

وزارة التعليم العالي والبحث العلمي

Université Badji Mokhtar
Annaba

Badji Mokhtar University -
Annaba



جامعة باجي مختار
عنابة

Faculté des Sciences

Année : 2013

Département de Mathématiques

MEMOIRE

Présenté en vue de l'obtention du diplôme
MAGISTER

LA MÉTHODE DU GRADIENT CONJUGUÉ AVEC LA RECHERCHE LINEAIRE NON MONOTONE

Specialité : Mathématique

Option: mathématique du développement
présenté Par:

CHERGUI AHMED

DIRECTEUR DE MEMOIRE : LASKRI YAMINA professeur U.B.M. ANNABA

Devant le jury

PRESIDENT :	Seddik Ameer Nacira	professeur	UBMA
RAPPORTEUR :	Laskri Yamina	professeur	UBMA
EXAMINATEUR :	Amiar Rachida	MCA	UBMA
EXAMINATEUR :	Mentri Sabria	MCA	UBMA

Table des matières

1	Optimisation sans contraintes	13
1.1	Rappel de quelques définitions	14
1.2	Conditions d'optimalité des problèmes de minimisation sans contraintes :	18
1.2.1	Minima locaux et globaux	18
1.3	Direction de descente	20
1.3.1	Schéma général des algorithmes d'optimisation sans contraintes	23
1.3.2	Conditions nécessaires d'optimalité	23
1.3.3	Conditions suffisantes d'optimalité	24
2	Recherche linéaire inexacte	26
2.1	Recherche linéaire	26
2.1.1	Introduction	26
2.1.2	Objectifs de la recherche linéaire	27
2.1.3	But de la recherche linéaire	28
2.1.4	Intervalle de sécurité	28
2.1.5	Algorithme de base	28
2.2	Recherches linéaires "exactes"	29
2.3	Recherches linéaires inexactes	30
2.3.1	Caractérisation de l'intervalle de sécurité	31
2.4	Convergence des méthodes à directions de descente	38
2.4.1	Condition de Zoutendijk	38
2.5	Méthodes itératives d'optimisation sans contraintes	39
2.5.1	Principe des méthodes de descente	40
2.5.2	Principe des méthodes de gradient (méthodes de la plus forte descente)	41
2.5.3	Les méthodes utilisant des directions conjuguées	41

3	la convergence de quelques méthodes du gradient conjugué non linéaire avec la recherche linéaire inexacte	51
3.1	méthode de Fletcher-Reev	53
3.1.1	Algorithme de la méthode de Fletcher-Reeves	53
3.1.2	. La propriété de descente de la méthode de FR	54
3.1.3	Convergence de la méthode de Fletcher-Reeves	
3.2	Méthode de Polak-Ribière-Polyak	59
3.2.1	Algorithme 3.2 de la méthode de Polak-Ribière-Polyak	60
3.2.2	Convergence de la méthode de PRP	60
3.3	Méthode de la descente conjugué	62
3.3.1	Algorithme 3.2 de la méthode de descente conjugué	62
3.3.2	La propriété de descente de la méthode de la descente conjuguée	62
3.3.3	Convergence de la méthode de la descente conjuguée	64
3.4	Méthode de Dai-Yuan	65
3.4.1	Algorithme 3.3 de la Méthode de Dai-Yuan avec la règle de Wolfe faible	65
3.4.2	La propriété de descente de la méthode de Dai-Yuan	66
3.4.3	Convergence de la méthode de Dai-Yuan	67
4	la méthode du gradient conjugué de Wei-Yao-Liu sous la recherche linéaire non monotone de Wolfe-Powell forte	71
4.1	1Introduction	71
4.2	2 Condition suffisante de descente	74
4.3	3 Méthode du gradient conjugué (WYL) et ses applications	76
4.3.1	Les formules modifiées	77
4.4	2,2. propriétés de convergence	79
4.5	2,3. Les résultats numériques	82
4.6	Conclusion	88

Remerciements

*Je tiens à adresser mes remerciements les plus chaleureux et ma profonde gratitude à mon encadreur madame **LASKRI Yamina**, professeur à l'Université Badji Mokhtar 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.*

*Mes remerciements vont également à madame **Seddik Ameer Nasira** professeur à l'Université Badji Mokhtar Annaba, pour avoir bien voulu me faire l'honneur d'accepter de présider le jury.*

*De même je remercie madame **Amiare Rchida**, maître de conférence classe A, à l'Université Badji Mokhtar Annaba, et madame **Mentri Sabria**, maître de conférence classe A, à l'Université Badji Mokhtar Annaba, pour l'honneur qu'ils m'ont fait de bien vouloir accepter de faire partie du jury.*

Je remercie également tout ceux qui ont contribué de près ou de loin à l'élaboration de ce mémoire, en particulier, ma femme pour ses encouragements.

*Enfin, Merci à toute ma famille, en particulier, ma **mère**, mon **père**, mes enfants, et à tous mes amis et tous mes collègues sans exception.*

Résumé

Dans ce mémoire, nous proposons une nouvelle technique de recherche linéaire non monotone. Cette technique est proposée pour garantir la convergence globale ainsi que la condition suffisante de descente des méthodes du gradient conjugué en particulier la méthode du gradient conjugué Wei-Yao Liu (WYL) sous la recherche linéaire de Wolfe-Powell forte.

Mots clés : Gradient conjugué, Algorithme, Convergence globale, Recherche linéaire inexacte, condition suffisante de descente.

Méthode de Hestenes-Stiefel, Méthode de Fletcher-Reeves, Méthode Polak-Ribière-Polyak, Méthode de la descente conjuguée, Méthode de Liu et Storey.

Abstract

In this paper, we propose A new no monotone line search technique . It is proposed to guarantee the global convergence of these conjugate gradient methods with sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search

Mots clés : Gradient conjugué, Algorithme, Convergence globale, Recherche linéaire inexacte, condition suffisante de descente.

Méthode de Hestenes-Stiefel, Méthode de Fletcher-Reeves, Méthode Polak-Ribière-Polyak, Méthode de la descente conjuguée, Méthode de Liu et Storey.

ملخص

في هذه الرسالة نقتراح تقنية جديدة للبحث الخطي غير الرتيب وذلك لضمان التقارب الكلي لطرق التدرج المترافقة

gradient conjugué

مع الشرط الكافي للانحدار لطريقة التدرج المترافقة

(Wei-Yao-Liu)

باستعمال البحث لخطي لوولف بوال فورة

(Wolfe-Powell fort)

Introduction

Optimiser : rendre optimal, donner à quelque chose les meilleures conditions d'utilisation, de fonctionnement ou de rendement au regard de certaines circonstances. (Déf. du Larousse).

Les gens optimisent :

Des compagnies aériennes établissent les horaires des équipages et des avions avant de minimiser le coût. Des investisseurs cherchent à constituer des portefeuilles boursiers qui évitent des risques excessifs tout en donnant un bon rendement. Des fabricants visent le maximum d'efficacité dans la conception et l'opération de leurs processus de production.

La nature optimise :

Des systèmes physiques tendent vers un état d'énergie minimale. Des molécules dans un système chimique réagissent entre eux jusqu'à ce que l'énergie potentielle totale de leurs électrons soit minimisée. Des rayons de lumière suivent des trajectoires qui minimisent leur temps de trajet. L'optimisation est un outil important dans la théorie des décisions et dans l'analyse des systèmes physiques. Avant de l'utiliser, nous devons d'abord identifier un objectif. Ceci est une mesure quantitative de la performance du système. Cet objectif pourrait être : bénéfice, temps, énergie potentielle, ou toute autre quantité ou combinaison des quantités qui peut être représentée par un seul nombre. L'objectif dépend de certaines caractéristiques du système, appelées des variables ou des inconnues. Notre but est de trouver des valeurs de variables qui optimisent l'objectif. Souvent les variables sont restreintes, ou contraintes, d'une certaine manière. Par exemple, la densité ou l'énergie ou le taux d'intérêt ne peuvent pas être négatifs. Le processus de l'identification de l'objectif, variables et contraintes pour un problème donné est appelé la modélisation. La construction d'un modèle approprié est la première étape. Parfois c'est l'étape la plus importante du processus d'optimisation. Si le modèle est simpliste, il ne donnera pas d'informations utiles concernant le problème pratique. Par contre, si le modèle est trop complexe, il pourra devenir trop difficile à résoudre. Une fois que le modèle a été formulé, un algorithme d'optimisation peut être utilisé avant de trouver sa solution. Ceci est fait, normalement, par un ordinateur. Il n'existe pas un algorithme universel d'optimisation. Plutôt, il existe de nombreux algorithmes, chacun étant adapté à un type particulier de problème d'optimisation. Il est souvent de la responsabilité de l'utilisateur/programmeur/modéleur de choisir

un algorithme approprié pour l'application spécifique. Ce choix est très important puisqu'il déterminera si le problème est résolu rapidement ou pas et effectivement, même si la solution est déjà trouvée. Après l'application d'un algorithme d'optimisation au problème, on doit être capable de reconnaître si l'on a réussi à trouver une solution. Dans certains cas, il existe des expressions mathématiques, les conditions d'optimalité, pour vérifier si l'ensemble des variables obtenu est bien la solution du problème. Si les conditions d'optimalité ne sont pas vérifiées, elles peuvent donner des indications sur la manière d'améliorer notre estimation actuelle. Finalement, le modèle peut être amélioré par l'application de techniques comme l'analyse de sensibilité qui révèle la sensibilité de la solution aux changements dans le modèle et dans les données. En résumé, voici les étapes principales :

- 1- identifier un objectif et des variables.
- construire un modèle approprié ;
- choisir un algorithme d'optimisation ;
- vérifier les conditions d'optimalité ;
- faire éventuellement une analyse de sensibilité..

Nous divisons les problèmes d'optimisation en plusieurs familles et catégories (voir la Figure 0.0.1) :

- 1- optimisation continue.
- optimisation discrète ;
- optimisation stochastique.

Parmi les problèmes pratiques en optimisation continue, nous pouvons citer les suivants :

- trouver le trajectoire optimal pour un avion, une navette spatiale ou le bras d'un robot ;

- identifier les propriétés sismiques de l'écorce terrestre en ajustant un modèle à un ensemble d'enregistrements pris sur le terrain (en fait, un problème inverse) ; composer une porte-feuille d'investissements afin de maximiser le retour tout en gardant un niveau de risque acceptable ;

- contrôler un processus chimique ou un dispositif mécanique afin d'optimiser la performance ;

- calculer la forme optimale d'une voiture, un voilier ou un avion

On s'intéresse dans ce mémoire au problème (P) suivant :

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

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

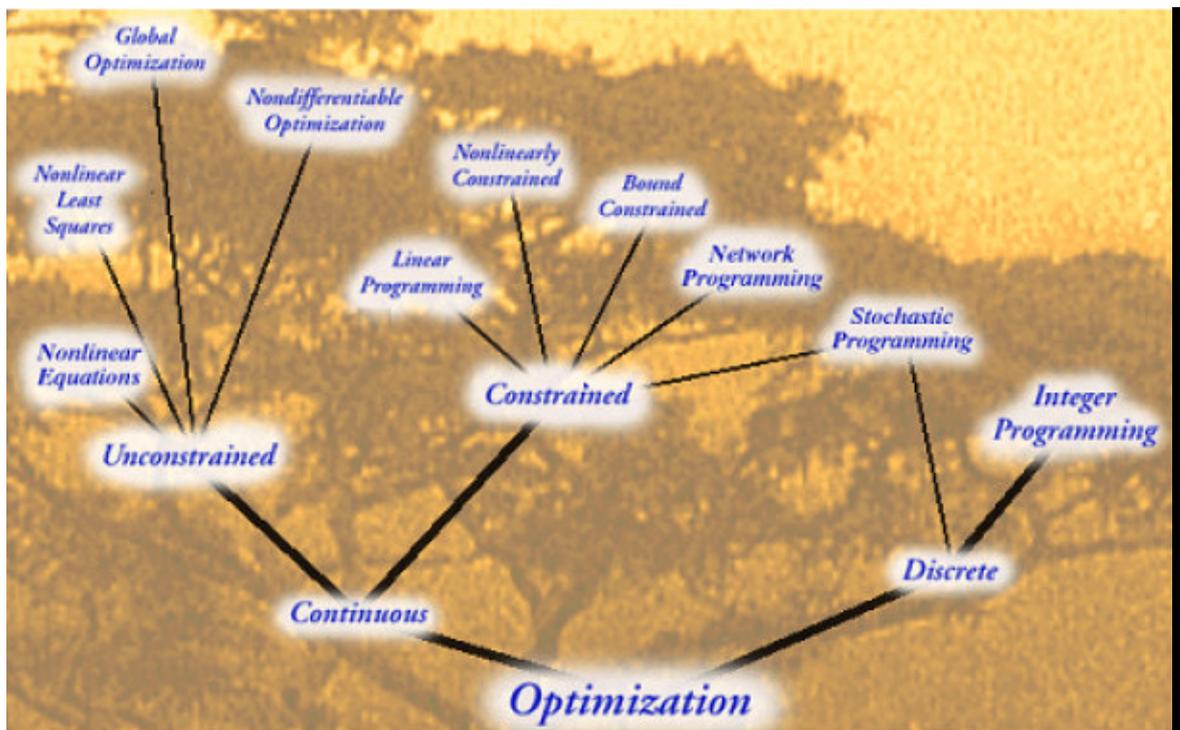


FIGURE 0.0.1. L'arbre d'optimisation

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 de fonctions quadratiques strictement convexes. Plusieurs mathématiciens ont étendu cette méthode pour le cas non linéaire. Ceci a été réalisé pour la première fois, en 1964 par Fletcher et Reeves (*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 :

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

Le pas $t_k \in \mathbb{R}$ 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 β_k définissent les différentes formes du gradient conjugué.

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \quad \text{Hestenes-Stiefel [14], 1952.} \quad (0.4)$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \quad \text{Fletcher Reeves [11], 1964} \quad (0.5)$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad \text{Polak-Ribière-Polyak [18, 19], 1969} \quad (0.6)$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \quad \text{descente conjugué. [10], 1987.} \quad (0.7)$$

$$\beta_k^{LS} = -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}, \quad \text{Liu-Storey [15], 1991} \quad (0.8)$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}, \quad \text{Dai-Yuan [7], 1999} \quad (0.9)$$

Dans ce travail, on s'intéresse à la méthode de *GC* de *wyl* qui'a été proposée par *wyl* en 2006.ont mis au point une nouvelle méthode du gradient conjugué et ont étudié sa convergence en utilisant la recherche linéaire non monotone de wolf powelle fort (*SWP*).

