

وزارة التعليم العالي والبحث

جامعة باجي مختار
- عنابة -

 **BADJI MOKHTAR -ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR
ANNABA**

Faculté des Sciences

Année : 2013

Département de Mathématiques

MEMOIRE

Présenté en vue de l'obtention du diplôme de **MAGISTER**

**SUR QUELQUES RESULTATS NOUVEAUX DE LA
METHODE DU GRADIENT CONJUGUE**

Option
Optimisation

Par
Boumaraf badredine

DIRECTEUR DE MEMOIRE : R. BENZINE Prof . U.B.M. ANNABA

Devant le jury

PRESIDENT : Y.LASKRI Prof U.B.M. ANNABA

EXAMINATEUR : F.REBANI Prof U.B.M. ANNABA

EXAMINATEUR : H.ZEGHDOUDI M.C. U.B.M. ANNABA

فليكن:

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

مسألة التجويد من دون نظام تقييد بموجب:

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

في هذا العمل . نقدم طريقة الجديدة للممال متزاوج . تم دراستها و تطويرها من طرف ريفال . مصطفى . إسماعيل و لونوغ في 2012 .

الكتاب استعملوا البحوث العلمانيات التامة و برهنوا أن الأشعة المولدة لهذه الطريقة الجديدة هم أشعة نزول كافية . و قد أظهروا أيضا تقارب الكلي لهذه الطريقة . و أختموا العمل بتجارب رقمية .

RESUME

Soit

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

et (P) le problème d'optimisation sans contraintes suivant:

$$\min \{f(x), x \in \mathbb{R}^n\} \quad (P) \quad (0.1)$$

Dans ce travail, on présente une nouvelle méthode du gradient conjugué, étudiée et développée par Rivaie, Mustafa, Ismail, Lelong en 2012 ([]).

Les auteurs ont utilisé des recherches linéaires exactes et ont montré que les directions qu'engendrent cette nouvelle méthode sont des directions de descente suffisante. Ils ont aussi montré la convergence globale de cette méthode. Ils ont terminé leur travail par des tests numériques.

Mots clés : Gradient conjugué, Algorithme, Convergence globale, Recherche linéaire inexacte, Règle d'Armijo, Règle de Wolfe (forte et faible), Méthode de Hestenes-Stiefel, Méthode de Fletcher-Reeves, Méthode de Polak-Ribière-Polyak, Méthode de la descente conjuguée, Méthode de Dai-Yuan, Méthode de Rivaie, Mustafa, Ismail, Lelong .

ABSTRACT

Let

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

and (P) the following unconstrained optimization problem :

$$\min \{f(x), x \in \mathbb{R}^n\} \quad (P) \quad (0.1)$$

In this work a new class of conjugate gradient coefficients that possess global convergence properties is presented. The global convergence result is established using exact line searches. Numerical results shows that proposed formula is superior and more efficient when compared to other gradient conjugate methods.

Key words : Conjugate gradient, Algorithm, Global convergence, Inexact line search, Armijo line search, Strong Wolfe line search, Weak Wolfe line search, Hestenes-Stiefel Method, Fletcher-Reeves Method, Polak-Ribière-Polyak Method, Conjugate descent Method, Dai-Yuan Method. Rivaie, Mustafa, Ismail, Lelong Method

Table des matières

| | | |
|----------|---|-----------|
| 1 | OPTIMISATION SANS CONTRAINTES | 7 |
| 1.1 | Définitions | 7 |
| 1.2 | Direction de descente | 7 |
| 1.3 | Schémas général des algorithmes d'optimisation sans contraintes | 8 |
| | 1.3.1 Exemples de choix de directions de descente | 9 |
| | 1.3.2 Exemple de choix de pas λ_k | 9 |
| 1.4 | Conditions nécessaires d'optimalité | 9 |
| | 1.4.1 Condition nécessaire d'optimalité du premier ordre . . | 9 |
| | 1.4.2 Condition nécessaire d'optimalité du second ordre . . . | 10 |
| 1.5 | Conditions suffisantes d'optimalité | 10 |
| 2 | Optimisation unidimensionnelle ou recherche linéaire | 12 |
| 2.1 | Position du Problème | 12 |
| 2.2 | Recherches linéaires exactes | 13 |
| | 2.2.1 L'intervalle d'incertitude | 13 |
| | 2.2.2 Deux méthodes d'optimisation unidimensionnelle sans calcul de dérivées. La méthode de dichotomie et la mé- thode du nombre d'or (golden section). | 14 |
| 2.3 | recherches linéaires inexactes ou économiques | 16 |
| | 2.3.1 Introduction. | 16 |
| | 2.3.2 Recherche linéaire d'Armijo(1966). | 17 |
| | 2.3.3 Recherches linéaires de Goldstein et de Wolf | 19 |
| | 2.3.4 Algorithme des recherches linéaires de Goldstein et de Wolfe | 20 |

| | | |
|----------|--|-----------|
| 3 | Méthodes du gradient conjugué. Etude algébrique | 22 |
| 3.1 | Méthode des directions conjuguées : Cas quadratique | 22 |
| 3.1.1 | Aperçu général | 22 |
| 3.1.2 | Définition et propriétés des directions conjugués | 23 |
| 3.2 | Méthode du gradient conjugué : Cas quadratique | 25 |
| 3.3 | Méthode du gradient conjugué : Cas des fonctions non quadratiques | 28 |
| 3.3.1 | Caractérisation des méthodes du GC | 28 |
| 3.3.2 | Algorithmes | 29 |
| 4 | Méthodes du gradient conjugué. Etude de la convergence | 33 |
| 4.1 | Résultats de Convergence du Gradient Conjugué, version Fletcher-Reeves | 36 |
| 4.1.1 | Algorithme 4.1 de la méthode de Fletcher-Reeves | 37 |
| 4.2 | Résultats de Convergence du Gradient Conjugué, version Polak Ribière | 39 |
| 4.2.1 | Algorithme 4.2 de la méthode de Polak-Ribière-Polyak | 39 |
| 4.3 | Résultats de Convergence du Gradient Conjugué, version (Dai-Yuan) | 40 |
| 4.3.1 | Descente de la méthode de Dai-Yuan | 40 |
| 4.3.2 | Convergence de la méthode de Dai-Yuan | 40 |
| 4.3.3 | Algorithme 4.3 de la Méthode de Dai-Yuan avec la règle de Wolfe faible | 41 |
| 5 | Une nouvelle classe de méthodes de gradients conjugués non linéaires | 43 |
| 5.1 | Introduction et présentation de la méthode | 43 |
| 5.2 | Algorithme de la nouvelle méthode du gradient conjugué version Rivaie, Mustafa, Ismail et Lelong | 45 |
| 5.3 | Etude de la convergence | 45 |
| 5.3.1 | Descente suffisante de la nouvelle méthode du gradient conjugué version <i>Rivaie, Mustafa, Ismail et Lelong</i> | 45 |
| 5.3.2 | Convergence globale de la nouvelle méthode du gradient conjugué version <i>Rivaie, Mustafa, Ismail et Lelong</i> | 46 |
| 5.4 | Résultats numériques | 52 |
| 5.5 | Conclusion et perspectives | 52 |
| 5.5.1 | Conclusions | 52 |
| 5.5.2 | Perspectives et problèmes ouverts | 52 |

INTRODUCTION

Soit

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

et (P) le problème d'optimisation sans contraintes suivant :

$$\min \{f(x), x \in \mathbb{R}^n\} \quad (P) \quad (0.1)$$

La méthode du gradient conjugué est l'une des méthodes les plus célèbres et les plus utilisées pour résoudre les problèmes du type (0.1). N'utilisant que les dérivées de premier ordre, elle est surtout utilisée pour les problèmes de grande taille. Cette méthode a été découverte en 1952 par Hestenes et Steifel pour la minimisation de fonctions quadratiques strictement convexes. Plusieurs mathématiciens ont étendu cette méthode pour le cas non linéaire, non quadratique. Ceci a été réalisé pour la première fois, en 1964, par Fletcher et Reeves, puis en 1969 par Polak, Ribière et Poyak. Une autre variante a été étudiée en 1987 par Fletcher.

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et (P) le problème de minimisation non linéaire (0.1), sans contraintes. Les différentes méthodes du gradient conjugué génèrent une suite $\{x_k\}_{k \in \mathbb{N}}$ de la façon suivante :

$$x_{k+1} = x_k + \lambda_k d_k \quad (0.2)$$

Le pas $\lambda_k \in \mathbb{R}$ est déterminé par une optimisation unidimensionnelle ou recherche linéaire exacte ou inexacte. On s'intéressera particulièrement dans ce mémoire aux recherches linéaires inexactes du type armijoo.

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

$$d_k = \begin{cases} -g_1 & \text{si } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 2 \end{cases} \quad (0.3)$$

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

Les différentes valeurs attribuées à β_k définissent les différentes formes du gradient conjugué. Si on note $y_{k-1} = g_k - g_{k-1}$, on obtient les variantes suivantes :

$$\beta_k^{PRP} = \frac{g_k^\top y_k}{\|g_{k-1}\|^2} \quad \text{Gradient conjugué variante Polak-Ribière-Polyak}$$

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad \text{Gradient conjugué variante Fletcher-Reeves}$$

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^\top g_{k-1}} \quad \text{Gradient conjugué variante descente conjugué}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^\top y_{k-1}} \quad \text{Gradient conjugué variante Dai-Yuan.}$$

Dans le cas, où f est une fonction quadratique avec une recherche linéaire exacte toutes ces variantes de β_k ont la même valeur :

$$\beta_k^{PRP} = \beta_k^{FR} = \beta_k^{CD} = \beta_k^{DY}.$$

Si f est quelconque, il n'en est plus de même et on parle respectivement de méthode de Polak-Ribière-Polyak ([36,1969])-([37,1969]), méthode de Fletcher-Reeves, méthode de descente conjugué, méthode de Dai-Yuan, selon que l'on utilise β_k^{PRP} , β_k^{FR} , β_k^{CD} ou β_k^{DY} à la place de β_k dans (0.3).

Rappelons que les premiers résultats de convergence de la méthode du gradient conjugué ont d'abord été établis avec des recherches linéaires inexactes de wolfe fortes, c'est à dire que les pas λ_k dans (0.2), doivent vérifier aussi les deux relations suivantes :

$$f(x + \lambda_k d_k) \leq f(x_k) + \rho \lambda_k \nabla f(x_k)^\top d_k \quad (0.4)$$

$$|\nabla f(x + \lambda_k d_k)^\top \cdot d_k| \leq -\sigma \nabla f(x_k)^\top d_k \quad (0.5)$$

avec

$$0 < \rho < \sigma < 1$$

Ces résultats furent par la suite généralisés en considérant des recherches linéaires inexacts du type Wolfe faibles, c'est à dire que les pas λ_k dans (0.2), doivent vérifier les deux relations plus faibles que (0.4) et (0.5), suivantes :

$$f(x + \lambda_k d_k) \leq f(x_k) + \rho \lambda_k \nabla f(x_k)^T d_k \quad (0.6)$$

$$\nabla f(x + \lambda_k d_k)^T \cdot d_k \geq \sigma \nabla f(x_k)^T d_k \quad (0.7)$$

Dans ce travail, on présente une nouvelle expression de β_k qu'on note β_k^{RMIL} , étudiée et développée par Rivaie, Mustafa, Ismail, Lelong en 2012 ([]). L'expression de β_k^{RMIL} est la suivante :

$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2} \quad (\text{gradient conjugué version Rivaie, Mustafa, Ismail et Lelong} [])$$

Les auteurs ont utilisé des recherches linéaires exactes et ont montré que les directions qu'engendrent cette nouvelle méthode sont des directions de descente suffisante. Ils ont aussi montré la convergence globale de cette méthode. Ils ont terminé leur travail par des tests numériques.

On présentera à la fin du mémoire un problème ouvert et des conjectures.

Le mémoire est divisé en quatre chapitres :

Le premier chapitre est un rappel de quelques notions préliminaires, on insiste surtout sur les conditions d'optimalité des problèmes de minimisation sans contrainte.

Le second chapitre traite l'optimisation unidimensionnelle et les notions de recherche linéaire exacte et inexacte (Armijo, Goldstein et Wolfe).

