}), \ \gamma_k > 0.$$

Dans les conditions ci–dessus et que $g(x)$ est lipschitz continue, et

$$\liminf \|g_k\|_{k \to +\infty} = 0.$$

4.1 Propriétés de la direction

Nous adoptons l'hypothèse suivante sur la fonction f qui est couramment utilisé dans la littérature.

Proposition 4.1.1 *la fonction f est LC^1 et fortement convexe dans $\mathbb{R}^n$, i.e, il existe une constante $\tau > 0$ tels que*

$$\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq \tau \|x_{k+1} - x_k\|, \ \forall x_{k+1}, x_k \in \mathbb{R}^n, \tag{4.5}$$

et

$$L(x_1) = \{x \in \mathbb{R}^n / f(x) \leq f(x_1)\} \text{ est bornée.} \tag{4.6}$$

Pour la méthode modifiée de descent avec des erreurs, nous laissons

$$x_{k+1} = x_k + \alpha_k(s_k - w_k), \tag{4.7}$$

où α_k est le pas optimal, w_k est une erreur accumultative et la direction principale s_k est déterminer par

$$s_k = \begin{cases} -g_k + w_k, \ k = 1, \\ -g_k + \beta_k d_{k-1} + w_k, \ k \geq 2, \end{cases} \tag{4.8}$$

et

$$d_{k-1} = \begin{cases} s_{k-1} - w_{k-1}, \ g_{k-1}^{\intercal}(s_{k-1} - w_{k-1}) \leq 0 \\ -s_{k-1} + w_{k-1}, \ g_{k-1}^{\intercal}(s_{k-1} - w_{k-1}) > 0, \end{cases} \tag{4.9}$$

où et β_k est déterminer par (4.3), s_k et w_k satisfaisant les hypothèses suivantes :

$$\begin{cases} (A) : \|w_k\| \leq \gamma_k(q + p\|g_k\|), \ p, q > 0. \\ (B) : \qquad\qquad 0 \leq \gamma_k \leq c_1/k, \ c_1 > 0. \end{cases}$$

En utilisant une recherche linéaire inéxacte de Wolfe forte.

$$\begin{cases} f(x_k + \lambda_k d_k) \leq f(x_k) + \rho\lambda_k g_k^{\intercal} d_k; \\ |g^T(x_k + \lambda_k d_k)d_k| \leq -\sigma g_k^{\intercal} d_k. \end{cases} \tag{4.10}$$

Tout d'abord, nous donnons l'algorithme suivant :

Algorithme 4.1.1

Soit $\sigma \in (0, 1/2)$ et $\rho \in (0, \sigma/(\sqrt{3} + 2\sigma))$, $x_1 \in \mathbb{R}^n$; $k := 1$

Etape 0 : soit x_1 le point de départ, $g_1 = \nabla f(x_1)$, poser $d_1 = -g_1$,

poser k = 0 et aller à l'étape 1.

Etape 1 : si $g_k = 0$: alors stop ($x^* = x_k$)."Test d'arrêt", si non passez à l'étape 2.

Etape 2 : soit d_k est déterminé par (4.8).

Etape 3 : soit $x_{k+1} = x_k + \alpha_k d_k$.Poser $k := k + 1$, aller à l'étape 1.

Lemme 4.1.1 *Pour tout $k \geq 1$, on a $g_k^{\intercal} s_k \leq (\rho - 1)\|g_k\|^2$.*

Preuve. Si $k = 1$ on a $g_1^\intercal s_1 = -\|g_1\|^2$.