Le but de ce mémoire est la Recherche linéaire non monotone est proposé pour garante la convergence globale de la méthode du gradient conjugué sous la condition suffisante de décent de la méthode W Y L gradient conjugué sous la forte Wolfe de recherche linéaire.

Ce mémoire comporte quatre chapitres.

Dans le **chapitre premier** on rappelle les notions préliminaires de base concernant l'optimisation sans contraintes.

Dans le **chapitre deux** on expose les grandes lignes des méthodes d'optimisation sans contraintes basées sur les directions de descente et les recherches linéaires. Ces méthodes génèrent des suites $\{x_k\}_{k \in \mathbb{N}}$ de la façon suivante : $x_{k+1} = x_k + t_k d_k$

On suppose connaître la direction de descente d_k au point x_k . La recherche linéaire consiste à trouver t_k de façon à diminuer la fonction f *suffisamment* le long de cette direction. On insistera dans ce chapitre sur les recherches linéaires non monotone avec lesquelles nous allons établir des résultats de convergence pour certaines méthodes.

Dans le **chapitre trois** on étudie la convergence de quelques méthodes de gradient conjugué avec la recherche linéaire inexacte de wolf. On s'intéressera spécialement sur les méthodes de FR. CD. PRP. DY.

Le **chapitre quatre** est la partie la plus importante de ce travail, car il contient La méthode de gradient conjugué de wei-yao-liu(**wyl**) sous la recherche linéaire non monotone de wolfe-powell fort.

Chapitre 1

Optimisation sans contraintes

Nous allons exposer dans ce premier chapitre quelques concepts clés concernant l'optimisation, l'effort définitionnel portera surtout sur la méthode du gradient conjugué, où les questions suivantes seront prises en considération 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 :

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

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

La majorité des méthodes de résolution d'optimisation qu'on étudie par la suite sont de nature itérative, c'est à dire qu'à partir d'un point initial x_1 , elles engendrent une suite infinie $x_1, x_2, \dots, x_k, \dots$ 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 :

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

où d_k détermine la direction de déplacement à partir du point x_k et t_k est solution optimale du problème d'optimisation unidimensionnel suivant $\min_{t>0} f(x_k + t d_k)$, c'est à dire que t_k vérifié :

$$f(x_k + t_k d_k) \leq f(x_k + t d_k), \forall t > 0 \quad (1.3)$$

1.1 Rappel de quelques définitions

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

Convergence des algorithmes.

Définition 1.1 *Un algorithme de résolution est un procédé qui permet, à partir d'une 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 associée 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 *Nous dirons qu'un algorithme décrit par une application multivoque A , est globalement convergent (ou encore : possède la propriété de convergence globale) si, quelque 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).*

Les 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 τ .

◆ 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

Ensemble convexe

Définition 1.3 Soit l'ensemble $C \subseteq \mathbb{R}^n$. C est convexe si et seulement si $\forall x, y \in C, \forall t \in [0, 1], tx + (1 - t)y \in C$

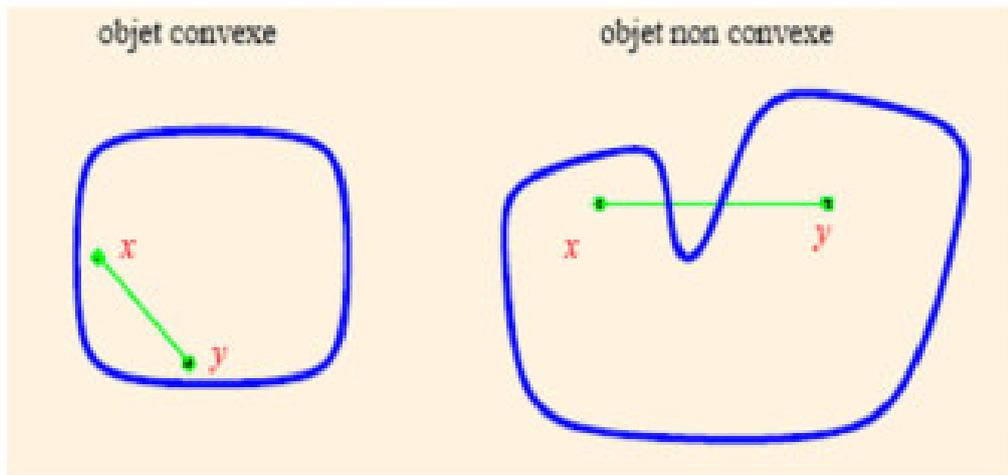


FIG. 1.1 – Objet convexe et non convexe

Fonction convexe

Définition 1.4 Soit $C \subseteq \mathbb{R}^n$ un ensemble convexe non vide. Une fonction $f : C \rightarrow \mathbb{R}$ est convexe si et seulement si

$$\forall x, y \in C, \forall t \in [0, 1], f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$$

Une fonction f est concave si $-f$ est convexe. On dira que f est strictement convexe dans C si et seulement si

$$\forall x, y \in C, x \neq y, \forall t \in [0, 1], f(tx + (1 - t)y) < tf(x) + (1 - t)f(y)$$

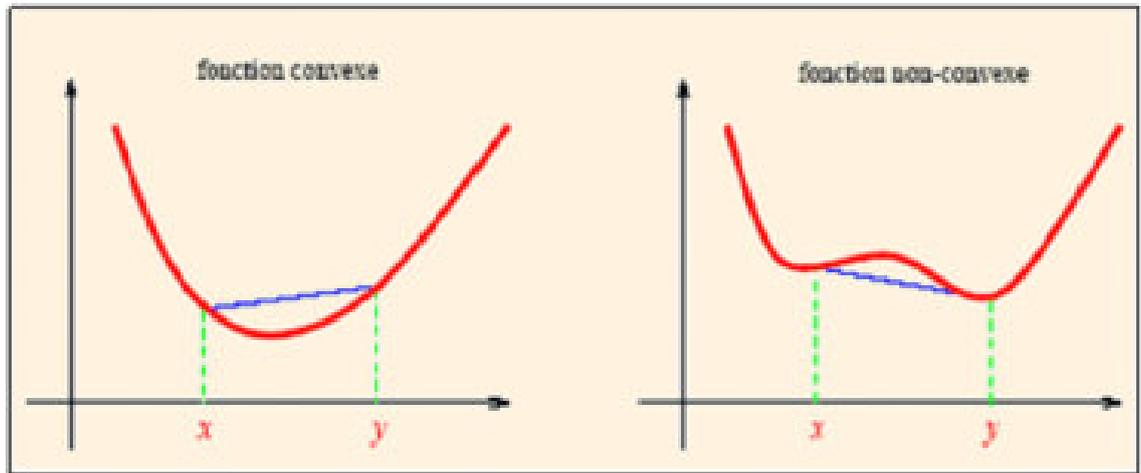


FIG. 1.2 – Fonction convexe et non convexe

Fonction convexe deux fois différentiable

Définition 1.5 Soit $C \subseteq \mathbb{R}^n$ non vide et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite deux fois différentiable au point $\hat{x} \in \text{int}(C)$ s'il existe un vecteur $\nabla f(\hat{x})$ et une matrice symétrique $H(\hat{x})$ d'ordre (n, n) appelée matrice hessienne, et une fonction $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ telles que

$$\forall x \in C : f(x) = f(\hat{x}) + \nabla f(\hat{x})^T (x - \hat{x}) + \frac{1}{2} (x - \hat{x})^T H(\hat{x}) (x - \hat{x}) + \|x - \hat{x}\|^2 \alpha(\hat{x}, x - \hat{x}), \quad (1.4)$$

ou $\alpha(\hat{x}, x - \hat{x}) \xrightarrow{x \rightarrow \hat{x}} 0$. On peut écrire :

$$H(\hat{x}) = \begin{bmatrix} \frac{\partial^2 f(\hat{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 f(\hat{x})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\hat{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\hat{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\hat{x})}{\partial x_2 \partial x_2} & \dots & \frac{\partial^2 f(\hat{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 f(\hat{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\hat{x})}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\hat{x})}{\partial x_n \partial x_n} \end{bmatrix}$$

◆ La matrice $H(\hat{x})$ est dite semi-définie positive si

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

◆ La matrice $H(\hat{x})$ est dite définie positive si

$$\forall x \in \mathbb{R}^n, x \neq 0 : x^T H(\hat{x}) x > 0$$

Théorème 1.1 soit $H(\hat{x})$ la matrice hessienne, on note par $\{t_i\}_{i=1 \dots n}$ ses valeurs propres (réelles) on a les équivalences suivantes :

- ◆ $H(\hat{x}) \geq 0 \iff t_i \geq 0$
- ◆ $H(\hat{x}) > 0 \iff t_i > 0$

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

Considérons le problème d'optimisation sans contraintes (P)

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

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

1.2.1 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

$$f(\hat{x}) \leq f(x) : \forall x \in V_\varepsilon(\hat{x}) \tag{1.5}$$

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

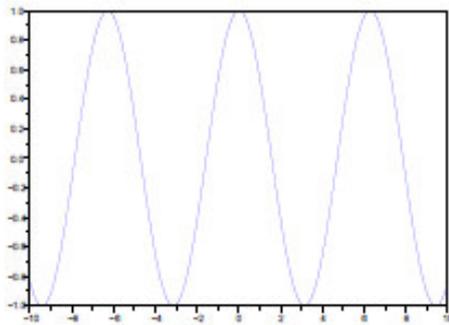
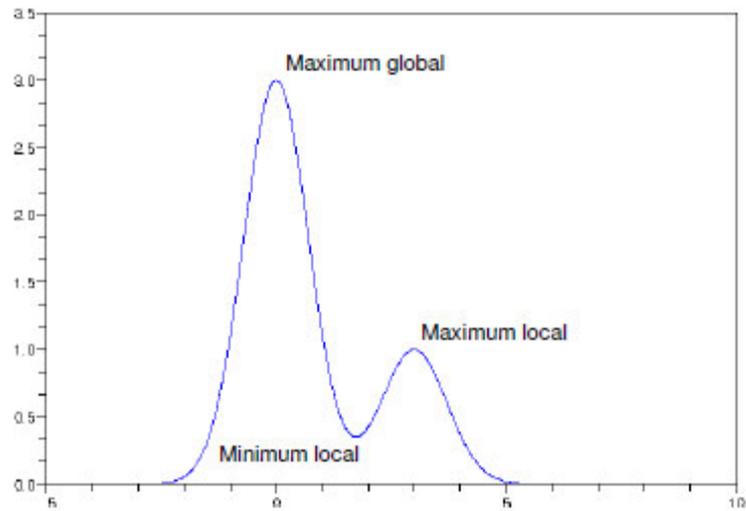
$$f(\hat{x}) < f(x) : \forall x \in V_\varepsilon(\hat{x}), x \neq \hat{x} \tag{1.6}$$

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

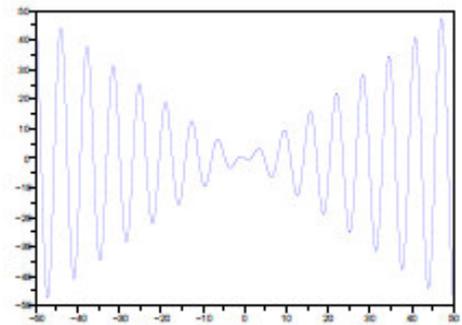
$$f(\hat{x}) \leq f(x) : \forall x \in \mathbb{R}^n \tag{1.7}$$

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 1.2 . 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



Pour la fonction $f : x \mapsto \cos(x)$, il existe une infinité de minima et maxima globaux.



Pour la fonction $f : x \mapsto x \cos(x)$, il existe une infinité de minima et maxima locaux mais aucun minimum ou maximum global.

1.3 Direction de descente

Principes généraux

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

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

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

On note aussi $\nabla f(x)$ et $\nabla^2 f(x)$ le gradient et le hessien de f en x pour le produit scalaire euclidien sur \mathbb{R}^n .

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

Définition 1.7 Soit d un vecteur non nul de \mathbb{R}^n . On dit que d est une direction de descente de f en $x \in \mathbb{R}^n$ si

$$\nabla f(x)^T d < 0. \tag{1.8}$$

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

$$\theta = \arccos \frac{-\nabla f(x)^T \cdot d}{\|\nabla f(x)\| \|d\|} \in \left[0, \frac{\pi}{2} \right[. \tag{1.9}$$

L'ensemble des directions de descente de f en x est le demi espace suivant

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

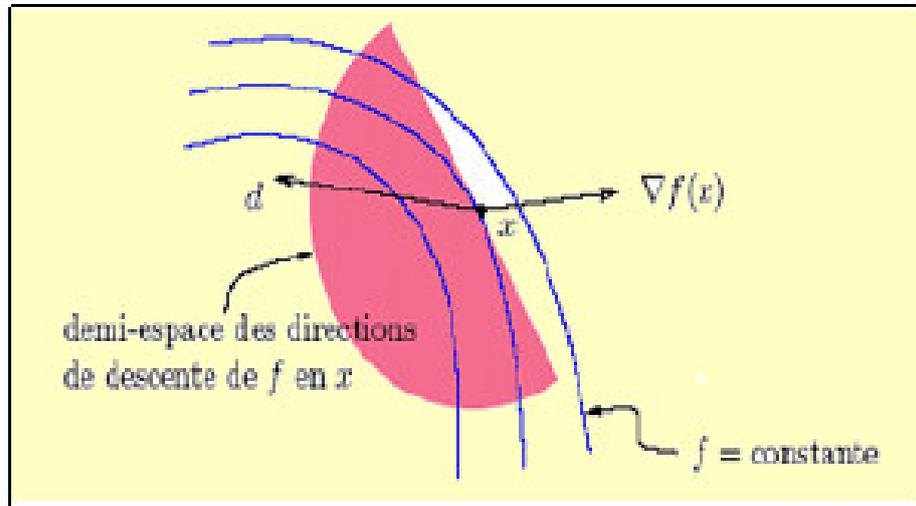


FIG. 1.3 – Demi-espace des directions de descente d de f en x .

On voit que si d est une direction de descente, $f(x + td) < f(x)$, pour tout $t > 0$ 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 :

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

Où t_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 t_k , c'est ce que l'on appelle la recherche linéaire.

