



الجمهورية الجزائرية الديمقراطية الشعبية
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
وزارة التعليم العالي و البحث العلمي
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
جامعة عباس لغرور خنشلة
ABBES LAGHROUR- KHENCHELA UNIVERSITY



Faculty of Sciences and Technology

Department of Mathematics and Computer Science

N° de série :.....

Mémoire de fin d'études

Pour l'obtention du diplôme de Master

Filière: **Mathématiques**
Spécialité: **Mathématiques Appliquées**

Intitulé par :

La convergence de quelques méthodes du gradient conjugué non linéaire.

Réalisé par : **Mahtallah Nassima**

Tlaes Rayenne

Dirigé par : **Dr. Chergui Ahmed**

Présenté le **18/06/2023**

Membres de jury :

M. Y. Meftah **Président**

M. F.Tebessi **Examineur**

Dr. A.Chattouh **Invité**

2022-2023

Remerciements

Nous tenons à remercier tout d'abord

DIEU ALLAH

qui nous a donné durant toutes ces années la santé, le courage et la patience pour en arriver ici.

Je tiens à exprimer ma profonde et sincère gratitude, à notre encadreur

M.CHERGUI Ahmed

Docteur à l'Université Abbes laghrour Khenchela.

Qui a accepté de m'encadrer avec disponibilité et bienveillance. Pour son enseignement, sa patience, et ses précieux conseils.

Nous voudrions aussi remercier tous les membres du jury pour nous avoir honoré par leur présence et pour avoir accepté d'évaluer ce travail de mémoire.

Aussi nous adressons nos sincères remerciements à l'ensemble des enseignants et tout le personnel du département de Mathématique et Informatique.

Enfin, nous remercions nos parents, nos frères et nos sœurs, sans oublier nos amis, qui de près ou loin nous ont soutenus et encouragés tout au long de ces années.

Dédicaces

Je dédie ce modeste travail :

A mes chers parents, pour leurs sacrifices, leur tendresse, leur soutien et leur amour qui m'ont donné toute ma vie, et n'ont pas cessé de m'encourager et de me soutenir et qui ont fait de moi ce que je suis aujourd'hui.

A ma chère sœur "**Ghoufran**" pour sa confiance, son soutien, ses encouragements et ses précieux conseils qui ont contribué à l'amélioration de ce travail .

A mes chers frères "**Djalil**" et "**Rafik**" pour leurs encouragements permanents, et leurs soutien moral.

A mon cher oncle "**MEDFAA AL-Hemami** " pour son soutien tout au long de mon parcours universitaire .

A tous ma famille et ceux qui me sont chers ,et tous mes enseignants .

Mahtallah Nassima

Dédicaces

A ma très chère mère,

Quoi que je fasse ou que je dise, je ne saurai point te remercier comme il se doit. Ton affection me couvre, ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force pour affronter les différents obstacles

A mon cher père,

Tu as toujours été à mes côtés pour me soutenir et m'encourager, que ce travail traduit ma gratitude et mon affection.

A ma chère soeur et mes frères qui m'ont soutenus jusqu'à la fin et ainsi qu'à toute ma famille.

Merci à l'équipe pédagogique de l'université Abbes Laghrour de kenchela et ainsi que toutes les personnes qui m'ont aidés et soutenus lors de mon parcours.

Puisse Dieu vous donne santé, bonheur, courage et surtout réussite.

Tlales Rayenne

Résumé

Dans ce mémoire, nous étudions la convergence de quelques méthode du gradient conjugué non linéaire avec la recherche linéaires inexacte de Wolfe-fort pour résoudre les problèmes d'optimisation sans contraintes.

Nos simulations numériques illustrent l'efficacité des méthodes de gradient conjugué, à travers le temps qu'il faut pour atteindre le minimum, ainsi que le nombre d'itérations.

Mots clés: Optimisation sans contraintes, méthodes gradient conjugué, recherche linéaires inexacte.

ملخص

في هذه المذكرة ندرس تقارب بعض طرق للتدرج المترافق غير الخطي من خلال تحديد الخطوة بواسطة البحث الخطي غير الدقيق القوي لولف لحل مسائل التحسين غير المقيد وبعد اجراء التجارب العددية لبعض المسائل تبين لنا كفاءة طرق التدرج المترافق وذلك من خلال الزمن المستغرق للوصول للنهاية الحدية كذلك عدد التكرارات.