Si $k \geq 2$ et $g_{k-1}^\intercal d_{k-1} = 0$, alors $x_k = x_{k-1}$,

$$\begin{aligned} g_k^\intercal s_k &= g_k^\intercal(-g_k + \beta_k d_{k-1}) \\ &= g_{k-1}^\intercal(-g_{k-1} + \beta_k d_{k-1}) \\ &= -\|g_{k-1}\|^2 + \beta_k g_{k-1}^\intercal d_{k-1} \\ &= -\|g_{k-1}\|^2 < 0. \end{aligned}$$

Si $g_{k-1}^\intercal d_{k-1} < 0$ on a :

$$\begin{aligned} g_k^\intercal s_k &= g_k^\intercal(-g_k + \beta_k d_{k-1}) \\ &= -\|g_k\|^2 + \beta_k g_k^\intercal d_{k-1} \\ &\leq -\|g_k\|^2 + \beta_k(-\sigma g_{k-1}^\intercal d_{k-1}) \\ &= -\|g_k\|^2 + (\frac{-\rho\|g_k\|^2}{\sigma g_{k-1}^\intercal d_{k-1}})(-\sigma g_{k-1}^\intercal d_{k-1}) \\ &= (\rho - 1)\|g_k\|^2. \end{aligned} \tag{4.11}$$

Ceci termine la preuve. ■

4.2 La convergence globale de la méthode de descente conjuguée

Lemme 4.2.1 *Si $\{x_k\}$ est une suite infinie générée par l'algorithme (4.1), alors l'ensemble $I = \{k \ / \ g_k^\intercal(s_k - w_k) \geq 0\}$ est infini et*

$$\lim \|g_k\|_{k \to +\infty} = 0.$$

Preuve. Il découle de la définition de I :

$$g_k^\intercal s_k \geq g_k^\intercal w_k.$$

D'après le lemme (4.1.1), de (4.11) on a :

$$g_k^\intercal s_k \geq g_k^\intercal w_k \Rightarrow g_k^\intercal w_k \leq (\rho - 1)\|g_k\|^2. \tag{4.12}$$

En utilisant (4.12), on aura :

$$\begin{aligned}(1-\rho)\left\|g_k\right\|^2 &\leq -g_k^{\intercal} w_k \\ &\leq \left\|g_k\right\| \left\|w_k\right\| \\ &\leq \left\|g_k\right\| \left(\gamma_k(q+p\left\|g_k\right\|)\right).\end{aligned}$$

Donc

$$\begin{aligned}(1-\rho-\gamma_k p)\left\|g_k\right\| &\leq \gamma_k q \\ &\Rightarrow \left\|g_k\right\| = 0.\end{aligned}$$

■

Dans la suite, nous supposons que I est un ensemble fini et

$$d_k = s_k - w_k,\ g_k^{\intercal} d_k < 0,\ \forall k > 0.$$

Lemme 4.2.2 *[?] Sous les hypothèses (4.5), (4.6), la méthode définie par (4.2), (4.3) et (4.8) engendre une suite $\{x_k\}$ telle que :*

$$\sum_{k=1}^{\infty} \frac{(g_k^{\intercal} d_k)^2}{\left\|d_k\right\|^2} < \infty. \tag{4.13}$$

Lemme 4.2.3 *Sous les hypothèses (4.5), (4.6), la méthode définie par (4.2), (4.3) et (4.8) engendre une suite $\{x_k\}$ telle que,*

$$\text{il existe une canstante positive } \varepsilon_0 : \left\|g_k\right\| \geq \varepsilon_0, \forall k \geq 1 \tag{4.14}$$

et

$$\sum_{k=1}^{\infty} \frac{\left\|g_k\right\|^4}{\left\|d_k\right\|^2} < +\infty. \tag{4.15}$$

Preuve. *On a*

$$\begin{aligned}g_k^{\intercal} d_k &= g_k^{\intercal}(-g_k + \beta_k d_{k-1} - w_k) \\ &= -\left\|g_k\right\|^2 + \beta_k g_k^{\intercal} d_{k-1} - g_k^{\intercal} w_k.\end{aligned}$$

Donc

$$\begin{aligned}|g_k^{\intercal} d_k| &= |-\left\|g_k\right\|^2 + \beta_k g_k^{\intercal} d_{k-1} - g_k^{\intercal} w_k| \\ &\geq \left\|g_k\right\|^2 - |\beta_k g_k^{\intercal} d_{k-1} - g_k^{\intercal} w_k|.\end{aligned}$$

D'autre part on a

$$-|\beta_k g_k^\intercal d_{k-1} - g_k^\intercal w_k| \geq -|\beta_k g_k^\intercal d_{k-1}| - |g_k^\intercal w_k|,$$