Le chapitre 3 est consacré à l'étude de la méthode du gradient conjugué dans le cas linéaire et non linéaire. On va aborder d'abord le principe général d'une méthode à directions conjuguées et le théorème fondamental qui garantit la convergence d'un algorithme à directions conjuguées appliqué à une fonction quadratique à n variables. On s'intéresse ensuite à la méthode du gradient conjugué dans le cas linéaire. On passe enfin à la méthode du gradient conjugué dans le cas non linéaire.

Dans le quatrième chapitre on fait une synthèse des résultats de convergence des différentes méthodes du gradient conjugué.

Le cinquième chapitre est consacré à la nouvelle méthode du gradient conjugué étudiée et développée par Rivaie, Mustafa, Ismail, Lelong en 2012 ([22]).

On finit le mémoire par une conclusion, un problème ouvert est des remarques sur l'article de Rivaie, Mustafa, Ismail, Lelong ([22]).

Chapitre 1

OPTIMISATION SANS CONTRAINTES

1.1 Définitions

Définition 1.1 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, on appelle problème de minimisation sans contraintes le problème (P) suivant :

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

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

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

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

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

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

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

1.2 Direction de descente

Définition 1.2 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\hat{x} \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ est dite direction de descente au point \hat{x} si et seulement si il existe $\delta > 0$ tel que

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

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

Théorème 1.1 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $\hat{x} \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ une direction vérifiant la condition suivante :

$$f'(\hat{x}, d) = \nabla f(\hat{x})^t \cdot d < 0.$$

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

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

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

avec

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

ceci implique que

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

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

$$\frac{f(\hat{x} + \lambda d) - f(\hat{x})}{\lambda} < 0, \quad \forall \lambda \in]-\delta, +\delta[. \quad (1.1)$$

La relation (1.1) est particulièrement vraie pour tout $\lambda \in]0, +\delta[$. On obtient le résultat cherché en multipliant la relation (4.1) par $\lambda > 0$. ■

1.3 Schémas général des algorithmes d'optimisation sans contraintes

Supposons que d_k soit une direction de descente au point x_k . Ceci nous permet de considérer le point x_{k+1} , successeur de x_k , de la manière suivante :

$$x_{k+1} = x_k + \lambda_k d_k, \quad \lambda_k \in]0, +\delta[.$$

Vu la définition de direction de descente, on est assuré que

$$f(x_{k+1}) = f(x_k + \lambda_k d_k) < f(x_k).$$

Un bon choix de d_k et de λ_k permet ainsi de construire une multitude d'algorithmes d'optimisation.

1.3.1 Exemples de choix de directions de descente

Par exemple si on choisit $d_k = -\nabla f(x_k)$ et si $\nabla f(x_k) \neq 0$, on obtient la méthode du gradient. La méthode de Newton correspond à $d_k = -(H(x_k))^{-1} \cdot \nabla f(x_k)$. Bien sur $-\nabla f(x_k)$ est une direction de descente ($\nabla f(x_k)^t \cdot d_k = -\nabla f(x_k)^t \cdot \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0$). Pour la deuxième direction si la matrice hessienne $H(x_k)$ est définie positive alors $d_k = -(H(x_k))^{-1} \cdot \nabla f(x_k)$ est aussi une direction de descente.

1.3.2 Exemple de choix de pas λ_k

On choisit en général λ_k de façon optimale, c'est à dire que λ_k doit vérifier

$$f(x_k + \lambda_k d_k) \leq f(x_k + \lambda d_k) : \quad \forall \lambda \in [0, +\infty[.$$

En d'autres termes on est ramené à étudier à chaque itération un problème de minimisation d'une variable réelle. C'est ce qu'on appelle recherche linéaire.

1.4 Conditions nécessaires d'optimalité

1.4.1 Condition nécessaire d'optimalité du premier ordre

Théorème 1.2 *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $\hat{x} \in \mathbb{R}^n$. Si \hat{x} est un minimum local de (P) alors $\nabla f(\hat{x}) = 0$.*

Preuve : C'est une conséquence directe du théorème 4.1 et de la remarque 4.1. En effet, supposons que $\nabla f(\hat{x}) \neq 0$. Puisque la direction $d = -\nabla f(\hat{x})$ est une direction de descente, alors il existe $\delta > 0$ tel que :

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

Ceci est contradiction avec le fait que \hat{x} est une solution optimale locale de (P) . ■

1.4.2 Condition nécessaire d'optimalité du second ordre

Théorème 1.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable au point $\hat{x} \in \mathbb{R}^n$. Si \hat{x} est un minimum local de (P) alors $\nabla f(\hat{x}) = 0$ et la matrice hessienne de f au point \hat{x} , qu'on note $H(\hat{x})$, est semi définie positive.

Preuve : Soit $x \in \mathbb{R}^n$ quelconque, f étant deux fois différentiable au point \hat{x} , on aura pour tout $\lambda \neq 0$

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

Ceci implique

$$\frac{f(\hat{x} + \lambda x) - f(\hat{x})}{\lambda^2} = \frac{1}{2} x^t H(\hat{x}) x + \|x\|^2 \alpha(\hat{x}, \lambda x). \quad (1.2)$$

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

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

Si on prend en considération (1.2) et on passe à la limite quand $\lambda \rightarrow 0$, $\lambda \neq 0$, on obtient

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

1.5 Conditions suffisantes d'optimalité

Théorème 1.4 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable au point $\hat{x} \in \mathbb{R}^n$. Si $\nabla f(\hat{x}) = 0$ et $H(\hat{x})$ est définie positive alors \hat{x} est un minimum local strict de (P) .

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

$$f(x) = f(\hat{x}) + \frac{1}{2} (x - \hat{x})^t H(\hat{x}) (x - \hat{x}) + \|(x - \hat{x})\|^2 \alpha(\hat{x}, (x - \hat{x})), \quad (1.3)$$

$$\alpha(\hat{x}, (x - \hat{x})) \xrightarrow{x \rightarrow \hat{x}} 0, \quad (\nabla f(\hat{x}) = 0).$$

Supposons que \hat{x} n'est pas un optimum local strict. Alors il existe une suite $\{x_k\}_{k \in \mathbb{NN}^*}$ telle que $x_k \neq \hat{x} : \forall k$ et

$$x_k \neq \hat{x} : \forall k, \quad x_k \xrightarrow[k \rightarrow \infty]{} \hat{x} \quad \text{et} \quad f(x_k) \leq f(\hat{x}). \quad (1.4)$$

Dans (1.3) prenons $x = x_k$, divisons le tout par $\|(x_k - \hat{x})\|^2$ et notons $d_k = \frac{(x_k - \hat{x})}{\|(x_k - \hat{x})\|}$, on obtient

$$\frac{f(x_k) - f(\hat{x})}{\|(x_k - \hat{x})\|^2} = \frac{1}{2} d_k^t H(\hat{x}) d_k + \alpha(\hat{x}, (x_k - \hat{x})), \quad \alpha(\hat{x}, (x_k - \hat{x})) \xrightarrow[k \rightarrow \infty]{} 0. \quad (1.5)$$

(1.4) et (1.5) impliquent

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

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

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

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

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

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

Chapitre 2

Optimisation unidimensionnelle ou recherche linéaire

2.1 Position du Problème

Soit : $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et (P) le problème d'optimisation sans contraintes suivant :

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

Comme on l'a souligné dans les chapitres précédents, la majorité des algorithmes d'optimisation non linéaire procèdent comme suit. Etant donné un point x_k , on cherche une certaine direction $d_k \in \mathbb{R}^n$ et un pas $\lambda_k \in \mathbb{R}$. Ceci nous permet de trouver le successeur x_{k+1} de x_k par la formule : $x_{k+1} = x_k + \lambda_k d_k$ et le processus est ainsi répété.

Le pas λ_k est une solution optimale d'un problème de minimisation unidimensionnelle, c'est à dire la minimisation d'une fonction d'une variable réelle et à valeurs réelles de la forme suivante :

$$R.L. \quad \{ \text{Minimiser } \varphi(\lambda) = f(x_k + \lambda d_k), \lambda \in I, \quad I \subset \mathbb{R}^+, \quad I \text{ fermé, borné ou non} \}$$

Donc pour les problèmes de minimisation sans contraintes, on résout à chaque itération des sous problèmes de minimisation de fonctions d'une variable réelle. C'est ce qu'on appelle recherche linéaire. On donne dans cette partie un aperçu sur les recherches linéaires qu'on utilisera plus tard. On les a classées en deux catégories :

Les recherches linéaires exactes. Dans ce cas la solution optimale λ_k est calculée de façon exacte (d'un point théorique car pratiquement on

n'obtient en général qu'une approximation). On donnera en détail le principe et l'algorithme de la recherche linéaire du nombre d'or.

Les recherches linéaires inexactes. Dans ce cas la solution optimale λ_k est calculée de façon qu'elle vérifie des inégalités, assurant la descente de la fonction économique f . Il s'agit des recherches linéaires de Goldstein-Wolf et celle d'Armijo. Le principe et l'algorithme des recherches linéaires de Goldstein de Wolfe et d'Armijo seront détaillées.

2.2 Recherches linéaires exactes

2.2.1 L'intervalle d'incertitude

Définition 2.1 *Considérons le problème unidimensionnel suivant :*

$$\min \{ \varphi(\lambda) : \lambda \in [a, b] \}$$

L'intervalle $[a, b]$ est dit intervalle d'incertitude si le minimum $\bar{\lambda}$ de $\varphi(\lambda)$ appartient à $[a, b]$, mais sa valeur exacte n'est pas connue.

Théorème 2.1 ([27]) *Soit $\varphi : R \rightarrow R$ strictement quasi-convexe sur $[a, b]$. Soient $\lambda, \mu \in [a, b]$ tels que $\lambda < \mu$.*

- 1- *Si $\varphi(\lambda) > \varphi(\mu)$, alors $\varphi(z) \geq \varphi(\mu)$; $\forall z \in [a, \lambda[$.*
- 2- *Si $\varphi(\lambda) \leq \varphi(\mu)$, alors $\varphi(z) \geq \varphi(\lambda)$; $\forall z \in]\mu, b]$.*

Conséquence importante du théorème 2.1. Le théorème 2.1 nous permet après les deux observations $\varphi(\lambda)$ et $\varphi(\mu)$ de diminuer l'intervalle d'incertitude. En effet :

Si la condition 1) est vérifiée, le nouveau intervalle d'incertitude sera l'intervalle : $[\lambda, b]$.

Si la condition 2) est vérifiée, le nouveau intervalle d'incertitude sera l'intervalle : $[a, \mu]$.

Ceci est l'idée de base pour la construction d'algorithmes d'optimisation unidimensionnelle sans calcul de dérivées. A chaque itération on fait diminuer l'intervalle d'incertitude jusqu'à ce qu'on arrive à un intervalle final de longueur inférieure à une tolérance fixée à l'avance. Bien sur une valeur quelconque de ce dernier intervalle conviendrait comme approximation de notre solution optimale.

2.2.2 Deux méthodes d'optimisation unidimensionnelle sans calcul de dérivées. La méthode de dichotomie et la méthode du nombre d'or (golden section).

La méthode de dichotomie.

Algorithme de la méthode

Initialisation : Choisir $2\varepsilon > 0$ et choisir la longueur finale de l'intervalle d'incertitude l . $[a_1, b_1]$ étant l'intervalle initial.

Poser $k = 1$ et aller à l'étape 1.

Etape 1 :

Si $b_k - a_k < l$ stop. Le minimum appartient à $[a_k, b_k]$.

Sinon poser :

$$\begin{aligned}\lambda_k &= \frac{a_k + b_k}{2} - \varepsilon \\ \mu_k &= \frac{a_k + b_k}{2} + \varepsilon\end{aligned}$$

et aller à l'étape 2

Etape 2 :

Si $\varphi(\lambda_k) > \varphi(\mu_k)$ poser $a_{k+1} = a_k, b_{k+1} = \mu_k$

Sinon $a_{k+1} = \lambda_k, b_{k+1} = b_k$. Remplacer k par $k + 1$ et aller à l'étape 1.

La méthode du nombre d'or.

La méthode du nombre d'or améliore la méthode de dichotomie, en diminuant le nombre d'observations, à chaque itération.

A l'itération k supposons que l'on a l'intervalle $[a_k, b_k]$ avec les points λ_k et μ_k . Le théorème 4.1 nous indique comment obtenir le nouveau intervalle d'incertitude $[a_{k+1}, b_{k+1}]$.

Si $\varphi(\lambda_k) > \varphi(\mu_k) : [a_{k+1}, b_{k+1}] = [\lambda_k, b_k]$.