الكلمات المفتاحية

التحسن غير المقيد، التدرج المترافق، البحث الخطي غير الدقيق.

Abstract

In this thesis, we study the convergence of some nonlinear gradient methods combined with Strong Wolf inexact linear search to solve unconstrained optimization problems.

Our numerical simultions iluustrate the efficiency of conjugate gradient methods, through the time it take to reach the minimum, as well as the number of iterations.

Keywords: Unconstrained optimization, conjuguate gradient methods, inexact linear search.

Table des matières

Introduction	1
1 Notions préliminaires	4
1.1 Définitions et notions de base	4
1.2 Problème d'optimisation sans contraintes	7
1.3 Direction de descente	8
1.4 Condition nécessaire d'optimalité du premier ordre :	9
1.5 Condition nécessaire d'optimalité du second ordre :	10
2 Recherches linéaires inexactes	12
2.1 Recherche linéaire	12
2.2 Recherche linéaire inexacte d'Armijio	14
2.3 Recherche linéaire inexacte de Wolfe	15
2.4 Condition de Zoutendijk	16
3 Méthode du gradient conjugué	18
3.1 Méthode du gradient	18
3.2 La méthode du gradient conjugué linéaire	18
3.3 Exemple	21
3.4 La méthode du gradient conjugué non linéaire	23
4 La convergence de quelques méthodes du gradient conjugué non linéaire	26
4.1 Introduction	26
4.2 Partie théorique	27
5 Simulation numérique	33
5.1 Résultats numérique et discussions	33
Conclusion	37

Table des figures

1.1	Extrema d'une fonction à une variable	8
2.1	Règle d'Armijo	14
2.2	Règle de Wolfe	15
5.1	Bouth Fonction	36

Introduction

Problématique Cadre un problème d'optimisation consiste, étant donnée une fonction $f : S \rightarrow \mathbb{R}$, à trouver :

1. Son minimum ν (resp. son maximum) dans S .
2. Un point $x_0 \in S$ qui réalise ce minimum (resp. maximum) i.e $f(x_0) = \nu$.

Vocabulaire

- f est la fonction objectif.
- ν est la valeur optimale.
- x_0 est la solution optimale.
- Écriture du problème : $\min_{x \in S} f(x)$ (resp. $\max_{x \in S} f(x)$).

Remarque

1. L'optimisation est une branche des mathématiques. Dans la pratique, on part d'un problème concret, on le modélise et on le résout mathématiquement (analytiquement : problème d'optimisation, numériquement : programme mathématique).
2. Lien minimum / maximum : soit f une fonction dont on veut trouver le maximum. Le problème $\max_{x \in S} f(x)$ renvoi (x_0, ν) alors que le problème $\min_{x \in S} f(x)$ renvoi $(x_0, -\nu)$. D'où ce lien. Ainsi la recherche d'un maximum peut toujours se ramener à la recherche d'un minimum.

Applications L'optimisation intervient dans de nombreux domaines :

- En recherche opérationnelle (problème de transport, économie, gestion de stocks, ...)
- En analyse numérique (approximation / résolution de systèmes linéaires et non linéaire, ...)
- En automatique (modélisation de systèmes, filtrage, ...)
- En ingénierie (dimensionnement de structures, conception optimale de systèmes (réseaux, ordinateur,...),...)

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

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

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

Parmi les plus anciennes méthodes utilisées pour résoudre les problèmes de types (P), on peut citer la méthode du Gradient conjugué. Cette méthode est surtout utilisée pour les problèmes de grande taille. Cette méthode a été découverte en 1952 par Hestenes et Steifel, pour la minimisation des fonctions quadratiques strictement convexes. Plusieurs mathématiciens ont étendu cette méthode pour le cas non linéaire. Ceci a été réalisé pour la première fois, en 1964 par Fletcher et Reeves (*Méthode de Fletcher-Reeves*) puis en 1969 par Polak, Ribière et Poyak (*Méthode de Polak-Ribière-Poyak*). 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 :

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

Le pas $\alpha_k \in \mathbb{R}$ 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 :

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

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

Récemment plusieurs mathématiciens ont développé des nouvelles méthodes du gradient conjugué non linéaire efficaces afin de résoudre des problèmes d'optimisation sans contraintes. Le but de ce mémoire, nous étudions la convergence de quelques méthodes de gradient conjugué non linéaire avec la recherche linéaire inexacte de Wolfe-fort pour résoudre les problèmes d'optimisation sans contraintes.