Posons $t_k = \frac{|g_k^\intercal d_k|}{\|d_k\|}$, *en utilisant la règle de Wolfe (4.10) :*

$$\begin{aligned} |g_k^\intercal d_k| &\geq \|g_k\|^2 - |\beta_k g_k^\intercal d_{k-1}| - |g_k^\intercal w_k| \\ &\geq \|g_k\|^2 - \beta_k \sigma |g_{k-1}^\intercal d_{k-1}| - \|g_k\| \|w_k\| \\ &= \|g_k\|^2 - \beta_k \sigma t_{k-1} \|d_{k-1}\| - \|g_k\| \|w_k\| . \end{aligned}$$

D'après (4.8), on aura

$$\beta_k d_{k-1} = s_k - w_k + g_k,$$

ce qui implique

$$-\sigma t_{k-1} \beta_k \|d_{k-1}\| \geq -\sigma t_{k-1}(\|s_k\| + \|w_k\| + \|g_k\|).$$

Donc

$$|g_k^\intercal d_k| \geq \|g_k\|^2 - \sigma t_{k-1}(\|s_k\| + \|w_k\| + \|g_k\|) - \|g_k\| \|w_k\| . \tag{4.16}$$

En utilisant (4.14), on aura :

$$\frac{\|g_k\|}{\varepsilon_0} \geq 1 \Rightarrow -\frac{\|g_k\|^2}{\varepsilon_0} \leq -\|g_k\| , \tag{4.17}$$

$$-\frac{\|g_k\|^2 \|w_k\|}{\varepsilon_0} \leq -\|g_k\| \|w_k\| , \tag{4.18}$$

et

$$\frac{\|g_k\|^2}{\varepsilon_0^2} \|w_k\| \geq \|w_k\| \frac{\|g_k\|}{\varepsilon_0} \geq \|w_k\| . \tag{4.19}$$

Par conséquent, en substituant (4.17), (4.18) et (4.19), dans (4.16), nous avons

$$\begin{aligned} |g_k^\intercal d_k| &\geq \|g_k\|^2 - \sigma t_{k-1} \|d_k\| - \sigma t_{k-1} \frac{\|g_k\|^2}{\varepsilon_0^2} \|w_k\| - \sigma t_{k-1} \frac{\|g_k\|^2}{\varepsilon_0} - \|g_k\| \|w_k\| \\ &\geq (1 - \sigma t_{k-1} \frac{\|w_k\|}{\varepsilon_0^2} - \frac{\sigma t_{k-1} + \|w_k\|}{\varepsilon_0}) \|g_k\|^2 - \sigma t_{k-1} \|d_k\| , \end{aligned}$$

ce qui conduit à

$$\frac{|g_k^{\intercal} d_k|}{\|d_k\|} \geq (1 - \frac{\sigma t_{k-1} \|w_k\|}{\varepsilon_0^2} - \frac{\sigma t_{k-1} + \|w_k\|}{\varepsilon_0}) \frac{\|g_k\|^2}{\|d_k\|} - \sigma t_{k-1},$$

donc

$$t_k + \sigma t_{k-1} \geq (1 - \frac{\sigma t_{k-1} \|w_k\|}{\varepsilon_0^2} - \frac{\sigma t_{k-1} + \|w_k\|}{\varepsilon_0}) \frac{\|g_k\|^2}{\|d_k\|}.$$

D'après le lemme (4.2.1) on a

$$t_k + \sigma t_{k-1} \geq \frac{\|g_k\|^2}{2\|d_k\|},$$

et

$$4(t_k + \sigma t_{k-1})^2 \geq \frac{\|g_k\|^4}{\|d_k\|^2},$$

de

$$(t_k + \sigma t_{k-1})^2 \leq 4t_k^2 + 4\sigma^2 t_{k-1}^2 \ \textit{et} \sum_{k=1}^{\infty} t_k^2 < +\infty,$$

on a

$$\sum_{k=1}^{\infty} [t_k + \sigma t_{k-1}]^2 < +\infty,$$

d'où

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

Ceci achève la démonstration. ■

Théorème 4.2.1 *Sous les hypothèses (4.5), (4.6), la méthode définie par (4.2), (4.3) et (4.8) engendre une suite* $\{x_k\}$ *telle que :*