Si $\varphi(\lambda_k) \leq \varphi(\mu_k) : [a_{k+1}, b_{k+1}] = [a_k, \mu_k]$.

Comment choisir les points λ_{k+1} et μ_{k+1} de façon à avoir le moins d'observations possibles.

Choix des points λ_{k+1} et μ_{k+1} : Les points λ_{k+1} et μ_{k+1} seront choisis de sorte qu'ils vérifient les conditions suivantes :

a) La longueur de $[a_{k+1}, b_{k+1}]$ ne dépend pas de l'itération, c'est à dire que dans les deux cas d'observations $\varphi(\lambda_k) > \varphi(\mu_k)$ ou $\varphi(\lambda_k) \leq \varphi(\mu_k)$ on a :

$$b_k - \lambda_k = \mu_k - a_k$$

Plus précisément si λ_k est de la forme

$$\lambda_k = a_k + (1 - \alpha)(b_k - a_k); \quad \alpha \in]0, 1[$$

Alors μ_k s'écrit :

$$\mu_k = a_k + \alpha(b_k - a_k)$$

Ces deux relations donnent :

$$b_{k+1} - a_{k+1} = \alpha(b_k - a_k).$$

b) Le choix de λ_{k+1} et μ_{k+1} se fait de façon que ou bien λ_{k+1} coïncide avec μ_k ou bien μ_{k+1} coïncide avec λ_k , ou de façon plus explicite :

Cas 1 Si $\varphi(\lambda_k) > \varphi(\mu_k)$, on choisira

$$a_{k+1} = \lambda_k, \quad b_{k+1} = b_k, \quad \lambda_{k+1} = \mu_k$$

et

$$\mu_{k+1} = a_{k+1} + \alpha(b_{k+1} - a_{k+1})$$

Cas 2 Si $\varphi(\lambda_k) \leq \varphi(\mu_k)$, on choisira

$$a_{k+1} = a_k, \quad b_{k+1} = \mu_k, \quad \mu_{k+1} = \lambda_k$$

et

$$\lambda_{k+1} = a_{k+1} + (1 - \alpha)(b_{k+1} - a_{k+1})$$

Remarque : Dans le cas 1 on avait $\lambda_{k+1} = \mu_k$, et dans le cas 2 : $\mu_{k+1} = \lambda_k$. On ne fera qu'une seule observation à chaque itération puisque l'autre observation a été faite à l'itération précédente.

Calcul du coefficient α Il n'est difficile en prenant en considération les relations précédentes de voir que α est nécessairement solution de l'équation du second degré suivante

$$\alpha^2 + \alpha - 1 = 0$$

dont les solutions approximatives sont $\alpha \simeq 0.618$ ou $\alpha \simeq -1.618$. Vu que $\alpha \in [0, 1]$. Donc $\alpha \simeq 0.618$.

Algorithme de la méthode du nombre d'or.

Etape initiale :

Choisir $l > 0$ la longueur finale de l'intervalle d'incertitude et $[a_1, b_1]$ l'intervalle initial. Calculer :

$$\begin{aligned}\lambda_1 &= a_1 + (1 - \alpha)(b_1 - a_1) \\ \mu_1 &= a_1 + \alpha(b_1 - a_1) \\ \text{avec } \alpha &= 0.618\end{aligned}$$

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

Etape principale :

1- Si $b_k - a_k < l$ stop. La solution optimale appartient à $[a_k, b_k]$.

Sinon

Si $\varphi(\lambda_k) > \varphi(\mu_k)$ aller à 2, et si $\varphi(\lambda_k) \leq \varphi(\mu_k)$ aller à 3.

2- Poser $a_{k+1} = \lambda_k$ et $b_{k+1} = b_k$, $\lambda_{k+1} = \mu_k$ et $\mu_{k+1} = a_{k+1} + \alpha(b_{k+1} - a_{k+1})$.

Evaluer $\varphi(\mu_{k+1})$ et aller à 4.

3- Poser $a_{k+1} = a_k$ et $b_{k+1} = \mu_k$ et $\mu_{k+1} = \lambda_k$ et $\lambda_{k+1} = a_{k+1} + (1 - \alpha)(b_{k+1} - a_{k+1})$.

Evaluer $\varphi(\lambda_{k+1})$ et aller à 4.

4- Remplacer k par $k + 1$ et aller à 1.

2.3 recherches linéaires inexactes ou économiques

2.3.1 Introduction.

Les recherches linéaires exactes, malgré qu'elles n'aboutissent qu'à une solution optimale avec une tolérance fixée à l'avance, elles nécessitent beaucoup d'observations à chaque itération de l'algorithme principal. Des mathématiciens ([12], [14], [20]) ont réussi dans les années 60,70 et 80 à élaborer des recherches linéaires qui demandent peu d'observations, mais respectent en même temps la descente de la fonction économique. On a même réussi à avoir une vitesse de convergence superlinéaire en les appliquant à des méthode quasi-Newtoniennes, comme la B.F.G.S ([28]).

2.3.2 Recherche linéaire d'Armijo(1966).

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\hat{x} \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ telle que $\nabla f(\hat{x})^t \cdot d < 0$.

Définissons les fonctions d'une variable réelle θ , θ_1 et $\hat{\theta}$ de la façon suivante :

$$\theta : \mathbb{R} \rightarrow \mathbb{R} : \theta(\lambda) = f(\hat{x} + \lambda d), \quad \lambda \geq 0$$

$$\theta_1 : \mathbb{R} \rightarrow \mathbb{R} : \theta_1(\lambda) = \theta(0) + \lambda \theta'(0), \quad \lambda \geq 0$$

$$\hat{\theta} : \mathbb{R} \rightarrow \mathbb{R} : \hat{\theta}(\lambda) = \theta(0) + \lambda \cdot \varepsilon \cdot \theta'(0), \quad \lambda \geq 0, \quad \varepsilon = 0.2$$

Remarquons que

$$\theta(0) = f(\hat{x})$$

et que

$$\theta'(0) = f'(\hat{x}, d) = \nabla f(\hat{x})^t \cdot d < 0.$$

$\bar{\lambda} > 0$ est acceptable s' il vérifie :

$$\mathbf{1-} \theta(\bar{\lambda}) \leq \hat{\theta}(\bar{\lambda}).$$

Cette inégalité assure la descente de la fonction économique.

En effet, l'inégalité **1** est équivalente à

$$f(\hat{x} + \bar{\lambda}d) \leq f(\hat{x}) + \bar{\lambda} \varepsilon \nabla f(\hat{x})^t \cdot d,$$

ceci implique nécessairement que

$$f(\hat{x} + \bar{\lambda}d) < f(\hat{x}).$$

$$\mathbf{2-} \theta(\alpha \bar{\lambda}) > \hat{\theta}(\alpha \bar{\lambda}).$$

Cette inégalité assure que la valeur de $\bar{\lambda}$ ne soit pas très petite car si tel est le cas $f(\hat{x} + \bar{\lambda}d)$ devient très proche de $f(\hat{x})$ et cela va influencer sur la vitesse de convergence de l'algorithme.

Algorithme de la recherche linéaire d'Armijo

Etape initiale : choisir un point initial λ_1 , et un $\varepsilon \in (0, 1)$

Etapes principales :

Etape1 : Si $\theta(\lambda_1) \leq \theta(0) + \lambda_1 \varepsilon \theta'(0)$, aller à l'étape 2

sinon

aller à l'étape 3

Etape2 : Soit t le plus petit entier positif avec $\lambda = \lambda_1 2^t$ qui vérifie

$$\theta(\lambda_1 2^t) < \theta(0) + \varepsilon \lambda_1 2^t \theta'(0)$$

Etape3 : Soit t le plus petit entier positif avec $\lambda = \lambda_1/2^t$ qui vérifie

$$\theta(\lambda_1/2^t) < \theta(0) + \varepsilon(\lambda_1/2^t)\theta'(0)$$

Programme en Fortran 90 de la recherche linéaire d'Armijo

```
!-----  
program armijo  
!-----  
implicit none  
real eps,l  
integer t  
print*, 'entrer eps'  
read*, eps  
print*, 'entrer l'  
read*, l  
t=0  
if(f(l)<=f(0.)+eps*l*df(0.))then  
do  
if(f(l*2**t)>f(0.)+eps*(l*2**t)*df(0.))exit  
t=t+1  
enddo  
l=l*2**(t-1)  
else  
t=t+1  
do  
if(f(l/2**t)<f(0.)+eps*(l/2**t)*df(0.))exit  
t=t+1  
enddo  
l=l/2**t  
endif  
print*, 'la longueur du pas l=', l  
contains  
function f(x)  
real f,x  
f=...  
endfunction
```

```

function df(x)
real df, x
df=...
endfunction
end

```

2.3.3 Recherches linéaires de Goldstein et de Wolf

La recherche linéaire de Goldstein ([29]) s'applique lorsque le gradient de la fonction ne peut être évalué (ou trop coûteux à obtenir).

La recherche linéaire de Wolfe (Wolfe ([30]) et Powell ([28])), nécessite l'évaluation du gradient chaque fois que la fonction est calculée.

Le principe commun à toutes ces méthodes est le suivant :

i) λ ne doit pas être choisi trop grand (sinon l'algorithme risque d'avoir un comportement oscillatoire).

ii) λ ne doit pas être choisi trop petit (sinon l'algorithme risque de converger prématurément).

Dans la règle de goldstein la condition **i)** est vérifiée si on a :

$$\theta(\lambda) \leq \theta(0) + c_1\theta'(0) \quad 2.1$$

où c_1 est un coefficient choisi dans $]0, 1[$; et l'on s'assure de la condition **ii)** si la relation suivante est vérifiée :

$$\theta(\lambda) \geq \theta(0) + c_2\theta'(0) \quad (2.2)$$

où c_2 est un coefficient choisi dans $]c_1, 1[$.

Notons que la condition (2.2) assure que le nouveau point $x_{k+1} = x_k + \lambda d_k$ obtenu vérifie $f(x_{k+1}) < f(x_k)$ (propriété de descente).

Dans la règle de Wolf, c'est encore la relation (2.2) qui permet de s'assurer que λ n'est pas choisi trop grand. Mais comme on suppose que le calcul du gradient de la fonction f ne nécessite pas beaucoup plus de calcul que l'évaluation de la fonction elle-même, la relation (2.2) peut être remplacée par la condition :

$$\theta'(\alpha) \geq c_3\theta'(0) \quad (2.3)$$

où c_3 est un coefficient choisi dans $]c_1, 1[$. Donc la procédure est la suivante :

2.3.4 Algorithme des recherches linéaires de Goldstein et de Wolfe

Etape initiale :

On dispose de $\lambda_g = 0$, $\lambda_d =$ une valeur maximale quelconque

Poser $c_1 = 0.1$, $c_2 = 0.7$.

Etapes principales :

Etape 1. Calculer $\theta(\lambda) = f(x_k + \lambda d_k)$

Si $\theta(\lambda) \leq \theta(0) + m\lambda\theta'(0)$ alors aller à l'étape 2

Sinon prendre $\lambda_d = \lambda$ et aller à l'étape 4.

Etape 2. (Goldstein) Si $\theta(\lambda) \geq \theta(0) + c_2\lambda\theta'(0)$ stop

Sinon aller à l'étape 3.

(Wolfe) Calculer $\theta'(\lambda) = \nabla f(x_k + \lambda d_k)^t d_k$

Si $\theta'(\lambda) \geq c_2\theta'(0)$ stop

Sinon aller à l'étape 3.

Etape 3. Poser $\lambda_g = \lambda$

Etape 4. Rechercher un nouveau $\lambda \in]\lambda_g, \lambda_d[$ et retourner à l'étape 2.

Programme en Fortran 90 de la recherche linéaire de Goldstein

```
!-----  
program Goldstein  
!-----  
implicit none  
real l0,lg,ld ,w,b  
b=0.7 ;w=0.1  
lg=0 ;ld=100 ;l0=1  
do  
if(f(l0)<=f(0.)+w*10*df(0.))then  
if(f(l0)>=f(0.)+b*10*df(0.))then  
exit  
else  
lg=l0 ;l0=(lg+ld)/2.  
endif  
else  
ld=l0 ;l0=(lg+ld)/2.  
endif  
enddo  
print*,'la longueur du pas l0=',l0
```

```
contains
function f(x)
real f,x
f=...
endfunction
function df(x)
real df, x
df=...
endfunction
end
```

Remarque :

Le programme précédent est celui de Goldstein, si on veut utiliser la condition de Wolf il suffit de remplacer la condition $\text{if}(f(10) \geq f(0.) + b \cdot 10 \cdot df(0.))$, par $\text{if}(df(10) \geq b \cdot 10 \cdot df(0.))$.

