

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

Université 8 Mai 1945 – Guelma

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



## Mémoire

Présenté en vue de l'obtention du diplôme de

**MAGISTER en Mathématiques**

Option : **Equations aux dérivées partielles**

Par :

**Monsieur : GHIAT mourad**

**Intitulé**

**Une famille de trois paramètres du gradient  
conjugué avec la recherche linéaire inexacte  
de wolf**

Dirigé par : Professeur LASKRI Yamina

Devant le jury

<b>PRESIDENT</b>	<b>: AISSAOUI Mohamed Zine</b>	<b>(M.C.A)</b>	<b>Univ-Guelma</b>
<b>RAPPORTEUR</b>	<b>: LASKRI Yamina,</b>	<b>Professeur</b>	<b>Univ-Annaba</b>
<b>EXAMINATEUR</b>	<b>: REBBANI Faouzia</b>	<b>Professeur</b>	<b>Univ-Annaba</b>
<b>EXAMINATEUR</b>	<b>: ELLAGGOUNE Fateh</b>	<b>(M.C.A )</b>	<b>Univ- Guelma</b>

# Table des matières

<b>1</b>	<b>Préliminaires :Optimisation sans contraintes</b>	<b>9</b>
1.1	<b>Rappel de quelques définitions</b> . . . . .	10
1.2	Conditions d'optimalité des problèmes de minimisation sans contraintes . . . . .	11
1.2.1	Pourquoi avons-nous besoin de conditions d'optimalité?	12
1.2.2	Minima locaux et globaux . . . . .	12
1.2.3	Direction de descente . . . . .	13
1.2.4	Schémas général des algorithmes d'optimisation sans contraintes . . . . .	15
1.2.5	Condition nécessaires d'optimalité . . . . .	15
1.2.6	Conditions suffisantes d'optimalité . . . . .	17
<b>2</b>	<b>Recherche linéaire inexacte</b>	<b>19</b>
2.1	Recherche linéaire . . . . .	19
2.1.1	Introduction . . . . .	19
2.1.2	Objectifs de la recherche linéaire . . . . .	20
2.1.3	But de la recherche linéaire . . . . .	21
2.1.4	Algorithme de base . . . . .	21
2.2	Recherches linéaires "exactes" . . . . .	22
2.3	Recherches linéaires "inexactes ou économiques" . . . . .	23
2.3.1	Caractérisation de l'intervalle de sécurité . . . . .	23
2.4	Convergence des méthodes à directions de descente . . . . .	29
2.4.1	Condition de Zoutendijk . . . . .	29
2.5	Méthodes itératives d'optimisation sans contraintes . . . . .	31
2.5.1	Principe des méthodes de descente . . . . .	32
2.5.2	Principe des méthodes du gradient (méthode de la plus forte descente) . . . . .	33
2.5.3	<b>Les méthodes utilisant des directions conjuguées</b>	<b>37</b>

<b>3 Synthèse sur la convergence de quelques méthodes du gradient conjugué non linéaire avec la recherche linéaire exacte</b>	<b>42</b>
3.1 méthodes du gradient conjugué non linéaire . . . . .	42
3.2 Méthode de Fletcher-Reeves . . . . .	44
3.2.1 Algorithme de la méthode de Fletcher-Reeves . . . . .	45
3.2.2 La propriété de descente de la méthode de FR . . . . .	45
3.2.3 Convergence de la méthode de Fletcher-Reeves . . . . .	48
3.3 Méthode de Polak-Ribière-Polyak . . . . .	50
3.3.1 Convergence de la méthode de PRP . . . . .	51
3.4 Méthode de la descente conjuguée . . . . .	53
3.4.1 La propriété de descente de la méthode de la descente conjuguée . . . . .	54
3.4.2 Convergence de la méthode de la descente conjuguée . . . . .	55
3.5 Méthode de Dai-Yuan . . . . .	56
3.5.1 La propriété de descente de la méthode de Dai-Yuan . . . . .	57
3.5.2 Convergence de la méthode de Dai-Yuan . . . . .	58
<b>4 Une famille de trois paramètres du gradient conjugué non linéaire</b>	<b>61</b>
4.1 Introduction . . . . .	61
4.2 Une famille de trois paramètres des méthodes du gradient conjugué . . . . .	64
4.3 Critère de relancement de Powell et propriété de descente . . . . .	65
4.4 Propriétés de descente . . . . .	67
4.5 Propriétés de convergence . . . . .	68
4.6 Tests numériques . . . . .	76
4.7 Conclusions et discussions . . . . .	78

## Résumé

Dans ce mémoire, nous présentons une famille de trois paramètres du gradient conjugué non Linéaire qui inclue les six méthodes du gradient conjugué non linéaire, lesquelles sont : Fletcher-Reeves ( $\beta_k^{FR}$ ), Polak-Ribiere-Polyak ( $\beta_k^{PRP}$ ) et Hestenes-Stiefel ( $\beta_k^{HS}$ ), la descente conjuguée ( $\beta_k^{DC}$ ), Liu et Storey ( $\beta_k^{LS}$ ), et enfin Dai et Yuan ( $\beta_k^{DY}$ ).

On étudie la descente et la convergence de cette méthode, pour cela on utilise critère de relancement de powel.

**Mots clés :** Gradient conjugué, Algorithme, Convergence globale, Recherche linéaire inexacte, Règle d'Armijo, Méthode de Hestenes-Stiefel, Méthode de Fletcher-Reeves, Méthode de Polak-Ribière-Polyak, Méthode de la descente conjuguée, méthode de Liu et Storey, Méthode de Dai et Yuan.

# Abstract

In this paper, we propose a three-parameter family of conjugate gradient methods for unconstrained optimization. The three-parameter family of methods not only includes the

already-existing six practical nonlinear conjugate gradient methods, but subsumes some other families of nonlinear conjugate gradient methods as its subfamilies. With Powell's restart

criterion, the three-parameter family of methods with the strong Wolfe line search is shown to ensure the descent property of each search direction. Some general convergence results are

also established for the three-parameter family of methods. This paper can also be regarded as a brief review on nonlinear conjugate gradient methods.

## Remerciements

Je tiens à adresser mes remerciements les plus chaleureux et ma profonde gratitude à mon encadreur madame Laskri Yamina, Professeur à l'université de Annaba pour m'avoir proposer le sujet de ce mémoire. C'est grâce à sa grande disponibilité, ses conseils, ses orientations et ses encouragements que j'ai pu mener à bien ce travail.

Je remercie également les membres du jury, notamment monsieur Aissaoui Mohamed Zine, Maître de conférence classe A, à l'université du 08 Mai 1945 Guelma pour avoir accepté de présider ce jury.

Je remercie également madame Rebbani Faouzia, Professeur à l'université Badji Mokhtar Annaba. et Monsieur Ellaggoune Fateh Maître de Conférences classe A, à l'université du 08 Mai 1945 Guelma, pour avoir accepté de juger ce mémoire.

Je tiens à remercier enfin tous ceux qui m'ont aidé à réaliser ce mémoire et persévérer dans la voie qui m'était destinée : ma chère famille et mes amis.

# Introduction

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . On cherche à résoudre le problème de minimisation sans contraintes suivant :

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

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 Steifel, pour la minimisation des fonctions quadratiques strictement convexes.

Plusieurs mathématiciens ont étendu cette méthode pour le cas non linéaire. Ceci a été réalisé pour la première fois, en 1964 par Fletcher et Reeves (*méthode de Fletcher-Reeves*) puis en 1969 par Polak, Ribière et Ployak (*méthode de Polak-Ribière-Ployak*). Une autre variante a été étudiée en 1987 par Fletcher (*Méthode de la descente conjuguée*). Une nouvelle variante a été proposée en 1991 par Liu et Storey (*Méthode de Liu et Storey*). Et enfin une dernière variante qui a été étudiée en 1999 par Dai et Yuan (*Méthode de Dai et Yuan*).

Toutes ces méthodes 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.

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  $\beta_k$  définissent les différentes formes du gradient conjugué.

Si on note  $y_{k-1} = g_k - g_{k-1}$ , on obtient les variantes suivantes :

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \text{ Gradient conjugué variante Polak-Ribière-Polyak} \quad (0.4)$$

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \text{ Gradient conjugué variante Fletcher Reeves.} \quad (0.5)$$

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}, \text{ Gradient conjugué variante descente conjugué.} \quad (0.6)$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \text{ Gradient conjugué variante de } Dai\text{-}Yuan. \quad (0.7)$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \text{ Gradient conjugué variante Hestenes-Stiefel.} \quad (0.8)$$

$$\beta_k^{LS} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}}, \text{ Gradient conjugué variante de } Liu\text{-}Storey. \quad (0.9)$$

Le but de ce mémoire est d'étudier la descente et la convergence des méthodes du gradient conjugué de trois paramètres incluant les méthodes de gradient conjuguées non linéaires mentionnées ci-dessus.

Ce mémoire comporte quatre chapitres.

**Chapitre 1** On introduit dans ce chapitre les notions préliminaires de base concernant l'optimisation sans contraintes. On commence par quelques éléments de Topologie et de calcul différentiel. On termine ce chapitre par quelques notions sur les problèmes de minimisation sans contraintes et leurs algorithmes.

**Chapitre 2** On expose dans ce chapitre les grandes lignes des méthodes d'optimisation sans contraintes basées sur les directions de descente et les recherches linéaires. Ces méthodes génèrent une suite  $\{x_k\}_{k \in \mathbb{N}}$  de la façon suivante :

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

On suppose connaître la direction de descente  $d_k$  au point  $x_k$ . La recherche linéaire consiste à trouver  $\lambda_k$  de façon à diminuer la fonction  $f$  suffisamment le long de cette direction.

On insistera dans ce chapitre sur les recherches linéaires monotone avec lesquelles nous allons établir des résultats de convergence pour certaines méthodes.

**Chapitre 3** Dans ce chapitre on fait une synthèse de différents résultats de convergence de différentes variantes de la méthode du gradient conjugué avec la recherche linéaire inexacte de wolfe. On s'intéressera spécialement aux méthodes de Fletcher Reeves(FR), Descente conjugué(CD), Polak-Ribière-Polyak(PRP), *Dai-Yuan*(DY).

**Chapitre 4** Ce chapitre est la partie la plus importante de ce travail, car il contient les résultats de descente et de convergence de la méthode du



## *TABLE DES MATIÈRES*

---

gradient conjugué de trois paramètres avec la recherche de wolfe et le critère de relancement de powel.

On pose à la fin du mémoire deux problèmes ouverts et leur conjecture.

# Chapitre 1

## Préliminaires : Optimisation sans contraintes

nous allons exposer dans ce premier chapitre quelques concepts clés concernant l'optimisation. On commence par se poser les questions suivantes :

- Les problèmes proposés comportent-ils des contraintes ?
- Les fonctions en jeu sont-elles quadratiques ? sont-elles convexes ?
- Les domaines de définition des fonctions sont-ils continus ou discrets ?

Chaque problème possède sa structure spécifique et demande donc d'être traité de manière particulière.

Le présent projet ne se focalisera que sur les problèmes d'optimisation sans contraintes et non linéaires. Ce problème sera formulé comme suit :

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

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

La majorité des méthodes de résolution du problème (P) sont de nature itérative, c'est à dire qu'à partir d'un point initial  $x_1$ , elles engendrent une suite infinie  $x_1, x_2, \dots, x_k$ , dont on espère qu'elle converge vers la solution optimale.

Pour construire des algorithmes de minimisation sans contraintes on fait appel à des processus itératifs du type :

$$x_{k+1} = x_k + \lambda_k d_k, \quad (1.2)$$

où  $d_k$  détermine la direction de déplacement à partir du point  $x_k$  et  $\lambda_k$  est le pas lequel est solution optimale du problème d'optimisation unidimensionnel suivant :

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

c'est à dire que  $\lambda_k$  vérifié :

$$f(x_k + \lambda_k d_k) \leq f(x_k + \lambda d_k), \quad \forall \lambda > 0 \quad (1.3)$$

Le type d'algorithme permettant de résoudre le problème (1. 1) sera déterminé dès qu'on définit les procédés de construction du vecteur  $d_k$  et de calcul de  $\lambda_k$  à chaque itération.

La façon avec laquelle on construit les vecteurs  $d_k$  et les scalaires  $\lambda_k$  détermine directement les propriétés du processus et spécialement en ce qui concerne la convergence de la suite  $\{x_k\}$ , et la vitesse de la convergence. Pour s'approcher de la solution optimale du problème (1. 1) (dans le cas général, c'est un point en lequel ont lieu peut être avec une certaine précision les conditions nécessaires d'optimalité de  $f$ ) on se déplace naturellement à partir du point  $x_k$  dans la direction de la décroissance de la fonction  $f$ .

## 1.1 Rappel de quelques définitions

Nous introduisons ici les définitions de base qui seront utilisées par la suite.

### Définition 1. 1 (Convergence des algorithmes)

Un algorithme de résolution est un procédé qui permet, à partir de la donnée du point initial  $x_1$ , d'engendrer la suite  $x_1, x_2, \dots, x_k, \dots$

Un algorithme est parfaitement défini par la donnée de l'application  $A$  qui associe à  $x_k$  le point  $x_{k+1} \in A(x_k)$ . Ceci permettra de confondre un algorithme et l'application  $A$  qui lui est associée.

### Définition 1. 2. (Convergence globale)

Nous dirons qu'un algorithme décrit par une application multivoque  $A$ , est globalement convergent (ou encore : possède la propriété de convergence globale) si, quel que soit le point de départ  $x_1$  choisi, la suite  $\{x_k\}$  engendrée par  $x_{k+1} \in A(x_k)$  (ou une sous suite) converge vers un point satisfaisant les conditions nécessaires d'optimalité (ou solution optimale).

### Définition 1. 3 (Modes de convergence)

Soit  $\{x_k\}_{k \in \mathbb{N}}$  une suite dans  $\mathbb{R}^n$  convergeant vers  $x^*$ .

◆ Si

$$\limsup \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \tau < 1,$$

On dit que  $\{x_k\}_{k \in \mathbb{N}}$  converge vers  $x^*$  linéairement avec le taux  $\tau$ .

◆ Si

$$\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \longrightarrow 0 \text{ quand } k \longrightarrow \infty,$$

on dit que la convergence est superlinéaire.

Plus précisément si  $\exists p > 1$  tel que :

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^p} < +\infty,$$

on dit que la convergence est superlinéaire d'ordre  $p$ .

En particulier si

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^2} < +\infty,$$

on dit que la convergence est quadratique (superlinéaire d'ordre 2).

**Définition 1. 4**

Soit  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ , on dit que  $f$  est continue au point  $a \in \mathbb{R}^n$  si pour tout réel  $\varepsilon > 0$ , Il existe  $\eta > 0$

tel que :  $\|x - a\| < \eta \implies \|f(x) - f(a)\| < \varepsilon$ .

**Définition 1. 5. (Fonction convexe différentiable).**

Soit  $C \subseteq \mathbb{R}^n$ ,  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$  et  $\hat{x} \in \text{int}(C)$ .  $f$  est dite différentiable au point  $\hat{x}$ , s'il existe un vecteur  $A \in \mathbb{R}^n$  et une fonction  $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$  telle que

$$f(x) = f(\hat{x}) + A(x - \hat{x}) + \|x - \hat{x}\| \alpha(\hat{x}, x - \hat{x}),$$

ou'  $\alpha(\hat{x}, x - \hat{x}) \xrightarrow{x \rightarrow \hat{x}} 0$ . On peut noter le vecteur  $A$  comme suit :

$$A = \nabla f(\hat{x}) = \left( \frac{\partial f(\hat{x})}{\partial x_1}, \dots, \frac{\partial f(\hat{x})}{\partial x_n} \right). \tag{1.4}$$

## 1.2 Conditions d'optimalité des problèmes de minimisation sans contraintes

Considérons le problème d'optimisation sans contraintes ( $P$ )

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

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

### 1.2.1 Pourquoi avons-nous besoin de conditions d'optimalité ?

Afin d'analyser ou de résoudre de manière efficace un problème d'optimisation, il est fondamental de pouvoir disposer de conditions d'optimalité. En effet, celles-ci nous servent non seulement à vérifier la validité des solutions obtenues, mais souvent l'étude de ces conditions aboutit au développement des algorithmes de résolution eux-mêmes. Des conditions équivalentes peuvent être obtenues de diverses manières, en procédant à des analyses suivant différentes "lignes directrices". L'approche considérée ici pour l'obtention de conditions est basée sur la notion de descente.

### 1.2.2 Minima locaux et globaux

#### Définition 1. 6.

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

$$f(\hat{x}) \leq f(x) : \forall x \in V_\varepsilon(\hat{x}) \quad (1.6)$$

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

$$f(\hat{x}) < f(x) : \forall x \in V_\varepsilon(\hat{x}), x \neq \hat{x} \quad (1.7)$$

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

$$f(\hat{x}) \leq f(x) : \forall x \in \mathbb{R}^n \quad (1.8)$$

#### Remarque 1. 1.

Dans le cas d'une fonction objectif convexe, il n'y a pas de distinction entre minimum local et global : tout minimum local est également global, comme l'établit le théorème suivant.

#### Théorème 1. 1.

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction convexe définie sur l'ensemble convexe  $X$ . Alors, tout minimum local de  $f$  est également un minimum global. Si  $f$  est strictement convexe, alors il existe au plus un minimum global de  $f$ .

### 1.2.3 Direction de descente

#### Principes généraux

Considérons le problème d'optimisation sans contrainte :

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

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

On note par  $\nabla f(x)$  et  $\nabla^2 f(x)$  respectivement le gradient et le hessien de  $f$  en  $x$ .

On s'intéresse ici à une classe d'algorithmes qui sont fondés sur la notion de direction de descente.

On dit que  $d$  est une direction de descente de  $f$  en  $x \in \mathbb{R}^n$  si

$$\nabla f(x)^T \cdot d < 0. \tag{1.9}$$

Ou encore que  $d$  fait avec l'opposé du gradient  $-\nabla f(x)$  un angle  $\theta$  strictement plus petit que  $90^\circ$  :

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

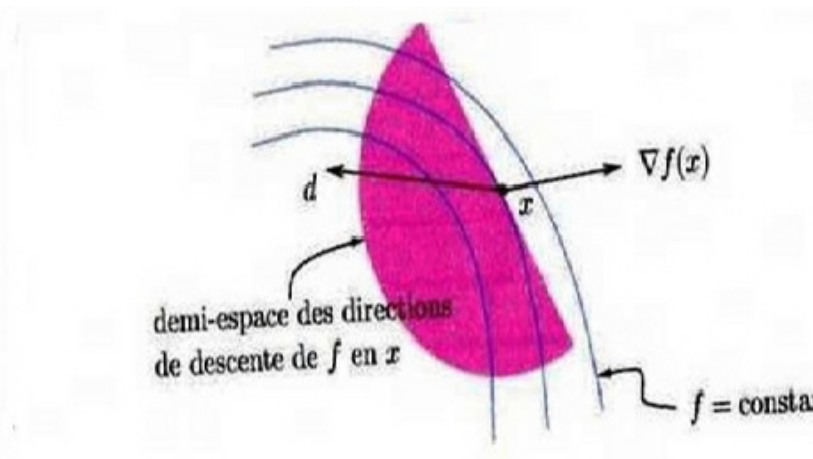


FIG 2.4 Demi espace des direction de descente de f en x

L'ensemble des directions de descente de  $f$  en  $x$ ,

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

forme un demi-espace ouvert de  $\mathbb{R}^n$ .