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{4.20}$$

Preuve. Supposons au contraire que $\|g_k\| \geq \varepsilon_0$ pour tout k, d'aprés le lemme (4.1.1) et (A), on a

$$\begin{aligned} g_k^T d_k &= g_k^T (s_k - w_k) \\ &= g_k^T s_k - g_k^T w_k \\ &\leq (\rho - 1) \|g_k\|^2 + \|g_k\| \|w_k\| \\ &\leq (\rho - 1) \|g_k\|^2 + \|g_k\| (\gamma_k (q + p \|g_k\|)) \\ &= (\rho - 1) \|g_k\|^2 + \gamma_k p \|g_k\|^2 + \gamma_k q \|g_k\| \\ &= [(\rho - 1) + \gamma_k p] \|g_k\|^2 + \gamma_k q \|g_k\| \\ &\leq (\rho - 1 + \gamma_k p) \|g_k\|^2 + \gamma_k q \frac{\|g_k\|^2}{\varepsilon_0} \\ &= (\rho - 1 + \gamma_k p + \gamma_k \frac{q}{\varepsilon_0}) \|g_k\|^2 \\ &= -\left(1 - \rho - \gamma_k \left(p + \frac{q}{\varepsilon_0}\right)\right) \|g_k\|^2, \end{aligned}$$

ainsi, de (B), pour k est suffisamment grand, nous avons

$$\begin{aligned} g_k^T d_k &\leq -(1 - \rho - \rho) \|g_k\|^2 \\ &= -(1 - 2\rho) \|g_k\|^2. \end{aligned} \tag{4.21}$$

D'aprés (4.2), (4.7) et (4.8), on obtient

$$\begin{aligned} d_k &= s_k - w_k \\ &= -g_k + \beta_k d_{k-1} - w_k. \end{aligned}$$

Il s'ensuit que

$$\begin{aligned} \|d_k\|^2 &\leq (\|g_k\| + \beta_k \|d_{k-1}\| + \|w_k\|)^2 \\ &\leq 3 \|g_k\|^2 + 3\beta_k^2 \|d_{k-1}\|^2 + 3 \|w_k\|^2 \\ &\leq 3 \|g_k\|^2 + \frac{3\rho^2 \|g_k\|^4}{\sigma^2 \left(g_{k-1}^T d_{k-1}\right)^2} \|d_{k-1}\|^2 + 3\gamma_k^2 (q + p \|g_k\|)^2. \end{aligned}$$

D'autre part on a

$$g_{k-1}^T d_{k-1} \leq -(1 - 2\rho) \|g_{k-1}\|^2,$$

ce qui emplique

$$\frac{1}{\left(g_{k-1}^T d_{k-1}\right)^2} \leq \frac{1}{(1 - 2\rho)^2 \|g_{k-1}\|^4}.$$

Donc

$$\|d_k\|^2 \leq 3\|g_k\|^2 + \frac{3\rho^2\|g_k\|^4}{\sigma^2(1-2\rho)^2\|g_{k-1}\|^4}\|d_{k-1}\|^2 + 3\gamma_k^2(q+p\|g_k\|)^2.$$

En divisons par$\|g_k\|^4$

$$\begin{aligned}\frac{\|d_k\|^2}{\|g_k\|^4} &\leq \frac{3}{\|g_k\|^2} + \frac{3\rho^2}{\sigma^2(1-2\rho)^2\|g_{k-1}\|^4}\|d_{k-1}\|^2 + \frac{3\gamma_k^2(q+p\|g_k\|)^2}{\|g_k\|^4} \\ &\leq \frac{3}{\|g_k\|^2} + \frac{3\rho^2}{\sigma^2(1-2\rho)^2}\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + 3\gamma_k^2\left(\frac{q}{\|g_k\|^2}+\frac{p}{\|g_k\|}\right)^2.\end{aligned}$$