Remarque 1.2 Par la suite l'angle θ_k entre la direction d_k et $-g_k$ jouera un rôle important dans le processus de la convergence. Pour θ_k on a la relation classique suivante : $-g_k^T d_k = \|g_k\| \|d_k\| \cos \theta_k$

Définition 1.8 Soit $f : x \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, $\hat{x} \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ est une direction de descente au point \hat{x} si et seulement si il existe $\delta > 0$ tel que

$$f(\hat{x} + td) < f(\hat{x}) : \quad \forall t \in]0, \delta[, \quad (1.9)$$

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

Théorème 1.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $\hat{x} \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ une direction vérifiant la condition suivante :

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

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

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

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

avec $\alpha(\hat{x}, d) \xrightarrow{t \rightarrow 0} 0$ ceci implique que

$$f'(\hat{x}, d) = \lim_{t \rightarrow 0} \frac{f(\hat{x} + td) - f(\hat{x})}{t} = \nabla f(\hat{x})^T \cdot d < 0. \quad (1.10)$$

La limite étant strictement négative, alors il existe un voisinage de zéro $V(0)$ tel que

$$\frac{f(\hat{x} + td) - f(\hat{x})}{t} < 0, \quad \forall t \in V(0). \quad (1.11)$$

La relation (1.11) est particulièrement vraie pour tout $t \in]-\delta, +\delta[$. On obtient le résultat cherché en multipliant la relation (1.11) par $t > 0$ ■

1.3.1 Schéma 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 :

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

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

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

Un bon choix de d_k et t_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 t_k :

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

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

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

1.3.2 Conditions nécessaires d'optimalité

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

Condition nécessaire d'optimalité du premier ordre

Théorème 1.4 *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 $\nabla f(\hat{x}) = 0$.*

Proof. 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 :

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

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 global. Dans le cas où f est deux fois différentiable, une autre condition nécessaire est donnée par le théorème (1.4). Elle est 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)*

Conditions nécessaires d'optimalité du second ordre

Théorème 1.5 *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 $H(\hat{x})$ est semi définie positive.*

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

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

Ceci implique

$$\frac{f(\hat{x} + tx) - f(\hat{x})}{t^2} = \frac{1}{2}x^T H(\hat{x})x + \|x\|^2 \alpha(\hat{x}, tx). \quad (1.12)$$

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

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

Si on prend en considération (1.13) et on passe à la limite quand $t \rightarrow 0$, $t \neq 0$, on obtient :

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

■

1.3.3 Conditions suffisantes d'optimalité

Les conditions données précédemment sont nécessaires (si f n'est pas convexe), c'est-à-dire qu'elles doivent être satisfaites pour tout minimum local ; cependant, tout vecteur vérifiant ces conditions n'est pas nécessairement un minimum local. Le théorème (1.5) établit une condition suffisante pour qu'un vecteur soit un minimum local, si f est deux fois différentiable.

Condition suffisante d'optimalité du second ordre

Théorème 1.6 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) .

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

$$f(x) = f(\hat{x}) + \frac{1}{2}(x - \hat{x})^t H(\hat{x})(x - \hat{x}) + \|x - \hat{x}\|^2 \alpha(\hat{x}, x - \hat{x}), \quad \alpha(\hat{x}, (x - \hat{x})) \xrightarrow{x \rightarrow \hat{x}} 0, \quad (\nabla f(\hat{x}) = 0) \quad (1.13)$$

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

$$x_k \neq \hat{x} : \forall k, \quad x_k \xrightarrow{k \rightarrow \infty} \hat{x} \quad \text{et} \quad f(x_k) \leq f(\hat{x}). \quad (1.14)$$

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

$$\frac{f(x_k) - f(\hat{x})}{\|x_k - \hat{x}\|^2} = \frac{1}{2} d_k^T H(\hat{x}) d_k + \alpha(\hat{x}, (x_k - \hat{x})), \quad \alpha(\hat{x}, (x_k - \hat{x})) \xrightarrow{k \rightarrow \infty} 0. \quad (1.15)$$

(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 $d_k \xrightarrow{k \rightarrow \infty} d$. 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 $H(\hat{x})$ n'est pas définie positive. Ceci est en contradiction avec l'hypothèse. ■

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.

Chapitre 2

Recherche linéaire inexacte

2.1 Recherche linéaire

2.1.1 Introduction

Considérons le problème d'optimisation sans contraintes P .

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

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

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

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

où t_k est solution optimale du problème d'optimisation unidimensionnel suivant :

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

c'est à dire que t_k vérifie

$$f(x_k + t_k d_k) \leq f(x_k + t d_k), \forall t > 0,$$

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

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

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

Dans ce chapitre nous allons décrire les différentes manières de déterminer un pas $t_k > 0$ 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 de 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 :

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

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

Le second objectif, on choisit le pas $t_k > 0$ 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 $t_k > 0$ arbitrairement petit. Or ceci peut entraîner une “fausse convergence”, c'est-à-dire la convergence des itérés vers un point non stationnaire, comme le montre l'observation suivante

Si on prend

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

la suite $\{x_k\}$ générée par (2.4) est de Cauchy, puisque pour $1 \leq l < k$ on a

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

lorsque $l \rightarrow \infty$.

Donc $\{x_k\}$ converge, disons 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 $\overline{B}(x_1, \varepsilon)$. On a donc arbitrairement forcé la convergence de $\{x_k\}$ en prenant des pas très petits.

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

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

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 $t_k > 0$ optimale ou non, vérifiant :

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

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

$$\varphi'(t^*) = 0,$$

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

autrement dit, t^* est un minimum local de φ 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 t^* signifie qu'il va falloir calculer un grand nombre de fois la fonction φ , et cela peut être dissuasif du point de vue du temps de calcul. En pratique, on recherche plutôt une valeur de t qui assure une décroissance suffisante de f . Cela conduit à la notion d'intervalle de sécurité.

2.1.4 Intervalle de sécurité

Définition 2.1 *On dit que $[a, b]$ est un intervalle de sécurité s'il permet de classer les valeurs de t de la façon suivante :*

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

Le problème est de traduire de façon numérique sur φ 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 une bonne stratégie pour itérativement réduire cet intervalle.

2.1.5 Algorithme de base

Etape 0 : (initialisation)

$a = b = 0$, choisir $t_1 > 0$, poser $k = 1$ et aller à l'étape 1 ;

Etape 1 :

Si t_k convient, poser $t^* = t_k$ et on s'arrête.

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

Si t_k est trop grand on prend $b_{k+1} = t_k$, $a_{k+1} = a$,
et on va à l'étape 2 .

Etape 2 :

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

Si $b_{k+1} \neq 0$ déterminer $t_{k+1} \in]a_{k+1}, b_{k+1}[$,
remplacer k par $k + 1$ et aller à l'étape 1.

Il faut maintenant préciser quelles sont les relations sur φ qui va nous permettre de caractériser les valeurs de t 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 t_k comme solution du problème

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

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 φ qui fait décroître par cette fonction :

$$t_k = \inf \{t > 0 : \varphi'(t) = 0, \varphi(t) < \varphi(0)\}.$$

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. Elles ne sont utilisées que

dans des cas particuliers, par exemple lorsque φ est quadratique, la solution de la recherche linéaire s'obtient de façon exacte et dans un nombre fini d'itérations.

Inconvénients : Pour une fonction non linéaire arbitraire,

- ▶ il ne peut pas exister de pas de Cauchy ou de Curry,
- ▶ la détermination de ces pas demande en général beaucoup de temps de calcul et ne peut de toutes 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 t_k minimise φ , 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

On considère la situation qui est typique pour l'application de la technique de recherche linéaire à l'intérieur de la méthode principale multidimensionnelle. Sur une itération k de la dernière méthode nous avons l'itération courante $x_k \in \mathbb{R}^n$ et la direction de recherche $d_k \in \mathbb{R}^n$ qui est direction de descente pour notre objectif : $f : \mathbb{R}^n \rightarrow \mathbb{R}$

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

Le but est de réduire "de façon importante" la valeur de l'objectif par un pas $x_k \rightarrow x_{k+1} = x_k + t_k d_k$ de x_k dans la direction d_k . Pour cela de nombreux mathématiciens (Armijo, Goldstein, Wolfe, Albaali, Lemaréchal, Fletcher...) ont élaboré plusieurs règles.

Schéma des recherches linéaires inexactes

Elles reviennent à déterminer, par tâtonnement un intervalle $[a, b]$, où $t^* \in [a, b]$, dans lequel :

$$f(x_k + t_k d_k) < f(x_k).$$

Le schéma de l'algorithme est donc :

Algorithme (Schéma général des recherches linéaires inexactes)

Etape 0 : (initialisation)

$a_1 = b_1 = 0$, choisir $t_1 > 0$, poser $k = 1$ et aller à l'étape 1 :

Etape 1 :

Si t_k est satisfaisant (suivant un certain critère) : STOP ($t^* = t_k$).

Si t_k est trop petit (suivant un certain critère) : nouvel intervalle : $[a_{k+1} = t_k, b_{k+1} = b]$, et aller à l'étape 2.

Etape 2 :

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

Si $a_{k+1} \neq 0$ déterminer $t_{k+1} \in]a_{k+1}, b_{k+1}[$,
remplacer k par $k + 1$ et aller à l'étape 1.

Il nous reste donc à décider selon quel(s) critère(s) est trop petit ou trop grand ou satisfaisant

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 $\delta \in]0, 1[$ 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 :

$$f(x_k + t_k d_k) \leq f(x_k) + \delta t_k \nabla f(x_k)^T d_k. \quad (2.6)$$

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

$$t \rightarrow f(x_k) + \delta t \nabla^T f(x_k) d_k. \quad (2.7)$$

Règle d'Armijo

- ◆ Si $\varphi(t) \leq \varphi(0) + \delta \varphi'(0)t$, alors t convient.
- ◆ Si $\varphi(t) > \varphi(0) + \delta \varphi'(0)t$, alors t est trop grand.

On peut noter que l'on a

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

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

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

Algorithme 2.1 (Règle d'Armijo)

Etape 0 : (initialisation)

$a_1 = b_1 = 0$, choisir $t_1 > 0$, $\omega_1 \in]0, 1[$ poser $k = 1$ et aller à l'étape 1.

Etape 1 :

Si $\varphi(t_k) \leq \varphi(0) + \delta \varphi'(0)t_k$: STOP ($t^* = t_k$).

Si $\varphi(t_k) > \varphi(0) + \delta \varphi'(0)t_k$, alors

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

Etape 2 :

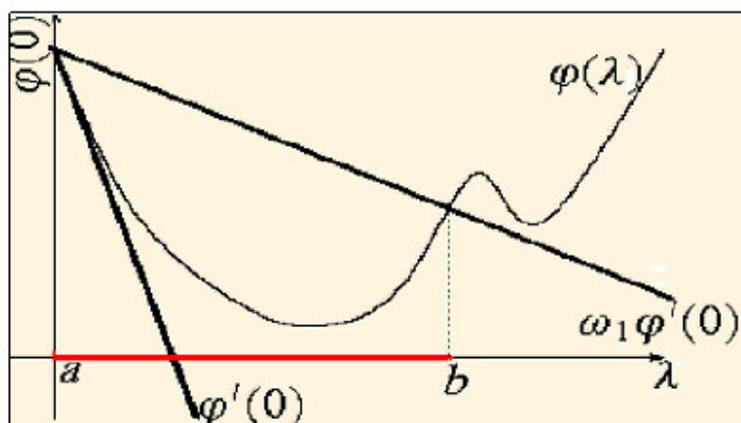


FIG. 2.1 - La règle d'Armijo

Si $b_{k+1} = 0$ déterminer $t_{k+1} \in] a_{k+1}, +\infty [$.
 Si $b_{k+1} \neq 0$ déterminer $t_{k+1} \in] a_{k+1}, b_{k+1}[$,
 remplacer k par $k + 1$ et aller à l'étape 1.

Remarque 2.1

1) En pratique, la constante δ est prise très petite, de manière à satisfaire (2.6) le plus facilement possible. Typiquement, $\delta = 10^{-4}$. 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 $\delta < \frac{1}{2}$ pour que le pas t_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 $t_k > 0$ est suffisamment petit.

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* [60, 1966] ou celle de *Goldstein*.

La règle de Goldstein

La règle de *Goldstein* remédie à cet inconvénient (le pas t_k doit être trop petit). Dans celle-ci, en ajoutant une deuxième inégalité à la règle d'*Armijo* on obtient la règle de *Goldstein*.

$$f(x_k) + \delta t_k \nabla^T f(x_k) d_k \geq f(x_k + t_k d_k) \geq f(x_k) + \sigma t_k \nabla^T f(x_k) d_k, \quad (2.8)$$

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

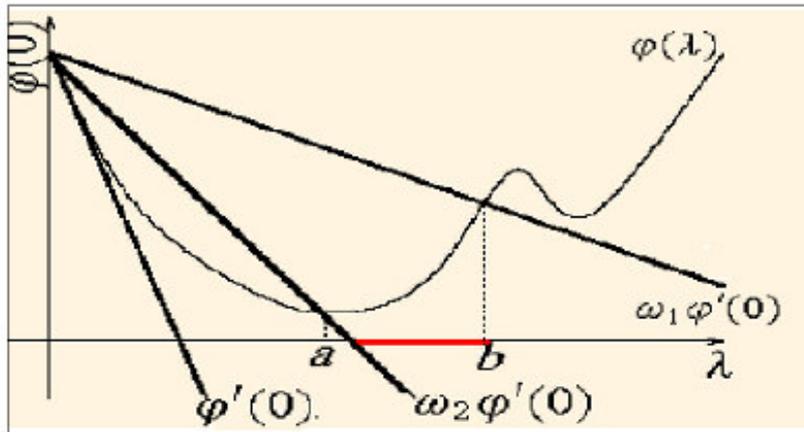


FIG. 2.2 – La règle de Goldstein

la règle de Wolf :

La règle de *Wolfe* fait appel au calcul de $\varphi'(t)$, 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 $t > 0$:

Règle de wolf faible

$$f(x_k + td_k) \leq f(x_k) + \delta t \nabla f(x_k)^T \cdot d_k.$$

$$\nabla f(x_k + td_k)^T \cdot d_k \geq \sigma \nabla f(x_k)^T \cdot d_k.$$

avec $0 < \delta < \sigma < 1$.

Pour expliquer ces conditions posons

$$l(t) = \phi(0) + (\sigma \phi'(0))t.$$

Règle de Wolfe faible

- ◆ Si $\phi(t) \leq \phi(0) + \delta \phi'(0)t$ et $\varphi'(t) \geq \sigma \varphi'(0)$, alors t convient.
- ◆ Si $\phi(t) > \phi(0) + \delta \phi'(0)t$, alors t est trop grand.
- ◆ Si $\varphi'(t) < \sigma \varphi'(0)$, alors t est trop petit.