On voit que si  $d$  est une direction de descente,  $f(x + \lambda d) < f(x)$ , pour tout  $\lambda > 0$  suffisamment petit, et donc que  $f$  décroît strictement dans la direction  $d$ . De telles directions sont intéressantes en optimisation car, pour faire décroître  $f$ , il suffit de faire un déplacement le long de  $d$ . Les méthodes à directions de descentes utilisent cette idée pour minimiser une fonction. Elles construisent la suite des itérés  $\{x_k\}_{k \geq 1}$ , approchant une solution  $x_k$  du problème (1.1) par la récurrence :

$$x_{k+1} = x_k + \lambda_k d_k, \text{ pour } k \geq 1.$$

Où  $\lambda_k$  est appelé le pas et  $d_k$  la direction de descente de  $f$  en  $x_k$ .

Pour définir une direction de descente il faut donc spécifier deux choses :

◆ dire comment la direction  $d_k$  est calculée, la manière de procéder donne le nom à l'algorithme ;

◆ dire comment on détermine le pas  $\lambda_k$ , c'est ce que l'on appelle la recherche linéaire.

**Remarque 1. 2.**

Par la suite l'angle  $\theta_k$  entre la direction  $d_k$  et  $-g_k$  jouera un rôle important dans le processus de la convergence. Pour  $\theta_k$  on a la relation classique suivante :

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

**Définition 1. 7.**

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

$$f(\hat{x} + \lambda d) < f(\hat{x}) : \quad \forall \lambda \in ]0, \delta[, \tag{1.10}$$

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

**Algorithme 2. 1(méthode à directions de descente- une itération)**

Etape 0 :(initialisation)

On suppose qu'au début de l'itération  $k$ , on dispose d'un itéré  $x_k \in \mathbb{R}^n$

Etape1 :

Test d'arrêt : si  $\|\nabla f(x_k)\| \simeq 0$ , arrêt de l'algorithme ;

Etape2 :

Choix d'une direction de descente  $d_k \in \mathbb{R}^n$ ;

Etape3 :

Recherche linéaire : déterminer un pas  $\lambda_k > 0$  le long de  $d_k$  de manière à "faire décroître  $f$  suffisamment" ;

Étape 4 :

Si la recherche linéaire réussie  $x_{k+1} = x_k + \lambda_k d_k$  ;  
remplacer  $k$  par  $k + 1$  et aller à l'étape 1.

### 1.2.4 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, est assuré que

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

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

#### Exemple de choix de direction 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. Dans le cas  $d_k = -(H(x_k))^{-1} \cdot \nabla f(x_k)$ , on obtient la méthode de Newton, où la matrice hessienne  $H(x_k)$  est définie positive.

#### Exemple de choix de pas $\lambda_k$

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

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

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

### 1.2.5 Condition nécessaires d'optimalité

Étant donné un vecteur  $\hat{x}$ , nous souhaitons déterminer si ce point est un minimum local ou global de  $(P)$ . La propriété de différentiabilité de  $f$  fournit une première manière de caractériser une solution optimale. Énonçons tout d'abord un théorème pour établir une première condition nécessaire d'optimalité.



### 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 (1. 2). 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 :

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

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

#### Remarque 1. 3.

Si  $f$  est convexe, la condition nécessaire du premier ordre est également suffisante pour que  $\hat{x}$  soit un minimum globale. Dans le cas où  $f$  est deux fois différentiable, une autre condition nécessaire est donnée par le théorème (1. 5) elle appelée condition nécessaire du second ordre car elle fait intervenir la matrice hessienne de  $f$  (que nous noterons  $\nabla^2 f(x)$ , dont les éléments sont ses secondes dérivées partielles)

### 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  $\hat{x}$  est semi définie positive.

#### Preuve

Soit  $x \in \mathbb{R}^n$  quelconque, étant  $f$  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

$$\frac{f(\hat{x} + \lambda x) - f(\hat{x})}{\lambda^2} = \frac{1}{2} x^T H(\hat{x}) x + \|x\|^2 \alpha(\hat{x}, \lambda x). \quad (1.13)$$

$\hat{x}$  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. 13) et on passe à la limite quand  $\lambda \rightarrow 0$ ,  $\lambda \neq 0$ , on obtient :

$$x^T H(\hat{x})x \geq 0, \forall x \in \mathbb{R}^n. \square$$

### 1.2.6 Conditions suffisantes d'optimalité

Les conditions données précédemment sont nécessaires, c'est-à-dire qu'elle doivent être satisfaites pour tout minimum local. Cependant tout vecteur vérifiant ces conditions n'est pas nécessairement un minimum local. Le théorème (1. 6) établit une condition suffisante pour qu'un point soit un minimum local, si  $f$  est deux fois différentiable.

#### Condition suffisante d'optimalité du second ordre

##### **Théorème 1. 4**

Soient  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  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  $x \in \mathbb{R}^n$

$$f(x) = f(\hat{x}) + \frac{1}{2}(x - \hat{x})^t H(\hat{x})(x - \hat{x}) + \|x - \hat{x}\|^2 \alpha(\hat{x}, x - \hat{x}), \quad \alpha(\hat{x}, (x - \hat{x})) \xrightarrow{x \rightarrow \hat{x}} 0, \quad (1.14)$$

$$(\nabla f(\hat{x}) = 0).$$

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

$$x_k \neq \hat{x} : \forall k, \quad x_k \xrightarrow{k \rightarrow \infty} \hat{x} \text{ et } f(x_k) \leq f(\hat{x}). \quad (1.15)$$

Dans (1. 14) 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

$$\frac{f(x_k) - f(\hat{x})}{\|x_k - \hat{x}\|^2} = \frac{1}{2} d_k^T H(\hat{x}) d_k + \alpha(\hat{x}, (x_k - \hat{x})), \quad \alpha(\hat{x}, (x_k - \hat{x})) \xrightarrow{k \rightarrow \infty} 0. \quad (1.16)$$

(1. 15) et (1. 16) 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  $\{d_k\}_{k \in \mathbb{N}^*}$  est bornée ( $\|d_k\| = 1, \forall n$ ). Donc il existe une sous suite  $\{d_k\}_{k \in \mathbb{N}_1 \subset \mathbb{N}}$  telle que  $d_k \xrightarrow[k \rightarrow \infty]{} \tilde{d}_{k \in \mathbb{N}_1}$ .

Finalement lorsque  $k \rightarrow \infty$  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  $H(x)$  n'est pas définie positive. Ceci est en contradiction avec l'hypothèse.  $\square$

**Remarque 1. 4.**

Dans le cas où  $f$  est convexe, alors tout minimum local est aussi global. De plus si  $f$  est strictement convexe, alors tout minimum local devient non seulement global mais aussi unique ([4]).

## Chapitre 2

# Recherche linéaire inexacte

### 2.1 Recherche linéaire

#### 2.1.1 Introduction

Considérons le problème d'optimisation sans contraintes  $(P)$ .

$$(P) : \min \{f(x) : x \in \mathbb{R}^n\} \quad (2.1)$$

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

Les algorithmes qu'on étudie par la suite suivent les schémas générales suivants.

$$x_{k+1} = x_k + \lambda_k d_k \quad (2.2)$$

où  $\lambda_k$  est solution optimale du problème d'optimisation unidimensionnel suivant :

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

c'est à dire que  $\lambda_k$  vérifie

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

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

$$\lambda \rightarrow \varphi(\lambda) = f(x_k + \lambda d_k). \quad (2.3)$$

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

Dans ce chapitre nous allons décrire les différentes manières de déterminer un pas  $\lambda_k > 0$  le long d'une direction de descente  $d_k$ . C'est ce que l'on appelle faire de la recherche linéaire. Il existe deux grandes classes des méthodes qui s'intéressent à l'optimisation unidimensionnelle :

- ▶ les recherches linéaires exactes.
- ▶ les recherches linéaires inexactes.

### 2.1.2 Objectifs de la recherche linéaire

Il s'agit de réaliser deux objectifs :

Le premier objectif, faire décroître  $f$  suffisamment, et pour cela on cherche à vérifier l'inégalité

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \text{“un terme négatif”} \quad (2.4)$$

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

Le second objectif, on choisi le pas  $\lambda_k > 0$  d'être trop petit, pour assurer la convergence d'algorithme au point stationnaire.

Le premier objectif n'est en effet pas suffisant car l'inégalité (2. 4) est en général satisfaite par des pas  $\lambda_k > 0$  arbitrairement petit. Or ceci peut entraîner une “fausse convergence”, c'est- dire la convergence des itérés vers un point non stationnaire, comme le montre l'observation suivante.

Si on prend

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

la suite  $\{x_k\}$  générée par (2. 4) est de Cauchy, puisque pour  $1 \leq l < k$  on a

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

lorsque  $l \rightarrow \infty$ .

Donc  $\{x_k\}$  converge vers un point  $\bar{x}$ . En prenant  $l = 1$  et  $k \rightarrow \infty$  dans l'estimation ci-dessus, on voit que  $\bar{x} \in \overline{B}(x_1, \varepsilon)$  et donc  $\bar{x}$  ne saurait être solution s'il n'y a pas de solution dans  $B(x_1, \varepsilon)$ . On a donc arbitrairement forcer la convergence de  $\{x_k\}$  en prenant des pas très petits.

Pour simplifier les notations, on définit la restriction de  $f$  à la droite  $\{x_k + \lambda d_k : \lambda \in \mathbb{R}\}$  comme la fonction

$$\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}; \lambda \mapsto \varphi(\lambda) = f(x_k + \lambda d_k).$$

### 2.1.3 But de la recherche linéaire

Dans le cas non-quadratique les méthodes de descente (2.2), nécessitent la recherche d'une valeur de  $\lambda_k > 0$  optimale ou non, vérifiant :

$$f(x_k + \lambda_k d_k) \leq f(x_k). \quad (2.5)$$

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

$$\varphi'(\lambda^*) = 0,$$

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

autrement dit,  $\lambda^*$  est un minimum local de  $\varphi$  qui assure de plus la décroissance de  $f$ . En fait, dans la plupart des algorithmes d'optimisation modernes, on ne fait jamais de recherche linéaire exacte, car trouver  $\lambda^*$  signifie qu'il va falloir calculer un grand nombre de fois la fonction  $\varphi$ , et cela peut être dissuasif du point de vue du temps de calcul. En pratique, on recherche plutôt une valeur de  $\lambda$  qui assure une décroissance suffisante de  $f$ . Cela conduit à la notion d'intervalle de sécurité.

#### Définition 2. 1

On dit que  $[a, b]$  est un intervalle de sécurité s'il permet de classer les valeurs de  $\lambda$  de la façon suivante :

- ◆ Si  $\lambda < a$  alors  $\lambda$  est considéré trop petit.
- ◆ Si  $b \geq \lambda \geq a$  alors  $\lambda$  est satisfaisant.
- ◆ Si  $\lambda > b$  alors  $\lambda$  est considéré trop grand.

Le problème est de traduire de façon numérique sur  $\varphi$  les trois conditions précédentes, ainsi que de trouver un algorithme permettant de déterminer  $a$  et  $b$ . L'idée est de partir d'un intervalle suffisamment grand pour contenir  $[a, b]$ , et d'appliquer un bonne stratégie pour itérativement réduire cet intervalle.

### 2.1.4 Algorithme de base

Etape 0 : (initialisation)

$a = b = 0$ , choisir  $\lambda_1 > 0$ , poser  $k = 1$  et aller à l'étape 1 ;

Etape 1 :

Si  $\lambda_k$  convient, poser  $\lambda^* = \lambda_k$  et on s'arrête.

Si  $\lambda_k$  est trop petit on prend  $a_{k+1} = \lambda_k$ ,  $b_{k+1} = b$ ,  
et on va à l'étape 2.

Si  $\lambda_k$  est trop grand on prend  $b_{k+1} = \lambda_k$ ,  $a_{k+1} = a$ ,

et on va à l'étape 2 .

Etape 2 :

Si  $b_{k+1} = 0$  déterminer  $\lambda_{k+1} \in ]a_{k+1}, +\infty[$ .

Si  $b_{k+1} \neq 0$  déterminer  $\lambda_{k+1} \in ]a_{k+1}, b_{k+1}[$ ,

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

Il faut maintenant préciser quelles sont les relations sur  $\varphi$  qui va nous permettre de caractériser les valeurs de  $\lambda$  convenables, ainsi que les techniques utilisées pour réduire l'intervalle.

## 2.2 Recherches linéaires "exactes"

Comme on cherche à minimiser  $f$ , il semble naturel de chercher à minimiser le critère le long de  $d_k$  et donc de déterminer le pas  $\lambda_k$  comme solution du problème

$$\min_{\lambda > 0} \varphi(\lambda)$$

C'est ce que l'on appelle la règle de Cauchy et le pas déterminé par cette règle est appelé pas de Cauchy ou pas optimal dans certains cas, on préférera le plus petit point stationnaire de  $\varphi$  qui fait décroître cette fonction :

$$\lambda_k = \inf \{ \lambda > 0 : \varphi'(\lambda) = 0, \varphi(\lambda) < \varphi(0) \}.$$

On parle alors de règle de Curry et le pas déterminé par cette règle est appelé pas de Curry. De manière un peu imprécise, ces deux règles sont parfois qualifiées de recherche linéaire exacte. Ils ne sont utilisés que dans des cas particuliers, par exemple lorsque  $\varphi$  est quadratique, la solution de la recherche linéaire s'obtient de façon exacte et dans un nombre fini d'itérations.

On aurait aimé restreindre notre étude sur ce domaine malheureusement. C'est très rare de rencontrer des problèmes quadratiques.

► la détermination de ces pas demande en général beaucoup de temps de calcul et ne peut de toutes les façons pas être faite avec une précision infinie,

► l'efficacité supplémentaire éventuellement apportée à un algorithme par une recherche linéaire exacte ne permet pas, en général, de compenser le temps perdu à déterminer un tel pas.

► les résultats de convergence autorisent d'autres types de règles, moins gourmandes en temps de calcul.

Au lieu de demander que  $\lambda_k$  minimise  $\varphi$ , on préfère imposer des conditions moins restrictives, plus facilement vérifiées, qui permettent toutefois de contribuer à la convergence des algorithmes. En particulier, il n'y aura plus un unique pas (ou quelques pas) vérifiant ces conditions mais tout un intervalle de pas (ou plusieurs intervalles), ce qui rendra d'ailleurs leur recherche plus facile.

C'est ce que l'on fait avec les règles d'Armijo, de Goldstein et de Wolfe décrites ci-dessous.

## 2.3 Recherches linéaires "inexactes ou économiques"

Les recherches linéaires exactes, malgré qu'elles n'aboutissent à une solution optimale qu'avec une tolérance fixée à l'avance, elles nécessitent beaucoup d'observations à chaque itération de l'algorithme principal. Des mathématiciens ont réussi (années 60, 70, 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 super linéaire.

L'objectif de cette section consiste à présenter les principales tests.

### 2.3.1 Caractérisation de l'intervalle de sécurité

#### la règle d'Armijo

Une condition naturelle est de demander que  $f$  décroisse autant qu'une portion  $\omega_1 \in ]0, 1[$  de ce que ferait le modèle linéaire de  $f$  en  $x_k$ . Cela conduit à l'inégalité suivante, parfois appelée condition d'Armijo ou condition de décroissance linéaire :

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \omega_1 \lambda_k \nabla f(x_k)^T d_k. \quad (2.6)$$

Elle est de la forme (2.4), car  $\omega_1$  devra être choisi dans  $]0, 1[$ . On voit bien à la figure (2.1) ce que signifie cette condition. Il faut qu'en  $\lambda_k$ , la fonction prenne une valeur plus petite que celle prise par la fonction affine

$$\lambda \rightarrow f(x_k) + \omega_1 \lambda \nabla f(x_k)^T d_k. \quad (2.7)$$

#### Règle d'Armijo

- ◆ Si  $\varphi(\lambda) \leq \varphi(0) + \omega_1 \varphi'(0)\lambda$ , alors  $\lambda$  convient.
- ◆ Si  $\varphi(\lambda) > \varphi(0) + \omega_1 \varphi'(0)\lambda$ , alors  $\lambda$  est trop grand.

On peut noter que l'on a

$$\varphi(\lambda) = f(x_k + \lambda d_k)$$

$$\varphi(0) = f(x_k),$$



$$\varphi'(0) = \nabla f(x_k)^T d_k.$$

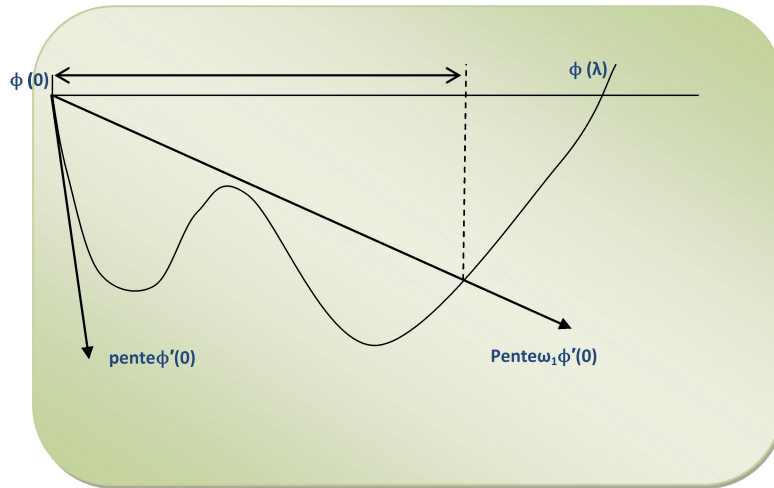


FIG 2.1 Règle d'armijo

**Algorithme 2. 1 (Règle d'Armijo)**

Etape 0 : (initialisation)

$a_1 = b_1 = 0$ , choisir  $\lambda_1 > 0$ ,  $\omega_1 \in ]0, 1[$  poser  $k = 1$  aller à l'étape 1.

Etape 1 :

Si  $\phi(\lambda_k) \leq \phi(0) + \omega_1 \phi'(0) \lambda_k$  : STOP ( $\lambda^* = \lambda_k$ ).

Si  $\phi(\lambda_k) > \phi(0) + \omega_1 \phi'(0) \lambda_k$ , alors

$b_{k+1} = b$ ,  $a_{k+1} = \lambda_k$  et aller à l'étape 2.

Etape 2 :

Si  $b_{k+1} = 0$  déterminer  $\lambda_{k+1} \in ] a_{k+1}, +\infty [$ .

Si  $b_{k+1} \neq 0$  déterminer  $\lambda_{k+1} \in ] a_{k+1}, b_{k+1}[$ ,

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

**Remarque 2. 1**

1) En pratique, la constante  $\omega_1$  est prise très petite, de manière à satisfaire (2. 6) le plus facilement possible. Typiquement,  $\omega_1 = 10^{-4}$ . Notons que cette constante ne doit pas être adaptée aux données du problème et donc que l'on ne se trouve pas devant un choix de valeur délicat.

2) Dans certains algorithmes, il est important de prendre  $\omega_1 < \frac{1}{2}$  pour que le pas  $\lambda_k$  soit accepté lorsque  $x_k$  est proche d'une solution.

3) Il est clair d'après la figure (2. 1) que l'inégalité (2. 6) est toujours vérifiée si  $\lambda_k > 0$  est suffisamment petit.