Chapitre 3

Méthodes du gradient conjugué. Etude algébrique

3.1 Méthode des directions conjuguées : Cas quadratique

3.1.1 Aperçu général

La méthode des directions conjuguées et spécialement les méthodes du gradient conjugué (GC) peut être considérée comme une méthode entre la méthode de la plus grande pente (steepest descent) et la méthode de Newton. Cette méthode est motivée par le fait qu'elle essaye d'accélérer la convergence de la méthode de la plus grande pente, qui est une méthode lente et de surmonter le calcul et le stockage de la matrice hessienne comme dans le cas de la méthode de Newton.

La méthode des directions conjuguées a été inventée et analysée pour les problèmes quadratiques de type :

$$\text{Minimiser } c^t x + \frac{1}{2} x^t H x, \quad x \in \mathbb{R}^n$$

où H est une matrice d'ordre $(n \times n)$ symétrique et définie positive. Ces techniques ont été par la suite étudiées au cas général non quadratique par approximation.

Spécialement c'est en 1952 et en 1956 que R. Hestenes et E. Stiefel ([18, 19]) ont proposé cette méthode pour des problèmes d'algèbre linéaire.

C'est R. Fletcher et M. Reeves ([12]) qui en 1964 ont généralisé la méthode du gradient conjugué au cas général non quadratique.

Les méthodes du gradient conjugué minimisent la fonction objective, d'abord suivant la direction de la plus grande pente $-\nabla f(x_1)$ à partir d'un point de départ donné x_1 puis elles construisent des directions qui ont la propriété suivante : *la nouvelle direction est une combinaison linéaire de la direction de la plus grande pente et la direction précédente.*

Il y a trois propriétés significatives pour ces méthodes : la première est la convergence après n itérations au plus pour les fonctions quadratiques (la matrice hessienne est définie positive), la deuxième propriété est la simplicité de la formule par laquelle on calcule la nouvelle direction et enfin la troisième propriété concerne sa progression uniforme car elles se basent sur le gradient, par contre pour des directions conjuguées quelconques, la progression se trouve probablement dans les dernières itérations.

3.1.2 Définition et propriétés des directions conjugués

Définition 3.1. Soit H une matrice symétrique. On dit que les vecteurs d_1, d_2, \dots, d_k sont H -conjugués si :

$$\begin{cases} d_i^t H d_j = 0 & \text{pour } i \neq j \text{ et } i, j \leq k \\ d_1, d_2, \dots, d_k & \text{sont linéairement indépendantes.} \end{cases}$$

Théorème 3.1. Soit $f(x) = c^t x + \frac{1}{2} x^t H x$ où H est une matrice carrée d'ordre $(n \times n)$ et $x \in \mathbb{R}^n$. Soient d_1, d_2, \dots, d_n , H -conjugués. Soit x_1 un point arbitraire de départ. Pour $k = 1, 2, \dots, n$, on suppose que α_k est la solution optimale du problème suivant :

$$\text{Minimiser } f(x_k + \alpha d_k), \alpha \geq 0.$$

Si on pose $x_{k+1} = x_k + \alpha_k d_k$, alors pour $k = 1, 2, \dots, n$, on a :

(i) $\nabla f(x_{k+1})^t d_j = 0, j = 1, 2, \dots, k$

(ii) $\nabla f(x_1)^t d_k = \nabla f(x_k)^t d_k$

(iii) x_{k+1} est la solution optimale du problème suivant : Minimiser $f(x)$, $x - x_1 \in L(d_1, d_2, \dots, d_k)$ où

$$\text{Minimiser } f(x), x - x_1 \in L(d_1, d_2, \dots, d_k) \text{ où}$$

$$L(d_1, d_2, \dots, d_k) = \left\{ \sum_{j=1}^k \mu_j d_j, \mu_j \in \mathbb{R}, \forall j \right\}.$$

En particulier, x_{k+1} minimise la fonction f sur \mathbb{R}^n .

Preuve : On note que $f(x_j + \alpha d_j)$ atteint le minimum α_j si :

$$\nabla f(x_j + \alpha_j d_j)^t d_j = 0 \Rightarrow \nabla f(x_{j+1})^t d_j = 0$$

donc (i) est vérifiée si $j = k$, mais pour $j < k$ on a :

$$\nabla f(x_{k+1}) = c + Hx_{k+1} = c + Hx_{j+1} + H\left(\sum_{i=j+1}^k \alpha_i d_i\right) = \nabla f(x_{j+1}) + H\left(\sum_{i=j+1}^k \alpha_i d_i\right). \quad (3.1)$$

On sait que $d_i^t H d_j = 0$ pour $i = j + 1, \dots, k$, d'où

$$\nabla f(x_{k+1})^t d_j = 0, \quad j < k$$

et par conséquent, (i) est vérifiée.

Pour (ii), si on remplace dans (i) k par $(k - 1)$ et j par 0, on obtient :

$$\nabla f(x_k) = \nabla f(x_1) + H\left(\sum_{i=1}^{k-1} \alpha_i d_i\right) \text{ pour } k \geq 2.$$

En multipliant cette égalité par d_k^t on obtient :

$$d_k^t \nabla f(x_k) = d_k^t \nabla f(x_1) + d_k^t H \left(\sum_{i=1}^{k-1} \alpha_i d_i\right)$$

d'où

$$d_k^t \nabla f(x_k) = d_k^t \nabla f(x_1).$$

Maintenant montrons (iii), on sait que $d_i^t H d_j = 0$ pour $i \neq j$, d'autre part on a :

$$f(x_{k+1}) = f(x_1 + (x_{k+1} - x_1)) = f\left(x_1 + \sum_{j=1}^k \alpha_j d_j\right)$$

d'où

$$f(x_{k+1}) = f(x_1) + \nabla f(x_1)^t \left(\sum_{j=1}^k \alpha_j d_j\right) + \frac{1}{2} \sum_{j=1}^k \alpha_j^2 d_j^t H d_j. \quad (3.2)$$

On suppose que $x - x_1 \in L(d_1, d_2, \dots, d_k)$, alors

$$x = x_1 + \sum_{j=1}^k \mu_j d_j, \quad \mu_j \in \mathbb{R}.$$

De (3.2), on obtient :

$$f(x) = f(x_1) + \nabla f(x_1)^t \left(\sum_{j=1}^k \mu_j d_j \right) + \frac{1}{2} \sum_{j=1}^k \mu_j^2 d_j^t H d_j. \quad (3.3)$$

Il reste à montrer que $f(x) \geq f(x_{k+1})$. Par contre, on suppose que $f(x) < f(x_{k+1})$. De (3.2) et (3.3), on obtient :

$$\nabla f(x_1)^t \left(\sum_{j=1}^k \mu_j d_j \right) + \frac{1}{2} \sum_{j=1}^k \mu_j^2 d_j^t H d_j < \nabla f(x_1)^t \left(\sum_{j=1}^k \alpha_j d_j \right) + \frac{1}{2} \sum_{j=1}^k \alpha_j^2 d_j^t H d_j \quad (3.4)$$

et par définition de α_j on a :

$$f(x_j + \alpha_j d_j) \leq f(x_j + \mu_j d_j), \quad \forall j.$$

De plus on a :

$$f(x_j) + \alpha_j \nabla f(x_j)^t d_j + \frac{1}{2} \alpha_j^2 d_j^t H d_j \leq f(x_j) + \mu_j \nabla f(x_j)^t d_j + \frac{1}{2} \mu_j^2 d_j^t H d_j$$

et comme $\nabla f(x_j)^t d_j = \nabla f(x_1)^t d_j$, on obtient :

$$\alpha_j \nabla f(x_1)^t d_j + \frac{1}{2} \alpha_j^2 d_j^t H d_j \leq \mu_j \nabla f(x_1)^t d_j + \frac{1}{2} \mu_j^2 d_j^t H d_j. \quad (3.5)$$

En sommant (3.5) pour $j = 1, 2, \dots, k$, on obtient la contradiction avec (3.4). Donc x_{k+1} minimise $f(x)$ sur le suos espace qui contient x_1 et engendré par d_1, d_2, \dots, d_k . En particulier, comme d_1, d_2, \dots, d_n sont linéairement independantes alors $L(d_1, d_2, \dots, d_n) = \mathbb{R}^n$, d'où x_{n+1} est la solution optimale du problème posé.

3.2 Méthode du gradient conjugué : Cas quadratique

La méthode de Hestenes-Stiefel (HS)

Soient $f(x) = c^t x + \frac{1}{2} x^t H x$ où H est une matrice symétrique d'ordre $(n \times n)$ définie positive et $x \in \mathbb{R}^n$. Comme H est définie positive alors le pas suivant la direction d_j est le suivant :

$$\alpha_j = -g_j^t d_j / d_j^t H d_j \text{ où } g_j = \nabla f(x_j).$$

L'algorithme suivant est l'algorithme de Hestnes et Stiefel (HS) pour la minimisation de la fonction quadratique f .

Algorithme de Hestenes et Stiefel (HS)

Etape 0. Initialisation

Donner x_0, ε et poser $d_0 = -g_0$.

Etape 1. Si $\|g_j\| > \varepsilon$ alors faire

$$\begin{cases} \alpha_j = -g_j^t d_j / d_j^t H d_j \\ x_{j+1} = x_j + \alpha_j d_j \\ g_{j+1} = g_j + \alpha_j H d_j \\ \beta_j = g_{j+1}^t H d_j / d_j^t H d_j \\ d_{j+1} = -g_{j+1} + \beta_j d_j \end{cases}$$

Sinon stop.

Vérification de la méthode

Pour assurer que l'algorithme précédent est un algorithme de direction conjuguée, il suffit de montrer que les d_k soient H -conjugués. D'abord, montrons quelques autres propriétés concernant cet algorithme.

Théorème 3.2. *L'algorithme de HS est un algorithme de direction conjugué. S'il ne s'arrête pas à x_j alors :*

(a) $L(g_0, g_1, \dots, g_j) = L(g_0, H g_0, \dots, H^j g_0)$

(b) $L(d_0, d_1, \dots, d_j) = L(g_0, H g_0, \dots, H^j g_0)$

(c) $d_j^t H d_i = 0$ pour $i \leq j - 1$

(d) $\alpha_j = g_j^t g_j / d_j^t H d_j$

(e) $\beta_j = g_{j+1}^t g_{j+1} / g_j^t g_j$.

Preuve. Pour $j = 0$, (a),(b),(c) sont évidentes. Supposons quelle sont vérifiées pour j et on les montre pour $(j + 1)$. On a par récurrence :

$$g_j, H d_j \in L(g_0, H g_0, \dots, H^{j+1} g_0).$$

Comme

$$g_{j+1} = g_j + \alpha_j H d_j,$$

on obtient :

$$g_{j+1} \in L(g_0, Hg_0, \dots, H^{j+1}g_0).$$

De plus on a :

$$g_{j+1} \notin L(g_0, Hg_0, \dots, H^jg_0) = L(d_0, d_1, \dots, d_j)$$

sinon, $g_{j+1} = 0$ car il est orthogonale à $L(g_0, Hg_0, \dots, H^jg_0)$. Alors

$$L(g_0, Hg_0, \dots, H^{j+1}g_0) = L(g_0, g_1, \dots, g_{j+1}).$$

Ce qui montre (a).

Comme

$$d_{j+1} = -g_{j+1} + \beta_j d_j$$

alors (b) est obtenue immédiatement à partir de (a).

Pour montrer (c), on a :

$$d_{j+1}^t H d_i = -g_{j+1}^t H d_i + \beta_j d_j^t H d_i.$$

Pour $i = j$, on obtient (par la définition de β_j) :

$$d_{j+1}^t H d_j = 0.$$

Pour $i < j$, on a :

$$g_{j+1}^t H d_i = 0$$

car $H d_i \in L(d_0, d_1, \dots, d_{i+1})$ par récurrence et $\beta_j d_j^t H d_i = 0$ (par récurrence sur (c)), d'où

$$d_{j+1}^t H d_i = 0, \text{ pour tout } i \leq j.$$

Ce qui montre (c). Et par conséquent l'algorithme de HS est un algorithme de direction conjuguée.