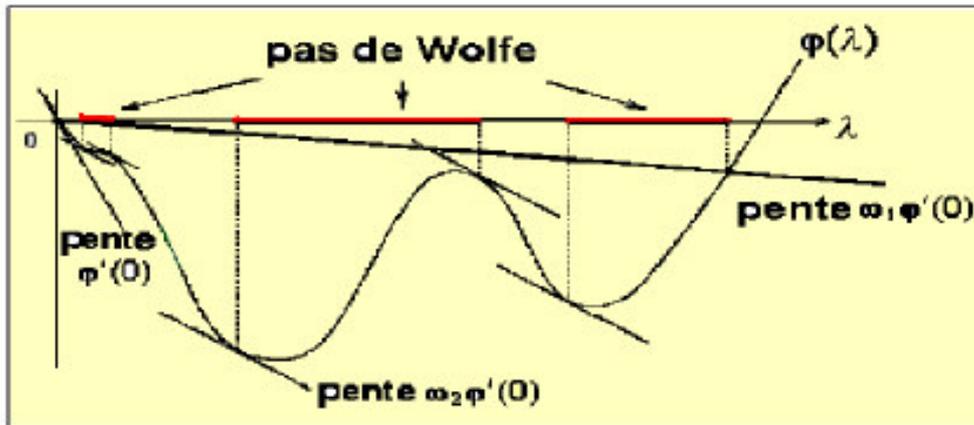


FIG. 2.3 – La règle de Wolfe

Algorithme 2.3 (Règle de Wolfe)

Etape 0 : (initialisation)

$a_1 = b_1 = 0$, choisir $t_1 > 0$ $\delta \in]0, 1[$, $\sigma \in]\delta, 1[$, poser $k = 1$ et aller à l'étape 1.

Etape 1 :

Si $\phi(t_k) \leq \phi(0) + \delta\phi'(0)t_k$ et $\phi'(t_k) \geq \sigma\phi'(0)$: STOP($t^* = t_k$).

Si $\phi(t_k) > \phi(0) + \delta\phi'(0)t_k$, alors

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

Si $\phi'(t_k) < \sigma\phi'(0)$, alors

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

Etape 2 :

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

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

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

$$|\nabla f(x_k + td_k)^T \cdot d_k| \leq -\sigma \nabla f(x_k)^T \cdot d_k$$

Les (W1) et (W3) sont les conditions de *Wolfe fortes*. La contrainte (W3) entraîne que $\sigma\phi'(0) \leq \phi'(t) \leq -\sigma\phi'(0)$ c-à-d. $\phi'(t)$ n'est pas "trop" positif.

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

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

2) L'existence de valeurs t 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 $0 < \delta < \sigma < 1$, 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 :

$$\begin{aligned} f(x_k + t d_k) &\leq f(x_k) + \delta t \nabla^T f(x_k) d_k \\ \acute{\sigma} \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + t_k d_k) d_k \leq -\ddot{\sigma} \nabla^T f(x_k) d_k, \end{aligned}$$

où $0 < \delta < \acute{\sigma} < 1$ et $\ddot{\sigma} > 0$.

Remarque 2.2 .

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 $\acute{\sigma} = \ddot{\sigma} = \sigma$, (w5) est équivalente à (w3). En effet :

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

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

$$\begin{aligned} \acute{\sigma} &= \nabla^T f(x_k) d_k \leq \nabla^T f(x_k + t_k d_k) d_k \leq -\ddot{\sigma} \nabla^T f(x_k) d_k \\ \Rightarrow \sigma \nabla^T f(x_k) d_k &\leq \nabla^T f(x_k + t_k d_k) d_k. \end{aligned} \quad (w2)$$

Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search

CHAPITRE 2. RECHERCHE LINÉAIRE INEXACTE

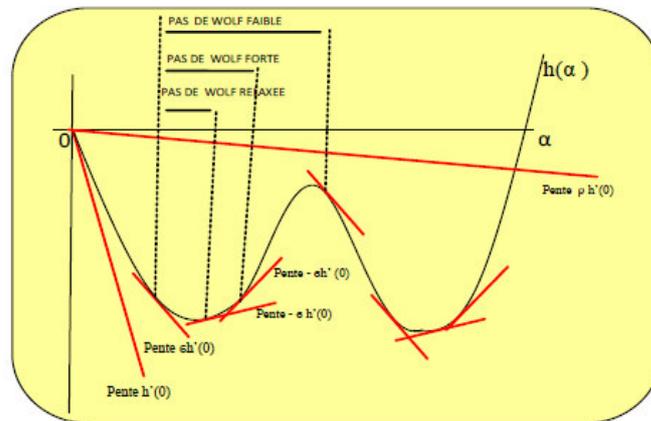


FIG . comparaison entre les trois règles de wolf

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

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

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*

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

Proof. 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 , $f(x_k) \geq C'$. Il ya des propositions précisent 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*).

qui concernant 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 continument différentiable dans un voisinage de $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$. 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.

Proof. Noton $g_k = \nabla f(x_k)$ et $g_{k+1} = \nabla f(x_k + t_k d_k)$. D'après (W2)

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

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

et du fait que f est continument différentiable :

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

en utilisant (w1), on aura :

$$\begin{aligned} f(x_k + t_k d_k) &\leq f(x_k) + \delta t_k g_k^T d_k \\ \Rightarrow & f(x_k + t_k d_k) \leq f(x_k) + \delta t g_k^T d_k \leq f(x_k) + |\delta t g_k^T d_k| \\ \Rightarrow & f(x_k + t_k d_k) \leq f(x_k) + \delta t |g_k^T d_k| \leq f(x_k) - \delta t g_k^T d_k \\ \Rightarrow & f(x_k + t_k d_k) \leq f(x_k) - \delta t \|g_k\| \|d_k\| \cos \theta_k \\ \Rightarrow & f(x_k + t_k d_k) \leq f(x_k) - \frac{\delta(1 - \sigma)}{L} \|g_k\|^2 \cos^2 \theta_k. \end{aligned}$$

On en déduit (2.10). ■

2.5 Méthodes itératives d'optimisation sans contraintes

Il convient de souligner que la plupart des algorithmes d'optimisation sans ou avec contrainte 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, soit le problème (P)

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

pour lequel nous allons commencer par décrire les méthodes suivantes :

- ◆ Les méthodes de gradient
- ◆ Les méthodes utilisant des directions conjuguées
- ◆ Les méthodes de Newton et quasi-Newton.

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.

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

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

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

$$x_{k+1} = x_k + t_k d_k \quad t_k > 0, \text{ tout en assurant la propriété}$$

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

Le vecteur d_k est la direction de descente en x_k . Le scalaire t_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 de gradient (méthodes de la plus forte descente)

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

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

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 $t_k = t > 0, \forall k$. On obtient alors la méthode du gradient simple :

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

Sous certaines hypothèses de régularité (f deux fois différentiable) cette méthode converge si t est choisi assez petit. On bien consiste à faire les itérations suivantes

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

Où t_k est choisi de manière à ce que

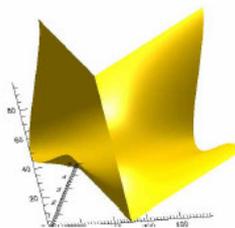
$$f(x_k + t_k d_k) \leq f(x_k + t d_k), \forall t > 0. \quad (2.16)$$

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

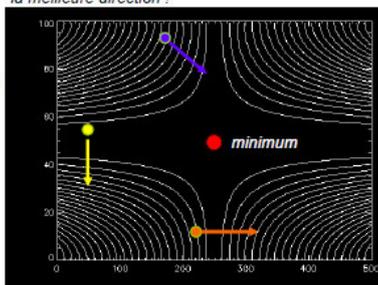
CHAPITRE 2. RECHERCHE LINÉAIRE INEXACTE

Soit à minimiser $F(x,y)$ (dessin ci-dessous)

Imaginons que nous sommes dans une « vallée » étroite.



La flèche est la direction de - gradient.
Pour les points jaune et orange le gradient n'est pas
la meilleure direction !



=> gradient conjugué

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

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

Proof. Si on introduit la fonction

$$\varphi(t) = f(x_k + td_k),$$

on a

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

et puisque φ est dérivable on a nécessairement

$$\varphi'(t) = 0,$$

donc

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

■

► 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 $\varphi(t) = f(x_k + td_k)$. Le pas optimal t_k est caractérisé par

$$\varphi'(t) = 0,$$

on a donc

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

soit

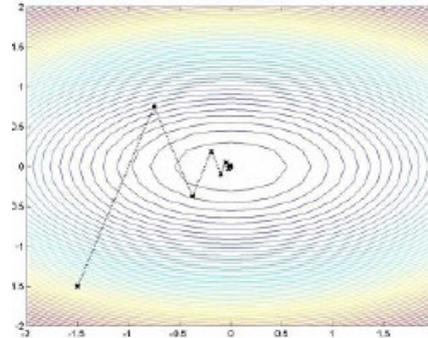
$$(\nabla f(x_k) + t_k H d_k)^T d_k = 0,$$

on obtient donc

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

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)

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

Exemple 2.1

► 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

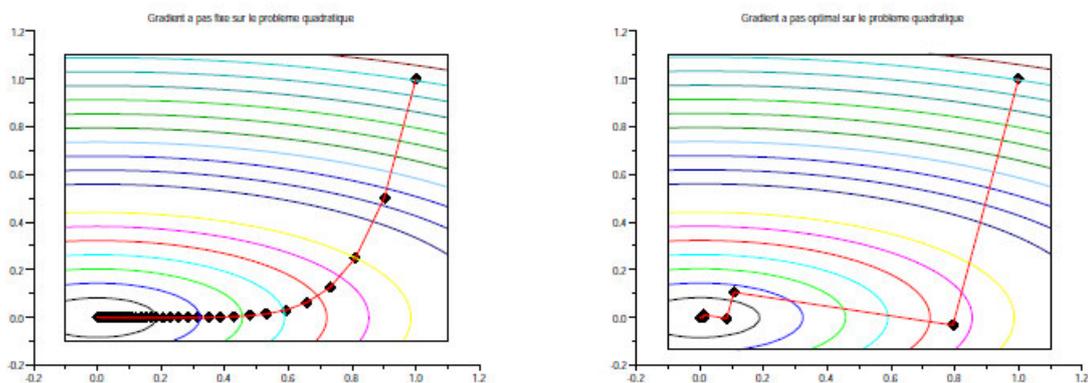
$$\begin{aligned} d_k &= b - Hx_k, \\ x_{k+1} &= x_k + t d_k, \end{aligned}$$

où $t > 0$ est fixé a priori. Il existe bien sûr des conditions sur pour que la méthode converge.

► 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 :

CHAPITRE 2. RECHERCHE LINÉAIRE INEXACTE



Algorithmes de gradient à pas fixe et à pas optimal.

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

Algorithme de la méthode de la plus forte pente

Etape initiale :

Choisir un $\varepsilon > 0$.

Choisir un point initial x_1 .

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

Etape principale :

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

Si non poser $d_k = -\nabla f(x_k)$ et soit t_k la solution optimale de la recherche linéaire

$$\min \{f(x_k + td_k); t > 0\}.$$

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

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

$$f(x_k + td_k) = f(x_k) + t\nabla f(x_k)^T d + t\|d\| \alpha(x_k; td),$$

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

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

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

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 t , 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} .

2.5.3 Les méthodes utilisant des directions conjuguées

Description de la méthode

Ces méthodes 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 de direction conjuguées peuvent résoudre les problèmes de cette forme en au plus n itérations et, contrairement aux méthodes présentées jusqu'à présent, elles 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 des directions A -conjuguées sont linéairement indépendantes, alors l'espace vectoriel engendré par un ensemble de n directions A -conjuguées est de dimension n .

Étant 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

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

où x_0 est un vecteur de départ choisi arbitrairement et où les t_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. À la n ^{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 assure qu'à l'itération n , l'espace vectorielle générée 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 methode 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 à 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 ($\beta_k \in \mathbb{R}$)

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

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

$$x_{k+1} = x_k + t_k d_k,$$

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

Ces méthodes sont des extensions de la méthode du gradient conjuguée si β_k prend l'une des valeurs

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \quad \text{Hestenes-Stiefel [14], 1952} \quad (2.21)$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \quad \text{Fletcher Reeves [11], 1964} \quad (2.22)$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad \text{Polak-Ribière-Polyak [18, 19], 1969} \quad (2.23)$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \quad \text{descente conjugué. [10], 1987} \quad (2.24)$$

$$\beta_k^{LS} = -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}, \quad \text{Liu-Storey [15], 1991} \quad (2.25)$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}, \quad \text{Dai-Yuan [7], 1999} \quad (2.26)$$

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 ces variantes de β_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* ([35, 1969])-(*[36, 1969]*) *méthode de Fletcher-Reeves* ([18, 1964]), *méthode de la descente conjuguée* ([55, 1987]) ou de *méthode de Dai-Yuan* ([54, 1999]) selon que l'on utilise β_k^{PRP} , β_k^{FR} , β_k^{CD} ou β_k^{DY} à la place de β_k dans (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-à-dire si le pas t_{k-1} est un point stationnaire de $t \rightarrow f(x_{k-1} + td_{k-1})$. 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

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

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

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

où f est régulière (continûment différentiable) et g est son gradient. Notons par g_k le gradient de f au point x_k .

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

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

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

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

Le coefficient β_k détermine la méthode du gradient conjugué en question (Fletcher Reeves, Polak Ribière Polyak,...) (voir Chapitre 2). Le pas $t_k \in \mathbb{R}^+$ étant déterminé par une recherche linéaire.

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

Supposition 3.1.

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

(ii) Sur un voisinage N de Ω , la fonction objectif f est continûment différentiable et son gradient est lipschitzien i.e

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

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 .([40, 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 :

$$f(x_k + td_k) \leq f(x_k) + \delta t \nabla f(x_k)^T \cdot d_k. \quad (3.7)$$

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

avec $0 < \delta < \sigma < 1$:

Les conditions de Wolfe fortes :

$$f(x_k + td_k) \leq f(x_k) + \delta t \nabla f(x_k)^T \cdot d_k. \quad (3.9)$$

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

Les conditions de Wolfe relaxée :

$$f(x_k + td_k) \leq f(x_k) + \delta t \nabla^T f(x_k) d_k \quad (3.11)$$

$$\acute{\sigma} \nabla^T f(x_k) d_k \leq \nabla^T f(x_k + t_k d_k) d_k \leq -\acute{\sigma} \nabla^T f(x_k) d_k, \quad (3.12)$$

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

où $0 < \delta < \sigma < 1$ et $\ddot{\sigma} > 0$

Présentons maintenant un théorème fondamental qui assure la condition de Zoutendijk pour toute méthode du type (3.2)-(3.3), dans laquelle le pas t_k est déterminé par la règle de Wolfe faible (3.7)-(3.8). Ce théorème était démontré par Zoutendijk ([52, 1970]) et Powell ([53, 1971]).