4) On a vu qu'il était dangereux d'accepter des pas trop petits, cela pouvait conduire à une fausse convergence. Il faut donc un mécanisme supplémentaire qui empêche le pas d'être trop petit. On utilise souvent la technique de rebroussement due à Armijo [9, 1966] ou celle de Goldstein.

### La règle de Goldstein

La règle de Goldstein remédie à cet inconvénient (le pas  $\lambda_k$  doit être trop petit).

Dans celle-ci, en ajoutant une deuxième inégalité à la règle d'Armijo on obtient la règle de Goldstein.

$$f(x_k) + \omega_1 \lambda_k \nabla^T f(x_k) d_k \geq f(x_k + \lambda_k d_k) \geq f(x_k) + \omega_2 \lambda_k \nabla^T f(x_k) d_k, \quad (2.8)$$

où  $\omega_1$  et  $\omega_2$  sont deux constantes vérifiant  $0 < \omega_1 < \omega_2 < 1$ , cette inégalité qui empêche le pas d'être trop petit.

#### Règle de Goldstein

- ◆ Si  $\varphi(\lambda) < \varphi(0) + \omega_2 \varphi'(0)\lambda$ , alors  $\lambda$  est trop petit.
- ◆ Si  $\varphi(\lambda) > \varphi(0) + \omega_1 \varphi'(0)\lambda$ , alors  $\lambda$  est trop grand.
- ◆ Si  $\varphi(0) + \omega_1 \varphi'(0)\lambda \geq \varphi(\lambda) \geq \varphi(0) + \omega_2 \varphi'(0)\lambda$ , alors  $\lambda$  convient.

On peut noter que l'on a

$$\varphi(\lambda) = f(x_k + \lambda d_k)$$

$$\varphi(0) = f(x_k),$$

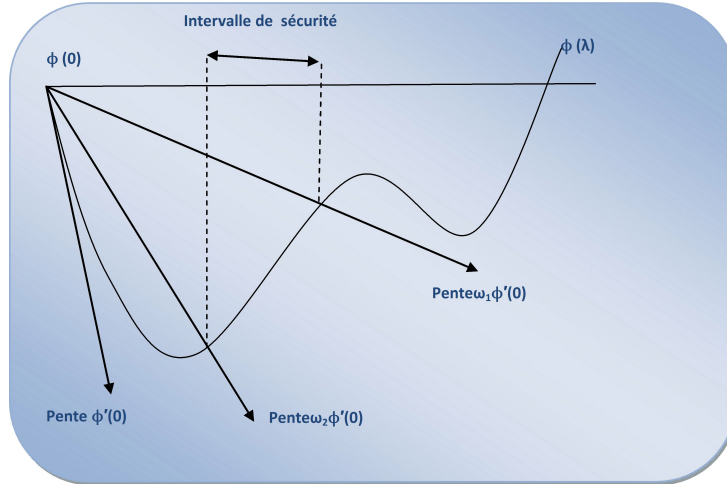


FIG 2.2 Règle de Goldstein

**Algorithme 2. 2 (Règle de Goldstein&Price)**

Etape 0 : (initialisation)

$a_1 = b_1 = 0$ , choisir  $\lambda_1 > 0$   $\omega_1 \in ]0, 1[$ ,  $\omega_2 \in ]\omega_1, 1[$ , poser  $k = 1$  et aller à l'étape 1.

Etape 1 :

Si  $\phi(0) + \omega_2 \phi'(0)\lambda \leq \phi(\lambda_k) \leq \phi(0) + \omega_1 \phi'(0)\lambda_k$  : STOP( $\lambda^* = \lambda_k$ ).

Si  $\phi(\lambda_k) > \phi(0) + \omega_1 \phi'(0)\lambda_k$ , alors

$b_{k+1} = \lambda_k$ ,  $a_{k+1} = a_k$ , et aller à l'étape 2.

Si  $\phi(\lambda_k) < \phi(0) + \omega_2 \phi'(0)\lambda_k$ , alors

$b_{k+1} = b_k$ ,  $a_{k+1} = \lambda_k$ , et aller à l'étape 2.

Etape 2 :

Si  $b_{k+1} = 0$  déterminer  $\lambda_{k+1} \in ]b_{k+1}, +\infty[$

Si  $b_{k+1} \neq 0$  déterminer  $\lambda_{k+1} \in ]a_{k+1}, b_{k+1}[$ .

Remarque 2. 2 il n'est pas facile de voir comment on peut trouver un pas de Goldstein en un nombre fini d'étapes.

**Règle de Wolfe**

La règle de Wolfe fait appel au calcul de  $\varphi'(\lambda)$ , elle est donc en théorie plus coûteuse que la règle de Goldstein. Cependant dans de nombreuses 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 très utilisée.

Nous allons présenter les conditions de Wolfe faibles sur  $\lambda > 0$  :

**Règle de wolfe faible**

$$f(x_k + \lambda d_k) \leq f(x_k) + \omega_1 \lambda \nabla f(x_k)^T . d_k.$$

$$\nabla f(x_k + \lambda d_k)^T . d_k \geq \omega_2 \nabla f(x_k)^T . d_k.$$

avec  $0 < \omega_1 < \omega_2 < 1$  .

Pour expliquer ces conditions posons

$$l(\lambda) = \phi(0) + (\omega_2 \phi'(0)) \lambda.$$

**Règle de Wolfe faible**

- ◆ Si  $\phi(\lambda) \leq \phi(0) + \omega_1 \phi'(0) \lambda$  et  $\phi'(\lambda) \geq \omega_2 \phi'(0)$ , alors  $\alpha$  convient.
- ◆ Si  $\phi(\lambda) > \phi(0) + \omega_1 \phi'(0) \lambda$ , alors  $\lambda$  est trop grand.
- ◆ Si  $\phi'(\lambda) < \omega_2 \phi'(0)$ , alors  $\lambda$  est trop petit.

On voit bien à la figure 2. 3 ce que signifie cette condition.

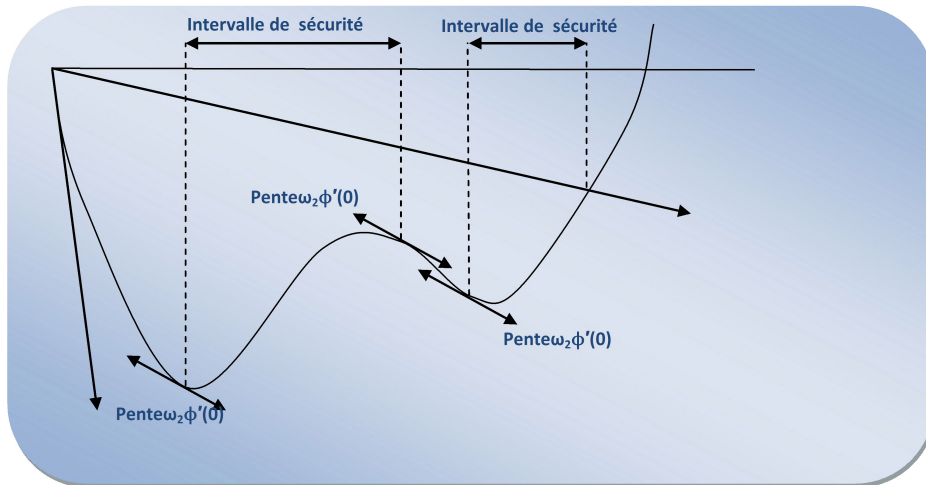


FIG 2.3 Règle de Wolfe faible

On remarque que  $\phi'(0) < 0$  car  $d_k$  est une direction de descente.

La condition (W1) impose que la réduction de  $f$  est proportionnelle à  $\lambda$  et  $\phi'(0)$ . On ne retiendra que les valeurs de  $\lambda$  pour lesquelles le graphe de  $\phi$  est en dessous de la droite  $l$ , comme  $0 < \omega_1 < 1$  ceci est possible au moins pour  $\lambda$  petit. La condition (W2) implique que  $\phi'(\lambda) \geq \omega_2 \phi'(0) \geq \phi'(0)$  : si  $\phi'(\lambda)$  est “très” négatif (i. e.  $< \omega_2 \phi'(0)$ ), on va chercher plus loin, si non on peut s’arrêter. De de cette façon le pas  $\lambda$  ne sera pas trop petit.

**Algorithme 2. 3 (Règle de Wolfe)**

Etape 0 : (initialisation)

$a_1 = b_1 = 0$ , choisir  $\lambda_1 > 0$   $\omega_1 \in ]0, 1[$ ,  $\omega_2 \in ]\omega_1, 1[$ , poser  $k = 1$  et aller à l'étape 1.

Etape 1 :

Si  $\phi(\lambda_k) \leq \phi(0) + \omega_1\phi'(0)\lambda_k$  et  $\phi'(\lambda) \geq \omega_2\phi'(0)$  : STOP( $\lambda^* = \lambda_k$ ).

Si  $\phi(\lambda_k) > \phi(0) + \omega_1\phi'(0)\lambda_k$ , alors

$b_{k+1} = \lambda_k$ ,  $a_{k+1} = a_k$ , et aller à l'étape 2.

Si  $\phi'(\lambda) < \omega_2\phi'(0)$ , alors

$b_{k+1} = b_k$ ,  $a_{k+1} = \lambda_k$  et aller à l'étape 2.

Etape 2 :

Si  $b_{k+1} = 0$  déterminer  $\lambda_{k+1} \in ]a_{k+1}, +\infty[$ .

Si  $b_{k+1} \neq 0$  déterminer  $\lambda_{k+1} \in ]a_{k+1}, b_{k+1}[$ .

**Règle de Wolfe Forte** On obtient des contraintes plus fortes si l'on remplace (W2) par

$$|\nabla f(x_k + \lambda d_k)^T \cdot d_k| \leq -\omega_2 \nabla f(x_k)^T \cdot d_k$$

Les (W1) et (W3) sont les conditions de Wolfe fortes. La contrainte (W3) entraîne que  $\omega_2\phi'(0) \leq \phi'(\lambda) \leq -\omega_2\phi'(0)$  c-à-d.  $\phi'(\lambda)$  n'est pas "trop" positif.

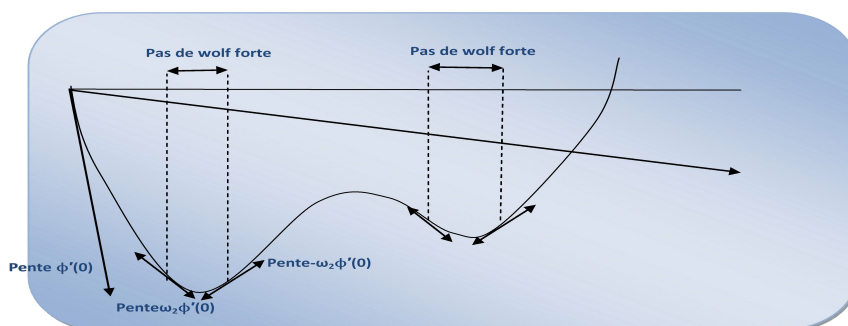


FIG 2.4 Rgle de Wolfe forte

**Remarque 2. 3.**

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

$$\begin{aligned} |\nabla^T f(x_k + \lambda_k d_k) d_k| &\leq -\omega_2 \nabla^T f(x_k) d_k \\ \Leftrightarrow \omega_2 \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + \lambda_k d_k) d_k \leq -\omega_2 \nabla^T f(x_k) d_k \\ \Rightarrow \omega_2 \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + \lambda_k d_k) d_k. \end{aligned} \quad (W2)$$

2) L'existence de valeurs  $\lambda$  de vérifiant les conditions de Wolfe faibles et fortes. est donnée par la Proposition suivante.

**Proposition 2. 1.** [20]

Soit  $f \in C^1$  et  $d$  une direction de descente de  $f$  en  $x$ , on suppose que  $f$  est minorée. Alors, si  $0 < \omega_1 < \omega_2 < 1$ , 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 [1996], cette règle consiste à choisir le pas satisfaisant aux conditions :

$$\begin{aligned} f(x_k + \lambda d_k) &\leq f(x_k) + \omega_1 \lambda \nabla^T f(x_k) d_k \\ \omega_2 \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + \alpha_k d_k) d_k \leq -\omega_2 \nabla^T f(x_k) d_k, \end{aligned}$$

où  $0 < \omega_1 < \omega_2 < 1$  et  $\omega_2 > 0$ .

**Remarque 2. 4.**

1) On voit bien que les conditions de Wolfe relaxée impliquent les conditions de Wolfe fortes. Effectivement (w4) est équivalente à (w1), tandis que pour le cas particulier  $\omega_2 = \omega_2 = \omega_2$ , (w5) est équivalente à (w3). En effet :

$$\begin{aligned} \omega_2 \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + \lambda_k d_k) d_k \leq -\omega_2 \nabla^T f(x_k) d_k \\ \Rightarrow \omega_2 \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + \lambda_k d_k) d_k \leq -\omega_2 \nabla^T f(x_k) d_k \\ \Rightarrow |\nabla^T f(x_k + \lambda_k d_k) d_k| &\leq -\omega_2 \nabla^T f(x_k) d_k. (w3) \end{aligned}$$

2) Les conditions de Wolfe relaxée impliquent les conditions de Wolfe faibles. Effectivement (w4) est équivalente à (w1), tandis que pour le cas particulier  $\omega_2 = \omega_2$  et  $\omega_2 = +\infty$ , (w5) est équivalente à (w2). En effet :

$$\begin{aligned} \omega_2' &= \nabla^T f(x_k) d_k \leq \nabla^T f(x_k + \lambda_k d_k) d_k \leq -\omega_2 \nabla^T f(x_k) d_k \\ \Rightarrow \omega_2 \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + \lambda_k d_k) d_k. (w2) \end{aligned}$$

## 2.4 Convergence des méthodes à directions de descente

### 2.4.1 Condition de Zoutendijk

Dans cette section on va étudier la contribution de la recherche linéaire inexacte à la convergence des algorithmes à directions de descente.

On dit qu'une 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

$$f(x_{k+1}) \leq f(x_k) - C \|\nabla f(x_k)\|^2 \cos^2 \theta_k, \quad (2.10)$$

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

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

Voici comment on se sert de la condition de Zoutendijk.

**Proposition 2. 2.**

Si la suite  $\{x_k\}$  générée par un algorithme d'optimisation vérifie la condition de Zoutendijk (2. 10) et si la suite  $\{f(x_k)\}$  est minorée, alors

$$\sum_{k \geq 1} \|\nabla f(x_k)\|^2 \cos^2 \theta_k < \infty. \quad (2.11)$$

**Preuve**

En sommant les inégalités (2. 10), on a

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

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

Il ya des propositions précises les circonstances dans lesquelles la condition de Zoutendijk (2. 10) est vérifiée avec les règles de la recherche linéaire exacte (Cauchy, Curry) et aussi les règles de la recherche linéaire inexacte (Armijo, Wolfe).

Ce qui concerne la proposition qui est vérifiée la condition de Zoutendijk avec la règle de Wolfe.

**Proposition 2. 3.**

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continuellement différentiable dans un voisinage de  $\Gamma = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ .

On considère un algorithme à directions de descente  $d_k$ , qui génère une suite  $\{x_k\}$  en utilisant la recherche linéaire de Wolfe (w1)-(w2).

Alors il existe une constante  $C > 0$  telle que, pour tout  $k \geq 1$ , la condition de Zoutendijk (2. 10) est vérifiée.

**Preuve**

Noton  $g_k = \nabla f(x_k)$  et  $g_{k+1} = \nabla f(x_k + \lambda_k d_k)$ .

D'après (W2)

$$g_{k+1}^T d_k \geq \omega_2 g_k^T d_k$$

$$\begin{aligned} \Rightarrow & (g_{k+1} - g_k)^T d_k \geq (\omega_2 - 1) g_k^T d_k \\ = & -(1 - \omega_2) g_k^T d_k = (1 - \omega_2) |g_k^T d_k| \\ \Leftrightarrow & (1 - \omega_2) |g_k^T d_k| \leq (g_{k+1} - g_k)^T d_k, \end{aligned}$$

et du fait que  $f$  est contiuement différentiable :

$$\begin{aligned} (1 - \omega_2) |g_k^T d_k| &= (1 - \omega_2) \|g_k\| \|d_k\| \cos \theta_k \\ &\leq \|g_{k+1} - g_k\| \|d_k\| \\ \Rightarrow & (1 - \omega_2) \|g_k\| \cos \theta_k \leq L \lambda_k \|d_k\| \\ \Rightarrow & \lambda_k \|d_k\| \leq \frac{(1 - \omega_2)}{L} \|g_k\| \cos \theta_k, \end{aligned}$$

en utilisant (w1), on aura :

$$\begin{aligned} f(x_k + \lambda_k d_k) &\leq f(x_k) + \omega_1 \lambda_k g_k^T d_k \\ \Rightarrow & f(x_k + \lambda d_k) \leq f(x_k) + \omega_1 \lambda g_k^T d_k \leq f(x_k) + |\omega_1 \lambda g_k^T d_k| \\ \Rightarrow & f(x_k + \lambda d_k) \leq f(x_k) + \omega_1 \lambda |g_k^T d_k| \leq f(x_k) - \omega_1 \lambda g_k^T d_k \\ \Rightarrow & f(x_k + \lambda d_k) \leq f(x_k) - \omega_1 \lambda \|g_k\| \|d_k\| \cos \theta_k \\ \Rightarrow & f(x_k + \lambda d_k) \leq f(x_k) - \frac{\omega_1(1 - \omega_2)}{L} \|g_k\|^2 \cos^2 \theta_k. \end{aligned}$$

On en déduit (2. 10).  $\square$

## 2.5 Méthodes itératives d'optimisation sans contraintes

A travers ce chapitre et du suivant, on s'intresse à present à la description des algorithmes itératifs (ou méthodes itératives) qui ont été implémentés et qui permettent la résolution des problèmes d'optimisation non linéaire. Il convient de souligner que la plupart des algorithmes d'optimisation, avec contrainte ou non, fonctionnent selon un schéma général consistant, à chaque itération, à se rapprocher du minimum par la résolution d'un sous-problème de minimisation.



Nous considérons ici les méthodes permettant de résoudre un problème d'optimisation sans contraintes (appelées aussi parfois méthodes d'optimisation directe), soit le problème

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

pour lequel nous commencerons par décrire les méthodes suivantes :

- ▶ Les méthodes de gradient
- ▶ Les méthodes de Newton
- ▶ Les méthodes utilisant des directions conjuguées.

Ces méthodes utilisent des dérivées (et donc la propriété de différentiabilité de  $f$ ) à l'exception des méthodes de directions conjuguées (sauf dans le cas particulier de la méthode du gradient conjugué) basée, elle, sur des propriétés plus géométriques.

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.

Après les avoir décrites, nous donnons la définition d'une forme quadratique et Principe des méthodes de descente et de gradient.

**Définition 2. 2.**

Soit  $H$  une matrice symétrique  $n \times n$  et  $b \in \mathbb{R}^n$ . On appelle forme quadratique la fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  définie par

$$f(x) = \frac{1}{2}x^T H(x)x - b^T x. \tag{2.12}$$

Lorsque la matrice  $H$  est définie positive (resp. semi-définie positive), on dira que  $f(x)$  est une forme quadratique définie positive (resp. semi-définie positive).