Pour montrer (d), on a :

$$-g_j^t d_j = g_j^t g_j - \beta_{j-1} g_j^t d_{j-1} = g_j^t g_j \Rightarrow \alpha_j = g_j^t g_j / d_j^t H d_j.$$

Enfin, pour montrer (e) on a :

$$g_{j+1}^t g_j = 0$$

car $g_j \in L(d_0, d_1, \dots, d_j)$, et par conséquent :

$$H d_j = \frac{1}{\alpha_j} (g_{j+1} - g_j) \Rightarrow g_{j+1}^t H d_j = \frac{1}{\alpha_j} g_{j+1}^t g_{j+1} \Rightarrow \beta_j = g_{j+1}^t g_{j+1} / g_j^t g_j. \square$$

3.3 Méthode du gradient conjugué : Cas des fonctions non quadratiques

3.3.1 Caractérisation des méthodes du GC

Etant donné une fonction différentiable $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Les méthodes du gradient conjugué engendrent une suite d'itérés x_j comme suit :

$$x_{j+1} = x_j + \alpha_j d_j \quad (3.6)$$

où d_j est la direction de recherche, α_j est le pas suivant la direction d_j à partir de x_j .

Ces méthodes prennent d_j comme suit :

$$\begin{cases} d_1 &= -\nabla f(x_1) \\ d_{j+1} &= -\nabla f(x_{j+1}) + \beta_j d_j, \quad j \geq 1 \end{cases} \quad (3.7)$$

où β_j est le paramètre qui caractérise les différentes versions des méthodes du gradient conjugué en considérant que les directions générées sont H -conjuguées.

D'abord, supposons que f est une fonction quadratique telle que la matrice hessienne H est définie positive. Si d_{j+1} et d_j sont H -conjuguées, on obtient :

$$d_{j+1}^t H d_j = 0 \Rightarrow d_{j+1}^t H p_j = d_{j+1}^t q_j = 0$$

où $p_j = \alpha_j d_j$ et $q_j = \nabla f(x_{j+1}) - \nabla f(x_j)$. En utilisant (3.11), on obtient :

$$-\nabla f(x_{j+1})^t q_j + \beta_j d_j^t q_j = 0$$

et par conséquent on trouve le choix de Hestenes et Stiefel ([14]) pour β_j (qui peut être aussi utilisé pour les fonctions non quadratiques en considérant l'approximation quadratique de f au voisinage du point x_j) comme suit :

$$\beta_j^{HS} = \frac{\nabla f(x_{j+1})^t q_j}{d_j^t q_j} = \frac{\nabla f(x_{j+1})^t H d_j}{d_j^t H d_j}. \quad (3.8)$$

Dans le cas où on fait la recherche linéaire exacte on a :

$$d_j^t \nabla f(x_{j+1}) = 0 = d_{j-1}^t \nabla f(x_j),$$

ce qui implique :

$$d_j^t q_j = -d_j^t \nabla f(x_j) = (\nabla f(x_j) - \beta_{j-1} d_{j-1})^t \nabla f(x_j) = \|\nabla f(x_j)\|^2.$$

En substituant ceci dans (3.8), on obtient le choix de Polak et Ribiere ([4, 10]) qui est :

$$\beta_j^{PR} = \frac{\nabla f(x_{j+1})^t q_j}{\|\nabla f(x_j)\|^2}. \quad (3.9)$$

De plus, si la fonction est quadratique on obtient :

$$\begin{aligned} \nabla f(x_{j+1})^t \nabla f(x_j) &= \nabla f(x_{j+1})^t (\beta_{j-1} d_{j-1} - d_j) \\ &= \beta_{j-1} \nabla f(x_{j+1})^t d_{j-1} = \beta_{j-1} (\nabla f(x_j) + \alpha_j H d_j)^t d_{j-1}, \end{aligned}$$

d'où

$$\nabla f(x_{j+1})^t \nabla f(x_j) = \alpha_j \beta_{j-1} d_j^t H d_{j-1} = 0.$$

En substituant ceci dans (3.9), on obtient le choix de Fletcher et Reeves ([12]) qui est :

$$\beta_j^{FR} = \frac{\|\nabla f(x_{j+1})\|^2}{\|\nabla f(x_j)\|^2}. \quad (3.10)$$

Et par conséquent si la fonction est quadratique et la recherche linéaire est exacte alors tous les choix précédents sont identiques.

3.3.2 Algorithmes

Citons deux algorithmes, l'un est l'algorithme de l'approximation quadratique et l'autre est l'algorithme de Fletcher et Reeves.

L'algorithme de l'approximation quadratique

Dans l'algorithme de Hestenes et Stiefel (HS), on fait les associations suivantes :

$$g_j \leftrightarrow \nabla f(x_j) \text{ et } H \leftrightarrow H(x_j) = H_j.$$

Etape 0. Initialisation

Donner x_0 et poser $g_0 = \nabla f(x_0)$, $d_0 = -g_0$, $j = 0$.

Etape 1. Poser

$$\begin{aligned} \alpha_j &= -g_j^t d_j / d_j^t H_j d_j \\ x_{j+1} &= x_j + \alpha_j d_j. \end{aligned}$$

Etape 2. Si $j < n - 1$ aller à l'étape 3, sinon aller à l'étape 4.

Etape 3. Poser

$$\beta_j = g_{j+1}^t H_j d_j / d_j^t H_j d_j$$

$$d_{j+1} = -g_{j+1} + \beta_j d_j$$

et remplacer j par $(j + 1)$ et aller à l'étape 1.

Etape 4. Poser

$$x_0 = x_n$$

$$j = 0$$

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

et aller à l'étape 1.

Comme la H -conjugance entre les directions de recherche dépend à la direction initiale qui est $-\nabla f(x_k)$, il est important de redémarrer l'algorithme de l'approximation quadratique chaque n ou $(n + 1)$ itérations par une étape pour laquelle la direction de recherche est moins le gradient, sinon les avantages des directions conjuguées (où l'approximation proche d'être exacte) peuvent être manqués.

La bonne chose dans cet algorithme est le fait qu'il n'exige plus la recherche linéaire, cependant il a l'inconvénient d'être ne pas convergent en générale car $d_j^t H d_j$ peut être négative ou nul.

L'algorithme de Fletcher-Reeves

Cet algorithme utilise la recherche linéaire exacte pour avoir le pas de déplacement.

Etape 0. Initialisation

Choisir $\varepsilon > 0$, donner x_0 et poser $d_0 = -\nabla f(x_0)$, $k = 0$.

Etape 1. Si $\|\nabla f(x_k)\| < \varepsilon$ stop, sinon poser α_k la solution optimale du sous

problème suivant :

$$\text{Minimiser } f(x_k + \alpha d_k), \alpha \geq 0.$$

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

Etape 2. Si $k < n - 1$ alors aller à l'étape 3, sinon aller à l'étape 4.

Etape 3. Poser

$$\beta_k = \|\nabla f(x_{k+1})\|^2 / \|\nabla f(x_k)\|^2$$

$$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$$

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

Etape 4. Poser

$$\begin{aligned}x_0 &= x_n \\d_0 &= -\nabla f(x_0) \\k &= 1 \text{ et aller à l'étape 1.}\end{aligned}$$

L'algorithme de Fletcher et Reeves est globalement convergent (sans la reinitialisation) avec une recherche linéaire non exacte si la condition suivante (du à AL-baali(1985)[1]) est vérifiée :

$$|\nabla f(x_{j+1})^t d_j| \leq -\sigma \nabla f(x_j)^t d_j, \quad \sigma \in]0, \frac{1}{2}] \quad (3.11)$$

Proposition 3.1 ([10] **exercice 4.7 p.94**). *Supposons qu'on a la méthode du GC de Fletcher et Reeves avec une recherche linéaire non exacte qui vérifie (2.11). Si $d_1 = -\nabla f(x_1)$ alors :*

$$-\sum_{i=0}^{j-1} \sigma^i \leq \frac{\nabla f(x_j)^t d_j}{\|\nabla f(x_j)\|^2} \leq -2 + \sum_{i=0}^{j-1} \sigma^i.$$

Preuve. Pour $k = 1$ est triviale. Pour $j = 2$, on a :

$\nabla f(x_2)^t d_2 = \nabla f(x_2)^t (-\nabla f(x_2) + \beta_1 d_1) = -\|\nabla f(x_2)\|^2 + \beta_1 \nabla f(x_2)^t d_1$,
et en utilisant (2.11) on obtient :

$$\begin{aligned}\frac{-\|\nabla f(x_2)\|^2 + \sigma \beta_1 \nabla f(x_1)^t d_1}{\|\nabla f(x_2)\|^2} &\leq \frac{\nabla f(x_2)^t d_2}{\|\nabla f(x_2)\|^2} \leq \frac{-\|\nabla f(x_2)\|^2 - \sigma \beta_1 \nabla f(x_1)^t d_1}{\|\nabla f(x_2)\|^2} \\ \Rightarrow & \\ -1 - \sigma &\leq \frac{\nabla f(x_2)^t d_2}{\|\nabla f(x_2)\|^2} \leq -1 + \sigma \\ \Rightarrow & \\ -\sum_{i=0}^1 \sigma^i &\leq \frac{\nabla f(x_2)^t d_2}{\|\nabla f(x_2)\|^2} \leq -2 + \sum_{i=0}^1 \sigma^i.\end{aligned}$$

Supposons que ceci est vrai pour j et on le montre pour $(j + 1)$.

$$\begin{aligned}\nabla f(x_{j+1})^t d_{j+1} &= \nabla f(x_{j+1})^t (-\nabla f(x_{j+1}) + \beta_j d_j) \\ &= -\|\nabla f(x_{j+1})\|^2 + \beta_j \nabla f(x_{j+1})^t d_j \\ \Rightarrow &\end{aligned}$$

$$\frac{-\|\nabla f(x_{j+1})\|^2 + \sigma \beta_j \nabla f(x_j)^t d_j}{\|\nabla f(x_{j+1})\|^2} \leq \frac{\nabla f(x_{j+1})^t d_{j+1}}{\|\nabla f(x_{j+1})\|^2} \leq \frac{-\|\nabla f(x_{j+1})\|^2 - \sigma \beta_j \nabla f(x_j)^t d_j}{\|\nabla f(x_j)\|^2}$$

\Rightarrow

$$-1 + \sigma \left(-\sum_{i=0}^{j-1} \sigma^i \right) \leq \frac{\nabla f(x_{j+1})^t d_{j+1}}{\|\nabla f(x_{j+1})\|^2} \leq -1 + \sigma \left(\sum_{i=0}^{j-1} \sigma^i \right)$$

\Rightarrow

$$-\sum_{i=0}^j \sigma^i \leq \frac{\nabla f(x_{j+1})^t d_{j+1}}{\|\nabla f(x_{j+1})\|^2} \leq -2 + \sum_{i=0}^j \sigma^i$$

d'où le résultat, et par conséquent la méthode du gradient conjugué de Fletcher et Reeves construit des directions de descente.

Remarques

(a) L'avantage de redémarrage des méthodes du GC à chaque n itérations par un gradient pure est la suivante : "Si les itérés progressent d'une region n'est pas quadratique vers un voisinage de la solution dans lequel l'approximation quadratique de f est bonne alors la méthode converge rapidement.

(b) Dans le cas où la recherche linéaire est exacte, les différents choix de β_j (PR, HS, FR) sont identiques. Cependant pour les fonctions non quadratiques, β_j^{PR} apparaît qu'il est mieux surtout pour les problèmes de grande taille, en particulier Powell(1977) a observé que si la méthode progresse lentement alors $g_{j+1} \simeq g_j$ et par conséquent $\beta_j^{PR} = 0$, ce qui implique que $d_{j+1} \simeq -g_{j+1}$, c.à.d la méthode se redémarre automatiquement. Par contre, $\beta_j^{FR} \simeq 1$ pour la méthode du GC de Fletcher et Reeves.

(c) Dans le cas où la recherche linéaire n'est pas exacte, β_j^{HS} est bon (par construction). Cependant même si la fonction est quadratique, on n'obtient que la conjuguance entre les directions successives.

Chapitre 4

Méthodes du gradient conjugué. Etude de la convergence

On expose dans ce chapitre une synthèse concernant la convergence des différentes variantes du gradient conjugué avec la recherche linéaire inexacte de *Wolfe forte* et de *Wolfe faible*.