Considérons une méthode du type (3.2) et (3.3) dans laquelle d_k est une direction de descente et le pas t_k est déterminé par la règle de Wolfe faible (3.7)-(3.8) avec $0 < \delta < 1/2$: Considérons aussi que la supposition 3.1 soit satisfaite. Alors pour une telle méthode la condition de Zoutendijk suivante :

$$\sum_{k \geq 1} \cos^2 \theta_k \|g_k\|^2 \leq \infty \quad (3.13)$$

3.1 méthode de Fletcher-Reev

La méthode de Fletcher-Reeves ([11, 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 (chapitre 2) que pour la méthode de Fletcher-Reeves la variante β_k est :

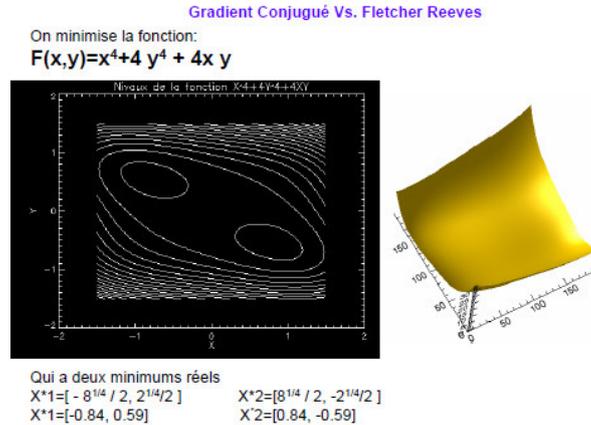
$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad \text{Gradient conjugué variante Fletcher-Reeves} \quad (3.14)$$

Lorsque la fonction à minimiser est quadratique, le coefficient B_K assure la conjugaison entre d_k et d_{k-1} et la méthode est assurée de converger en au plus n itérations, où n est la dimension du vecteur d'inconnues. Dans le cas non quadratique, il existe de nombreuses variantes à la méthode de Fletcher et Reeves, différant par le choix du paramètre B_K .

3.1.1 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 (essentiellement trois vecteurs de dimension n) ; d'autre part, par sa vitesse de convergence très supérieure à celle des algorithmes du gradient classique

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



Algorithme 3.1 de la méthode de Fletcher-Reeves

Etape0 : (initialisation)

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

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

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

Si non aller à l'étape 2.

Etape 2 :

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

où

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

3.1.2 . La propriété de descente de la méthode de FR

Powell ([42, 1984]) a démontré la satisfaction de la propriété de descente de la fonction objective pour la méthode de Fletcher-Reeves avec recherche linéaire exacte.

Al-Baali ([50, 1985]) a démontré la satisfaisse 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([40, 1992]) ont généralisé ce résultat pour toute méthode du type (3. 2)-(3. 3) dont

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

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

Théorème 3.1 *On considère que la supposition (3.1) est satisfaite. Considérons une méthode du type (3.2) et (3.3) avec β_k satisfaisant à (3.14) et le pas t_k vérifiant la règle de Wolfe forte (3.10) où $\sigma \in]0, \frac{1}{2}[$*

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

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

Proof. . ([40, 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 :

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

2) Supposons que (3.16) est satisfaite pour k et démontrons qu'elle le sera pour $k+1$: Supposons que :

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

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 aura :

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

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

Remplaçant ceci dans (**):

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

De (*) on aura :

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

et de (3.15)

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

On aura :

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

Ce qui achève la démonstration. ■

D'après (4.18) on déduit que pour tout $k \geq 1$: d_k est une direction de descente avec $C = \frac{2\sigma - 1}{1 - \sigma}$

En effet :

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

Remarque 3.1 .

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

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

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

Puisque

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

Donc t_k vérifie la condition nécessaire d'optimalité :

$$\dot{\varphi}_k(t_k) = \nabla f(x_k + t_k d_k)^T . d_k = g_{k+1}^T d_k = 0 , \quad \forall k \geq 1 .$$

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

3.1.3 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 $\sigma < \frac{1}{2}$) était démontré par Al-Baali ([50, 1985]). Touati Ahmed et Story ([51,1990]) ont généralisé ce résultat pour

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

Gilbert et Nocedal ([48, 1992]) ont généralisé ce résultat pour

$$|\beta_k| \leq \beta_k^{FR} \quad (3.17)$$

Théorème 3.2 *Supposons que l'hypothèse 3.1 soit satisfaite. Considérons une méthode du type (3.2) et (3.3) dont β_k satisfait à (3.14) et le pas t_k satisfait aux conditions de Wolfe fortes (3.9)-(3.10) où $\sigma \in]0, \frac{1}{2}[$: Alors cette méthode est globalement convergente, dans le sens suivant :*

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.18)$$

Proof. . ([49, 1992])

Puisque les conditions du théorème 3.3.1 sont satisfaites alors on a :

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

D'autre part de (3.10)

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

d'où

$$|d_{k-1}^T g_k| \leq -\sigma d_{k-1}^T g_{k-1} \leq \frac{\sigma}{1-\sigma} \|g_{k-1}\|^2. \quad (3.19)$$

De (3.2), (3.15) et (3.19) :

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

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

Posons

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

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

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 (3.5) on a :

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

d'où

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} \geq \frac{\omega^2}{\hat{\omega} \gamma^4} \sum_{k \geq 1} \frac{1}{k} > \infty. \quad (3.20)$$

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

D'autre part, puisque les conditions du Théorème 3.3.1 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. 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 (3.20), d'où le résultat :

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

■

3.2 Méthode de Polak-Ribière-Polyak

Cette méthode fut découverte par Polak, Ribière ([18, 1969]) et Polyak ([19, 1969]). Rappelons que pour la méthode de Polak-Ribière-Polyak la variante β_k est :

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \quad (3.21)$$

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

**3.2.1 Algorithme 3.2 de la méthode de Polak-Ribière-
Polyak**

Etape0 : (initialisation)

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

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

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

Si non aller à l'étape 2.

Etape 2

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

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

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

où

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

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

3.2.2 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.

Théorème 3.3 *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 .*

Proof. ([33]) 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

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

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

La forte convexité de f implique que

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

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

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

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

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

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

ou encore

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

La condition de Zoutendijk est vérifiée. Donc $\{f(x_k)\}$ est bornée inférieurement (car f est fortement convexe).

On en déduit que $g_k \rightarrow 0$.

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^* . ■

3.3 Méthode de la descente conjugué

Cette méthode fût proposée en 1987 par Fletcher et Reeves([37, 1987]),
Rappelons que pour la méthode de la descente conjuguée la variante β_k
est :

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}} \quad (3.22)$$

3.3.1 Algorithme 3.2 de la méthode de descente conjugué

Etape0 : (initialisation)

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

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

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

Si non aller à l'étape 2.

Etape 2 :

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

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

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

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.

3.3.2 La propriété de descente de la méthode de la descente conjuguée

Fletcher ([37, 1987]) a démontré que la méthode de la descente conjuguée est une méthode de descente si le pas t_k est déterminé par la règle forte de Wolfe(3.9) et (3.10) avec $\sigma \leq \frac{1}{2}$. Dai et Yuan ([35, 1996]) ont démontré que cette méthode avec la règle de Wolfe relaxé ou $0 < \delta < \sigma < 1$ et $0 \leq \sigma' \leq 1$ génère des directions de descente a chaque itération $k \geq 1$

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

Théorème 3.4

Supposons que l'hypothèse 3.1 est satisfaite.

Pour toute méthode du type (3.2) et (3.3) dont β_k satisfait à (3.19)

et le pas t_k satisfait aux conditions de Wolfe relaxée (3.11) et (3.12) :

$$\begin{aligned} f(x_k + t_k d_k) &\leq f(x_k) + \delta t d_k^T g_k \\ \text{et } \sigma d_k^T g_k &\leq d_k^T g_{k+1} \leq -\acute{\sigma} d_k^T g_k \end{aligned}$$

où $0 < \delta < \sigma < 1$ et $0 \leq \acute{\sigma} \leq 1$

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

Proof. ([35, 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)

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

d'où

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

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

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

et donc d_k est une direction de descente suffisante ■

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

3.3.3 Convergence de la méthode de la descente conjuguée

Yuan ([36, 1993]) a démontré la convergence au sens (3.18) de cette méthode avec un pas satisfaisant aux conditions (3.11)-(3.12) si $\sigma < \frac{1}{2}$ et $\acute{\sigma} = 0$. Dai et Yuan ([35,1996]) ont démontré ce résultat pour $\sigma < 1$ et $\acute{\sigma} = 0$,

Théorème 3.5 *Supposons que la l'hypothèse 3.1 est satisfaite. Toute méthode du type (3.2) et (3.3) dans laquelle β_k vérifie (3.19) et le pas t_k est déterminé par la règle de Wolfe relaxée (3.11)-(3.12) où $0 < \delta < \sigma < 1$ et $\acute{\sigma} = 0$; est de descente convergente, dans le sens où*

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

Proof. ([35,1996]) Du théorème 3.4 on a :

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

Donc β_{k+1}^{CD} vérifie l'inégalité (3.15). D'après le théorème 3.2 on a :

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

■

■

3.4 Méthode de Dai-Yuan

Cette méthode était découverte par Dai et Yuan ([33, 1999]),
Rappelons (chapitre2) que pour la méthode de la descente conjuguée

la variante β_k est :

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \quad (3.23)$$

Remarquons que cette variante on a les mêmes numérateur et dénominateur des 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

3.4.1 Algorithme 3.3 de la Méthode de Dai-Yuan avec la règle de Wolfe faible

Etape0 : (initialisation)

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

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

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

Si non aller à l'étape 2.

Etape 2 :

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

t_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.

3.4.2 La propriété de descente de la méthode de Dai-Yuan

Dai et Yuan ([34, 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.

Ils ([33, 1999]) s'ont généralisé ce résultat pour toute fonction régulière

Théorème 3.6 *Supposons que L'hypothèse 3.1 soit satisfaite. Pour toute méthode du type (3.2) et (3.3) dont β_k satisfait à (3.20) et le pas t_k satisfait aux conditions de Wolfe faible :*

$$\begin{aligned} f(x_k + td_k) &\leq f(x_k) + \delta t \nabla f(x_k)^T \cdot d_k \\ \text{et } d_k^T g_{k+1} &\geq \sigma d_k^T g_k \end{aligned}$$

où $0 < \delta < \sigma < 1$

toutes les directions générées sont de descente, autrement dit $d_k^T g_k < 0$; $\forall k \geq 1$

Proof. ([33, 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 :

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

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

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 achève la démonstration. ■

3.4.3 Convergence de la méthode de Dai-Yuan

Dai-Yuan ([33, 1999]) ont démontré la convergence de la méthode de Dai-Yuan au sens (3.17) si le pas t_k est déterminé par la règle de Wolfe faible.

et la fonction objective satisfait les conditions qui situent dans la proposition suivantes.

Théorème 3.7 *Supposons que la proposition 3.1 est satisfaite. La suite $\{x_k\}$ générée par l'algorithme 3.1 converge dans le sens*

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

Proof. ([33, 1999]) En utilisant le théorème 3.1, on aura :

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

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

d'autre part on a :

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

De (3.20) :

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

remplaçant ceci dans (**), on aura :

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

d'où

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

Supposons maintenant que (4.18) n'est pas satisfaite, autrement dit :

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

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

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. ■

Exemple 3.8 Minimiser $f(x) = \frac{1}{2}x^T \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} x + \begin{pmatrix} -4 \\ -8 \end{pmatrix} x$

par la méthode de Fletcher-Reeves à partir de

$$x^0 = (0 \ 0)^T.$$

Vérifier que :

- les directions de déplacement sont mutuellement conjuguées
- les gradients sont orthogonaux
- $d_k^T g_k = -g_k^T g_k < 0$ (Condition de descente)

f(x) est quadratique

⇒ F.R. ≡ algo des gradients conjugués

$$\nabla f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} x + \begin{pmatrix} -4 \\ -8 \end{pmatrix}$$

Première itération :

$$x^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad d_0 = -\nabla f(x^0) = \begin{pmatrix} 4 \\ 8 \end{pmatrix} = -g_0$$

$$\alpha_k ? \quad \alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}$$

$$\Rightarrow \alpha_0 = \frac{(4 \ 8) \begin{pmatrix} 4 \\ 8 \end{pmatrix}}{(4 \ 8) \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 4 \\ 8 \end{pmatrix}} = \frac{80}{544} = \frac{5}{34}$$

$$x^1 = x^0 + \alpha_0 d_0 = \frac{5}{34} \begin{pmatrix} 4 \\ 8 \end{pmatrix} = \begin{pmatrix} \frac{20}{34} \\ \frac{40}{34} \end{pmatrix}$$

$$g_1 = \nabla f(x^1) = \begin{pmatrix} -\frac{48}{17} \\ \frac{17}{17} \end{pmatrix}$$

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

$$\implies B_0 = \frac{g_1^T g_1}{g_0^T g_0} = \frac{36}{289}$$

$$d_1 = -g_1 + B_0 d_0 = \frac{120}{289} \begin{pmatrix} 8 \\ -1 \end{pmatrix}$$

Deuxième itération :

$$\alpha_1 = \frac{-\frac{24}{17} \frac{120}{289} (-2 \ 1) \begin{pmatrix} 8 \\ -1 \end{pmatrix}}{\frac{120}{289} \frac{120}{289} (8 \ -1) \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 8 \\ -1 \end{pmatrix}} = \dots = \frac{17}{40}$$

$$x^2 = x^1 + \alpha_1 d_1 = \begin{pmatrix} \frac{20}{34} \\ \frac{40}{34} \end{pmatrix} + \frac{17}{40} \frac{120}{289} \begin{pmatrix} 8 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$g_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\implies B_1 = 0 \implies \alpha_2 = 0; \quad x^2 = x^3 = \dots$$

\implies optimum atteint (pt stationnaire!)