### 2.5.1 Principe des méthodes de descente

Le principe d'une méthode de descente consiste à faire les itérations suivantes

$$x_{k+1} = x_k + \lambda_k d_k \quad \lambda_k > 0,$$

tout en assurant la propriété

$$f(x_{k+1}) < f(x_k). \tag{2.13}$$

Le vecteur  $d_k$  est la direction de descente en  $x_k$ . Le scalaire  $\lambda_k$  est appelé le pas de la méthode à l'itération  $k$ . On peut caractériser les directions de descente en  $x_k$  à l'aide du gradient.

### 2.5.2 Principe des méthodes du gradient (méthode de la plus forte descente)

On cherche à déterminer la direction de descente qui fait décroître  $\varphi(\lambda) = f(x_k + \lambda_k d_k)$  le plus vite possible (au moins localement). Pour cela on va essayer de minimiser la dérivée de  $\varphi(\lambda)$  en 0. On a  $\varphi'(0) = \nabla f(x)^T d$ , et on cherche  $d$  solution du problème

$$\min_{d \in \mathbb{R}^n, \|d\|=1} \varphi'(0).$$

La solution est bien sûr

$$d = -\frac{\nabla f(x)}{\|\nabla f(x)\|}.$$

En vertu de l'inégalité de Schwartz. Il y a ensuite de nombreuses façon d'utiliser cette direction de descente.

On peut par exemple utiliser un pas fixé a priori  $\lambda_k = \lambda > 0, \forall k$ . On obtient alors la méthode du gradient simple :

$$\begin{cases} d_k = -\nabla f(x_k), \\ x_{k+1} = x_k + \lambda d_k. \end{cases} \quad (2.14)$$

Sous certaines hypothèses de régularité ( $f$  deux fois différentiable) cette méthode converge si  $\lambda$  est choisi assez petit.

Où bien consiste à faire les itérations suivantes

$$\begin{cases} d_k = -\nabla f(x_k), \\ x_{k+1} = x_k + \lambda_k d_k. \end{cases} \quad (2.15)$$

Où  $\lambda_k$  est choisi de manière à ce que

$$f(x_k + \lambda_k d_k) \leq f(x_k + \lambda d_k), \forall \lambda > 0. \quad (2.16)$$

On obtient alors la méthode du gradient à pas optimal, cette méthode possède une propriété intéressante :

**Proposition 2. 4.**

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction différentiable. Les directions de descente  $d_k$  générées par la méthode (2. 15) et (2. 16) vérifient

$$d_{k+1}^T d_k = 0. \quad (2.17)$$

**Preuve**

Si on introduit la fonction

$$\varphi(\lambda) = f(x_k + \lambda d_k),$$

on a

$$\varphi'(\lambda) = \nabla f(x_k + \lambda d_k)^T d_k,$$

et puisque  $\varphi$  est dérivable on a nécessairement

$$\varphi'(\lambda) = 0,$$

donc

$$\nabla f(x_k + \lambda_k d_k)^T d_k = \nabla f(x_{k+1})^T d_k = -d_{k+1}^T d_k = 0. \quad \square$$

**Exemple 2. 1.**

► Calcul du pas optimal dans le cas quadratique :

On a  $f(x) = \frac{1}{2}x^T Hx - b^T x$  avec  $H > 0$  et on note  $\varphi(\lambda) = f(x_k + \lambda d_k)$ .

Le pas optimal  $\lambda_k$  est caractérisé par

$$\varphi'(\lambda) = 0,$$

on a donc

$$\nabla f(x_k + \lambda_k d_k)^T d_k = (H(x_k + \lambda_k d_k) - b)^T d_k = 0,$$

soit

$$(\nabla f(x_k) + \lambda_k H d_k)^T d_k = 0,$$

on obtient donc

$$\lambda_k = -\frac{\nabla f(x_k)^T d_k}{d_k^T H d_k}, \tag{2.18}$$

qui est bien positif car  $d_k$  est une direction de descente et

$$d_k^T H d_k > 0 \text{ (car } H > 0 \text{)}.$$

La méthode du gradient à pas optimal peut donc s'écrire (dans le cas quadratique)

$$\begin{aligned} d_k &= b - Hx_k, \\ \lambda_k &= \frac{d_k^T d_k}{d_k^T H d_k}, \\ x_{k+1} &= x_k + \lambda_k d_k. \end{aligned}$$

**Exemple 2. 2**

► Méthode du gradient simple dans le cas quadratique

Dans le cas où  $f(x) = \frac{1}{2}x^T H(x)x - b^T x$  la méthode du gradient simple peut s'écrire

$$\begin{aligned} d_k &= b - Hx_k, \\ x_{k+1} &= x_k + \lambda d_k, \end{aligned}$$

où  $\lambda > 0$  est fixé a priori. Il existe bien sûr des conditions sur pour que la méthode converge. Nous illustrons ici le fonctionnement de la méthode dans le cas  $n = 2$  sur une petite simulation.

► Méthode du gradient à pas optimal dans le cas quadratique

Dans le cas où  $f(x) = \frac{1}{2}x^T H(x)x - b^T x$  la méthode du gradient à pas optimal peut s'écrire :

$$\begin{aligned} d_k &= b - Hx_k, \\ \lambda_k &= \frac{d_k^T d_k}{d_k^T H d_k}, \\ x_{k+1} &= x_k + \lambda_k d_k. \end{aligned}$$

Nous illustrons ici le fonctionnement de la méthode dans le cas  $n = 2$  sur une petite simulation.

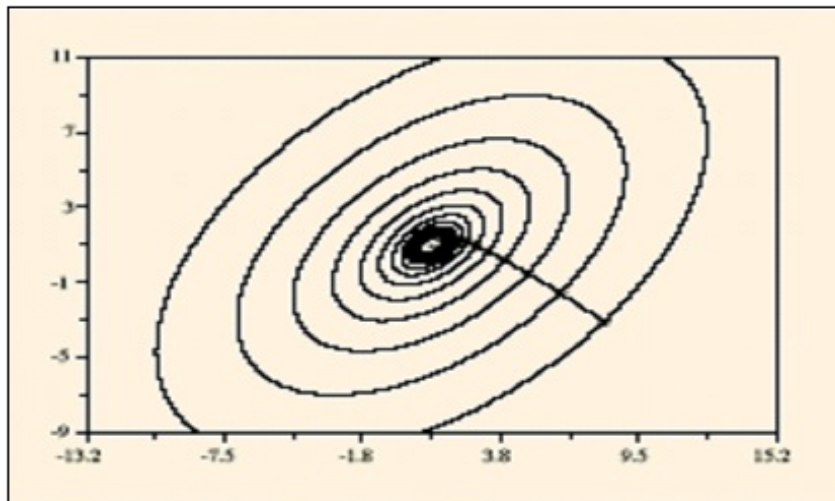


FIG 2.6 Illustration de la convergence plus rapide de la méthode du gradient

### Algorithme de la méthode de la plus forte pente

Etape initiale :

Choisir un  $\varepsilon > 0$ .

Choisir un point initiale  $x_1$ .

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

Etape principale :

Si  $\|\nabla f(x)\| < \varepsilon$  stop.

Sinon poser  $d_k = -\nabla f(x_k)$  et soit  $\lambda_k$  la solution optimale de la recherche linéaire

$$\min \{f(x_k + \lambda d_k); \lambda > 0\}.$$

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

Remplacer  $k$  par  $k + 1$  et répéter l'étape principale.

### Inconvénients de la méthode de la plus forte pente

- ▶ Lenteur de la méthode au voisinage des points stationnaires.
- ▶ Cette méthode travaille de façon performante dans les premières étapes de l'algorithme. Malheureusement, dès qu'on s'approche du point stationnaire,

La méthode devient très lente. On peut expliquer intuitivement ce phénomène par les considérations suivantes

$$f(x_k + \lambda d_k) = f(x_k) + \lambda \nabla f(x_k)^T d + \lambda \|d\| \alpha(x_k; \lambda d),$$

où  $\alpha(x_k; \lambda d) \rightarrow 0$  quand  $\lambda d \rightarrow 0$ .

Si  $d_k = -\nabla f(x_k)$ , on obtient :  $x_{k+1} = x_k - \lambda \nabla f(x_k)$  et par conséquent

$$f(x_{k+1}) - f(x_k) = \lambda \left[ -\|\nabla f(x_k)\|^2 + \|\nabla f(x_k)\| \alpha(x_k; \lambda \nabla f(x_k)) \right].$$

D'après l'expression précédente, on voit que lorsque  $x_k$  s'approche d'un point stationnaire, et si  $f$  est continument différentiable, alors  $\|\nabla f(x_k)\|$  est proche de zéro. Donc le terme à droite s'approche de zéro, indépendamment de  $\lambda$ , et par conséquent  $f(x_{k+1})$  ne s'éloigne pas beaucoup de  $f(x_k)$  quand on passe du point  $x_k$  au point  $x_{k+1}$ .

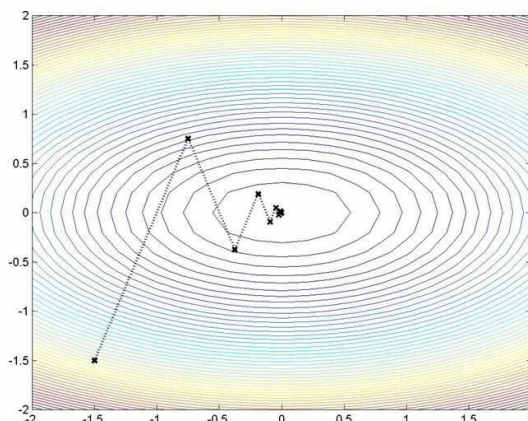


FIG 2.7 Exprime les direction de la mthode de la plus fort pente

### 2.5.3 Les méthodes utilisant des directions conjuguées

#### Description de la méthode

Ces méthode sont basées sur l'important concept de la conjugaison et ont été développées afin de résoudre le problème quadratique

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T A x + b^T x + c.$$

Où  $A \in \mathbb{R}^{n \times n}$  est symétrique et définie positive,  $b \in \mathbb{R}^n$  et  $c \in \mathbb{R}$ . Les méthodes à directions conjuguées peuvent résoudre les problèmes de cette forme en au plus  $n$  itérations et, contrairement aux méthodes présentées jusqu'à présent, elle n'utilisent pas de dérivées, sauf dans le cas particulier de la méthode du gradient conjugué. Donnons la définition de la notion de "conjugaison" :

**Définition 2. 3**

Soient  $A$  une matrice  $n \times n$  symétrique et définie positive et un ensemble de directions non nulles  $\{d_1, d_2, \dots, d_k\}$ . Ces directions sont dites  $A$ -conjuguées si

$$d_i^T A d_j = 0, \quad \forall i, j \text{ tels que } i \neq j. \quad (2.19)$$

**Propriété 2. 1**

Si  $d_1, \dots, d_k$  sont  $A$ -conjuguées, alors elles sont linéairement indépendantes.

**Propriété 2. 2**

Comme les directions  $A$ -conjuguées sont linéairement indépendantes, alors l'espace vectoriel engendré par un ensemble de  $n$  directions  $A$ -conjuguées est de dimension  $n$ .

Etant donné un ensemble de  $n$  directions  $A$ -conjuguées  $d_0, d_1, \dots, d_{n-1}$ , la méthode de directions conjuguées correspondante est donnée par

$$x_{k+1} = x_k + \lambda_k d_k, \quad k = 0, \dots, n - 1,$$

où  $x_0$  est un vecteur de départ choisi arbitrairement et où les  $\lambda_k$  sont obtenus par minimisation monodimensionnelle le long de  $d_k$ .

Le principal résultat concernant les méthodes utilisant des directions conjuguées est qu'à chaque itération  $k$ , la méthode minimise  $f$  sur le sous-espace généré par les  $k$  premières directions  $A$ -conjuguées utilisées par l'algorithme. A la  $n^{\text{ième}}$  itération au plus tard, ce sous-espace inclura alors le minimum global de  $f$  grâce à la propriété d'indépendance linéaire des directions  $A$ -conjuguées qui assurent qu'à l'itération  $n$ , l'espace vectoriel généré par les  $n$  directions  $A$ -conjuguées ne sera autre que  $\mathbb{R}^n$ .

**Remarque 2. 5**

La notion de conjugaison n'a pas de sens dans le cas non quadratique.

### La méthode du gradient conjugué non linéaire

L'idée de la méthode est de construire itérativement des directions  $d_0, \dots, d_k$  mutuellement conjuguées. A chaque étape  $k$  la direction  $d_k$  est obtenue comme combinaison linéaire du gradient en  $x_k$  et de la direction précédente  $d_{k-1}$ , les coefficients étant choisis de telle manière que  $d_k$  soit conjuguée avec toutes les directions précédentes. On s'intéresse ici à la minimisation d'une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , non nécessairement quadratique :

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

et on cherche à étendre la méthode du gradient conjugué a ce problème. Il y'a plusieurs manières de le faire et peu de critères permettant de dire laquelle est la meilleure. Une extension possible consiste simplement a reprendre les formules utilisées dans le cas quadratique. On se propose donc d'étudier les méthodes où la direction  $d_k$  est définie par la formule de récurrence suivante ( $\beta_k \in \mathbb{R}$ )

$$d_k = \begin{cases} -g_1 & \text{si } k = 1, \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 2, \end{cases} \quad (2.20)$$

ou  $g_k = \nabla f(x_k)$ ,  $\beta_k \in \mathbb{R}$  et  $\{x_k\}$  est générée par la formule :

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

le pas  $\lambda_k \in \mathbb{R}$  étant déterminé par une recherche linéaire.

Ces méthodes sont des extensions de la méthode du gradient conjuguée si  $\beta_k$  prend l'une des valeurs suivantes

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \text{ Gradient conjugué variante Polak-Ribière-Polyak (PRP)} \quad (2.21)$$

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \text{ Gradient conjugué variante Fletcher-Reeves (FR)} \quad (2.22)$$

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \text{ Gradient conjugué variante Hestenes-Stiefel (HS)} \quad (2.23)$$

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}} \text{ Gradient conjugué variante descente conjugué (CD)} \quad (2.24)$$



$$\beta_k^{LS} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}} \text{ Gradient conjugué variante de Liu-Storey (LS)} \quad (2.25)$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \text{ Gradient conjugué variante de Dai-Yuan (DY)} \quad (2.26)$$

où  $y_{k-1} = g_k - g_{k-1}$ .

Pour faciliter la présentation, nous appelons les méthodes qui correspondent à (2. 21) - (2. 26) la méthode PRP, la méthode FR, la méthode HS, la méthode CD, et la méthode LS, et la méthode DY, respectivement.

Dans le cas où  $f$  est une fonction quadratique strictement convexe avec une recherche linéaire exacte toutes les variantes de  $\beta_k$  ont la même valeur :

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

Si  $f$  est quelconque, il n'en est plus de même et on parle respectivement de *méthode de Polak-Ribière-Ployak n* ([35, 1969])-( [36, 1969]) *méthode de Fletcher-Reeves* ([18, 1964 ]), *méthode de la descente conjuguée* ([16, 1987]) ou de *méthode de Dai-Yuan* ([10, 1999]) selon que l'on utilise  $\beta_k^{PRP}$ ,  $\beta_k^{FR}$ ,  $\beta_k^{CD}$  ou  $\beta_k^{DY}$  à la place de  $\beta_k$  dans (2. 20).

Pour que les méthodes ainsi définies soient utilisables, il faut répondre aux deux questions suivantes.

► Les directions  $d_k$  définies par (2. 20) sont-elles des directions de descente de  $f$  ?

► Les méthodes ainsi définies sont-elles convergentes ?

En ce qui concerne la première question remarquons que quel que soit  $\beta_k \in \mathbb{R}$ ,  $d_k$  est une direction de descente si on fait de la recherche linéaire exacte, c'est- a-dire si le pas  $\lambda_{k-1}$  est un point stationnaire de  $\lambda \rightarrow f(x_{k-1} + \lambda d_{k-1})$ . En effet, dans ce cas  $g_k^T d_{k-1} = 0$  et on trouve lorsque  $g_k \neq 0$  :

$$\begin{aligned} d_k^T g_k &= (-g_k + \beta_k d_{k-1})^T g_k \\ &= -\|g_k\|^2 + \beta_k d_{k-1}^T g_k = -\|g_k\|^2 < 0 \end{aligned}$$

Cependant, il est fortement déconseillé de faire de la recherche linéaire exacte lorsque  $f$  n'est pas quadratique : le coût de détermination de  $k$  est excessif.

L'efficacité de la méthode du gradient conjugué repose essentiellement sur deux points :

- ▶ La recherche linéaire (détermination du pas optimal) doit être exacte,
- ▶ Les relations de conjugaison doivent être précises.

La recherche du pas optimal doit être réalisée à l'aide d'un algorithme spécifique, puisque  $f$  est quelconque. Par contre la notion de conjugaison n'a pas de sens dans le cas non quadratique.

L'étude des propriétés de convergence de quelques méthodes du gradient conjugué non linéaire est l'objectif du troisième chapitre.

## Chapitre 3

# Synthèse sur la convergence de quelques méthodes du gradient conjugué non linéaire avec la recherche linéaire inexacte

Dans ce chapitre on va essayer de présenter une synthèse sur les différents résultats de convergence des méthodes du gradient conjugué pour la minimisation des fonctions sans contraintes. Ces méthodes seront utilisées avec une recherche linéaire inexacte (Wolfe forte ou faible). L'analyse couvre quatre classes de méthodes qui sont globalement convergentes pour des fonctions régulières non nécessairement convexes. Dans la première famille, ce sont certaines propriétés de la méthode de Fletcher-Reeves qui jouent un rôle crucial, tandis que la seconde famille partage avec la méthode de Polak-Ribière-Polyak une propriété importante. La troisième concerne la méthode de la descente conjuguée et la dans dernière famille on va présenter quelques propriétés de la nouvelle méthode du gradient conjugué non linéaire dite de Dai-Yuan.

### 3.1 méthodes du gradient conjugué non linéaire

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

$$(P) : \min \{f(x) : x \in \mathbb{R}^n\} \quad (3.1)$$

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

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 méthodes du gradient conjugué génèrent des suites  $\{x_k\}$  de la forme suivante :

$$x_{k+1} = x_k + \lambda_k d_k \quad \lambda_k > 0, \quad (3.2)$$

où la direction recherchée est définie par la formule de récurrence suivante :

$$d_k = \begin{cases} -g_1 & \text{si } k = 1, \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 2, \end{cases} \quad (3.3)$$

Le coefficient  $\beta_k$  détermine la méthode du gradient conjugué en question (Fletcher Reeves, Polak Ribière Polyak, . . .) (voir Chapitre 2)

Le pas  $\lambda_k \in \mathbb{R}^+$  étant déterminé par une recherche linéaire.