Notre problème consiste à minimiser une fonction f de n variables à valeurs réelles

$$\min \{f(x) : x \in \mathbb{R}^n\} \quad (4.1)$$

où f est régulière (continûment différentiable) et g est son gradient. Notons par g_k le gradient de f au point x_k . Rappelons que les différentes variantes du gradient conjugué génèrent une suite $\{x_k\}$ de la forme suivante

$$x_{k+1} = x_k + \lambda_k d_k, \quad (4.2)$$

la direction d_k est définie par la formule de récurrence suivante ($\beta_k \in \mathbb{R}$)

$$d_k = -g_1 \quad \text{si } k = 1, -g_k + \beta_k d_{k-1} \dots \dots \dots \text{si } k \geq 2, \quad (4.3)$$

le pas $\lambda_k \in \mathbb{R}$ étant déterminé par une recherche linéaire. Rappelons les différentes variantes du gradient conjugué qui diffèrent selon les valeurs que prennent les constantes β_k . On y rencontre particulièrement les variantes

suivantes :

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad (\text{Fletcher} - \text{Reeves}) \quad (4.1)$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad (\text{Polak} - \text{Ribière}) \quad (4.2)$$

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}, \quad (\text{Hestenes} - \text{Stiefel}) \quad (4.3)$$

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}, \quad (\text{The Conjugate Descent Method}) \quad (4.4)$$

où $\|\cdot\|$ est la norme Euclidienne. Récemment Dai et yuan ont aussi introduit la forme suivante :

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})}. \quad (4.8)$$

La convergence globale de la méthodes (4.4)-(4.5)-(4.6)-(4.7)-(4.8) a été étudiée par beaucoup d'auteurs. Citons surtout *Al-Baali* ([1, 1985]), *Gilbert et Nocedal* ([13, 1992]), *Hestenes et Stiefel* ([14, 1952]), *Hu et Story* ([15,1991]) et *Zoutendijk* ([26,1970]). Un facteur clé dans l'étude de la convergence globale est comment sélectionner le paramètre λ_k . Le choix le plus naturel de λ_k est de faire une recherche linéaire exacte, c.-à-d poser $\lambda_k = \arg \min_{\lambda \geq 0} f(x_k + \lambda d_k)$. Cependant, quelque peu étonnamment, ce choix naturel pourrait résulter en une séquence non-convergente dans le cas de PR et méthodes de HS, comme montré par *Powell* ([28, 1986]). Inspiré par le travail de *Powell, Gilbert et Nocedal* ([13, 1992]) a mené une analyse élégante et a montré que la méthode PR est globalement convergente si β_k^{PR} est restreint pour être non-négatif et λ_k est déterminé en un pas de la recherche linéaire satisfaisant la condition suffisante de descente :

$$g_k^T d_k < -c \|g_k\|^2, \quad \text{où } c > 0, \quad (4.9)$$

en plus des conditions standard de *Wolfe* :

$$\begin{cases} f(x_k + \lambda_k d_k) \leq f(x_k) + \omega_1 \lambda_k d_k^T g_k, \\ d_k^T g_{k+1} \geq \omega_2 d_k^T g_k. \end{cases} \quad (4.10)$$

ou' $0 < \omega_1 < \omega_2 < 1$.

Récemment, *Dai et Yuan* [4,6] ont montré que la méthode CD et la méthode FR sont globalement convergentes si les conditions de la recherche de linéaire suivantes pour λ_k sont satisfaites :

$$\begin{cases} f(x_k + \lambda_k d_k) \leq f(x_k) + \omega_1 \lambda_k d_k^T g_k, \\ \omega_2' d_k^T g_k \leq d_k^T g_{k+1} \leq -\omega_2'' d_k^T g_k. \end{cases} \quad (4.11)$$

où $0 < \omega_1 < \omega_2' < 1$ et $\omega_2'' > 0.$, pour la méthode CD et de plus $\omega_2' + \omega_2'' \leq 1$ pour la méthode FR.

Par commodité, nous assumons que $g_k \neq 0$ pour tout k .

Nous adoptons la Supposition suivante sur fonction f qui est utilisée communément dans la littérature.

Supposition 4.1.

(i) L'ensemble $\mathcal{L} := \{x \in \mathbb{R}^n; f(x) \leq f(x_1)\}$ est borné; où $x_0 \in \mathbb{R}^n$ est le point initial.

(ii) Sur un voisinage N de \mathcal{L} , la fonction objectif f est continûment différentiable et son gradient est lipschitzien i.e

$$\exists L > 0 \text{ tel que } \|g(x) - g(\tilde{x})\| \leq L \|x - \tilde{x}\|, \forall x, \tilde{x} \in \mathcal{N} \quad (4.12)$$

Remarque 4.1 Ces suppositions impliquent qu'il existe $\gamma > 0$ tel que

$$\|g(x)\| \leq \gamma, \forall x \in \mathcal{L} \quad (4.13)$$

Rappelons maintenant le Théorème suivant obtenu essentiellement par *Zoutendijk* [26] et *Wolfe* [30]. Ce théorème assure la satisfaction de la condition de *Zoutendijk* (voir chapitre3), pour toute méthode du type (4.2)-(4.3), dans laquelle le pas λ_k est déterminé par la règle de *Wolfe* faible (4.10).

Théorème 4.1. [26]

Considérons la suite $(x_k)_k$, définie par (4.2), où d_k est une direction de descente et λ_k vérifie les conditions (4.10). Considérons aussi que la supposition (4.1) soit satisfaite, alors :

$$\sum_{k \geq 1} \cos^2 \theta \|g_k\|^2 < \infty \quad (4.14)$$

est vérifiée, avec

$$\cos \theta_k = -\frac{g_k^T d_k}{\|g_k\| \|d_k\|}.$$

Remarque 4.2

$$(4.14) \iff \sum_{k \geq 1, \|d_k\| \neq 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \quad (4.15)$$

En effet :

D'après la 2^{eme} condition de (4.10) nous avons

$$d_k^T (g_{k+1} - g_k) \geq (\omega_2 - 1) d_k^T g_k.$$

D'autre part

$$\begin{aligned} d_k^T (g_{k+1} - g_k) &\leq \|g_{k+1} - g_k\| \|d_k\| \\ &\leq \lambda_k L \|d_k\|^2, \end{aligned}$$

d'où

$$\lambda_k \geq \left(\frac{\omega_2 - 1}{L} \right) \frac{d_k^T g_k}{\|d_k\|^2}.$$

En remplaçant ceci dans la 1^{ere} condition de on aura :

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \mu \cos^2 \theta_k \|g_k\|^2,$$

où $\mu = \frac{\omega_1(\omega_2 - 1)}{L}$.

Or puisque f est bornée sur \mathcal{N} on a :

$$\sum_{k \geq 1} \cos^2 \theta_k \|g_k\|^2 \leq \infty.$$

Ce qui achève la démonstration. □

4.1 Résultats de Convergence du Gradient Conjugué, version Fletcher-Reeves

Avant d'exposer sans démonstration les différents résultats de convergence, donnons d'abord l'algorithme de la méthode de Fletcher Reeves :

4.1.1 Algorithme 4.1 de la méthode de Fletcher-Reeves

Etape0 : (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 $x_{k+1} = x_k + \lambda_k d_k$ avec : $\lambda_k = \min_{\alpha > 0} f(x_k + \lambda d_k)$
 $d_{k+1} = -g_{k+1} + \beta_{k+1}^{FR} d_k$.

où

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

Le premier résultat de convergence de la méthode du gradient conjugué non linéaire (version *Fletcher Reeves*) avec la recherche linéaire inexacte de *Wolfe forte* :

$$\begin{cases} f(x_k + \lambda_k d_k) \leq f(x_k) + \omega_1 \lambda_k d_k^T g_k, \\ |d_k^T g_{k+1}| \leq -\omega_2 d_k^T g_k. \end{cases} \quad (4.16)$$

où ($\omega_2 < \frac{1}{2}$) a été démontré par *Al-Baali* ([2, 1985]).

- *Touati Ahmed* et *Story* ([24, 1990]) ont généralisé ce résultat pour

$$0 \leq \beta_k \leq \beta_k^{FR}. \quad (4.17)$$

- *Gilbert* et *Nocedal* ([13, 1992]) ont généralisé ce résultat pour

$$|\beta_k| \leq \beta_k^{FR}. \quad (4.18)$$

- Récemment, *Dai et Yuan* ([4, 1996]) a montré que la méthode *FR* es globalement convergente avec la recherche linéaire inexacte de *Wolfe relaxée*.

On donne dans ce paragraphe les principaux résultats de convergence de la méthode du gradient conjugué non linéaire avec la recherche linéaire inexacte de *Wolfe forte* et aussi avec la recherche linéaire inexacte de *Wolfe relaxée*.

proposition 4.1. [13]

On considère que la supposition (4.1) est satisfaite. Considérons une méthode du type (4.2) et (4.3) avec β_k satisfaisant à (4.18) et le pas λ_k vérifiant la règle de Wolfe forte (4.16) où $\omega_2 \in]0, \frac{1}{2}[$.

Alors cette méthode génère des directions de descente. De plus on a :

$$\frac{-1}{1 - \omega_2} \leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\omega_2 - 1}{1 - \omega_2}; \quad k = 1, \dots$$

Théorème 4.2. [13] *On considère que la supposition (4.1) est satisfaite. Toute méthode du type (4.2) et (4.3) dans laquelle β_k vérifie*

$$|\beta_k| \leq \beta_k^{FR}, \quad \forall k \geq 1,$$

et le pas λ_k est déterminé par la règle de Wolfe forte (4.16) avec $0 < \omega_1 < \omega_2 < \frac{1}{2}$ est une méthode de descente ($g_k^t d_k < 0, \forall k \geq 1$) convergente, dans le sens où

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (4.19)$$

Théorème 4.3 [13] *Soit x_1 un point de départ pour lequel la supposition (4.1) soit satisfaite. Considérons une méthode du type (4.2) et (4.3) avec β_k satisfaisant à (4.4). On suppose aussi que*

◆ *Chaque direction d_k vérifie*

$$g_k^T d_k \leq 0. \quad (4.20)$$

◆ *Le pas λ_k est déterminé par la règle de Wolfe relaxée (4.11) avec $\omega''_2 < \infty$, alors :*

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

Remarque 4.3. D'un point de vue théorique, la méthode de *Fletcher-Reeves*

est une bonne méthode car :

- ▶ elle ne nécessite pas de stocker une matrice .
- ▶ elle converge pour une classe assez large de fonctions f , et sa vitesse de convergence est meilleure que celle de la méthode du gradient.

Cependant, en pratique, Il est préférable d'utiliser la méthode de *Polak-Ribière* dont les performances moyennes dépassent de loin celles de la méthode de *Fletcher-Reeves*.

4.2 Résultats de Convergence du Gradient Conjugué, version Polak Ribière

La variante dite de *Polak-Ribière* consiste à définir β_k par la formule (4.5). Cette méthode fut découverte par *Polak, Ribière* ([19, 1969]) et *Polyak* ([20, 1969]). L'algorithme de la méthode de Polak Ribière- Poliak est le suivant :

4.2.1 Algorithme 4.2 de la méthode de Polak-Ribière-Polyak

Etape0 : (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"
Si non aller à l'étape 2.

Etape 2 :

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

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

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

où

$$\beta_{k+1}^{PRP} = \frac{g_{k+1}^T [g_{k+1} - g_k]}{\|g_k\|^2} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}$$

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

Le résultat suivant est du à *Polak et Ribière* ([19, 1969]). Il établit la convergence de cette méthode pour une fonction fortement convexe avec une recherche linéaire exacte.

Proposition 4.2 [19] *Si f est fortement convexe, de classe C^1 avec un gradient lipschitzien, alors la méthode de Polak-Ribière avec recherche linéaire exacte génère une suite $\{x_k\}$ convergeant vers l'unique point x_* réalisant le minimum de f .*

Remarque 4.4 Si f n'est pas convexe, la méthode de *Polak-Ribière-Polyak* peut ne pas converger.

4.3 Résultats de Convergence du Gradient Conjugué, version (Dai-Yuan)

Récemment *Dai* et *Yuan* ([3, 1998]) ont introduit une formule pour β_k sous la forme suivante :

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \quad (4.21)$$

où $y_{k-1} = g_k - g_{k-1}$.

Cette méthode possède plusieurs propriétés, par exemple elle possède la propriété de descente à chaque itération et la convergence, si le pas est déterminé par la règle de (*Wolfe faible, Armijo et Goldstein*), si f est strictement convexe ou régulière.