La méthode converge bien en n (=2) étapes , au plus pour une fct quadratique.

$$\text{Le min est en } x^* = (2 \ 1)^T ; f(x^*) = -8$$

C'est un min global (cf. répet précédente)

Vérifications :

- Directions mutuellement conjuguées ?

$$d_1^T A d_0 = \frac{120}{289} (8 \ -1) \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 4 \\ 8 \end{pmatrix} = 0$$

$\implies d_0, d_1$ conjuguées par rapport à A

- Gradients orthogonaux ?

$$\nabla f(x^0) = g_0 = \begin{pmatrix} -4 \\ -8 \end{pmatrix}$$

$$\nabla f(x^1) = g_1 = -\frac{24}{17} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$$\implies g_0^T g_1 = 0 \text{ ok}$$

- Condition de descente ?

$$k=0 \quad d_0^T g_0 = -g_0^T g_0 < 0 \text{ vérifie car } d_0 = -g_0$$

$$k=1 \quad d_1^T g_1 = \frac{120}{289} (8 \ -1) \begin{pmatrix} 2 \\ -1 \end{pmatrix} \frac{24}{17} = \frac{-120 \cdot 24}{289}$$

$$-g_1^T g_1 = -\left(\frac{24}{17}\right)^2 (-2 \ 1) \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \frac{-120 \cdot 24}{289}$$

$\implies d_0$ et d_1 sont bien des directions de descente

Chapitre 4

la méthode du gradient conjugué de Wei-Yao-Liu sous la recherche linéaire non monotone de Wolfe-Powell forte

4.1 Introduction

On considère le problème d'optimisation sans contraintes suivants :

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

où $f : \tilde{\mathbb{R}}^n \rightarrow \mathbb{R}$ est une fonction non linéaire, lisse dont le gradient est g . La formule itérative de la méthode du gradient conjugué est donnée par :

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

où t_k est une longueur de pas obtenu par une recherche linéaire unidimensionnelle et d_k est le direction de la recherche définie par

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

où $g_k = \nabla f(x_k)$ et β_k est un scalaire, Il y'a au moins six variantes pour β_k , elles sont données ci-dessous par :

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \quad \text{Hestenes-Stiefel [14], 1952} \quad (4.4)$$

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \quad \text{Fletcher Reeves [11], 1964} \quad (4.5)$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad \text{Polak-Ribière-Polyak [18, 19], 1969} \quad (4.6)$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \quad \text{descente conjugué. [10], 1987} \quad (4.7)$$

$$\beta_k^{LS} = -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}, \quad \text{Liu-Storey [15], 1991.} \quad (4.8)$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}, \quad \text{Dai-Yuan [7], 1999} \quad (4.9)$$

De considérables efforts ont été fournis pour l'étude de la convergence globale de ces méthodes.

Zoutendijk [22] a prouvé que la méthode FR avec la recherche linéaire exacte est globalement convergente. Al-Baali [1] a étendu ce résultat à la recherche linéaire de Wolfe-Powell forte. Powell [20] a montré que la suite des normes du gradient $\|g_k\|$ pourrait tendre vers zéro seulement lorsque la relation suivante est satisfaite :

$$\sum \frac{1}{\|d_k\|} < +\infty. \quad (4.10)$$

Donc on peut prouver que la méthode FR est globalement convergente pour les fonctions générales, en utilisant (4.10). Toutefois la convergence globale n'a pas été établie pour la méthode PRP avec les conditions de la recherche linéaire de Wolf- Powell forte. En réalité Powell a prouvé que même si le pas a été choisi pour minimiser la fonction d'une variable $\varphi_k(t) = f(x_k + td_k)$, $t \in R$ la méthode PRP ne converge vers aucune solution la condition suffisante de descente

$$g_k^T d_k \leq -C \|g_k\|^2 \quad (4.11)$$

Pour une certaine constante $c \in (0, 1)$, Gilbert et Nocedal [12] ont donné une autre façon de prouver la convergence globale de la méthode PRP avec la recherche linéaire de Wolfe-Powell faible. Dans [12], le paramètre B_k dans (4,6) ne doit pas être négatif, à savoir :

$$\beta_k = \max \{ \beta_k^{PRP}, 0 \} \quad (4.12)$$

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

En utilisant une recherche linéaire complexe, Gilbert et Nocedal [12] a pu établir le résultat de convergence globale des méthodes PRP et HS , en imposant au scalaire β_k d'être positif ou nul. Pour les algorithmes liés aux méthodes de PRP, Gilbert et Nocedal [12] ont étudié l'ensemble des choix de β_k qui assurent la convergence globale. Grippo et Lucidi [13] ont proposé des nouvelles conditions de recherche linéaire qui ont été conçus pour répondre aux exigences de la théorie de la convergence pour s'assurer que la méthode PRP est globalement convergente pour des problèmes non convexes, mais l'étude de convergence globale ne pourrait pas donner un meilleur algorithme du gradient conjugué ni de meilleures résultats numériques.

ont conclu que pour la méthode PRP, la bornitude de l'ensemble et la restriction (4,12) sont indispensables.

la recherche linéaire inexacte d'Armijo , Dai [3], Dai et Yuan [8] ont discuté la convergence globale des méthodes FR et PRP d'une manière systématique. Ils ont conclu que la méthode PRP est globalement convergente si la fonction objective f est strictement convexe et la condition suffisante de descente (4.11) est vérifiée. Deux nouvelles formules (4,8) et (4.9) ont été données dans [15,7]. La nouvelle formule non linéaire du gradient conjugué (4,9) a été prouvée ainsi que la convergence globale [7-9]. Récemment, Wei et al. [23] ont proposé la nouvelle formule suivante :

$$B_k^{wyl} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{g_{k-1}^T g_{k-1}} \quad (4.13)$$

cette formule n'a pas seulement de bons résultats numériques , mais elle est globalement convergentes avec la recherche linéaire exacte et les conditions de recherche linéaire de Grippo-Lucidi. Donc, il est évident qu'il faut s'assurer qu'elle vérifie la condition suffisante de descente. Dans ce travail , nous allons montrer que la formule (WYL) avec la recherche linéaire de Wolfe-Powell forte (SWP) assure la condition suffisante de descente. Les t_k vérifient les relations de wolfe fortes suivantes :

$$f(x_k + t_k d_k) - f(x_k) \leq \delta t_k g_k^T . d_k. \quad (4.14)$$

$$|g(x_k + t_k d_k)^T . d_k| \leq \sigma |g_k^T d_k| \quad (4.15)$$

Où $\delta \in (0, \frac{1}{2})$ et $\sigma \in (\delta, 1)$

4.2 2 Condition suffisante de descente

Le théorème suivant montre que la formule (WYL) avec la recherche linéaire de (SWP) assure la condition suffisante de descente

Théorème 4.1 .

Supposons que les suites $\{g_k\}$ et $\{d_k\}$ sont générées par la méthode de la forme (4,2),(4,3) et (4,13)

que le pas t_k est déterminé par la recherche linéaire de Wolf- Powell forte (4,14) et (4,15), si $\sigma < \frac{1}{4}$ donc la suite $\{d_k\}$ assure la condition suffisante de descente (4.11).

Proof. Nous avons d'abord prouver la propriété de descente de $\{d_k\}$. D'après (4.3) nous avons $\boxed{g_k^T d_k = -\|g_k\|^2 + B_k^{wyl} g_k^T d_{k-1}}$ en combinant

cette équation avec (B_k^{wyl}) on obtient $\frac{g_k^T d_k}{\|g_k\|^2} = -1 + B_k^{wyl} \frac{g_k^T d_{k-1}}{\|g_k\|^2}$

$$\begin{aligned}
 &= -1 + \left(\frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{g_{k-1}^T g_{k-1}} \right) \frac{g_k^T d_{k-1}}{\|g_k\|^2} \\
 &= -1 + \frac{1}{\|g_k\|^2} \left(g_k^T g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1} \right) \frac{g_k^T d_{k-1}}{g_{k-1}^T g_{k-1}} \\
 \frac{g_k^T d_k}{\|g_k\|^2} &= -1 + \frac{g_k^T d_{k-1}}{g_{k-1}^T g_{k-1}} \left(1 - \frac{g_k^T g_{k-1}}{\|g_k\| \|g_{k-1}\|} \right) \tag{4.16}
 \end{aligned}$$

de (4.15) et (4.16) on :

$$-1 + \frac{\sigma g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \left(1 - \frac{g_k^T g_{k-1}}{\|g_k\| \|g_{k-1}\|} \right) \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -1 + \frac{-\sigma g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \left(1 - \frac{g_k^T g_{k-1}}{\|g_k\| \|g_{k-1}\|} \right) \tag{4.17}$$

puisque $g_1^T d_1 = -\|g_1\|^2 < 0$ si $g_1 \neq 0$, Supposons maintenant que $d_i, i = 1, 2, 3, \dots, k-1$. sont toutes des directions de descente qui vérifient $g_k^T d_k < 0$. De l'inégalité de Cauchy-Schwarz, nous obtenons

$$0 \leq 1 - \frac{g_k^T g_{k-1}}{\|g_k\| \|g_{k-1}\|} \leq 2, \tag{4.18}$$

de (4,17) et (4.18) on a

$$-1 + 2\sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -1 - 2\sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \tag{4.19}$$

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

En répétant ce processus et le fait que $g_1^T d_1 = -\|g_1\|^2$, nous avons

$$-\sum_{j=0}^{k-1} (2\sigma)^j \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \sum_{j=0}^{k-1} (2\sigma)^j \quad (4.20)$$

puisque $\sum_{j=0}^{k-1} (2\sigma)^j < \sum_{j=0}^{\infty} (2\sigma)^j = \frac{1}{1-2\sigma}$, alors de (4,20) on peut écrire que :

$$-\frac{1}{1-2\sigma} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \frac{1}{1-2\sigma}, \quad (4.21)$$

En prenant $\sigma \in (0, \frac{1}{4})$, nous avons $g_k^T d_k < 0$, Donc $\forall k \in N$ $g_k^T d_k < 0$.

Maintenant, prouvons la propriété suffisante de descente de d_k . si $\sigma \in (0, \frac{1}{4})$ on pose $c = 2 - \frac{1}{1-2\sigma}$, $0 < c < 1$, et d'après(4,21) on obtient :

$$c - 2 \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -c, \quad (4.22)$$

$\Rightarrow g_k^T d_k \leq -C \|g_k\|^2$ ou $C = -2 + \frac{1}{1-2\sigma}$
ce qui signifie que (4,11) est vérifiée. ■

·
·

4.3 3 Méthode du gradient conjugué (WYL) et ses applications

On remarque que les six formules (3.2),(33),(3.4),(3.5),(3.6),(3.7) partagent les numérateurs $\|g_k\|^2$ et $g_k^T y_k$ entre-elles les méthodes FR, DY et CD ont toutes un numérateur commun $\|g_k\|^2$ les méthodes PRP, HS et LS ont le même dénominateur $g_k^T y_k$.

Une différence théorique entre ces méthodes et les autres c'est le choix de mise à jour des paramètres. Les théorèmes de convergence globale exigent seulement l'hypothèse de Lipschitz et non l'hypothèse de bornitude. Le premier résultat de convergence globale pour la méthode FR a été donnée par Zoutendijk [28]. Powell [26] a souligné que la méthode FR avec la recherche linéaire exacte a l'inconvénient de produire un effet de zigzaguing.

Les méthodes de DY et e CD se rapproche de la méthode de FR. Les méthodes PRP, HS et LS qui partagent le numérateur commun $g_k^T y_k$, ont l'inconvénient de produire un phénomène de zigzaguing et donc converge très lentement : Lorsque le $x_{k+1} - x_k$ est petit, le facteur y_k dans le numérateur de B_k tend vers zéro. Par conséquent, B_k devient petit et la prochaine recherche de direction d_{k+1} est essentiellement la direction de la plus grande pente $-g_k$. Les méthodes PRP, HS et LS doivent automatiquement ajuster B_k pour éviter le phénomène de zigzaguing. Ainsi la performance numérique de ces méthodes est meilleure que la performance des méthodes avec $\|g_k\|^2$ au numérateur dans B_k . Mais ces méthodes ont possèdent de faibles propriétés de convergence. a partir de des discussions ci-dessus on s'est intéressé à trouver des méthodes qui ont non seulement des propriétés de convergence globale, mais aussi de bonnes expériences numériques.

Récemment, Wei et al. [29] ont proposé une nouvelle formule.

$$B_k^{wyl} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{g_{k-1}^T g_{k-1}} \quad (4.23)$$

La formule WYL a non seulement de bons résultats numérique mais aussi assure la condition suffisante de descente et les propriétés de convergence globale sous la recherche linéaire Wolfe-Powell forte [29,30]. La recherche linéaire de Wolfe-Powell Forte (SWP) permet de trouver t_k satisfaisant

$$f(x_k + t_k d_k) - f(x_k) \leq \delta t_k g_k^T . d_k. \quad (4.24)$$

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

$$|g(x_k + t_k d_k)^T \cdot d_k| \leq \sigma |g_k^T d_k| \quad (4.25)$$

Où $\delta \in (0, \frac{1}{2})$ et $\sigma \in (\delta, 1)$,

de la formule *WYL*, on peut voir que la méthode *WYL* pouvait également éviter le phénomène de zingzagging : lorsque le pas $x_{k+1} - x_k$ est petit, $\frac{\|g_k\|}{\|g_{k-1}\|}$ tend vers 1 et le facteur y_k dans le numérateur de B_k tend vers zéro. Par conséquent, B_k devient petit et le sens de la recherche suivant d_{k+1} est essentiellement la plus forte pente direction. $-g_k$ et dans [30], l'auteur a prouvé qu'en restreignant le paramètre $\sigma < \frac{1}{4}$ sous la condition de recherche linéaire *SWP*, la Méthode *WYL* assure la condition suffisante de descente ainsi que de bons les résultats de convergence globale.