**Supposition 3. 1**

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

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

$$\exists L > 0 \text{ tel que } \|g(x) - g(\tilde{x})\| \leq L \|x - \tilde{x}\|, \forall x, \tilde{x} \in \mathcal{N} \quad (3.4)$$

**Remarque 3. 1**

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

$$\|g(x)\| \leq \gamma, \forall x \in \mathcal{L} \quad (3.5)$$

**Définition 3. 1([20, 1992])**

On dit que  $d_k$  est une direction de descente suffisante si

$$g_k^T d_k < -c \|g_k\|^2. \quad (3.6)$$

Rappelons les conditions de Wolfe faibles :

$$f(x_k + \lambda d_k) \leq f(x_k) + \omega_1 \lambda \nabla f(x_k)^T \cdot d_k. \quad (3.7)$$

$$\nabla f(x_k + \lambda d_k)^T \cdot d_k \geq \omega_2 \nabla f(x_k)^T \cdot d_k. \quad (3.8)$$

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

avec  $0 < \omega_1 < \omega_2 < 1$  :

Les conditions de Wolfe fortes :

$$f(x_k + \lambda d_k) \leq f(x_k) + \omega_1 \lambda \nabla f(x_k)^T \cdot d_k. \quad (3.9)$$

$$|\nabla f(x_k + \lambda d_k)^T \cdot d_k| \leq -\omega_2 \nabla f(x_k)^T \cdot d_k \quad (3.10)$$

Les conditions de Wolfe relaxée :

$$f(x_k + \lambda d_k) \leq f(x_k) + \omega_1 \lambda \nabla^T f(x_k) d_k \quad (3.11)$$

$$\omega_2 \nabla^T f(x_k) d_k \leq \nabla^T f(x_k + \alpha_k d_k) d_k \leq -\omega_2 \nabla^T f(x_k) d_k, \quad (3.12)$$

où

$$0 < \omega_1 < \omega_2 < 1 \text{ et } \omega_2 > 0.$$

Présentons maintenant un théorème fondamental qui assure la satisfaction de la condition de Zoutendijk (voir chapitre2), pour toute méthode du type (3. 2)-(3. 3), dans laquelle le pas  $d_k$  est déterminé par la règle de Wolfe faible (3. 7)-(3. 8).

Ce théorème a été démontré par Zoutendijk ([46, 1970]) et Powell ([38, 1971]).

**Théorème 3. 1**

Considérons une méthode du type (3. 2) et (3. 3) dans la quelle  $d_k$  est une direction de descente et le pas  $\lambda$  est déterminé par la règle de Wolfe faible (3. 7)-(3. 8) avec  $0 < \omega_1 < 1/2$  : Considérons aussi que la supposition 3. 1 soit satisfaite. Alors pour une telle méthode la condition de Zoutendijk suivante :

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

## 3.2 Méthode de Fletcher-Reeves

La méthode de Fletcher-Reeves ([17, 1964]) est une extension directe de la méthode du gradient conjugué linéaire au cas des fonctions quelconques.

Appliquée à une fonction quadratique, elle est identique au gradient conjugué linéaire.

Rappelons (chapitre2) que pour la méthode de Fletcher-Reeves la variante  $\beta_k$  est :

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \text{ Gradient conjugué variante Fletcher-Reeves} \quad (3.14)$$

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

### 3.2.1 Algorithme de la méthode de Fletcher-Reeves

Cette méthode est très intéressante, d'une part parce qu'elle nécessite le stockage de très peu d'informations, d'autre part, par sa vitesse de convergence très supérieure à celle des algorithmes du gradient classique.

**Algorithme 3. 1 de la méthode de Fletcher-Reeves**

**Etape 0 : (initialisation)**

Soit  $x_0$  le point de départ,  $g_0 = \nabla f(x_0)$ , poser  $d_0 = -g_0$   
Poser  $k = 0$  et aller à l'étape 1.

**Etape 1 :**

Si  $g_k = 0$  : STOP (  $x^* = x_k$ ). "Test d'arrêt"  
Sinon aller à l'étape 2.

**Etape 2 :**

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

où

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

### 3.2.2 La propriété de descente de la méthode de FR

Powell ([37, 1984]) a démontré la satisfaction de la propriété de descente de la fonction objectif pour la méthode de Fletcher-Reeves avec recherche linéaire exacte.

Al-Baali ([3, 1985]) a démontré la satisfaction de la propriété de descente de la fonction objectif pour la méthode de Fletcher-Reeves avec la recherche linéaire inexacte de Wolfe forte.

J. C. Gilbert et Nocedal([20, 1992]) ont généralisé ce résultat pour toute méthode du type (3. 2)-(3. 3) dont

$$|\beta_k| \leq \beta_k^{FR}. \tag{3.15}$$

On considère que la supposition (3. 1) est satisfaite. Considérons une méthode du type (3. 2) et (3. 3) avec  $\beta_k$  satisfaisant à (3. 14) et le pas  $\lambda_k$  vérifiant la règle de Wolfe forte (3. 10) où  $\omega_2 \in ]0, \frac{1}{2}[$ .

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

$$\frac{-1}{1 - \omega_2} \leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\omega_2 - 1}{1 - \omega_2}; \quad k = 1, \dots \tag{3.16}$$

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

**Preuve.** ([20, 1992])

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 :

$$0 < \omega_2 < \frac{1}{2} \Rightarrow \begin{cases} \frac{-1}{1-\omega_2} \leq -1 \\ \frac{2\omega_2-1}{1-\omega_2} \geq -1 \end{cases}$$

2) Supposons que (3. 16) est satisfaite pour  $k$  et démontrons qu'elle le sera pour  $k + 1$  :

Supposons que :

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

On a :

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

D'autre part de (3. 14) on aura :

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

d'où :

$$\frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} = -1 + \frac{\beta_{k+1}}{\beta_k^{FR}} \frac{d_k^T g_{k+1}}{\|g_k\|^2}. \quad (**)$$

En utilisant la condition de recherche linéaire (4. 10) on obtient :

$$|d_k^T g_{k+1}| \leq -\omega_2 d_k^T g_k \implies \omega_2 |\beta_{k+1}| d_k^T g_k \leq \beta_{k+1} d_{k+1}^T g_{k+1} \leq -\omega_2 |\beta_{k+1}| d_k^T g_k$$

Remplaçons ceci dans (\*\*)

$$-1 + \omega_2 \frac{|\beta_{k+1}|}{\beta_{k+1}^{FR}} \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \omega_2 \frac{|\beta_{k+1}|}{\beta_{k+1}^{FR}} \frac{d_k^T g_k}{\|g_k\|^2}.$$

De (\*) on aura

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

$$-1 - \frac{|\beta_{k+1}| \omega_2}{\beta_{k+1}^{FR} (1 - \omega_2)} \leq \frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{|\beta_{k+1}| \omega_2}{\beta_{k+1}^{FR} (1 - \omega_2)}.$$

et de (3.15)

$$\frac{-\omega_2}{1 - \omega_2} \leq \frac{|\beta_{k+1}|}{\beta_{k+1}^{FR}} \leq 1.$$

d'où :

$$\frac{-1}{1 - \omega_2} \leq \frac{d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^2} \leq \frac{2\omega_2 - 1}{1 - \omega_2}.$$

Ce qui découle la démonstration.  $\square$

**Corollaire 3. 1.**

D'après (4. 18) on déduit que pour tout  $k \geq 1$  :  $d_k$  est une directions de descente avec  $C = \frac{2\omega_2 - 1}{1 - \omega_2}$

**Preuve.**

$$\frac{-1}{1 - \omega_2} \leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\omega_2 - 1}{1 - \omega_2} \Rightarrow d_k^T g_k \leq -C \|g_k\|^2 .ou C = \frac{1 - 2\omega_2}{1 - \omega_2}$$

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 :

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

Puisque

$$\lambda_k = \arg \min_{\lambda > 0} f(x_k + \lambda_k d_k) = \arg \min_{\lambda > 0} \varphi_k(\lambda)$$

Donc  $\lambda_k$  vérifie la condition nécessaire d'optimalité  $\dot{\varphi}_k(\lambda_k) = \nabla f(x_k + \lambda_k d_k)^T \cdot d_k = g_{k+1}^T d_k = 0, \forall k \geq 1.$



**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

### 3.2.3 Convergence de la méthode de Fletcher-Reeves

Le premier résultat de convergence de la méthode du gradient conjugué non linéaire (version Fletcher Reeves) avec des recherches linéaires inexactes (recherche linéaire inexacte de Wolfe forte (3. 9)-(3. 10) où  $\omega_2 < \frac{1}{2}$ ) était démontré par Al-Baali ([3, 1985]).

Touati Ahmed et Story ([40, 1990]) ont généralisé ce résultat pour

$$0 \leq \beta_k \leq \beta_k^{FR}.$$

Gilbert et Nocedal ([20, 1992]) ont généralisé ce résultat pour

$$|\beta_k| \leq \beta_k^{FR}.$$

#### **Théorème 3. 3**

Supposons que l'hypothèse 3. 1 soit satisfaite. Considérons une méthode du type (3. 2) et (3. 3) dont  $\beta_k$  satisfait (3. 14) et le pas  $\lambda_k$  satisfait aux conditions de Wolfe fortes (3. 9)-(3. 10) où  $\omega_2 \in ]0, \frac{1}{2}[$  Alors

cette méthode est globalement convergente, dans le sens suivant :

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.17)$$

#### **Preuve. ([20, 1992])**

Puisque les conditions du théorème 3. 2 sont satisfaites alors on a :

$$\frac{-1}{1 - \omega_2} \leq \frac{d_k^T g_k}{\|g_k\|^2} \leq \frac{2\omega_2 - 1}{1 - \omega_2} \Rightarrow -\omega_2 d_{k-1}^T g_{k-1} \leq \frac{\omega_2}{1 - \omega_2} \|g_{k-1}\|^2.$$

D'autre part de (3. 10)

$$|d_k^T g_{k+1}| \leq -\omega_2 d_k^T g_k \Rightarrow |d_{k-1}^T g_k| \leq -\omega_2 d_{k-1}^T g_{k-1}$$

d'où

$$|d_{k-1}^T g_k| \leq -\omega_2 d_{k-1}^T g_{k-1} \leq \frac{\omega_2}{1 - \omega_2} \|g_{k-1}\|^2 \quad (*)$$

De (3. 2), (4. 15) et (\*) :

$$\begin{aligned} \|d_k\|^2 &= \left| \|g_k\|^2 - 2\beta_k d_{k-1}^T g_k + \beta_k^2 \|d_{k-1}\|^2 \right| \\ &\leq \|g_k\|^2 + |2\beta_k d_{k-1}^T g_k| + \beta_k^2 \|d_{k-1}\|^2 \\ &\leq \frac{1 + \omega_2}{1 - \omega_2} \|g_k\|^2 + (\beta_k^{FR})^2 \|d_{k-1}\|^2 \end{aligned}$$

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

Posons

$\hat{\omega} = \frac{1+\omega_2}{1-\omega_2}$  on aura :

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

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

$$\|g_k\| \geq \omega > 0; \forall k \Rightarrow \|g_k\|^{-2} \leq \omega^{-2}$$

de (4. 5) on a :

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

d'où

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} \geq \frac{\omega^2}{\hat{\omega} \lambda^4} \sum_{k \geq 1} \frac{1}{k} > \infty \quad (**)$$

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

D'autre part, puisque les conditions du Théorème 3. 1, et du théorème 3. 2 sont satisfaites on a :

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

et

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

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

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 (\*\*), d'où le résultat :

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

**Remarque 3. 3**

Lui, Han et Yuan ([29, 1995]) ont prouvé le résultat (4. 19) pour cette méthode si  $\omega_1 = \omega_2$

Dai et Yuan([10, 1996]) l'ont encore démontré mais par une approche plus simple.

La méthode de Fletcher-Reeves possède de bonne propriété théoriquement, 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'élargisse pour un grand nombre d'itérations, et c'est ce qu'il nous oblige à réinitialiser en posant  $\beta_k^{FR} = 0$ .

Powell ([39, 1977]) est le premier qui a observé ce comportement, ainsi il a donné un contre exemple avec une recherche linéaire exacte. On évite cet inconvénient en réinitialisant chaque  $n$  itération par exemple.

Nemirovsky et Yudin ([32, 1983]) ont démontré par un contre exemple que la méthode de Fletcher-Reeves converge plus lentement que la méthode la plus raide descente. (cas d'une fonction quadratique avec recherche linéaire exacte).

### 3.3 Méthode de Polak-Ribière-Polyak

Cette méthode fut découverte par Polak, Ribière ([34, 1969]) et Polyak ([35, 1969]).

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

Rappelons que pour la méthode de Polak-Ribière-Polyak la variante  $\beta_k$  est :

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \quad (3.18)$$

**Algorithme 3. 2 de la méthode de Polak-Ribière-Polyak**

**Etape 0 : (initialisation)**

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

Poser  $k = 0$  et aller à l'étape 1.

**Etape 1 :**

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

Si non aller à l'étape 2.

**Etape 2 :**

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

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

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

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.

### 3.3.1 Convergence de la méthode de PRP

La convergence de cette méthode est assurée pour une fonction fortement convexe avec recherche linéaire, mais si  $f$  n'est pas convexe elle ne converge pas.

Pour remédier a cet inconvénient Powell ([36, 1986]) a modifié le choix de  $\beta_k^{PRP}$

afin d'assurer la convergence avec une recherche linéaire exacte.

Gilbert et Nocedal ([20, 1992]) ont proposé une nouvelle méthode qui converge dans le sens (4. 19) avec ou bien une recherche linéaire exacte ou inexacte.

Le résultat suivant est dû à Polak et Ribière ([34, 1969]).

**Théorème 3. 4.**

Si  $f$  est fortement convexe, continûment différentiable 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  $f$ .

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

**Preuve.** ([19])

Montrons dans un premier temps que

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

est uniformément positif.

Grâce à la recherche linéaire exacte, on a

$$\begin{aligned} d_{k-1}^T y_{k-1} &= d_{k-1}^T (g_k - g_{k-1}) \\ &= -d_{k-1}^T g_{k-1} = -(-g_{k-1} + \beta_{k-1}^{PRP} d_{k-2})^T g_{k-1} \\ &= \|g_{k-1}\|^2 \end{aligned}$$

La forte convexité de  $f$  implique que

$$d_{k-1}^T y_{k-1} = \frac{1}{\alpha_{k-1}} (x_k - x_{k-1})^T y_{k-1} \geq \frac{\eta}{\alpha_{k-1}} \|x_k - x_{k-1}\|^2$$

où  $\eta > 0$  est le module de forte convexité de  $f$ .

On en déduit, en utilisant la constante de lipschitz  $L$  de  $g$  :

$$|\beta_k^{PRP}| = \frac{|g_k^T y_{k-1}|}{\|g_{k-1}\|^2} = \frac{|g_k^T y_{k-1}|}{d_{k-1}^T y_{k-1}} \leq \frac{\alpha_{k-1} L \|g_k\| \|x_k - x_{k-1}\|}{\eta \|x_k - x_{k-1}\|^2} = \frac{L}{\eta} \frac{\|g_k\|}{\|d_{k-1}\|}$$

On peut alors borner  $\|d_k\|$  par :

$$\begin{aligned} \|d_k\| &\leq \|g_k\| + |\beta_k^{PRP}| \|d_{k-1}\| \\ &\leq \|g_k\| + \frac{L}{\eta} \frac{\|g_k\|}{\|d_{k-1}\|} \|d_{k-1}\| \\ &\leq \left(1 + \frac{L}{\eta}\right) \|g_k\| \end{aligned}$$

Ensuite

$$d_k^T g_k = -\|g_k\|^2 \leq -\left(1 + \frac{L}{\eta}\right) \|g_k\| \|d_k\|$$

ou encore

$$\cos \theta_k = \frac{-d_k^T g_k}{\|g_k\| \|d_k\|} \geq \left(1 + \frac{L}{\eta}\right)^{-1}$$

D'après la proposition 2. 1 et la recherche linéaire exacte, la condition de Zoutendijk est vérifié. Mais  $f$  et donc  $\{f(x_k)\}$  est bornée inférieurement (car  $f$  est fortement convexe).

On en déduit que  $g_k \rightarrow 0$ .

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

D'autre part,  $\{x_k\}$  est bornée ( $f$  est fortement convexe) et possède donc des sous suites convergentes.

La limite de celles-ci ne peut être que l'unique minimum  $x^*$  de  $f$  (car  $g_k \rightarrow 0$ ).

Donc toute la suite  $\{x_k\}$  converge vers  $x^*$ .  $\square$

**Remarque 3. 4**

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

Powell ([37, 1984]) a donné un exemple de fonction (de 3 variables, deux fois continûment différentiable) pour laquelle l'algorithme génère une suite  $\{x_k\}$  dont aucun des points d'adhérence n'est stationnaire.

En 1986 Powell([36, 1986]), il a modifié la variante  $\beta_k^{PRP}$  en évitant les valeurs négatifs, autrement dit si à l'itération  $k$  on a : si  $\beta_k^{PRP} < 0$ , on redémarre en posant  $\beta_k^{PRP} = 0$  (prendre la direction de la plus profonde pente)

$$\beta_k = \max \{0, \beta_k^{PRP}\}$$

Ce choix assure la convergence si le pas  $\lambda_k$  est déterminé par la règle de Wolfe forte.

**Remarque 3. 5**

Gilbert et Nocedal ([20, 1992]) ont assuré la convergence avec une recherche linéaire exacte ou inexacte, en hybridant les méthodes de  $FR$  et de  $PRP$ .

Donc cette nouvelle méthode consiste à prendre la variante  $\beta_k$  comme suit :

$$\beta_k = \begin{cases} -\beta_k^{FR} & \text{si } \beta_k^{PRP} < -\beta_k^{FR} \\ \beta_k^{PRP} & \text{si } |\beta_k^{PRP}| \leq \beta_k^{FR} \\ \beta_k^{FR} & \text{si } \beta_k^{PRP} > \beta_k^{FR} \end{cases}$$

**Remarque 3. 6**

Grippo et Lucidi ([24, 1997]) et P. Armand ([1, 2005]) ont suggéré des modifications dans le choix de  $\lambda_k$  afin d'établir le résultat de la convergence.

### 3.4 Méthode de la descente conjuguée

Cette méthode fut proposée en 1987 par Fletcher et Reeves([15, 1987]),

Rappelons que pour la méthode de la descente conjuguée la variante  $\beta_k$  est :

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}} \tag{3.19}$$

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

**Algorithme 3. 2** de la méthode de descente conjuguée

**Etape 0 : (initialisation)**

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

Poser  $k = 0$  et aller à l'étape 1.

**Etape 1 :**

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

Si non aller à l'étape 2.

**Etape 2 :**

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

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

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

où

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

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

### 3.4.1 La propriété de descente de la méthode de la descente conjuguée

Fletcher ([15, 1987]) a démontré que la méthode de la descente conjuguée est une méthode de descente si le pas  $\lambda_k$  est déterminé par la règle forte de Wolfe (3. 9) et (3. 10) avec  $\omega_2 \leq \frac{1}{2}$ .

Dai et Yuan ([12, 1996]), ont démontré que cette méthode avec la règle de Wolfe relaxée où  $0 < \omega_1 < \omega_2 < 1$  et  $0 \leq \hat{\omega}_2 \leq 1$  génère des directions de descente. a chaque itération  $k \geq 1$ .

#### **Théorème 3. 5**

Supposons que l'hypothèse 3. 1 est satisfaite.

Pour toute méthode du type (3. 2) et (3. 3) dont  $\beta_k$  satisfait à (3. 19) et le pas  $\lambda_k$  satisfait aux conditions de Wolfe relaxées suivantes :

(3. 11) et (3. 12) :

$$\begin{aligned} f(x_k + \lambda_k d_k) &\leq f(x_k) + \omega_1 \lambda_k d_k^T g_k \\ \text{et } \omega_2 d_k^T g_k &\leq d_k^T g_{k+1} \leq -\hat{\omega}_2 d_k^T g_k \end{aligned}$$

où  $0 < \omega_1 < \omega_2 < 1$  et  $0 \leq \hat{\omega}_2 \leq 1$

Alors la méthode génère des directions de descente suffisante à chaque itération  $k \geq 1$ .