On pose $m_k = \frac{\|d_k\|^2}{\|g_k\|^4}$, alors

$$\begin{aligned}m_k &\leq \frac{3}{\|g_k\|^2} + \frac{3\rho^2}{\sigma^2(1-2\rho)^2}m_{k-1} + 3\gamma_k^2\left(\frac{q}{\|g_k\|^2}+\frac{p}{\|g_k\|}\right)^2 \\ &\leq \frac{3}{\varepsilon_0^2} + vm_{k-1} + 3\gamma_k^2\left(\frac{q}{\varepsilon_0^2}+\frac{p}{\varepsilon_0}\right)^2, \text{ tq} : v = \frac{3\rho^2}{\sigma^2(1-2\rho)^2} \\ &\leq \Psi_k + vm_{k-1} \qquad \text{tq } : \Psi_k = \frac{3}{\varepsilon_0^2} + 3\gamma_k^2\left(\frac{q}{\varepsilon_0^2}+\frac{p}{\varepsilon_0}\right)^2.\end{aligned}$$

D'après (B) et $0 < \rho < \sigma/\left(\sqrt{3}+2\sigma\right)$, il existe $\Psi > 0$ tq $\Psi_k < \Psi$, $0 < v < 1$.

Ainsi

$$\begin{aligned}m_k &\leq \Psi_k + vm_{k-1} \\ &\leq \Psi_k + v(\Psi_{k-1} + vm_{k-2}) \\ &\vdots \\ &\leq \Psi\left(1+v+v^2+v^3+\ldots+v^{k-2}\right) + m_1 \\ &\leq \Psi(k-1) + m_{1.}\end{aligned}$$

Donc

$$\sum_{k=1}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} = \sum_{k=1}^{\infty}\frac{1}{m_k} \geq \sum_{k=1}^{\infty}\frac{1}{m_1+\Psi(k-1)} = +\infty.$$

Ce qui contredit le lemme (4.2.3). Donc (4.20) est vrai. ■

4.3 Résultats numériques

Dans ce qui suit, nous rapportons les résultats numériques d'une nouvelle méthode du descente conjuguée pour deux problèmes de test standard. La nouvelle méthode est notée par **MCD**. Le terme d'erreur w_k est obtenu aléatoirement dans le cas où il vérifie les hypothèses (A) et (B). Pour chaque problème, l'unité de temps **CPU** est la seconde et **IT** désigne le numéro d'itération à la fin de l'algorithme.

Les paramètres utilisés dans l'algorithme sont définis comme suite : $p = 1, q = 0.1, c_1 = 1$. Tous les codes sont écrits en matlab 7.1 et exécutés sur un ordinateur portable (Intel(R) Celeron (R) CPU N3060@).

Problème 1.[12]

$$\begin{aligned} f(x) &= 10\left(x_1^2 - x_2\right)^2 + (1-x_1)^2 + 9\left(x_4 - x_3^2\right)^2 + (1-x_3)^2 + 10.1\left((x_2-1)^2 + (x_4-1)^2\right) \\ &\quad +19.8\,(x_2-1)\,(x_4-1)\,. \end{aligned}$$

Le point initiale $(-3,-1,-3,-1)^\intercal$.

Problème 2.[12]

$$f(x) = \sum_{i=1}^{N/2}\left(\left(x_{2i} - x_{2i-1}^2\right)^2 + (1-x_{2i-1})^2\right).$$

Le point initiale $(-1,2,1,\ldots,-1,2,1)^\intercal$.

Problème 3.[18]

$$f(x) = (x_1 + 10x_2)^4 + 5\,(x_3 - x_4)^4 + (x_2 - 2x_3)^4 + 10\,(x_1 - x_4)^4\,.$$

Le point initiale $(2,2,-2,-2)^\intercal$.

Pour chaque méthode, nous obtenons les résultats correspondants pour $\|g_k\| \leq 10^{-6}$.

Tableau 1 liste des résultats numériques.