4.3.1 Les formules modifiées

En fait, la formule de *WYL* est similaire à la formule *PRP*. On récrit les formules

PRP, *HS*, *LS* et *WYL* comme suit :

$$\beta_k^{HS} = \frac{g_k^T y_k}{(g_k - g_{k-1})^T d_{k-1}}, \quad (4.26)$$

$$\beta_k^{LS} = -\frac{g_k^T y_k}{d_{k-1}^T g_{k-1}}, \quad (4.27)$$

$$B_k^{wyl} = \frac{g_k^T \bar{y}_k}{g_{k-1}^T g_{k-1}} \quad (4.28)$$

où $y_k = g_k - g_{k-1}$ et $\bar{y}_k = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}$

Alors, une question simple se pose, c'est que si les numérateurs $g_k^T y_k$ des formules *HS* et *LS* ont été remplacé par $g_k^T \bar{y}_k$, alors il est bon de savoir si la formule déduite possède certaines propriétés de convergence qui sont similaires à la méthode de *WYL* ou non ? Cette question a été discutée dans [31,32]. Pour l'intégralité, nous allons montrer quelques propriétés de convergence et les résultats numériques des méthodes déduites puis nous allons comparer les méthodes modifiées avec leurs méthodes d'origines respectivement. d'après la discussion ci-dessus pour les formules de *HS* et de *LS* nous pouvons en déduire la formule modifiée de B_k , on obtient en conséquence :

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

$$\beta_k^{MHS} = \frac{g_k^T \bar{y}_k}{(g_k - g_{k-1})^T d_{k-1}} \quad (4.29)$$

$$\beta_k^{MLS} = -\frac{g_k^T \bar{y}_k}{d_{k-1}^T g_{k-1}}, \quad (4.30)$$

<i>la méthode origine</i>	<i>la méthode modifiée</i>
$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\ g_{k-1}\ ^2}$	$E_k^{wyl} = \frac{g_k^T \left(g_k - \frac{\ g_k\ }{\ g_{k-1}\ } g_{k-1} \right)}{g_{k-1}^T g_{k-1}}$
$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}$	$\beta_k^{MHS} = \frac{g_k^T \bar{y}_k}{(g_k - g_{k-1})^T d_{k-1}}$
$\beta_k^{LS} = -\frac{g_k^T y_k}{d_{k-1}^T g_{k-1}}$	$\beta_k^{MLS} = -\frac{g_k^T \bar{y}_k}{d_{k-1}^T g_{k-1}}$

4.4 2,2. propriétés de convergence

Dans ce paragraphe , nous allons réintroduire les propriétés de convergence de la formule (4.23), (4.29) et (4.30). Ces propriétés sont données dans [29-32]. Nous introduisons d'abord certaines hypothèses et des lemmes qui nous sont utiles. Pour l'étude des méthodes du gradient conjugué la condition suffisante de descente est la suivante :

$$g_k^T d_k \leq -C \|g_k\|^2, C > 0 \quad (4.31)$$

Malheureusement cette condition ne peut être satisfaite pour de nombreuses méthodes de gradient conjugué.

Les hypothèses dont on a besoin et que l'on note A et B sont suivantes :

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

Hypothèses B : Dans certains voisinages N de Ω , $f(x)$ est continûment différentiable et son gradient est lipschitz continu, c'est-à-dire, pour tout $x, y \in N$, il existe une constante $L \geq 0$ telle que

$$\|g(x) - g(y)\| \leq L \|x - y\| \quad (4.32)$$

Gilbert et Nocedal [12] ont introduit les propriétés suivantes (*), propres à la méthode *PRP* sous la condition suffisante de descente. Ces propriétés (*) jouent un rôle important dans l'étude des méthodes CG en général. Nous allons alors montrer que les trois nouvelles formules (4.23), (4.29) et (4.30) possèdent elles aussi à cette propriété.

propriété (*)

Considérons une méthode de forme (4.2) et (4.3). Supposons que

$$0 < \gamma \leq \|g_k\| \leq \bar{\gamma} \quad (4.33)$$

où γ et $\bar{\gamma}$ sont deux constantes positives. Nous disons que la méthode a la propriété (*),

si pour tout k , il existe des constantes $b > 1; k > 0$, telles que $\|B_k\| \leq b$ et $\|s_k\| \leq \lambda$ ce qui implique que $\|B_k\| \leq \frac{1}{2b}$ où $s_k = t_k d_k$

Lemme 4.1 *Supposons que les hypothèses A et B ont vérifiées. considérons la méthode de la forme (1.2) et (1.3), où d_k satisfait $g_k^T d_k < 0$, pour tout k , et t_k est obtenu par SWP (4.24) et (4.25) alors,*

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \quad (4.34)$$

Proof. nous avons $g_k^T d_k < 0$ pour tout $k \geq 1$ Nous avons aussi de (4.25) et (4.32) que

$$-(1 - \sigma)g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq Lt_k \|d_k\|^2 \quad (4.35)$$

d'où

$$t_k \geq -\frac{(1 - \sigma) g_k^T d_k}{L \|d_k\|^2} \quad (4.36)$$

avec (4.24) on obtient :

$$f(x_k) - f(x_k + t_k d_k) \geq -\delta t_k g_k^T d_k \geq -\delta \frac{(1 - \sigma) (g_k^T d_k)^2}{L \|d_k\|^2} \quad (4.37)$$

En outre, de l'hypothèse A nous avons que $\{f(x_k)\}$ est une suite décroissante et a une limite dans Ω , ce qui montre que $\lim_{k \rightarrow \infty} f(x_{k+1}) < +\infty$; et d'après (4.37) on a :

$$+\infty > f(x_1) - \lim_{k \rightarrow \infty} f(x_{k+1}) = \sum |f(x_k) - f(x_{k+1})| \geq \delta \frac{(1 - \sigma)}{L} \sum \frac{(g_k^T d_k)^2}{\|d_k\|^2} \quad (4.38)$$

Ce qui achève la démonstration. ■

Gilbert et Nocedal [12] ont introduit le théorème important suivant :

Théorème 4.2 *Considérons une méthode du CG de la forme (4.2) et(4.3) qui vérifie les conditions suivantes :*

- (1) $B_k > 0$
- (2) la recherche de directions vérifie la condition suffisante de descente (4.31).
- (3) La condition de Zoutendijk (4.34) est vérifiée.
- (4) la propriété (*) est vérifiée.

Si l' hypothèses A et B sont vérifiées alors la méthode converge globalement.

Les lemmes et les théorèmes suivants que nous allons établir montrent que la méthode du CG avec le paramètre (4.23), (4.29) ou

(4.30) satisfait les quatre conditions du théorème 2.1 sous la recherche linéaire de Wolfe-Powell forte (4.24) et (4.25). Dans [29] les auteurs montrent que $B_k^{wyl} \geq 0$ pour tout k et B_k^{wyl} vérifie aussi la propriété(*).

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

de 4,23 on a

$$\begin{aligned}
 B_k^{wyl} &= \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{g_{k-1}^T g_{k-1}} \\
 B_k^{wyl} &= \frac{g_k^T g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{g_{k-1}^T g_{k-1}} \\
 &= \frac{g_k^T g_k \cdot \|g_{k-1}\| - \|g_k\| g_k^T g_{k-1}}{\|g_{k-1}\| g_{k-1}^T g_{k-1}} \\
 &= \frac{\|g_k\|^2 \|g_{k-1}\| - \|g_k\| g_k^T g_{k-1}}{\|g_{k-1}\| \|g_{k-1}\|^2} \\
 &= \frac{\|g_k\|^2 \|g_{k-1}\| - \|g_k\| \|g_k\| \|g_{k-1}\| \cos \alpha_k}{\|g_{k-1}\| \|g_{k-1}\|^2} \cdot \frac{1}{\|g_{k-1}\|^2} \\
 &= \frac{\|g_k\|^2 \|g_{k-1}\| (1 - \cos \alpha_k)}{\|g_{k-1}\| \|g_{k-1}\|^2} \geq 0
 \end{aligned}$$

où α_k est l'angle entre g_k et g_{k-1} et $(1 - \cos \alpha_k) \geq 0$
 $\Rightarrow B_k^{wyl} \geq 0$

Dans [30], les auteurs montrent le théorème suivant :

Théorème 4.3 *Supposons que les suites $\{g_k\}$ et $\{d_k\}$ sont générées par la méthode de la forme (4.2), (4.3) et (4.23) et le pas t_k est déterminée par la recherche linéaire SWP (4.24) et (4., 25), si $\sigma < \frac{1}{4}$, alors Les suites $\{d_k\}$ vérifient la condition suffisante de descente (4.31).*

D'après les résultats de [31,32], nous pouvons déduire les lemmes suivants.

Lemme 4.2 [31,32] *Supposons que les suites $\{g_k\}$ et $\{d_k\}$ sont générées par la méthode de la forme (4.2), (4.3) et (4.29) et le pas t_k est déterminé par la recherche linéaire de SWP (4.24) et (4.25), si $\sigma < \frac{1}{3}$ alors la méthode satisfait toutes les conditions (1) – (4) du théorème 2.1.*

Lemme 4.3 [31,32] *Supposons que les suites $\{g_k\}$ et $\{d_k\}$ sont générées par la méthode de la forme (4.2), (4.3) et (2.30) et le pas t_k est déterminé par la recherche linéaire de SWP (4.24) et (4.25) si $\sigma < \frac{1}{2}$ alors la méthode satisfait à toutes les conditions (1) – (4) du théorème 2.1.*

En résumé, on peut donc dire que :

▲ Pour la méthode du CG avec la formule B_k^{wyl} sous la recherche linéaire de Wolfe-Powell forte si le paramètre $\sigma < \frac{1}{4}$ alors la méthode est globalement convergente.

**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**

▲ Pour la méthode du CG avec la formule B_k^{MHS} sous la recherche linéaire de Wolfe-Powell forte et si le paramètre $\sigma < \frac{1}{3}$, alors la méthode est globalement convergente.

▲ Pour la méthode du CG avec la formule B_k^{MLS} sous la recherche linéaire de Wolfe-Powell forte et si le paramètre $\sigma < \frac{1}{2}$, alors la méthode est globalement convergente..

4.5 2.3. Les résultats numériques

Nous testons les six méthodes du gradient conjugué suivantes.
PRPSWP, WYLSWP, HSSWP, MHSSWP, LSSWP, MLSSWP.
où $\delta = 0,01$ et $\sigma = 0,1$,
la condition d'arrêt est $\|g_k\| < 10$

probleme1

considérer le problème suivant

$$\begin{cases} \text{minimiser } f(x, y) & f(x, y) = 3x^2 + 4y^2 - 3x - 4y \\ & (x, y) \in \mathbb{R}^2 \end{cases}$$

où la solution exacte est (0.5, 0.5)

Les résultats numériques dans les tableaux 3,4 et les figures 1,2 et 3
tableau 3

point initiale	NIPRP	NIWYL	NIHS	NIMHS	NILS	NIMLS
(0, -11)	18	12	9	12		13
(-7, 0)	19	14	12	14	13	14
(1, 1)	10	6	12	14		7
(1, 100)	20	15	26	15		15
(100, 100)		9	11	9		9
(1000, 1000)	11	10	13	10		10
(1, 1000)	25	17	15	16		17
(-1, -1)	6	8	6	9		9

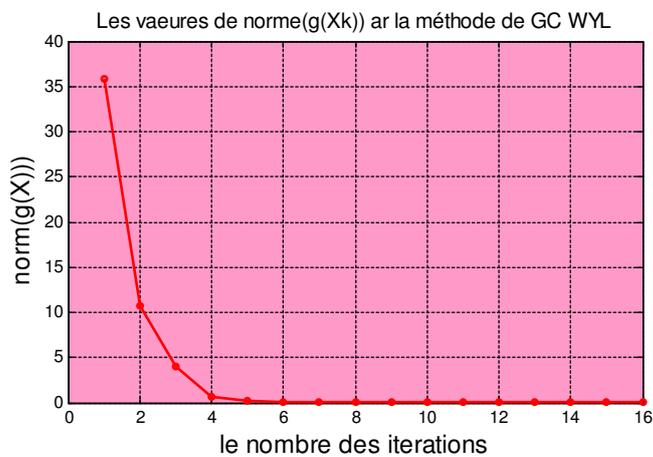
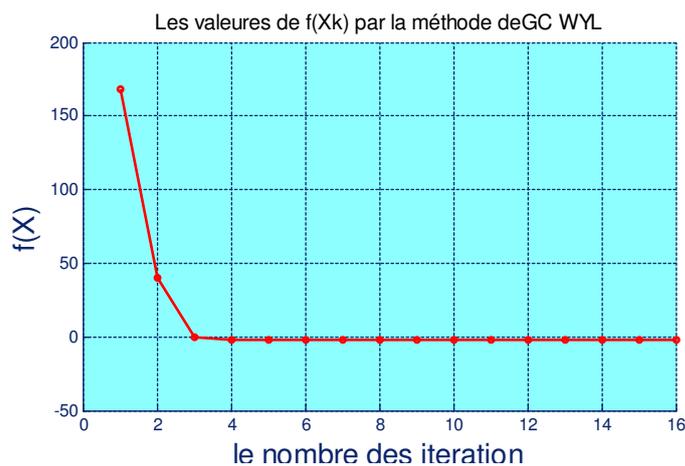
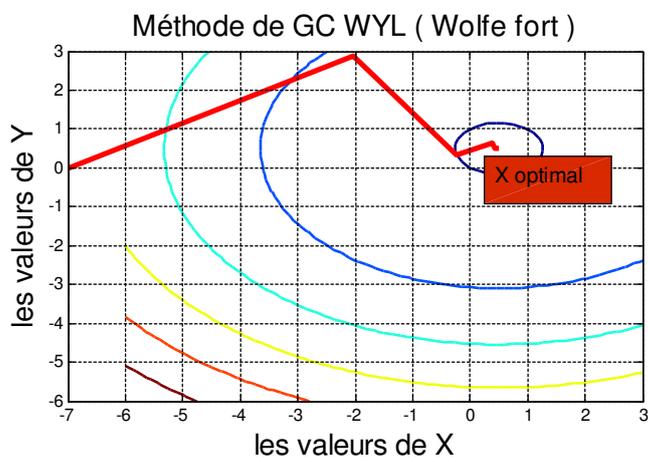
tableau 4

la méthode	PRP	NIWYL	HS	MHS	LS	MLS
la moyenne arithmétique de NI	16	11	13	12.37		11.75

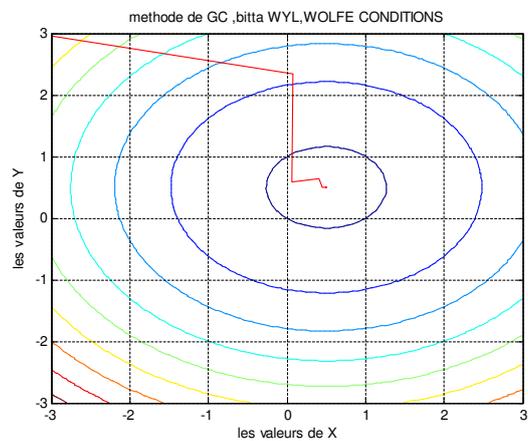
Remarque :

il est claire que la méthode de WYL est un peu mieux que les méthodes de MHS et MLS

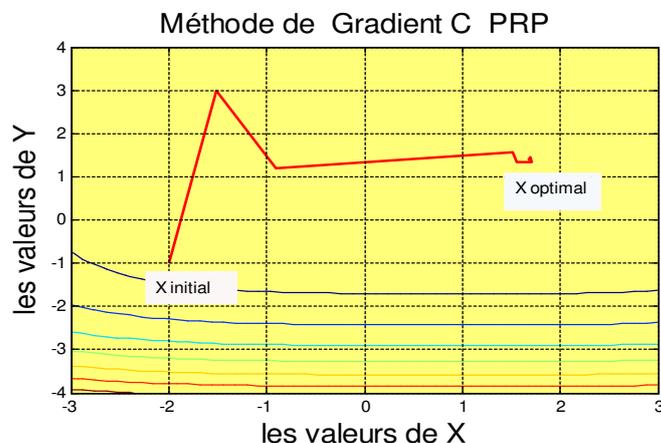
**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**