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

**Preuve.** ([12, 1996])

On a

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

D'autre part de (3. 12)

$$\begin{aligned} \omega_2 d_k^T g_k &\leq d_k^T g_{k+1} \leq -\omega_2 d_k^T g_k \\ \Rightarrow 1 - \omega_2 &\leq 1 + \frac{d_{k-1}^T g_k}{d_{k-1}^T g_{k-1}} \leq 1 + \omega_2 \end{aligned}$$

d'où

$$1 - \omega_2 \leq \frac{-d_k^T g_k}{\|g_k\|^2} \leq 1 + \omega_2$$

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

$$d_k^T g_k \leq -C \|g_k\|^2 \text{ où } C = 1 - \omega_2 > 0$$

et donc  $d_k$  est une direction de descente suffisante.  $\square$

### 3.4.2 Convergence de la méthode de la descente conjuguée

Yuan ([45, 1993]) a démontré la convergence au sens (3. 17) de cette méthode avec un pas satisfaisant aux conditions (3. 11)-(3. 12) si  $\omega_2 < \frac{1}{2}$  et  $\acute{\omega}_2 = 0$ .

Dai et Yuan ([12, 1996]) ont démontré ce résultat pour  $\omega_2 < 1$  et  $\acute{\omega}_2 = 0$ .

#### **Théorème 3. 6**

Supposons que l'hypothèse 3. 1 est satisfaite.

Toute méthode du type (3. 2) et (3. 3) dont  $\beta_k$  vérifie (3. 19) et le pas  $\lambda_k$  est déterminé par la règle de Wolfe relaxée (3. 11)-(3. 12) où  $0 < \omega_1 < \omega_2 < 1$  et  $\acute{\omega}_2 = 0$  est de descente convergente dans le sens où

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



**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

**Preuve.** ([12, 1996])

Du théorème 3. 4 on a :

$$\begin{aligned}
 1 - \omega_2 &\leq \frac{-d_k^T g_k}{\|g_k\|^2} \leq 1 + \omega_2 \\
 \Rightarrow 1 &\leq \frac{-d_k^T g_k}{\|g_k\|^2} \leq 1 + \omega_2 \\
 \Rightarrow (1 + \omega_2)^{-1} &\leq \frac{\|g_k\|^2}{-d_k^T g_k} \leq 1 \\
 \Rightarrow (1 + \omega_2)^{-1} &\leq \frac{\|g_{k+1}\|^2 \|g_k\|^2}{-d_k^T g_k \|g_{k+1}\|^2} \leq 1 \\
 \Rightarrow (1 + \omega_2)^{-1} &\leq \frac{\beta_{k+1}^{CD}}{\beta_{k+1}^{FR}} \leq 1 \\
 \Rightarrow \beta_{k+1}^{CD} &\leq \beta_{k+1}^{FR}
 \end{aligned}$$

Donc  $\beta_{k+1}^{CD}$  vérifie l'inégalité (3. 15).

D'après le théorème 3. 2 on a :

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

### 3.5 Méthode de Dai-Yuan

Cette méthode a été découverte par Dai et Yuan ([9, 1999]).

Rappelons (chapitre2) que pour la méthode de la descente conjugué la variante  $\beta_k$  est :

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \quad (3.20)$$

Remarquons que pour cette variante on a le même numérateur et la même dénominateur que pour les variantes de Fletcher-Reeves et Hestenes-Stiefel respectivement.

Cette méthode possède plusieurs propriétés, par exemple elle possède la propriété de descente à chaque itération, la convergence au sens (3. 17) si le pas est déterminé par la règle de Wolfe faible.

**Algorithme 3. 3 de la Méthode de Dai-Yuan avec la règle de Wolfe faible**

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

**Etape 0 : (initialisation)**

Soit  $x_0$  le point de départ,  $g_0 = \nabla f(x_0)$ , poser  $d_0 = -g_0$   
Poser  $k = 0$  et aller à l'étape 1.

**Etape 1 :**

Si  $g_k = 0$  : STOP (  $x^* = x_k$ ). "Test d'arrêt"  
Si non aller à l'étape 2.

**Etape 2 :**

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

$\lambda_k$  vérifie les conditions (3.10)

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

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.

### 3.5.1 La propriété de descente de la méthode de Dai-Yuan

Dai et Yuan ([13, 1998]) ont démontré que à chaque itération  $k \geq 1$ , la direction recherchée par cette méthode avec la recherche de Wolfe faible (3.7)-(3.8), est de descente si la fonction objectif  $f$  est strictement convexe.

En 1999, dans [9], ils s'ont généralisé ce résultat pour toute fonction régulière.

**Théorème 3.7**

Supposons que L'hypothèse 3.1 soit satisfaite. Pour toute méthode du type (3.2) et (3.3) dont  $\beta_k$  satisfait à (3.20) et le pas  $\lambda_k$  satisfait aux conditions de Wolfe faibles suivantes :

$$\begin{aligned} f(x_k + \lambda d_k) &\leq f(x_k) + \omega_1 \lambda \nabla f(x_k)^T \cdot d_k \\ \text{et } d_k^T g_{k+1} &\geq \omega_2 d_k^T g_k \end{aligned}$$

où

$0 < \omega_1 < \omega_2 < 1$  alors

toutes les directions générées sont de descente, autrement dit :

$$d_k^T g_k < 0 ; \forall k \geq 1 \tag{3.21}$$

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

**Preuve.** ([9, 1999])

La démonstration se fait par récurrence.

1) Pour  $k = 1$  :

$$d_1^T g_1 = -\|g_1\|^2 < 0$$

2) Supposons que (3. 21) est satisfaite pour  $k > 1$  et démontrons qu'elle le sera pour  $k + 1$  :

Supposons que :

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

En utilisant (3. 7), on aura :

$$d_k^T y_k = d_k^T (g_{k+1} - g_k) > d_k^T (g_{k+1} - g_k) = (\omega_2 - 1) d_k^T g_k > 0$$

D'autre part :

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

or puisque  $:d_k^T g_k < 0, d_k^T y_k > 0$ , il en résulte

$$d_{k+1}^T g_{k+1} < 0$$

Ce qui découle la démonstration.  $\square$

### 3.5.2 Convergence de la méthode de Dai-Yuan

Dai-Yuan ([9, 1999]) ont démontré la convergence au sens (3. 17) si le pas  $\lambda_k$  est déterminé par la règle de wolfe faible.

On a alors le théorème suivant.

**Théorème 3. 8**

Supposons que la proposition 3. 1 est satisfaite.

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

La suite  $\{x_k\}$  générée par l'algorithme 3. 1 converge dans le sens

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

**Preuve. ([9, 1999])**

En utilisant le théorème 3. 1, on aura :

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

d'autre part on a

$$\begin{aligned} \|d_{k+1} + g_{k+1}\|^2 &= \|\beta_{k+1}^{DY} d_k\|^2 \\ \Rightarrow \|d_{k+1}\|^2 &= (\beta_{k+1}^{DY})^2 \|d_k\|^2 - 2d_{k+1}^T g_{k+1} - \|g_{k+1}\|^2 (**)$$

De (3. 20)

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

remplaçons ceci dans (\*\*), on aura

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

d'où

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

**CHAPITRE 3. SYNTHÈSE SUR LA CONVERGENCE DE  
QUELQUES MÉTHODES DU GRADIENT CONJUGUÉ NON  
LINÉAIRE AVEC LA RECHERCHE LINÉAIRE INEXACTE**

---

Supposons maintenant que (4. 18) n'est pas satisfaite, autrement dit :

$$\exists \omega > 0 \text{ tel que } \|g_k\| > \omega; \forall k$$

On aura :

$$\begin{aligned} \frac{\|d_k\|^2}{(d_k^T g_k)^2} &\leq \frac{1}{\|g_i\|^2} \leq \frac{1}{\omega^2} \sum_{i=1}^k 1 = \frac{1}{\omega^2} k \\ \Rightarrow \sum_{k \geq 1} \frac{(d_k^T g_k)^2}{\|d_k\|^2} &\geq \omega^2 \sum_{k \geq 1} \frac{1}{k} \end{aligned}$$

d'où

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

ce qui contredit (\*)

Ce qui achève la démonstration.  $\square$

## Chapitre 4

# Une famille de trois paramètres du gradient conjugué non linéaire

### 4.1 Introduction

Considérons le problème d'optimisation sans contrainte suivant

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

où  $f$  est une fonction différentiable.

Les méthodes du gradient conjugué sont considérées comme des méthodes importantes pour résoudre le problème (4.1), particulièrement pour les problèmes à grande taille. Ces méthodes suivent le schéma itératif suivant :

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

la direction  $d_k$  est définie par la formule de récurrence suivante :

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

où  $g_k = \nabla f(x_k)$ ;  $\lambda_k$  est la longueur de direction obtenue par une recherche linéaire unidimensionnelle et  $\beta_k$  est un scalaire.

Les  $\lambda_k$  vérifient les relations de wolfe forte suivantes :

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \omega_1 \lambda_k d_k^T g_k. \quad 4.4$$

$$|d_k^T g_{k+1}| \leq -\omega_2 d_k^T g_k. \quad (4.5)$$

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

avec  $0 < \omega_1 < \omega_2 < 1$ .

Le scalaire  $\beta_k$  est choisi de telle sorte que la méthode (4. 2) (4. 3) est réduite à la méthode du gradient conjugué linéaire dans le cas quand ou  $f$  est quadratique convexe et la recherche linéaire est exacte.

Pour des fonctions générales, cependant, de différentes formules pour le scalaire  $\beta_k$  résultent des méthodes du gradient conjugué.

On peut citer par exemple Fletcher-Reeves (*FR*), Polak-Ribiere-Polyak (*PRP*), et Hestenes-Stiefel (*HS*) (voir [15, 27, 28, 18]) lesquelles sont données par

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}. \quad (4.6)$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \quad (4.7)$$

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \quad (4.8)$$

respectivement, où  $\|\cdot\|$  signifie la norme euclidienne et  $y_{k-1} = g_k - g_{k-1}$ .

Les propriétés de convergence de ces méthodes ont été étudiées dans beaucoup de références, par exemple [1, 5, 7, 16, 17, 21, 25, 28, 31].

Cependant, si la condition imposée à  $\omega_2$  dans (4. 5) est seulement  $\omega_2 < 1$ , ni l'une ni l'autre des trois méthodes du gradient conjugué non-linéaires citées ci-dessus ne peuvent assurer la descente avec la recherche linéaire de Wolfe forte (4. 5), même si  $f$  est quadratique (voir [3, 7]).

La méthode de descente conjuguée (*CD*) de Fletcher [14], où

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}} \quad (4.9)$$

assure une direction de descente pour des fonctions générales qui satisfont la recherche linéaire de Wolfe forte (4. 4) (4. 5) avec  $\omega_2 < 1$ .

Mais la convergence globale de la méthode de *CD* est prouvée (voir[8]) seulement pour le cas de la recherche linéaire qui satisfait la condition de wolfe (4. 4) et

$$\omega_2 d_k^T g_k \leq g(x_k + \lambda_k d_k) d_k \leq 0 \quad (4.10)$$

Pour toute constante positive  $\omega_2$ , un exemple dans [8] montre que la méthode de descente conjuguée avec  $\lambda_k$  satisfaisant (4. 1) et

$$\omega_2 d_k^T g_k \leq g(x_k + \lambda_k d_k) d_k \leq -\omega_2 d_k^T g_k, \quad (4.11)$$

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

ne converge pas.

Dai et Yuan [6] ont proposé une nouvelle méthode du gradient conjuguée, dans laquelle

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \quad (4.12)$$

Une propriété remarquable de la méthode de *DY* fournit une direction de descente à chaque itération et converge globalement à condition que  $\lambda$  satisfait la condition de Wolfe (4. 4), à savoir :

$$g(x_k + \lambda_k d_k) d_k \geq \omega_2 d_k^T g_k \quad (4.13)$$

Quelques autres propriétés de la méthode de *DY* ont été exposées dans les références [4, 9, 11, 12].

Dans [11], Dai et Yuan proposent une famille d'un paramètre entre deux méthodes (*FR*) et (*DY*) pour le scalaire  $\beta_k$  défini de la manière suivante :

$$\beta_k = \frac{\|g_k\|^2}{\lambda \|g_{k-1}\|^2 + (1 - \lambda) d_{k-1}^T y_{k-1}} \quad (4.14)$$

où  $\lambda \in [0; 1]$  est un paramètre. Ils établissent que la famille d'un paramètre converge globalement si les paramètres  $\omega_2$ ,  $\acute{\omega}_2$  dans les conditions de wolfe relaxée (4. 4) et (4. 11), et  $\lambda$  sont tels que :

$$\omega_2 - 1 \leq (\omega_2 + \acute{\omega}_2) \lambda \leq 1 \quad (4.15)$$

Nazareth a considéré les formules *FR*, *PRP*, *HS*, et *DY* comme les quatre compétiteurs principaux pour le scalaire  $\beta_k$ , et a proposé une famille de deux paramètres des méthodes du gradient conjugué de la manière suivante :

$$\beta_k = \frac{\lambda_k \|g_k\|^2 + (1 - \lambda_k) g_k^T y_{k-1}}{\mu_k \|g_{k-1}\|^2 + (1 - \mu_k) d_{k-1}^T y_{k-1}} \quad (4.16)$$

$$\lambda_k, \mu_k \in [0; 1]$$

Dans ce mémoire, on propose une famille de trois paramètres des méthodes du gradient conjugué, qui inclue les méthodes du gradient conjugué non linéaire mentionnés ci-dessus.

Dans la section 3, nous étudierons la propriété de descente de la famille de trois paramètres. Nous prouverons, si le critère de relancement de Powel [24] est utilisé, la famille de trois paramètres avec la recherche linéaire de Wolfe forte produit une direction de descente à chaque itération.

Dans la section 4, quelques résultats généraux de convergence pour la famille de trois paramètres sont donnés.



## 4.2 Une famille de trois paramètres des méthodes du gradient conjugué

Dans [22], Liu et Storey ont présenté la formule suivante pour le scalaire  $\beta_k$  :

$$\beta_k^{LS} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}} \quad (4.17)$$

On observe d'après les formules (4. 8) et (4. 17) que la prochaine direction  $d_{k+1}$  dans (4. 2) est indépendante de la longueur de  $d_k$  quand  $\beta_k$  prend la forme de (4. 8) ou (4. 17).

Beaucoup d'auteurs ont présenté d'autres choix pour le scalaire  $\beta_k$ , par exemple Buckley et Lenir[2], Daniel [13], Gilbert et Nocedal [ 16 ], Qi et Al [ 28 ], Shanno [ 29 ], et Touati-Ahmed et Storey [30]. On observe alors que les formules (4. 6) (4. 9) (4. 12) et (4. 17) partagent deux numérateurs et trois dénominateurs, nous pouvons utiliser les combinaisons de ces numérateurs et dénominateurs afin d'obtenir la famille de trois-paramètre suivante :

$$\beta_k = \frac{(1 - \lambda_k) \|g_k\|^2 + \lambda_k g_k^T y_{k-1}}{(1 - \mu_k - \omega_k) \|g_{k-1}\|^2 + \mu_k d_{k-1}^T y_{k-1} - \omega_k d_{k-1}^T g_{k-1}} \quad (4.18)$$

puisque  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$ ,  $\omega_k \in [0, 1 - \mu_k]$  sont des paramètres.

Puisque

$$g_k^T y_{k-1} = \|g_k\|^2 - g_k^T g_{k-1} \quad (4.19)$$

et

$$d_{k-1}^T g_{k-1} = \beta_{k-1} d_{k-2}^T g_{k-1} - \|g_{k-1}\|^2 \quad (4.20)$$

on peut écrire la formule (4. 18) ainsi

$$\beta_k = \frac{\|g_k\|^2 - \lambda_k g_k^T g_{k-1}}{\|g_{k-1}\|^2 + \mu_k g_k^T d_{k-1} - (\mu_k + \omega_k) \beta_{k-1} d_{k-2}^T g_{k-1}} \quad (4.21)$$

Si la fonction objective est quadratique convexe et  $\lambda$  la minimisation unidimensionnelle exacte, la formule ci-dessus pour  $\beta_k$  se réduit clairement à la formule de *FR* (4. 6), dans ce cas nous avons

$$g_k^T d_{k-1} = 0, \quad (4.22)$$

et

$$g_k^T g_{k-1} = 0. \quad (4.23)$$

Cependant, pour des fonctions générales, les relations (4. 22) et (4. 23) n'ont pas besoin d'être satisfaites.

Par conséquent les méthodes (4. 2), (4. 3) et (4. 21) avec des valeurs différentes de  $\lambda_k$ ,  $\mu_k$ ,  $\omega_k$  forment une famille de trois paramètres du gradient conjugué non-linéaire.

Il est facile de voir de (4. 18) que la famille de trois paramètres inclut six méthodes simples du gradient conjugué non-linéaire qui existes déjà.

Si  $\omega_k = 0$ , alors la famille se réduit à la famille des méthodes de deux paramètres du gradient conjugué [23].

En outre, si  $\lambda_k = 0$ ,  $\mu_k = \mu$ ,  $\omega_k = 0$ , alors la famille se réduit à la famille d'un paramètre [11].

Par conséquent la famille d'un paramètre et la famille de deux paramètres sont des cas particuliers de la famille de trois paramètres.

En outre, les méthodes hybrides dans [6, 16, 19] peuvent être également considérées comme des cas particuliers de familles de trois paramètres.

Par exemple, pour améliorer la propriétés de convergence de la Méthode de *FR* et les bonnes exécutions numériques de la méthode de *PRP*, Hu et Storey[ 19 ] ont proposé une méthode hybride de la manière suivante :

$$\beta_k = \max \{0; \min \{ \beta_k^{PRP}; \beta_k^{FR} \} \}. \quad (4.24)$$

On peut facilement voir cela, la formule (4. 24) correspond à (4. 21)

$$\lambda_k = \begin{cases} \frac{\|g_k\|^2}{g_k^T g_{k-1}} & \text{si } g_k^T g_{k-1} \geq \|g_k\|^2, \\ 1 & \text{si } g_k^T g_{k-1} \in (0, \|g_k\|^2) \\ 0 & \text{si } g_k^T g_{k-1} \leq 0, \end{cases} \quad (4.25)$$

$$\mu_k \equiv 0,$$

$$\omega_k \equiv 0.$$

### 4.3 Critère de relancement de Powell et propriété de descente

Comme mentionné dans la première section, si nous avons besoin seulement que  $\omega_2 < 1$ , quelques méthodes comme *FR*, *PRP*, et *HS* avec la recherche linéaire de Wolfe forte peuvent produire des directions de recherche

qui ne sont pas de descente même si la fonction objective est quadratique. Ainsi, de particulières attentions doivent être données au problème suivant : comment garder la propriété de descente des méthodes du gradient conjugué ?

Dans cette section, nous prouverons cela, si la recherche linéaire de Wolfe peut garantir la propriété de descente de chaque direction de recherche. Dans [24], Powell a suggéré le relancement  $d_k = -g_k$  si la condition suivante est satisfaite :

$$|g_k^T g_{k-1}| \leq \xi \|g_k\|^2, \quad (4.26)$$

pour traiter la méthode du gradient conjugué de trois paramètres de Beale's. où  $\xi > 0$  est une certaine constante positive.

En fait, pour les méthodes du gradient conjugué classique, si la fonction est quadratique convexe et la recherche linéaire est exacte alors la relation (4. 23) n'implique aucun relancement.