Ce mémoire comporte quatre chapitres :

PREMIER CHAPITRE :

On introduit dans ce chapitre les notions préliminaires de base concernant l'optimisation sans contraintes. et quelques notions sur les problèmes de minimisation sans contraintes.

DEUXIEME CHAPITRE :

Le deuxième chapitre est dédié à l'étude des recherches linéaires inexactes d'Armijo et Wolfe de plus l'étude de la convergence des algorithmes à directions de descente.

TROISIEME CHAPITRE :

Dans ce chapitre on étudie la méthode de gradient conjugué. cas des fonctions quadratiques strictement convexe. Et le cas des fonctions non quadratiques.

QUATRIEME CHAPITRE :

Dans ce chapitre on étudie la convergence de la méthode du gradient conjugué non linéaire de WYL et on la compare avec les méthodes de PRP et FR pour la connaissance de l'efficacité de ces méthodes de gradient conjugué, à travers le temps qu'il faut pour atteindre le minimum, ainsi que le nombre d'itérations.

CINQUIEME CHAPITRE : Dans ce chapitre ,nous avons étudié la convergence de la méthode du gradient et sa comparaison avec les méthode *PRP* et *FR* selon les résultat préliminaires numérique de la performance de la méthode *WYL*.

Ce mémoire se terminera par une conclusion générale qui récapitule le travaille effectué.

L'objectif de ce chapitre est de rappeler brièvement les définitions de quelques notions importantes sur lesquelles on va s'appuyer dans notre travail, ainsi que quelques propriétés utiles.

1.1 Définitions et notions de base

Définition 1.1.1 [17] Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

1. Le gradient d'une fonction de plusieurs variables en un certain point $x_0 \in \mathbb{R}^n$ est un vecteur qui caractérise la variabilité de cette fonction au voisinage de ce point. Le gradient est la généralisation à plusieurs variables de la dérivée d'une fonction d'une seule variable.

On le note par ∇f tel que :

$$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_0) \end{bmatrix}. \quad (1.1.1)$$

▷ Il est courant aussi de noter le gradient par : $\overrightarrow{\text{Grad}} f$.

2. On appelle la matrice hessienne (ou simplement la hessienne) de la fonction f en un certain point $x_0 \in \mathbb{R}^n$ la matrice carrée, notée $H(f)$ de ses dérivées partielles secondes.

Plus précisément en supposant que toutes les dérivées partielles secondes de f existent, le coefficient d'indice i, j de la matrice hessienne de f vaut :

$$H_{ij}(f(x_0)) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \text{ avec } i, j = (1, 2, \dots, n). \quad (1.1.2)$$

Ou, en d'autres termes :

$$H(f(x_0)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x_0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_2^2}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x_0) \end{bmatrix}. \quad (1.1.3)$$

Définition 1.1.2 [18] Soit A une matrice carrée, $A \in \mathbf{M}_{n \times n}$

— la matrice A est dite **Symétrique** si :

$$A^T = A. \quad (1.1.4)$$

— la matrice A est dite **Défini positive** si et seulement si elle est positive et inversible, autrement dit :

$$\forall x \in \mathbb{R}^n, x^T A x > 0. \quad (1.1.5)$$

Définition 1.1.3 [4] soit $A \in \mathbf{M}_{n \times n}$ une matrice symétrique définie positive, soit x et y deux vecteurs de \mathbb{R}^n .

On dit que deux vecteurs non nuls x et y sont **conjugués** par rapport à la matrice A ou (A -conjugués), s'ils vérifient :

$$x^T A y = \langle x, y \rangle_A = 0. \quad (1.1.6)$$

On dit que ces deux vecteurs sont orthogonaux pour ce produit scalaire.

▷ Une famille de vecteurs sont alors conjugués par rapport à A s'ils sont orthogonaux au sens de ce produit scalaire.

En particulier, ils forment donc une famille libre.

Définition 1.1.4 [18].

soit $A \in \mathbf{M}_{n \times n}$ une matrice symétrique, b un vecteur de \mathbb{R}^n et c une constante.