**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**



**CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE**



Test 2

probleme 2

considérer le problème suivant :

var dim fonction : $n = 2$

$$f(x, y) = (x - 1)^2 + (y - 1)^2 + (x - 3)^2 + (2y - 3)^2 + (x - 1)^4 + (2y - 3)^4;$$

où la solution exacte est (2.3177, 1.4054)

où $\delta = 0,01$ et $\sigma = 0,1$,

la condition d'arrêt est $\|g_k\| < 10^{-5}$

Les résultats numérique dans le tableau 5,6 et les figures 4, 5,6 et 7

tableau 5

point initial	NIPRP	NIWYL
(-2, -1)	37	28
(1, 1)	39	28
(1, 2)	37	17
(-1, -7)	41	19
(-21, -17)	45	23
(-21, -1)	42	28
(-21, 0)	49	27
(0, 10)	43	20

tableau 6

la méthode	PRP	WYL
la moyenne arithmétique de NI	42,625	23,74

Figure 4

*CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE*

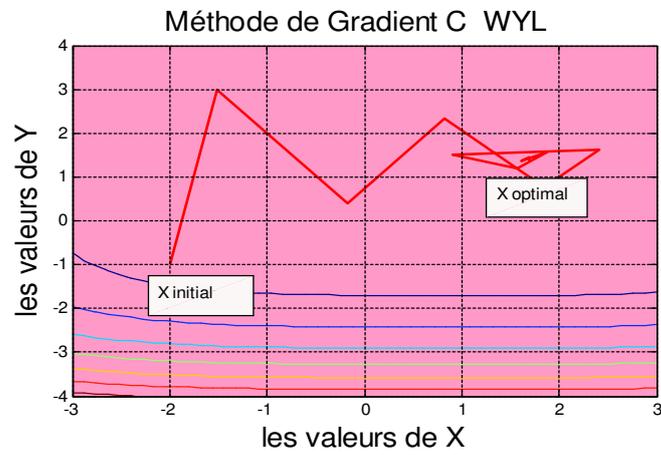


Figure 5

*CHAPITRE 4. LA MÉTHODE DU GRADIENT CONJUGUÉ
DE WEI-YAO-LIU SOUS LA RECHERCHE LINÉAIRE NON
MONOTONE DE WOLFE-POWELL FORTE*

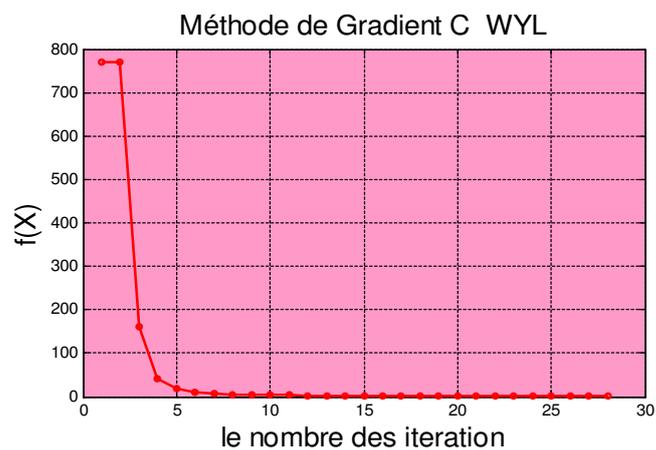
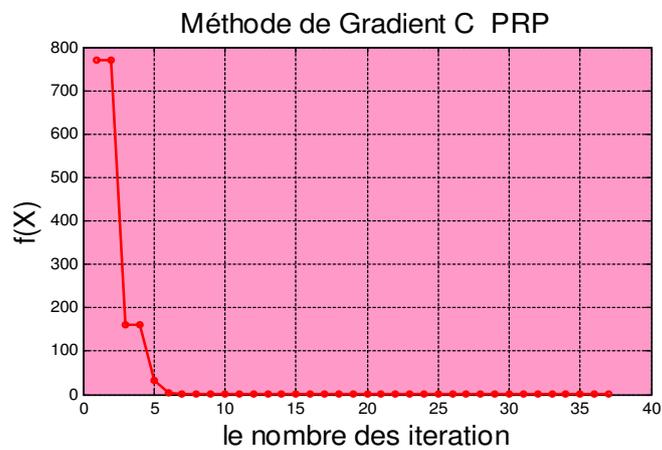


Figure 6
Figure 7

4.6 Conclusion

1- Dans ce travail, nous avons proposé un nouvel algorithme du gradient conjugué pour résoudre des problèmes d'optimisation sans contrainte, Les nouvelles méthodes sont modifiées *LSSWP*, *HSSWP* et *PRPSWP* de telle sorte que la direction générée par l'algorithme résultant répond à la conditions suffisante de descente . Les résultats de convergence globale de la méthode modifiée avec la recherche linéaire Wolfe-Powell forte est établie. Les résultats préliminaires numériques montrent que la performance de la méthode *WYLSWP* ET *MHSSWP* est un peu plus efficace que les méthodes de *MLSSWP*, *MLSSWP* . *PRPSWP*, *HSSWP*, et *LSSWP* pour les problèmes tests donnés

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] X. X. Chen, J. Sun, Global convergence of a two-parameter family of conjugate gradient methods without line search *J. Comput Appl.Math.* 146 (2002) 37–
- [3] Y. Dai, Convergence of nonlinear conjugate methods, *J. Comput Math.* 19 (2001) 539–549
- [4] Y. Dai, Convergence of Polak–Ribie‘re–Polyak conjugate gradient method with constant stepsizes, Manuscript, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 2001.
- [5] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin, Y. Yan, Convergence properties of nonlinear conjugate methods, *SIAM J. Optim.* 2 (1999) 345–358
- [6] Y. Dai, Y. Yan , Further studies on the Polak–Ribie‘re–Polyak method, Research Report ICM-95-040, Institute of Computational Mathematics and Scientific/ Engineering Computing, Chinese Academy of Sciences, 1995.
- [7] Y. Dai, Y. Yan , A nonlinear conjugate gradient with a strong global convergence properties, *SIAM J. Optim.* 10 (2000) 177–182.
- [8] Y. Dai, Y. Yan , A nonlinear Conjugate Gradient Methods, Science Press of Shanghai, Shanghai, 2000.
- [9] Y. Dai, Y. Yuan, An efficient hybrid conjugate gradient method for unconstrained optimization, *Ann..Oper. Res.* 103 (2001) 33–47
- [10] Practical Method of Optimization, second ed., Unconstrained Optimization, vol. I, Wiley, New York, 1997.
- [11] R. Fletcher, Reeves, Function minimization by conjugate gradients, *Compute. J.* 7 (1964) 149–154

-
- [12] , J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* 2 (1992) 21–42
- [13] L. Grippo, S. Lucidi, A globally convergent version of the Polak–Ribie‘re gradient method, *Math. Prog.* 78 (1997) 375–391
- [14] M.R. Hestenes, E. Stiefel, Method of conjugate gradient for solving linear equations, *J. Res. Nat. Bur. Stand.* 49 (1952) 409–436..
- [15] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms, part 1 : theory, *J. Optim. Theory Appl.* 69 (1992) 129–137.
- [16] J. Nocedal, Conjugate gradient methods and nonlinear optimization, in : L. Adams, J.L. Nazareth (Eds.), *Linear and Nonlinear Conjugate Gradient Related Methods*, SIAM, Philadelphia, 1995, pp. 9–23.
- [17] E. Polak, *Optimization : Algorithms and Consistent Approximations*, Springer, New York, 1997.
- [18] E. Polak, G. Ribie‘re, Note Sur la convergence de directions conjuge‘es, *Rev. Francaise Informat Recherche Operationelle*, 3e Anne‘e 16 (1969) 35–43
- [19] B.T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. Math. Phys.* 9(1969) 94–112.
- [20] M.J.D. Powell, Nonconvex minimization calculations and the conjugate gradient method, *Lecture Notes in Mathematics*, vol. 1066, Springer, Berlin, 1984, pp. 122–141.
- [21] J. Sun, J. Zhang, Convergence of conjugate gradient methods without line search, *Ann. Oper. Res.* 103(2001) 161–173
- [22] , G. Zoutendijk, Nonlinear programming computational methods, in : J. Abadie (Ed.), *Integer and Nonlinear Programming*, North-Holland, Amsterdam, 1970, pp. 37–86
- [23] Z. Wei et al., The convergence properties of some conjugate gradient methods, *Appl. Math. Nonlinear Programming*, North-Holland, Amsterdam, 1970, pp. 37–86
- [24] E. Polak, G. Ribie‘re, Note Sur la convergence de directions conjuge‘es, *Rev. Francaise Informat Recherche Operationelle*, 3e Anne‘e 16 (1969) 35–43.
- [25] B.T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. Math. Phys.* 9 (1969) 94–112
- [26] .Powell, Nonconvex minimization calculations and the conjugate gradient method, *Lecture Notes in Mathematics*, vol. 1066, Springer-Verlag, Berlin, 1984, pp. 122–141.

-
- [27] Zhang, Convergence of conjugate gradient methods without line search, *Annal. Oper. Res.*103 (2001) 161–173.
- [28] Nonlinear programming computational methods, in : J. Abadie (Ed.), *Integer and Nonlinear Programming*, North-Holland, Amsterdam, 1970, pp. 37–86.
- [29] Wei et al., The convergence properties of some conjugate gradient methods, *Appl. Math. Comput.* 183 (2006) 1341–1350.
- [30] The proof of the sufficient descent condition of the Wei–Yao–Liu conjugate gradient method under the strong Wolfe–Powell line search, *Appl. Math. Comput.* doi :doi :10.1016/j.amc.2006.12.006, 2007
- [31] Lin Suihua, Yao Shengwei, A new conjugate gradient method of modified LS formula, Department of Mathematics and Information Science, Guangxi University, 2006.
- [32] Shengwei, Lin Suihua, A new conjugate gradient method combined HS and DY formulas, Department of Mathematics and Information Science, Guangxi University, 2
- [33] Y.H. Dai and Y. Yuan (1999), A non linear conjugate gradient with a strong global convergence property, *SIAM J. Optimization*, Vol. 10(1), pp.177-182.
- [34] Y. H. Dai and Y. Yuan (1998), Some properties of a new conjugate gradient method, in : *Advances in Nonlinear Programming*, ed . Kluwer Academic, Boston, pp. 251-262.
- [35] Y. H. Dai and Y. Yuan (1996), Convergence properties of the conjugate descent method, *Advances in Mathematics*, 6, pp.552-562.Y
- [36] Y.X.Yuan (1993), *Numerical Methods for Non linear Programming*, Shanghai scienti. . . c & Technical Publishers (in Chinese)
- [37] .R. Fletcher (1987), *Practical methods of optimization*, John Wiley&Sons, Chichester.
- [38] P. Armand (2005), Modification of the Wolfe Line Search Rules to Satisfy the Descent Condition in the Polak-Ribière-Polyak Conjugate Gradient Method, *Journal of Optimization Theory and Applications*, Vol. 131, pp. 103-115.
- [39] L. Grippo and S. Lucidi (1997), A Globally Convergent Version of the Polak-Ribière Conjugate Gradient Method, *Mathematical Programming*, Vol. 78, pp. 375–391
- [40] J.C. Gilbert and J. Nocedal (1992), Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optimization*. Vol. 3, No.1, pp. 21-42.

-
- [41] M.J.D. Powell (1986), Convergence properties of algorithms for nonlinear optimization, *SIAM rev.*, 28, pp. 487-500
- [42] M.J.D. Powell (1984), Non convex minimization calculation and the conjugate gradient method, in : *Lecture Notes in Mathematics 1066* (Springer, Berlin), pp. 122-14
- [43] M.J.D. Powell (1986), Convergence properties of algorithms for nonlinear optimization, *SIAM rev.*, 28, pp. 487-500.
- [44] A.S. Nemirovsky et D.B. Yudin (1983), *Problem Complexity and Method Efficiency*, New York, Wiley.
- [45] M.J.D. Powell (1977), Restart procedures for the conjugate gradient method, *Math Programming*, 2 , pp 241-254.
- [46] Y. H. Dai and Y. Yuan (1996), Convergence properties of the Fletcher-Reeves method, *IMA J Numer. Anal.*, Vol.16(2), pp. 155-164.
- [47] G.H. Liu and J.Y. Han and H.X. Yin (1995), Global convergence of the Fletcher-Reeves algorithm with an inexact line search, *Appl. math. J. Chinese Univ. Ser. B*, 10, pp. 75-82.
- [48] J.C. Gilbert and J. Nocedal (1992), Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optimization*. Vol. 3, No.1, pp. 21-42.
- [49] 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
- [50] M. Al-Baali (1985), Descent property and global convergence of the Fletcher-Reeves method with inexacte line search. *IMA J. Num. Anal.*, Vol. 5, pp.121-124.
- [51] D. Touati-Ahmed and C. Storey (1990), Efficient hybrid conjugate gradient techniques, *JOTA*, 64, pp. 379-397.
- [52] G. Zoutendijk (1970), *Nonlinear Programming Computational Methods*, Integer and Nonlinear Programming, North Holland, Amsterdam.
- [53] M.J.D. Powell (1971), On the convergence of the variable metric algorithm ; *J. Inst. Math. Appl*, 7, pp. 21-36
- [54] Y.H. Dai and Y. Yuan (1999), A non linear conjugate gradient with a strong global convergence property, *SIAM J. Optimization*, Vol. 10(1),pp.177-182.
- [55] R. Fletcher (1987), *Practical methods of optimization*, John Wiley&Sons, Chichester

- [56] R. Fletcher and C. Reeves (1964), Function minimization by conjugate gradients. *Comput. J.*, 7, pp.149-154
- [57] B.T. Polyak (1969), The conjugate gradient method in extremem problems, *Comput. Math. Math. Phys.*, 9, pp. 94-112...
- [58] E. Polak and G. Ribière (1969), Note sur la convergence de directions conjuguées, *Rev. Française Informat. Recherche Operationelle*, 3e année, 16, pp. 35-43.
- [59] Y.H. Dai and Y. Yuan (1999), A non linear conjugate gradient with a strong global convergence property, *SIAM J. Optimization*, Vol. 10(1),pp.177-182
- [60] L. Armijo (1966), Minimzation of function having lipschitz continuous...rst partial derivatives, *Paci...c Journal of Mathematics*, Vol. 16(1), pp.1-3.