Dans les réalisations des méthodes du gradient conjugué, le critère de relancement de Powell a été employé par beaucoup d'auteurs, par exemple Buckley et Lenir [2] et Khoda et Al [20].

Pour montrer l'importance du critère de relancement de Powell en gardant la propriété de descente des méthodes du gradient conjugué, considérons d'abord la méthode de *HS* comme exemple d'illustration.

A cette fin, nous définissons :

$$r_k = \frac{-g_k^T d_k}{\|g_k\|^2}, \quad (4.27)$$

Il est évident que  $d_k$  soit une direction de descente si et seulement si  $r_k > 0$ . Pour la méthode de *HS* (4. 2), (4. 3) et (4. 8), des calculs directs donnent

$$r_k = \frac{-g_{k-1}^T d_{k-1}}{d_{k-1}^T y_{k-1}} \left[ 1 - \frac{g_k^T g_{k-1}}{\|g_k\|^2} \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \right]. \quad (4.28)$$

Supposons que  $d_{k-1}$  est une direction de descente dans l'étape  $(k-1)$  et la recherche linéaire satisfait les conditions de Wolfe forte (4. 4). (4. 5).

Alors nous avons  $g_{k-1}^T d_{k-1} < 0$  et  $d_{k-1}^T y_{k-1} > 0$ .

En outre, il suit de (4. 5) que :

$$\left| \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \right| \leq \omega_2. \quad (4.29)$$

Par conséquent d'après (4. 28), si le critère de relancement de Powell (4. 27) est employé, la méthode de *HS* peut assurer la propriété de descente de la prochaine direction  $d_k$ , à condition que le paramètre  $\xi$  et  $\omega_2$  satisfont

$$\xi \omega_2 < 1. \quad (4.30)$$

## 4.4 Propriétés de descente

Pour la famille de trois paramètres du gradient conjugué, nous pouvons prouver le théorème général suivant.

### **Théorème 4. 1**

Considérons n'importe quelle méthode de la forme (4. 2), (4. 3) et (4. 18) avec  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$ ,  $\omega_k \in [0, 1 - \mu_k]$  et la recherche de wolf forte (4. 5), ainsi que le critère de relancement de Powell (4. 26). Si de plus les paramètres  $\xi$  et  $\omega_2$  satisfont

$$(1 + \xi) \omega_2 \leq \frac{1}{2}, \quad (4.31)$$

alors pour tout  $k \geq 1$  on a

$$0 < r_k < \frac{1}{1 - (1 + \xi) \omega_2}. \quad (4.32)$$

### **Preuve**

Supposons que (4. 26) soit satisfaite pour tout  $k$ . Nous allons montrer (4. 32) par recurrence.

Notant  $d_1 = -g_1$  alors  $r_1 = 1$ , nous remarquons que (4. 32) est satisfaite pour  $k = 1$ .

Nous supposons maintenant que (4. 32) est satisfaite pour  $k - 1$ , on a

$$0 < r_{k-1} < \frac{1}{1 - (1 + \xi) \omega_2}. \quad (4.33)$$

d'après (4.3) et (4.21) et par un calcul direct on trouve

$$r_k = 1 - \left[ 1 - \lambda_k \frac{g_k^T g_{k-1}}{\|g_k\|^2} \right] b_k, \quad (4.34)$$

où

$$b_k = \frac{g_k^T d_{k-1}}{(1 - \mu_k - \omega_k) \|g_{k-1}\|^2 + \mu_k g_k^T d_{k-1} - (\mu_k + \omega_k) g_{k-1}^T d_{k-1}}. \quad (4.35)$$

On utilise (4.5), (4.33) et le fait que  $\mu_k, \omega_k \geq 0$  dans (4.34), on obtient

$$\begin{aligned}
 b_k &\leq \frac{-\omega_2 g_{k-1}^T d_{k-1}}{(1 - \mu_k - \omega_k) \|g_{k-1}\|^2 - \omega_2 \mu_k g_{k-1}^T d_{k-1} - (\mu_k + \omega_k) g_{k-1}^T d_{k-1}} \\
 &= \frac{\omega_2 r_{k-1}}{(1 - \mu_k - \omega_k) + [(1 + \omega_2) \mu_k + \omega_k] r_{k-1}} \\
 &< \frac{\omega_2 [1 - (1 + \xi) \omega_2]^{-1}}{(1 - \mu_k - \omega_k) + [(1 + \omega_2) \mu_k + \omega_k] [1 - (1 + \xi) \omega_2]^{-1}} \\
 &\leq \frac{\omega_2}{[1 - (1 + \xi) \omega_2] + (2 + \xi) \omega_2 \mu_k + (1 + \xi) \omega_2 \omega_k} \\
 &\leq \frac{\omega_2}{1 - (1 + \xi) \omega_2}. \\
 &\leq \frac{\omega_2}{1 - (1 + \xi) \omega_2} \tag{4.36}
 \end{aligned}$$

de la même manière on établit

$$b_k > -\frac{\omega_2}{1 - (1 + \xi) \omega_2} \tag{4.37}$$

donc de (4.34) . (4.36) (4.37) . (4.26) . (4.31) et le fait que  $\lambda_k \in [0, 1]$ , on obtient

$$0 \leq 1 - \frac{(1 + \xi) \omega_2}{1 - (1 + \xi) \omega_2} < r_k < 1 + \frac{(1 + \xi) \omega_2}{1 - (1 + \xi) \omega_2} = \frac{1}{1 - (1 + \xi) \omega_2} \tag{4.38}$$

donc (4.32) satisfaite pour tout  $k \geq 1$ .  $\square$

dans les calculs, Powel [24] a proposé que la valeur de  $\xi$  dans (4.26) est  $\xi = 0, 2$ . Nous devons préciser que la condition (4.31) permet des valeurs relativement grandes de  $\xi$  et par conséquent le point courant  $x_k$  est instable parce que le paramètre  $\omega_2$  dans (4.5) prend généralement une petite valeur, normalement  $\omega_2 = 0, 1$ .

## 4.5 Propriétés de convergence

Dans cette section, on étudie les propriétés de la convergence globale de la famille de trois paramètres du gradient conjugué non-linéaire.

Pour la convergence, nous proposons que  $g_k \neq 0$  pour tout  $k$ , car autrement on a un point stationnaire.

Nous supposons aussi que  $\beta_k \neq 0$  pour tout  $k$ . car la direction dans (4.3) est réduite à  $-g_k$  si  $\beta_k = 0$ , alors le nouveau point de départ reste stationnaire

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

et nous pouvons prendre  $x_{\bar{k}}$  où  $\bar{k}$  est le plus grand indice pour lequel  $\beta_k = 0$  et où la convergence faible suivante :

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

est satisfaite.

On suppose dans tout ce qui suit que l'hypothèse suivante est vérifiée.

**Hypothèse 4. 1**

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

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

$$\exists L > 0 \text{ tel que } \|g(x) - g(\tilde{x})\| \leq L \|x - \tilde{x}\|, \forall x, \tilde{x} \in \mathcal{N} \quad (4.39)$$

L'hypothèse ci-dessus implique qu'il existe une constante positif  $\bar{\gamma}$  telle que

$$\|g(x)\| \leq \bar{\gamma}, \forall x \in \mathcal{L} \quad (4.40)$$

Pour donner le premier résultat de convergence pour la famille de trois paramètres nous avons besoin du lemme suivant qui peut être prouvé par ailleurs que théorème 3. 3 dans [11].

**lemme 4. 1**

supposons que  $x_1$  est un point de départ pour que l'hypothèse 4. 1 soit satisfaite.

Considérons la méthode de la forme (4. 2). (4. 3) où  $d_k$  est une direction de descente et  $\lambda_k$  satisfait les conditions de Wolfe faible (4. 4) et (4. 13).

S'il existe une suite  $\phi_k$  positive telle que

$$|\beta_k| \leq \frac{\phi_k}{\phi_{k-1}} \quad (4.41)$$

et

$$\sum_{k \geq 1} \frac{\|g_k\|^2}{\phi_k^2} = +\infty. \quad (4.42)$$

alors la méthode converge dans le sens suivant

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (4.43)$$

Par le lemme 4. 1, nous pouvons trouver le résultat général suivant pour la famille de trois paramètres du gradient conjugué non-linéaire.

**Théorème 4. 2**

supposons que  $x_1$  est un point de départ pour que l'hypothèse 4. 1 soit satisfaite.

Pour toute méthode du type (4. 2), (4. 3) et (4. 18) avec  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  et  $\omega_k \in [0, 1 - \mu_k]$  et la recherche linéaire de Wolfe forte (4. 4) (4. 5) ainsi que le critère de relancement de Powell (4. 26).

Notons par  $\iota_k = |\beta_k / \beta_k^{FR}|$ .

Si (4. 31) est vérifiée et si les paramètres  $\lambda_k$ ,  $\mu_k$ , et  $\omega_k$  sont tels que

$$\prod_{j=2}^k \iota_j \leq r\sqrt{k}. \quad (4.44)$$

pour une certaine constante  $r > 0$  et tout  $k \geq 2$  alors la méthode converge dans le sens faible (4. 43).

**Preuve**

Puisque les paramètres  $\xi$  et  $\omega_2$  satisfont (4. 31), nous avons par le théorème 3. 1 que (4. 32) est satisfaite pour tout  $k$ , ceci implique que chaque  $d_k$  est une direction de descente.

Définissons

$$\phi_k = \left( \prod_{j=2}^k \iota_j \right) \|g_k\|^2. \quad (4.45)$$

D'après (4. 6) et la définition de  $\iota_k$ , on peut écrire que

$$|\beta_k| = \frac{\phi_k}{\phi_{k-1}} \quad (4.46)$$

D'autre part (4. 40), (4. 44) et (4. 45) impliquent que

$$\frac{\|g_k\|^2}{\phi_k^2} \geq \frac{1}{r^2 \bar{\gamma}^2 k}, \quad (4.47)$$

d'où (4. 42).

Par conséquent par le lemme 4. 1, on obtient (4. 43).  $\square$

Maintenant nous discutons quelques choix spéciaux qui satisfont la condition (4. 44).

Comme mentionné dans le dernier paragraphe de la section 2, les méthodes hybrides dans [16] et [19] peuvent être considérées comme des cas

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

particuliers de la famille des méthodes de trois paramètres du gradient conjugué.

En utilisant le théorème ci-dessus, nous pouvons déduire encore la convergence globale des méthodes hybrides.

Par exemple, pour la méthode hybride (4. 24), nous avons  $0 \leq \iota_k \leq 1$ , qui indique que la relation (4. 44) est satisfaite avec  $r = 1$ . Par conséquent considérons (4. 43) et Supposons que  $\prod_{j=2}^{k-1} \iota_k \leq r\sqrt{k-1}$  pour un certain  $k \geq 3$ .

Alors à l'itération de  $k$ -ième, (4. 44) est vérifiée :

$$\iota_k \leq \sqrt{1 + \frac{1}{k-1}} \quad (4.48)$$

Pour  $\lambda_k = \mu_k = \omega_k = 0$ ,  $\beta_k$  est réduit à  $\beta_k^{FR}$ , et par conséquent  $\iota_k = 1$ . Ainsi même si  $\iota_k = 1$  pour  $k$  grand, nous pouvons voir de la relation (4. 48) qu'il existe un certain intervalle dans  $[0, 1]$  pour chacun des paramètres  $\lambda_k$ ,  $\mu_k$ , et  $\omega_k$ . tels que (4. 44) soit vérifiée.

Généralement il existe les choix de  $\lambda_k$ ,  $\mu_k$ , et  $\omega_k$ . qui satisfont la condition (4. 44) (comme cité ci-dessus), cette restriction peut réduire les intervalles admissibles pour  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  et  $\omega_k \in [0, 1 - \mu_k]$ .

Généralement, pour un certain  $k$  la valeur de  $\iota_k$  peut être moins de 1, et en conséquence ceci permettra à  $\iota_k$  de prendre une grande valeur. Par exemple, si l'étape  $\|x_k - x_{k-1}\|$  est très petite et  $\lambda_k$  est proche de 1 à une certaine itération loin de la solution, alors  $\iota_k$  peut être plus petite que 1.

Un autre point qui devrait être précisé ici, nous pouvons également agrandir l'intervalle admissible pour  $\lambda_k$ ,  $\mu_k$ , et  $\omega_k$ . par la situation (réglage) de  $r$  dans (4. 44) a une grande valeur.

L'inégalité (4. 44) suggère une seule possibilité pour choisir  $\lambda_k$ ,  $\mu_k$ , et  $\omega_k$ , à savoir que la valeur absolue du côté droit de (4.18) est aussi petite que possible.

Par le lemme 4. 1, nous pouvons également prouver le résultat de convergence suivant.

**Théorème 4. 3**

Supposons que  $x_1$  est un point de départ pour l'hypothèse 4. 1 est satisfaite.

Considérons toute méthode sous la forme (4. 2), (4. 3) et (4. 18) avec  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  et  $\omega_k \in [0, 1 - \mu_k]$ , avec la recherche linéaire de Wolfe forte (4. 4) – (4. 5) et le critère de relancement de Powell (4. 26).

Si (4. 31) satisfait, et si les paramètres  $\lambda_k$ ,  $\mu_k$ , et  $\omega_k$  sont tels que :

$$0 \leq \lambda_k g_k^T g_{k-1} \leq \|g_k\|^2. \quad (4.49)$$



et

$$\mu_k d_{k-1}^T g_k - \omega_k \beta_{k-1} d_{k-2}^T g_{k-1} \geq -\lambda_{k-1} g_{k-1}^T g_{k-2} \quad (4.50)$$

pour tout  $k \geq 2$ , alors la méthode converge dans du sens faible (4. 43) .

**Preuve**

Puisque les paramètres  $\xi$  et  $\omega_2$  satisfont (4. 31), nous avons par le théorème 3. 1 que (4. 32) est satisfaite pour tout  $k$ , ce qui implique que chaque  $d_k$  est une direction de descente.

De (4. 49), (4. 50) avec  $k$  remplacé par  $k + 1$ , et la formule équivalente (4. 21) de  $\beta_k$ , on voit que la relation (4. 41) est satisfaite avec

$$\phi_k = (1 - \mu_{k+1} - \omega_{k+1}) \|g_k\|^2 + \mu_{k+1} d_k^T y_k - \omega_{k+1} d_k^T g_k. \quad (4.51)$$

en utilisant (4.5) et (4.32) dans (4.51), on peut prouver que

$$\begin{aligned} \phi_k &\leq (1 - \mu_{k+1} - \omega_{k+1}) + (1 + \omega_2) \mu_{k+1} r_k + \omega_{k+1} r_k \|g_k\|^2 \quad (4.52) \\ &\leq [1 + (2 + \omega_2) (1 - (1 + \xi) \omega_2)^{-1}] \|g_k\|^2 \\ &= \frac{3 - \xi \omega_2}{1 - (1 + \xi) \omega_2} \|g_k\|^2 \end{aligned}$$

Pour tout  $\mu_{k+1} \in [0, 1]$  et  $\omega_{k+1} \in [0, 1 - \mu_{k+1}]$ .

La relation ci-dessus et (4.43) implique

$$\sum_{k \geq 1} \frac{\|g_k\|^2}{\phi_k^2} = +\infty. \quad (4.53)$$

ainsi, par le lemme 4. 1, on a (4.46).  $\square$

Si  $\xi \leq 1$ , la deuxième inégalité dans (4.52) découle clairement, puisque (4.26) et le fait que  $\lambda_k \in [0, 1]$  donnent

$$\lambda_k g_k^T g_{k-1} \leq |g_k^T g_{k-1}| \leq \xi \|g_k\|^2 \leq \|g_k\|^2. \quad (4.54)$$

Il est facile de voir que la méthode  $FR$  ( $\lambda_k = \mu_k = \omega_k = 0$ ) satisfait les deux conditions (4. 49) et (4. 50).

Cependant, pour d'autres méthodes il n'est pas clair que ces conditions soient vérifiées tant que ces dernières dépendent de l'ordre des points produits par les méthodes. Par exemple, dans le cas extrême  $g_k^T g_{k-1} < 0$ ,  $g_k^T d_{k-1} < 0$  et  $\beta_{k-1} d_{k-2}^T g_{k-1} > 0$ , nous devons choisir  $\lambda_k = \mu_{k+1} = \omega_{k+1} = 0$ . Si  $g_k^T g_{k-1} > 0$ , Il existe toujours des intervalles admissibles pour  $\lambda_k$ ,  $\mu_{k+1}$ , et  $\omega_{k+1}$ .

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

Le lemme suivant est dû à Gilbert et de Nocedal [16].

**lemme 4. 2**

Supposons que  $x_1$  est un point de départ pour que l'hypothèse 4. 1 soit satisfaite.

considérons que la méthode sous la forme (4. 2) – (4. 3) avec les trois propriétés suivantes

(i)  $\beta_k > 0$

(ii) les conditions de Wolfe faible (4. 4) et (4. 13), et la condition de descente suffisante

$$g_k^T d_k < -c \|g_k\|^2, \text{ où } c > 0, \quad (4.55)$$

pour tout  $k$  et une certaine constante positive  $c$ .

(iii) Il existe des constantes  $b > 1$  et  $\lambda > 0$  tels que  $|\beta_k| \leq b$  pour tout  $k$ , et si  $\|x_k - x_{k-1}\| \leq \lambda$  alors  $|\beta_k| \leq (2b)^{-1}$ .

Alors la méthode converge au sens faible (4. 43).

Par lemme 4. 2, nous pouvons prouver le résultat général suivant pour la famille de trois paramètres du gradient conjugué non-linéaire.

**Théorème 4. 4**

Supposons que  $x_1$  est un point de départ pour que l'hypothèse 4. 1 soit satisfaite.

Considérons que la méthode dans la forme (4. 2), (4. 3) et (4. 18), où  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  et  $\omega_k \in [0, 1 - \mu_k]$ , où  $\lambda$  satisfait les conditions de Wolfe forte (4. 4) – (4. 5), et où le critère de relancement

$$-\xi \|g_k\|^2 \leq g_k^T g_{k-1} \leq \|g_k\|^2. \quad (4.56)$$

est utilisé.

Si les paramètres sont tels que

$$(1 + \xi) \omega_2 < \frac{1}{2}. \quad (4.57)$$

et

$$\lambda_k \geq 1 - C_1 \|x_k - x_{k-1}\|. \quad (4.58)$$

où  $C_1 > 0$  est constant, la méthode converge au sens faible (4. 43).

**Preuve**

Supposons que

$$\liminf_{k \rightarrow \infty} \|g_k\| \neq 0. \quad (4.59)$$

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

alors il existe une constante positive  $\gamma$ , telle que

$$\|g_k\| \geq \gamma, \quad \forall k \geq 1 \quad (4.60)$$

En utilisant (4. 56) , (4. 57) et (4. 38) on obtient pour tout  $k \geq 1$  :

$$r_k \geq \frac{1 - 2(1 + \xi)\omega_2}{1 - (1 + \xi)\omega_2} = C_2 \quad (4.61)$$

où  $r_k$  est donné dans (4. 27). Ainsi les conditions de descente suffisantes sont satisfaites.