— On dit que la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une **forme quadratique** si on peut l'écrire sous la forme suivante :

$$f(x) = \frac{1}{2} x^T A x - b^T x + c. \quad (1.1.7)$$

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

Définition 1.1.5 [6] Soit Ω une partie de \mathbb{R}^n , $f : \Omega \rightarrow \mathbb{R}$ une application, On dit que f est **localement lipschitzienne** si :

$$\forall M > 0, \forall (x, y) \in \Omega^2.$$

$$\|f(x) - f(y)\| \leq M \|x - y\|. \quad (1.1.8)$$

Définition 1.1.6 [6] Une partie Ω de \mathbb{R}^n est dite **convexe** si,

$$\forall t \in [0, 1], \forall (x, y) \in \Omega^2, (1-t)x + ty \in \Omega. \quad (1.1.9)$$

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

Définition 1.1.7 [18] soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application, on dit que :

1. l'application f est **convexe** sur \mathbb{R}^n si :

$$\forall t \in [0, 1], \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n : f(tx + (1-t)y) \leq tf(x) + (1-t)f(y). \quad (1.1.10)$$

2. l'application f est **strictement convexe** sur \mathbb{R}^n si :

$$\forall t \in]0, 1[, \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \text{ avec } x \neq y : f(tx + (1-t)y) < tf(x) + (1-t)f(y). \quad (1.1.11)$$

Définition 1.1.8 . [17] soit Ω un ouvert de \mathbb{R}^n et soit f une application à valeurs dans \mathbb{R} et soit x un point de Ω , f est dite **différentiable** au point x s'il existe une application linéaire continue L de \mathbb{R}^n dans \mathbb{R} telle que :

$$\forall h \in \mathbb{R}^n; f(x+h) = f(x) + L(h) + o|h|, \text{ avec } x+h \in \Omega. \quad (1.1.12)$$

Où, d'une façon équivalente :

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x) - L(h)}{\|h\|} = 0. \quad (1.1.13)$$

▷ Une telle application linéaire L si elle existe, est alors unique.

Elle s'appelle différentielle de f au point x et notée par : $L = Df(x)$. De plus, sa continuité assure la continuité de f en x .

Définition 1.1.9 . [21] Soit f une fonction définie sur \mathbb{R}^n à valeurs dans \mathbb{R} et $d \in \mathbb{R}^n$

1. On dit que f admet une dérivée directionnelle au point x suivant le vecteur d si l'application $t \in \mathbb{R} \mapsto f(x + td)$ est dérivable en 0 et se note par :

$$\frac{\partial f}{\partial d}(x) = \lim_{t \rightarrow 0} \frac{f(x + td) - f(x)}{t}. \quad (1.1.14)$$

▷ Si d est le vecteur nul, cette **dérivée directionnelle** existe et a une valeur nulle. alors on va supposer que d n'est pas le vecteur nul dans ce qui suit.

2. Dans ce cas, si $d = e_k$ (l'un des vecteurs de base de \mathbb{R}^n), on appelle cette dérivée directionnelle la K^{eme} **drivée partielle** de f au point x , que l'on note :

$$\frac{\partial f}{\partial e_k}(x) = \lim_{t \rightarrow 0} \frac{f(x_1, x_2, \dots, x_k + t, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{t}. \quad (1.1.15)$$

Définition 1.1.10 . [4] soit $\{x_k\}_{k \in \mathbb{N}}$ une suite dans \mathbb{R}^n convergeant vers \tilde{x} telle que $x_k \neq \tilde{x}$

◇ Si

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - \tilde{x}\|}{\|x_k - \tilde{x}\|} = \tau < 1.$$

on dit que $\{x_k\}_{k \in \mathbb{N}}$ converge vers \tilde{x} linéairement avec le taux τ .

◇ Si

$$\frac{\|x_{k+1} - \tilde{x}\|}{\|x_k - \tilde{x}\|} \rightarrow 0, \quad \text{quand } k \rightarrow \infty.$$

alors on dit que la convergence est super linéaire.

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

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - \tilde{x}\|}{\|x_k - \tilde{x}\|^p} < +\infty.$$

on dit que la convergence est super linéaire d'ordre p .