4.3.1 Descente de la méthode de Dai-Yuan

Théorème 4.4 ([9, 19]) *Soit x_1 un point de départ pour lequel la supposition (4.1) est satisfaite. Considérons une méthode du type (4.2) et (4.3) avec β_k satisfaisant à (4.21).*

◆ *Si f est strictement convexe sur l'ensemble convexe \mathcal{L} c-à-d :*

$$(g(x) - g(y))^T(x - y) > 0, \quad \text{pour tout } x, y \in \mathcal{L}. \quad (4.22)$$

Alors pour tout $k \geq 1$:

$$g_k^T d_k < 0 \quad (4.23)$$

En 1999 ils s'ont généralisé ce résultat pour toute fonction régulière avec la recherche de *Wolfe faible* (4.10).

Théorème 4.5 ([3, 1999]) *Supposons que Supposition (4.1) soit satisfaite. Pour toute méthode du type (4.2) et (4.3) dont β_k satisfait à (4.21) et le pas λ_k satisfait aux conditions de *Wolfe faible* (4.10), alors toutes les directions générées sont de descente, autrement dit :*

$$d_k^T g_k < 0; \quad \forall k \geq 1$$

4.3.2 Convergence de la méthode de Dai-Yuan

Les résultats suivant sont dus à *Dai* et *Yuan* ([9, 1998]). Ils assurent la convergence de cette méthode pour une fonction strictement convexe avec

une recherche linéaire inexacte d'Armijo et Goldstein, (voir théorème (4.5), (4.6)).

Théorème 4.6. [9] Soit x_1 un point de départ pour lequel la supposition (4.1) soit satisfaite. Considérons une méthode du type (4.2) et (4.3) avec β_k satisfaisant à (4.21).

◆ Si f est strictement convexe sur l'ensemble convexe \mathcal{L} et le pas λ_k satisfait aux conditions de Goldstein

$$\omega_2 \lambda_k \nabla^T f(x_k) d_k \leq f(x_k + \lambda_k d_k) - f(x_k) \leq \omega_1 \lambda_k \nabla^T f(x_k) d_k, \quad (4.24)$$

où $g_k = \nabla f(x_k)$ et ω_1 et ω_2 sont deux constantes vérifiant $0 < \omega_1 < 1/2 < \omega_2 < 1$.

Alors :

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

Théorème 4.7. [9] Soit x_1 un point de départ pour lequel la supposition (4.1) est satisfaite. Considérons une méthode du type (4.2) et (4.3) avec β_k satisfaisant à (4.21).

◆ Si f est uniformément convexe sur l'ensemble convexe \mathcal{L} c a d s'il existe une constante $\eta > 0$ tel que :

$$(g(x) - g(y))^T (x - y) \geq \eta \|x - y\|^2, \quad \text{pour tout } x, y \in \mathcal{L}, \quad (4.25)$$

et le pas λ_k satisfait aux conditions d'Armijo

$$f(x_k + \lambda_k d_k) - f(x_k) \leq \omega \lambda_k \nabla^T f(x_k) d_k, \quad (4.26)$$

où $g_k = \nabla f(x_k)$ et $0 < \omega < 1$, pour lequel la condition suffisante de descente (4.9) est satisfaite.

Alors

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

4.3.3 Algorithme 4.3 de la Méthode de Dai-Yuan avec la règle de Wolfe faible

Etape0 : (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"

Si non aller à l'étape 2.

Etape 2 :

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

λ_k vérifie les conditions (4.10)

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

où

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

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

Théorème 4.8. ([9, 1999]) *Soit x_1 un point de départ pour lequel la proposition (4.1) est satisfaite.*

La suite $\{x_k\}$ générée par l'algorithme 4.3 converge dans le sens

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

Chapitre 5

Une nouvelle classe de méthodes de gradients conjugués non linéaires

5.1 Introduction et présentation de la méthode

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

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

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

Cette méthode a été découverte en 1952 par Hestenes et Steihaug ([14, 1952]) pour la minimisation de fonctions quadratiques strictement convexes.

Dans le cas non linéaire les méthodes du gradient conjugué génèrent des suites $\{x_k\}_{k \in \mathbb{N}}$, de la forme suivante :

$$x_{k+1} = x_k + \alpha_k d_k : k = 0, 1, 2, \dots \quad (5.1)$$

où x_k est le point obtenu à l'itération k , d_k est une direction de descente qu'on définira plus tard et $\alpha_k > 0$ est obtenu par une recherche linéaire exacte ou inexacte. Dans le cas d'une recherche linéaire exacte α_k est solution du problème d'optimisation unidimensionnel suivant :

$$f(x_k + \alpha_k d_k) = \min\{f(x_k + \alpha d_k) : \alpha \in]0, +\infty[\} \quad (5.2)$$

Dans le cas d'une recherche linéaire inexacte α_k vérifie soit les conditions d'Armijo, soit celles de Goldstein ou celles de Wolfe ou d'autres plus récentes.

Les directions de recherche d_k sont définies de façon récurrente par les relations suivantes :

$$d_k = \begin{cases} -g_k & \text{si } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{si } k = 1, 2, \dots \end{cases} \quad (5.3)$$

avec $g_k = \nabla f(x_k)$ et $\beta_k \in \mathbb{R}$ est un scalaire appelé coefficient du gradient conjugué et son expression définit et caractérise les propriétés du gradient conjugué en question. Donnons quelques expressions de β_k assez connues. Citons particulièrement

$$\beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2} \quad (\text{gradient conjugué version Fletcher reeves []}) \quad (5.4)$$

$$\beta_k^{PR} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad (\text{gradient conjugué version Polak Ribière []}) \quad (5.5)$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \quad (\text{gradient conjugué version Hestenes et Steifel []}) \quad (5.6)$$

$$\beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}} \quad (\text{gradient conjugué version Liu et Storey []}) \quad (5.7)$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}} \quad (\text{gradient conjugué version Dai et Yuan []}) \quad (5.8)$$

$$\beta_k^{DC} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \quad (\text{gradient conjugué version Descente Conjuguée []}) \quad (5.9)$$

Dans le cas d'une fonction f quadratique strictement convexe, on a

$$\beta_k^{FR} = \beta_k^{PR} = \beta_k^{HS} = \beta_k^{LS} = \beta_k^{DY} = \beta_k^{DC} \quad (5.10)$$

Si α_k est obtenu par une recherche linéaire exacte, les différentes méthodes de gradients conjugués qu'engendrent ces coefficients, convergent en un nombre fini d'itérations vers la solution optimale.

Dans les autres cas (f non quadratique ou α_k obtenu par une recherche linéaire inexacte), ces coefficients sont différents et engendrent des suites

$\{x_k\}_{k \in \mathbb{N}}$, ayant des propriétés de convergence et de vitesse de convergence différentes.

Dans ce travail, on présente une nouvelle expression de β_k qu'on note β_k^{RMIL} , étudiée et développée par Rivaie, Mustafa, Ismail, Lelong. L'expression de β_k^{RMIL} est la suivante : (gradient conjugué version Rivaie, Mustafa, Ismail et Lelong[22])

$$\beta_k^{RMIL} = \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2} \quad (5.11)$$

5.2 Algorithme de la nouvelle méthode du gradient conjugué version Rivaie, Mustafa, Ismail et Lelong

Algorithme

Etape1 : Initialisation. Donner $x_0 \in \mathbb{R}^n$. Poser $k = 0$.

Etape2 : Calculer β_k en utilisant la formule suivante :

$$\beta_k = \begin{cases} 0 & \text{si } k = 0 \\ \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2} & \text{si } k \geq 1 \end{cases} \quad (5.12)$$

Etape3 : Calculer d_k en se basant sur la formule (5.3). Si $g_k = 0$, alors stop.

Etape4 : Calculer α_k en se basant sur la relation (5.2).

Etape5 : Calculer le nouveau point en se basant sur la relation (5.1)

Etape6 : test de convergence et critère d'arrêt.

Si $f(x_{k+1}) < f(x_k)$ and $\|g_k\| < \varepsilon$, alors stop.

Sinon poser $k = k + 1$ et aller à Etape2.

5.3 Etude de la convergence

5.3.1 Descente suffisante de la nouvelle méthode du gradient conjugué version Rivaie, Mustafa, Ismail et Lelong

Théoreme 5.1([22]) *Considérons la méthode du gradient conjugué utilisant les directions de recherche $\{d_k\}_{k=0,1,\dots}$ vérifiant la relation de récurrence*

(5.3) et les coefficients $\{\beta_k^{RMIL}\}_{k=0,1,\dots}$ définies par les relations (5.12). Alors on a la relation suivante dite condition de descente suffisante :

$$g_k^T d_k \leq -C \|g_k\|^2, \quad k = 0, 1, \dots, C > 0. \quad (5.13)$$

Preuve du théorème 5.1 ([22])

La preuve du théorème se fait par récurrence :

(5.13) est vraie pour $k = 0$. En effet.

$$g_0^T d_0 = g_0^T (-g_0) = -\|g_0\|^2$$

avec $C = 1$.

Pour $k \geq 0$, considérons la relation

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

Multiplions les deux membres de l'égalité précédente par g_{k+1} , on obtient

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

Puisque α_k est obtenu par une recherche linéaire exacte, alors

$$g_{k+1}^T d_k = 0.$$

Par conséquent

$$g_{k+1}^T \cdot d_{k+1} = -\|g_{k+1}\|^2, \quad k = 0, 1, 2, \dots$$

Donc la relation (5.13) est vraie pour tout $k \in \mathbb{N}$ (en prenant $C = 1$). ■

5.3.2 Convergence globale de la nouvelle méthode du gradient conjugué version Rivaie, Mustafa, Ismail et Lelong

Pour établir le résultat essentiel de la convergence globale, les hypothèses suivantes sur f sont nécessaires et sont citées en général dans tous les résultats d'optimisation sans contraintes.

HYPOTHESES1

Soit $x_0 \in \mathbb{R}^n$ un point initial quelconque. Notons par L l'ensemble suivant :

$$\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}. \quad (5.15)$$

- i) f est minorée dans un voisinage $V(L)$, contenant l'ensemble L
- ii) Il existe une constante $L > 0$, telle que

$$\|g(x) - g(y)\| \leq L \|x - y\| : \forall x \in V(\mathcal{L}), \forall y \in V(\mathcal{L}) \quad (5.16)$$

avec $g(x) = \nabla f(x)$.

Avant d'étudier la convergence globale de la nouvelle méthode, formulons et démontrons le lemme suivant qui spécifie une propriété des coefficients β_k^{RMIL} .

Lemme 5.1 Les coefficients β_k^{RMIL} vérifient la relation suivante

$$0 \leq \beta_{k+1}^{RMIL} \leq \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \quad (5.17)$$

Preuve du Lemme 5.1

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

d'où le résultat. ■

Rappelons le célèbre théorème de Zoutendijk ([26]). Avant cela notons par θ_k l'angle que fait la direction d_k et la direction de la méthode du gradient, i.e. $-g_k = -\nabla f(x_k)$. Par conséquent θ_k sera défini par

$$\cos(\theta_k) = \frac{-g_k^T d_k}{\|g_k\| \|d_k\|} \quad (5.18)$$

Théorème 5.2 (Zoutendijk) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et (P) le problème de minimisation sans contraintes suivant :

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

Soit $\{x_k\}_{k \in \mathbb{N}}$, de la forme suivante :

$$x_{k+1} = x_k + \alpha_k d_k : k = 0, 1, 2, \dots$$

où x_k est le point obtenu à l'itération k , d_k est une direction de descente et $\alpha_k > 0$ est obtenu par une recherche linéaire exacte ou une recherche linéaire inexacte de Wolfe ou de Goldstein. On suppose que f vérifie les hypothèses 1. Alors on a

$$\sum_{k=0}^{\infty} \cos^2(\theta_k) \|g_k\|^2 < \infty \quad (5.19)$$

Remarque 5.1

Le théorème de Zoutendijk est applicable dans la nouvelle méthode du gradient conjugué version Rivaie, Mustafa, Ismail et Lelong car les directions d_k engendrées par cette nouvelle méthode sont des directions de descente (voir le théorème1).