Problème	ρ	σ	IT	CPU	$f(x_k)$
1	0.05	0.1	306	0.1903	3.9468(-13)
	0.09	0.2	356	0.2203	2.9511(-13)
	0.1	0.3	474	0.3004	3.5701(-13)
2	0.09	0.2	35	0.2053	3.2782(-12)
	0.15	0.4	40	0.3292	1.5623(-12)
3	0.05	0.1	25	0.0200	7.0888(-6)
	0.09	0.2	24	0.0200	9.9293(-6)
	0.1	0.3	26	0.0200	9.9407(-6)
	0.15	0.4	22	0.0200	6.0837(-6)

D'après le tableau 1, la MCD est stable et efficace pour les problèmes testés. Certainement, plus de test devrait probablement être nécessaire.

4.4 Conclusion

Dans ce travail, nous avons présenté une nouvelle méthode du gradient qui donne toujours une direction de recherche de descente, et nous avons prouvé qu'elle converge globalement vers la solution.

Dans la méthode de descente conjugée avec des erreurs. Quand la direction de la descente est légèrement perturbée, la convergence globale est toujours valable et montrons la stabilité de nouvelle méthode. Pour rechercher plus poussées, nous devrions étudier la limite supérieure du terme d'erreur w_k et espère trouver une limite supérieure plus grande.

Bibliographie

[1] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA J. Numer. Anal. 5 (1985), pp. 121-124.

[2] Y. H. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. Optim., 10 (1999), pp.177–182.

[3] Y. H. Dai and Y. Yuan (1996), Convergence properties of the conjugate descent method, Advances in Mathematics, 6, pp.552-562.Y.

[4] A. V. Fiacco and G. P. McCormick (1968), Nonlinear Programming, John Wiley, New York.

[5] R. Fletcher, Practical Methods of Optimization vol. 1 : Unconstrained Optimization, John Wiley & Sons, New York, 1987.

[6] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, Comput. J. 7(1964), pp. 149–154.

[7] J. C. Gilbert (2007), Eléments d'Optimisation Di erentiable : Théorie et Algorithmes, Notes de cours, Ecole Nationale Supérieure de Techniques Avancées, Paris.

[8] A. A. Goldstein (1965), On steepest descent, SIAM J. on Control A, Vol. 3, No. 1, pp. 147-151.

[9] M.R. Hestenes and E.L. Stiefel (1952), Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards Sect.,5(49), pp. 409.436.

[10] B. Khelifa (2012,2013), (mémoire) Etude de la convergence d'une combinaison de familles du gradient conjugue, Souk-Ahras.

[11] G.H. Liu and J.Y. Han and H.X. Yin (1995), Global convergence of the Fletcher-Reeves algorithm with an inexact line search, Appl. math. J. Chinese Univ. Ser. B, 10, pp. 75-82.

[12] M. X. Li and C. Y. Wang, Dai-Yuan conjugate gradient method with linesearch in the presence of errors, Chinese J. Engin. Math. 23 (5) (2006), 891–900.

[13] Min sun, Convergence of a modified conjugate descent method in the presence of errors, Acta mathematica vietnamica, Volume 35, Number 3, 2010, pp. 371–378.

[14] J.J.Moré, B.S.Garbow, and K. E. Hillstorm,"Testing unconstrainedoptimization software," ACM Transactions on Mathematical Software, vol.7, pp.17-41, 1981.

[15] E. Polak and G. Ribière, Note sur la convergence de directions conjugées, Rev. Francaise Informat Recherche Opertionelle, 3e Année 16 (1969), pp. 35–43.

[16] B.T. Polyak (1969), The conjugate gradient method in extremem problems, Comput. Math. Math. Phys., 9, pp. 94-112.

[17] B. Rachid. Optimisation sans contraintes tome1, https ://www.fichier-pdf.fr/professeur-benzine-rachid-cours-optimisation.2016

[18] Z. J. Shi, A new memory gradient method under exact line search, Asian Pacific J. Oper. Res. 20 (2) (2003), 275–284.

[19] J. Y. Tang and Z. J. Shi, A class of global convergent memory gradient methods and its linear convergence rate, Advance in Math. 36 (1) (2007), 67–75.