De (4. 5), (4. 32), (4. 39), (4. 40), (4. 60) et le fait que  $\lambda_k \leq 1$ , on peut montrer que

$$\begin{aligned} |\beta_k| &\leq \frac{(C_1 \|g_k\|^2 + L \|g_k\|) \|x_k - x_{k-1}\|}{\|g_{k-1}\|^2 [(1 - \mu_k - \omega_k) + \mu_k (1 - \omega_2) C_2 + \omega_k C_2]} \\ &\leq \frac{(C_1 \|g_k\|^2 + L \|g_k\|) \|x_k - x_{k-1}\|}{C_2 \|g_{k-1}\|^2} \leq C_3 \|x_k - x_{k-1}\| \end{aligned} \quad (4.62)$$

où

$$C_3 = \frac{C_1^2 \gamma + L \bar{\gamma}}{\gamma^2}, \quad (4.63)$$

l'hypothèse 4. 1 implique qu'il existe une constante positive  $p$  telle que

$$\|x\| \leq p, \quad \forall x \in \mathcal{L}, \quad (4.64)$$

$b = 2C_3 P$  et  $\lambda = (4C_3^2 P)^{-1}$ , de (4. 62) et (4. 64) on a

$$|\beta_k| \leq b, \quad (4.65)$$

et si

$$\|x_k - x_{k-1}\| \leq \lambda,$$

alors

$$|\beta_k| \leq (2b)^{-1} \quad (4.66)$$

Ainsi la propriété (\*) est satisfaite.

En outre, (4. 32) et (4. 56) impliquent que  $\beta_k \geq 0$ .

Par conséquent les conditions du lemme 4. 5 sont toutes satisfaites et donc on obtient (4. 43).  $\square$

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

Les théorèmes 4. 3, 4. 4 et 4. 6 fournissent quelques résultats généraux de convergence pour la famille de trois paramètres du gradient conjugué non-linéaire.

Si les paramètres dans (4. 18) sont spécifiquement choisis, alors les résultats globaux de convergence peuvent être prévus.

Nous terminons cette section avec un exemple d'illustration.

Nous disons que la méthode converge globalement à condition que la norme de  $d_k$  n'augmente pas plus rapidement que linéairement.

Exactement, la méthode donne (4. 43) si la condition suivante est satisfaite :

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} = +\infty. \quad (4.67)$$

Par conséquent, dans la famille de trois paramètres du gradient conjugué non-linéaire , pour déminuer la longueur du  $d_k$ , il est raisonnable de choisir les paramètres tels que  $|\beta_k|$  atteint sa plus petite valeur. à savoir, on définit

$$(\bar{\lambda}_k, \bar{\mu}_k; \bar{\omega}_k) = \arg \min_{\lambda_k; \mu_k \in [0;1] \omega_k \in [0;1-\mu_k]} |\beta_k|. \quad (4.68)$$

où  $\beta_k$  est défini dans (4. 18). Dans ce cas, nous pouvons obtenir le résultat de convergence globale suivant.

**Théorème 4. 5**

Supposons que  $x_1$  est un point de départ pour que l'hypothèse 4. 1 soit satisfaite.

Considérons que la méthode de la forme (4. 2), (4. 3) et

$$\bar{\beta}_k = \frac{(1 - \bar{\lambda}_k) \|g_k\|^2 + \bar{\lambda}_k g_k^T y_{k-1}}{(1 - \bar{\mu}_k - \bar{\omega}_k) \|g_{k-1}\|^2 + \bar{\mu}_k d_{k-1}^T y_{k-1} - \bar{\omega}_k d_{k-1}^T g_{k-1}}, \quad (4.69)$$

où  $\bar{\lambda}_k, \bar{\mu}_k, \bar{\omega}_k$ , sont donnés dans (4. 68). Si  $\lambda$  satisfait les conditions de Wolfe faible (4. 4) et (4. 13), alors chaque  $d_k$  produit par la méthode est une direction de descente et de plus la méthode converge au sens faible (4. 43).

**preuve**

Des choix de  $\bar{\lambda}_k, \bar{\mu}_k$  et  $\bar{\omega}_k$ , il est facile de voir que

$$0 \leq \bar{\beta}_k \leq \beta_k^{DY} \quad (4.70)$$

où  $\beta_k^{DY}$  est donné dans (4. 12) [6].  $\square$

## 4.6 Tests numériques

Des tests numériques ont été effectués sur les méthodes du gradient conjugué afin de comparer l'efficacité de ces méthodes sur les problèmes de Dixon, Oren, Rosenbrock, et un autre, qui sont donnés respectivement par les formules suivantes.

$$f(x) = (x_1 - 1)^2 + \sum_{i=2}^n (2x_i^2 - x_{i-1})^2. \quad (4.71)$$

$$f(x) = \left[ \sum_{i=1}^n ix_i^2 \right]^2. \quad (4.72)$$

$$f(x) = \sum_{i=1}^{n-1} \{100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\}. \quad (4.73)$$

$$f(x) = 1/2 \sum_{i=1}^n x_i^2. \quad (4.74)$$

Un critère d'arrêt de l'algorithme est fixé à  $\varepsilon = 10^{-5}$ .

$n$  la dimension du problème.

Les résultats (nombre d'itération) sont illustrés dans les tableaux suivants.

La fonction	$B_k$	$B_k^{PRP}$	$B_k^{FR}$	$B_k^{HS}$	$B_k^{CD}$	$B_k^{LS}$	$B_k^{DY}$
	$n$						
Dixon	5	87	103	73	1000	90	46
	50	275	1000	177	1000	285	1000
	100	447	1000	362	1000	451	72

La fonction	$B_k$	$B_k^{PRP}$	$B_k^{FR}$	$B_k^{HS}$	$B_k^{CD}$	$B_k^{LS}$	$B_k^{DY}$
	$n$						
Oren	5	20	25	39	22	24	21
	50	50	54	60	46	48	51
	100	76	74	72	71	70	72

CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE

---

La fonction	$B_k$	$B_k^{PRP}$	$B_k^{FR}$	$B_k^{HS}$	$B_k^{CD}$	$B_k^{LS}$	$B_k^{DY}$
	$n$						
Rosenbrock	5	62	1000	44	319	68	186
	50	64	149	19	63	55	187
	100	66	715	47	69	63	187

La fonction	$B_k$	$B_k^{PRP}$	$B_k^{FR}$	$B_k^{HS}$	$B_k^{CD}$	$B_k^{LS}$	$B_k^{DY}$
	$n$						
Autre	05	15	06	03	06	15	02
	50	15	06	03	06	15	02
	100	15	07	03	07	15	02

**Conclusion**

► Aucune méthode à travers les méthodes précédentes n'est meilleure que l'autre.

► L'efficacité de ces méthodes s'améliore si  $n$  prend de grandes valeurs. et pour montrer l'efficacité de la méthode du gradient conjugué particulièrement pour les problèmes à grande taille, on prend le problème de Rosenbrock comme exemple d'illustration, avec quatre méthodes.

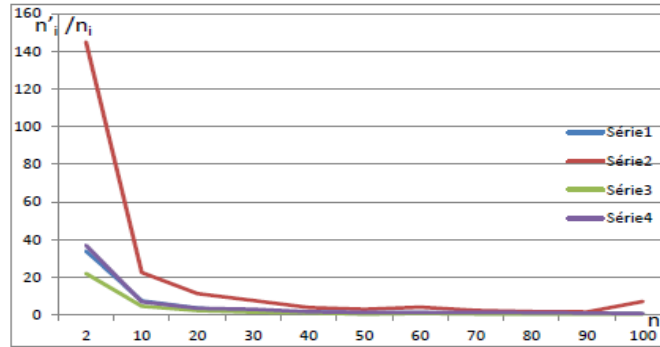
Fonction	$B_k$	$B_k^{PRP}$	$B_k^{FR}$	$B_k^{HS}$	$B_k^{CD}$
	$n$				
Rosenbrock	02	68	290	44	74
	10	75	225	46	70
	20	72	226	46	70
	30	57	230	46	91
	40	70	156	50	69
	50	64	149	19	63
	60	66	249	46	78
	70	64	169	38	1000
	80	77	142	40	1000
	90	70	146	41	88
	100	66	715	47	69

Et la figure ci-dessous explique le tableau précédent.

## CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU GRADIENT CONJUGUÉ NON LINÉAIRE

l'axe des ordonnées présente la valeur  $(\hat{n}_i/n_i)$ ,  $\hat{n}_i$  le nombre d'itérations,  $n_i$  le nombre des variables de la fonction.

l'axe des abscisses présente  $n_i$  le nombre des variables de la fonction.



► A partir de ce graphe on confirme que la méthode du gradient conjugué doit être utilisée de préférence pour les problèmes à grande taille parce que la fraction  $\hat{n}_i/n_i$  tend vers zéro.

### 4.7 Conclusions et discussions

Dans ce présent mémoire, nous avons proposé une famille des méthodes de trois paramètres du gradient conjugué non-linéaire, et étudié la convergence globale de ces méthodes. La famille de trois paramètres inclut non seulement les six méthodes simples déjà-existantes des méthodes du gradient conjugué classique, mais quelques autres familles des méthodes du gradient conjugué en tant que sous familles.

La famille de trois paramètres inclut également quelques méthodes hybrides en tant que des cas spéciaux.

Avec le critère de relancement de Powell nous avons montré que la famille de trois paramètres peut assurer la direction de descente à chaque itération. Alors dans des conditions appropriées, nous avons établi des résultats généraux de convergence, à savoir, les théorèmes 4. 2, 4. 3 et 4. 5, pour la famille de trois paramètres du gradient conjugué non-linéaire.

Si les paramètres sont spécifiquement choisis, de meilleurs résultats globaux de convergence pourraient être réalisés.

On a précisé que la condition (4. 31) n'est pas stricte, parce que le paramètre  $\omega_2$  dans (4. 5) est généralement choisi pour être une valeur relativement petite, ce qui implique que  $\xi$  dans (4. 26) pourrait être relativement grand. Cependant, il reste toujours à explorer comment trouver de nouvelles et efficaces familles de méthodes de trois paramètres du gradient conjugué.

*CHAPITRE 4. UNE FAMILLE DE TROIS PARAMÈTRES DU  
GRADIENT CONJUGUÉ NON LINÉAIRE*

---

Spécifiquement il serait intéressant de trouver une exécution pratique de la méthode (4. 68).



**Quelques problèmes ouverts et leurs conjectures**

**Problème (01) :**

Grâce aux études préalables, on choisit le scalaire  $\beta_k$  de la même façon mais incluant quatre compétiteurs lesquels sont  $\beta_k^{FR}$ ,  $\beta_k^{PRP}$ ,  $\beta_k^{CD}$ ,  $\beta_k^{LS}$  le scalaire  $\beta_k$  sera défini par :

$$\beta_k = \frac{\lambda_k \|g_k\|^2 + (1 - \lambda_k) g_k^T y_{k-1}}{\mu_k \|g_{k-1}\|^2 + (\mu_k - 1) d_{k-1}^T g_{k-1}}$$

$$\lambda_k \in [0, 1], \mu_k \in [0, 1]$$

et étudier la descente et la convergence de cette méthode de deux paramètres.

**Problème (02) :**

Dans le théorème 4. 5 , la valeur minimale de  $\beta_k$  assure la descente et la convergence, mais la question qui se pose toujours, quelles sont les relations établies entre les trois paramètres  $\lambda_k$  ,  $\mu_k$  ,  $\omega_k$  afin que  $\beta_k$  prenne cette valeur minimale ?

# Bibliographie

- [1] 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] A. Buckley and A. Lenir, QN-like variable storage conjugate gradients, *Math. Prog.* 27 (1983), 155-175.
- [3] 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.
- [4] Y. H. Dai, Some new properties of a nonlinear conjugate gradient method, Research report ICM-98-010, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998.
- [5] 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, Research report ICM-98-024, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998 (accepted by *SIAM J. Optimization*)
- [6] Y. H. Dai and Y. Yuan, A Nonlinear Conjugate Gradient Method with Nice Global Convergence Properties, Research report ICM-95-038, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1995 (accepted by *SIAM J. Optimization*).
- [7] Y. H. Dai and Y. Yuan, Convergence properties of the Fletcher-Reeves method, *IMA J. Numer. Anal.* Vol. 16 No. 2 (1996), 155-164.
- [8] Y. H. Dai and Y. Yuan, Convergence properties of the conjugate descent method, *Mathematical Advances* Vol. 25 No. 6 (1996), 552-562.
- [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] Y. H. Dai and Y. Yuan, Convergence properties of Beale-Powell restart method, *Sciences in China (series A)*, Vol. 28, No. 5, pp. 424-432.
- [11] 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. (submitted to *IMA J. Numer. Anal.* )
- [12] Y. H. Dai and Y. Yuan, Extension of a class of conjugate gradient methods, Research report ICM-98-049, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998. (submitted to *BIT*)
- [13] J. W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, *SIAM J. Numer. Anal.* , 4 (1967), 10-26.
- [14] R. Fletcher, *Practical Methods of Optimization vol. 1- Unconstrained optimization*, John Wiley & Sons (New York), 1987.
- [15] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, *Comput. J.* 7 (1964), pp. 149-154.
- [16] 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.
- [17] L. Grippo and S. Lucidi, A globally convergent version of the Polak-Ribière conjugate gradient method, *Math. Prog.* 78 (1997), pp. 375-391.
- [18] 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.
- [19] Y. F. Hu and C. Storey, Global convergence result for conjugate gradient methods, *JOTA* : Vol. 71 No. 2 (1991) 399-405.
- [20] K. M. Khoda, Y. Liu, and C. Storey, Generalized Polak-RiWe Algorithm, *Journal of Optimization Theory and Applications*, Vol. 75, No. 2 (1992), 345-354.
- [21] G. H. Liu, J. Y. Han, and H. X. Yin, Global convergence of the Fletcher-Reeves algorithm with an inexact line search, Report, Institute of Applied Mathematics, Chinese Academy of Sciences, 1993.
- [22] Y. Liu and C. Storey, Efficient Generalized Conjugate Gradient Algorithms, Part 1 - Theory, *Journal of Optimization Theory and Applications*, Vol. 69 (1991), 129-137.
- [23] L. Nazareth, Conjugate-gradient methods, to appear in : *Encyclopedia of Optimization* (C. Floudas and P. Pardalos, eds. ), Kluwer Academic Publishers, Boston, USA and Dordrecht, The Netherlands (1999).

- 
- [24] M. J. D. Powell, Restart procedures of the conjugate gradient method, *Math. Program.* 2 (1977), pp. 241-254.
- [25] M. J. D. Powell, Nonconvex minimization calculations and the conjugate gradient method, in : *Lecture Notes in Mathematics* vol. 1066, Springer-Verlag (Berlin) (1984), pp. 122-141.
- [26] E. Polak and G. Ribière, Note sur la convergence de directions conjuguées, *Rev. ev. Francaise Informat Recherche Operationelle*, 3e Année 16 (1969), pp. 35-43.
- [27] Andrei, N. : Scaled conjugate gradient algorithms for unconstrained optimization. *Comput. Optim. Appl.* 38, 401–416 (2007). doi :10.1007/s10589-007-9055-7
- [28] Andrei, N. : Another hybrid conjugate gradient algorithm for unconstrained optimization. *Numer. Algorithms* 47, 143–156 (2008). doi10.1007/s11075-007-9152-9
- [29] Barzilai, J. , Borwein, J. M. : Two point step size gradient methods. *IMA J. Numer. Anal.* 8, 141–148(1988)
- [30] Birgin, E. G. , Martínez, J. M. : A spectral conjugate gradient method for unconstrained optimization. *Appl. Math. Optim.* 43, 117–128 (2001)
- [31] Buckley, A. : Extending the relationship between the conjugate gradient and BFGS algorithms. *Math. Program.* 15, 343–348 (1978)
- [32] Dolan, E. D. , Moré, J. J. : Benchmarking optimization software with performance profiles. *Math. Program. Ser. A* 91, 201–213 (2002)
- [33] Hager, W. W. , Zhang, H. C. : A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. Optim.* 16, 170–192 (2005)
- [34] Hestenes, M. R. , Stiefel, E. : Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.* 49(6), 409–439 (1952)
- [35] Nocedal, J. , Wright, S. J. : *Numerical Optimization*, 2nd edn. Springer, Berlin (2006)
- [36] Perry, A. : A modified conjugate gradient algorithm. *Oper. Res. Tech. Notes* 26(6), 1073–1078 (1978)
- [37] Powell, M. J. D. : A new algorithm for unconstrained optimization. In : Rosen, J. B. , Mangasarian, O. L. , Ritter, K. (eds. ) *Nonlinear Programming*, pp. 31–66. Academic Press, New York (1970)
- [38] Raydan, M. : The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.* 7, 26–33 (1997)

- 
- [39] Shanno, D. F. : Conjugate gradient methods with inexact searches. *Math. Oper. Res.* 3, 244–256 (1978)
- [40] Shanno, D. F. , Phua, K. H. : Algorithm 500, minimization of unconstrained multivariate functions. *ACM Trans. Math. Softw.* 2, 87–94 (1976)
- [41] Touati-Ahmed, D. , Storey, C. : Efficient hybrid conjugate gradient techniques. *J. Optim. Theory Appl.* 64(2), 379–397 (1990)
- [42] Wei, Z. X. , Li, G. Y. , Qi, L. Q. : New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems. *Appl. Math. Comput.* 179, 407–430 (2006)
- [43] Yu, G. H. , Zhao, Y. L. , Wei, Z. X. : A descent nonlinear conjugate gradient method for large-scale un-constrained optimization. *Appl. Math. Comput.* 187, 636–643 (2007)
- [44] J. C. Gilbert (2007), *Eléments d'Optimisation Différentiable : Théorie et Algorithmes*, Notes de cours, ... cole Nationale Supérieure de Techniques avancées, Paris.
- [45] A. A. Goldstein (1967), *Constructive Real Analysis*, A Harper International Edition. 88 BIBLIOGRAPHIE
- [46] A. A. Goldstein (1965), On steepest descent, *SIAM J. on Control A*, Vol. 3, No. 1, pp. 147-151.
- [47] A. A. Goldstein and J. F. Price (1969), An effective algorithm for minimization, *Num. Math.* , 10, pp. 184-189.
- [48] L. Grippo and S. Lucidi (1997), A Globally Convergent Version of the Polak-Ribière Conjugate Gradient Method, *Mathematical Programming*, Vol. 78, pp. 375–391.
- [49] W. W. Hager et H. Zhang (2005), A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM J. Optim.* , 16, pp. 170-192.
- [50] 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.
- [51] M. Minoux (1983), *Programmation Mathématique, Théorie et Algorithmes*, tome 1 , Dunod.
- [52] J. J. Moré, B. S. Garbow and K. E. Hillstom, Testing unconstrained optimization software, *ACM Transactions on Mathematical Software* 7 (1981) 17–41.
- [53] P. Wolfe (1961), A duality theorem for nonlinear programming, *Quart. Appl. Math.* , 19, pp. 239-244.

- [54] P. Wolfe (1969), Convergence conditions for ascent methods, SIAM Review, 11, pp. 226-235.
- [55] P. Wolfe (1971), Conditions for ascent methods some corrections 2, SIAM Review, 13, pp. 185-188.
- [56] P. Wolfe (1963), Methods of Nonlinear Programming, in Recent Advances in Mathematical Programming (Eds R. L. Graves and P. Wolfe), McGraw-Hill, New York.
- [57] G. Zoutendijk (1970), Nonlinear Programming Computational Methods, Integer and Nonlinear Programming, North Holland, Amsterdam.
- [58] G. Zoutendijk (1960), Methods of Feasible Directions, Elsevier, Amsterdam