Remarque 5.2

Puisque

$$\cos(\theta_k) = \frac{-g_k^T d_k}{\|g_k\| \|d_k\|}$$

alors la relation (5.19) devient

$$\cos^2(\theta_k) \|g_k\|^2 = \frac{(g_k^T d_k)^2}{\|g_k\|^2 \|d_k\|^2} \|g_k\|^2 = \frac{(g_k^T d_k)^2}{\|d_k\|^2}$$

et la condition de Zoutendijk devient la suivante

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \quad (5.20)$$

Ceci étant on obtient le théorème suivant :

Théorème 5.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et (P) le problème de minimisation sans contraintes suivant :

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

Soit $\{x_k\}_{k \in \mathbb{N}}$, de la forme suivante :

$$x_{k+1} = x_k + \alpha_k d_k : k = 0, 1, 2, \dots$$

où x_k est le point obtenu à l'itération k , d_k est une direction de descente suffisante obtenue avec $\beta_k = \beta_k^{RMIL}$ et $\alpha_k > 0$ est obtenu par une recherche linéaire exacte. On suppose que f vérifie les hypothèses1. Alors on a

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad \text{ou} \quad \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \quad (5.21)$$

Preuve du théorème 5.3

On démontre le théorème 3 par l'absurde. En effet, supposons le contraire. Alors

$$\exists c > 0 \text{ telle que } \|g_k\| \geq c. \quad (5.22)$$

Rappelons l'expression de d_k

$$d_k = \begin{cases} -g_k & \text{si } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 1 \end{cases}$$

donc

$$d_{k+1} + g_{k+1} = \beta_{k+1} d_k \quad (5.23)$$

(5.23) implique

$$\|d_{k+1}\|^2 = (\beta_{k+1})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \quad (5.24)$$

Soit en divisant les deux membres de (5.24) par $(g_{k+1}^T d_{k+1})^2$, on obtient

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &= \frac{-\|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1} + \beta_{k+1}^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} \quad (5.1) \\ &= \frac{\beta_{k+1}^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} - \left(\frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \\ &\leq \frac{\beta_{k+1}^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2} \end{aligned}$$

Or d'après le lemme 1 et la relation (5.17), on a

$$0 \leq \beta_{k+1}^{RMIL} \leq \frac{\|g_{k+1}\|^2}{\|d_k\|^2}.$$

Donc (5.25) implique

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &\leq \frac{\|d_k\|^2 (\|g_{k+1}\|^2)^2}{(g_{k+1}^T d_{k+1})^2 (\|d_k\|^2)^2} + \frac{1}{\|g_{k+1}\|^2} \quad (5.2) \\ &= \frac{(\|g_{k+1}\|^2)^2}{(g_{k+1}^T d_{k+1})^2 (\|d_k\|^2)} + \frac{1}{\|g_{k+1}\|^2} \\ &= \frac{(\|g_{k+1}\|^2)^2}{\|g_{k+1}\|^2 \|d_{k+1}\|^2 \|d_k\|^2} + \frac{1}{\|g_{k+1}\|^2} \\ &= \frac{1}{\|d_k\|^2} \left(\frac{\|g_{k+1}\|}{\|d_{k+1}\|} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \end{aligned}$$

Notons que

$$\frac{1}{\|d_0\|^2} = \frac{1}{\|g_0\|^2}.$$

On obtient donc en tenant compte de (5.26)

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{1}{\|g_0\|^2} + \frac{1}{\|g_k\|^2}. \quad (5.27)$$

Par conséquent

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=0}^k \frac{1}{\|g_i\|^2} \quad (5.3)$$

Si (5.22) est vraie, alors on a

$$\frac{1}{\|g_i\|^2} \leq \frac{1}{c^2}, \quad i = 0, 1, \dots, k \quad (5.29)$$

(5.29) implique

$$\sum_{i=0}^k \frac{1}{\|g_i\|^2} \leq \frac{k}{c^2} \quad (5.30)$$

(5.28) et (5.30) donnent

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{c^2}{k} \quad (5.31)$$

Puisque la série $\sum_{k=0}^{\infty} \frac{c^2}{k}$ est divergente, la relation (5.31) implique que la série $\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2}$ est aussi divergente. Par conséquent

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

Ceci est en contradiction avec le théorème de Zoutendijk (théorème2 et remarque2). Par conséquent

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad \blacksquare$$

Théorème 5.4 ([22]) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ et (P) le problème de minimisation sans contraintes suivant :

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

Soit $\{x_k\}_{k \in \mathbb{N}}$, de la forme suivante :

$$x_{k+1} = x_k + \alpha_k d_k : k = 0, 1, 2, \dots$$

où x_k est le point obtenu à l'itération k , d_k est une direction de descente suffisante où x_k est le point obtenu à l'itération k , d_k est une direction de descente suffisante obtenue avec $\beta_k = \beta_k^{RMIL}$ et $\alpha_k > 0$ est obtenu par une recherche linéaire exacte. On suppose que f vérifie les hypothèses 1. Alors on a

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad \text{ou} \quad \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^4}{\|d_k\|^2} < \infty \quad (5.32)$$

Théorème 5.5 ([22]) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ et (P) le problème de minimisation sans contraintes suivant :

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

Soit $\{x_k\}_{k \in \mathbb{N}}$, de la forme suivante :

$$x_{k+1} = x_k + \alpha_k d_k : k = 0, 1, 2, \dots$$

où x_k est le point obtenu à l'itération k , d_k est une direction de descente suffisante où x_k est le point obtenu à l'itération k , d_k est une direction de descente suffisante obtenue avec $\beta_k = \beta_k^{RMIL}$ et $\alpha_k > 0$ est obtenu par une recherche linéaire exacte. On suppose que f vérifie les hypothèses 1. Notons par θ_k l'angle que fait la direction d_k et la direction de la méthode du gradient, i.e. $-g_k = -\nabla f(x_k)$. Alors il existe $\delta > 0$ pour tout $k \in \mathbb{N}$ on a

$$\cos(\theta_k) \geq \delta \quad (5.33)$$

et

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad (5.34)$$

Preuve

La preuve de $\cos(\theta_k) \geq \delta$ se trouve dans []. Pour démontrer que $\lim_{k \rightarrow \infty} \|g_k\| = 0$, on utilise le théorème de Zoutendijk. En effet, les hypothèses d théorèmes impliquent que

$$\sum_{k=0}^{\infty} \cos^2(\theta_k) \|g_k\|^2 < \infty. \quad (5.35)$$

La relation (5.33) implique que

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad \blacksquare$$

5.4 Résultats numériques

Pour étudier numériquement l'efficacité de la nouvelle méthode du gradient conjugué avec les coefficients β_k^{RMIL} , l'auteur [22] utilise les fonctions tests de N. Andrei qu'on trouve dans [22.2012]. On compare la nouvelle méthode avec les autres méthodes classiques du gradient conjugué : FR, PR, HS, LS, DY et CD.

On prend $\varepsilon = 10^{-6}$ (critère d'arrêt) pour la valeur du gradient conjugué ($\|g_k\| \leq 10^{-6}$).

Les résultats obtenus montrent que la nouvelle méthode du gradient conjugué est plus performante numériquement que les autres méthodes classiques du gradient conjugué, en l'occurrence : FR, PR, HS, LS, DY et CD.

5.5 Conclusion et perspectives

5.5.1 Conclusions

Lors de notre étude de cette nouvelle méthode du gradient conjugué : la méthode CG version *RMIL* nous avons remarqué dans l'article (Mohd Rivaie, MUSTAFA Mamat, Lelong Wah June, Ismail MOHD, *A new class of nonlinear conjugate gradient coefficients with global convergence properties*, Applied Mathematics and Computation, 218(2012), pp.11323-11332), les faits suivants :

- 1) La nouvelle méthode est performante numériquement
- 2) Nous avons remarqué beaucoup de lacunes et un manque flagrant de précision dans l'élaboration des théorèmes et les démonstrations (théorème2, preuve du théorème2, théorème3, preuve du théorème3,...)
- 3) L'auteur a étudié la convergence globale dans le cas de la recherche linéaire exacte.

5.5.2 Perspectives et problèmes ouverts

- 1) On pourra étudier le même problème en considérant des recherches linéaires inexacts de Wolfe ou Armijo ou Goldstein au lieu de la recherche linéaire exacte.

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] M. Al-Baali, *New property and global convergence of the Fletcher-Reeves method with inexact line searches*, IMA J. Numer. Anal. 5(1985) 122-124.
- [3] Y.H. Dai and Y. Yuan (1999), *A nonlinear conjugate gradient with a strong global convergence property*, SIAM J. Optimization, Vol. 10(1) pp.177-182.
- [4] Y. H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal. 16 (1996) 155-164.
- [5] Y. H. Dai, *Analyses of nonlinear conjugate gradient method*, Ph.D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.
- [6] Y. H. Dai and Y. Yuan, *Convergence properties of the conjugate descent method*, Mathematical Advances Vol. 25 No. 6 (1996), 552-562. CMP 97 :13
- [7] Y. H. Dai and Y. Yuan, *A class of globally convergent conjugate gradient methods*, Research report ICM-98-030, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998.
- [8] Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, H. X. Yin, and Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, 1998. (accepted by SIAM Journal of Optimization)
- [9] Y. H. Dai and Y. Yuan, *Some properties of a new conjugate gradient method*, in : Y.Yuan ed., Advances in Nonlinear Programming (Kluwer, Boston, 1998), pp. 251-262.

- [10] J. W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967), 10-26.
- [11] R. Fletcher (1987), *Practical methods of optimization*, John Wiley and Sons, Chichester.
- [12] R. Fletcher and C. Reeves (1964), *Function minimization by conjugate gradients. Comput. J.*, 7, pp.149-154.
- [13] J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods. for optimization*, SIAM. J. Optimization. Vol. 2 No. 1 (1992), pp. 21-42.
- [14] M. R. Hestenes and E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards Sect. 5, 49 (1952), 409-436.
- [15] Y. F. Hu and C. Storey, *Global convergence result for conjugate gradient methods*, J. Optim. Theory Appl. Vol. 71 No. 2 (1991) 399-405.
- [16] Y. Liu and C. Storey, *Efficient Generalized Conjugate Gradient Algorithms*, Part 1 : Theory, Journal of Optimization Theory and Applications, Vol. 69 (1991), 129-137.
- [17] J.J. Moreë , B.S. Garbow, K.E. Hillstrome, *Testing unconstrained optimization software*, ACM Trans. Math. Software 7 (1981) 17-41.
- [18] J. Nocedal, S.J. Wright, *Numerical optimization*, Springer Series in Operations Research, Springer, New York, 1999.
- [19] E. Polak and G. Ribière, *Note sur la convergence de directions conjuguées*, Rev. Francaise Informat Recherche Operationelle, 3e Ann ee 16 (1969), pp. 35-43.
- [20] B. T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comp. Math. and Math. Phys. 9 (1969), pp. 94-112.
- [21] D. Pu and W. Yu, *On the convergence properties of the DFP algorithms*, Annals of Operations Research 24 (1990) 175-184.
- [22] Mohd Rivaie, MUSTAFA Mamat, Lelong Wah June, Ismail MOHD, *A new class of nonlinear conjugate gradient coefficients with global convergence properties*, Applied Mathematics and Computation, 218(2012), pp.11323-11332
- [23] D. F. Shanno, *Conjugate gradient methods with inexact searches*, Math. Oper. Res. 3 (1978), 244-256.

- [24] D. Touati-Ahmed and C. Storey (1990), Efficient hybrid conjugate gradient techniques, *JOTA*, 64, pp. 379-397.
- [25] P.Wolfe, *Convergence conditions for ascent methods*, SIAM Rev. 11 (1969) 226–235.
- [26] G. Zoutendijk, *Nonlinear Programming, Computational Methods*, in : J. Abadie (Ed.), *Integer and Nonlinear Programming*, North-Holland, Amsterdam, 1970, pp.37–86.
- [27] M. S. Bazaraa, H. D. Sherali, et C. M. Shetty (1993), *Nonlinear Programming, Theory and Algorithms*, Wiley-Interscience.
- [28] R. Fletcher and M. J. D. Powell (1963), A rapidly convergent descent method for minimization, *Computer. J.*, 6, pp. 163-168.
- [29] A. A. Goldstein (1965), On steepest descent, *SIAM J. on Control A*, Vol. 3, No. 1, pp. 147-151.
- [30] P. Wolfe (1963), *Methods of Nonlinear Programming*, in *Recent Advances in Mathematical Programming* (Eds R. L. Graves and P.Wolfe), McGraw-Hill, New York.