En particulier si

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - \tilde{x}\|}{\|x_k - \tilde{x}\|^2} < +\infty.$$

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

1.2 Problème d'optimisation sans contraintes

soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ on appelle problème d'optimisation (minimisation) sans contraintes le problème (P) suivant :

$$(P) : \min f(x) : x \in \mathbb{R}^n. \quad (1.2.1)$$

▷ On considère le problème (1.2.1) :

Définition 1.2.1 soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, une fonction continûment différentiable :

1. Minimum global : $\tilde{x} \in \mathbb{R}^n$ est un minimum global de (1.2.1) si et seulement si :

$$f(\tilde{x}) \leq f(x) : \forall x \in \mathbb{R}^n. \quad (1.2.2)$$

2. Minimum local : $\tilde{x} \in \mathbb{R}^n$ est un minimum local strict de (1.2.1) si et seulement si il existe un voisinage $\nu_\varepsilon(\tilde{x})$ tel que :

$$f(\tilde{x}) \leq f(x) : \forall x \in \nu_\varepsilon(\tilde{x}). \quad (1.2.3)$$

3. Minimum local strict : $\tilde{x} \in \mathbb{R}^n$ est un minimum local strict de (1.2.1) si et seulement si il existe un voisinage $V_\varepsilon(\tilde{x})$ tel que :

$$f(\tilde{x}) < f(x) : \forall x \in V_\varepsilon(\tilde{x}) \text{ avec } x \neq \tilde{x}. \quad (1.2.4)$$

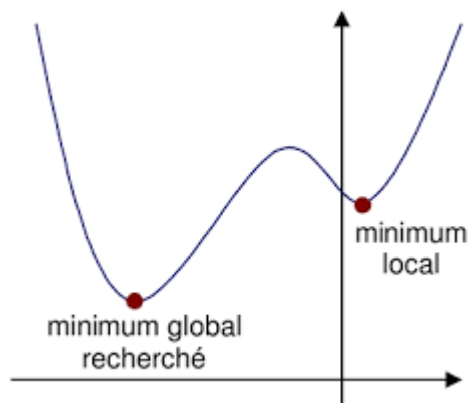


FIGURE 1.1 – Extrema d'une fonction à une variable

1.3 Direction de descente

Définition 1.3.1 soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, d est un vecteur non nul de \mathbb{R}^n est dites **direction de descente** de f en $\tilde{x} \in \mathbb{R}^n$ si et seulement si il existe un réel $\delta > 0$, tel que pour tout α dans l'intervalle $]0; \alpha[$ on a :

$$f(\tilde{x} + \alpha d) < f(\tilde{x}). \quad (1.3.1)$$

Théorème 1.3.1 [21] soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable au point $\tilde{x} \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ une direction vérifiant la condition suivante :

$$\nabla^T f(\tilde{x})d < 0. \quad (1.3.2)$$

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

Preuve 1.3.1 f est différentiable au point \tilde{x} . Donc :

$$f(\tilde{x} + \alpha d) = f(\tilde{x}) + \alpha \nabla f(\tilde{x})^T d + \alpha \|d\| o(\tilde{x}, \alpha d). \quad (1.3.3)$$

avec

$$o(\tilde{x}, \alpha d) \rightarrow 0_{\alpha \rightarrow 0}.$$

ceci implique que

$$\lim_{\alpha \rightarrow 0} \frac{f(\tilde{x} + \alpha d) - f(\tilde{x})}{\alpha} = \nabla^T f(\tilde{x})d < 0. \quad (1.3.4)$$

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

$$\frac{f(\tilde{x} + \alpha d) - f(\tilde{x})}{\alpha} < 0, \quad \forall \alpha \in]-\delta; +\delta[. \quad (1.3.5)$$

La relation (1.3.5) est particulièrement vraie pour tout $\alpha \in]0; +\delta[$. On obtient le résultat cherché en multipliant la relation (1.3.5) par $\alpha > 0$.

C'est à dire :

$$f(\tilde{x} + \alpha d) < f(\tilde{x}), \quad \forall \alpha \in]-\delta; +\delta[.$$

L'ensemble des directions de descente de f en x est donné par :

$$\{d \in \mathbb{R}^n : \nabla f(x)^T d < 0\}. \quad (1.3.6)$$

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

$$\Theta = \arccos \frac{-\nabla^T f(x)d}{\|\nabla^T f(x)d\| \|d\|} \in [0, \frac{\pi}{2}]. \quad (1.3.7)$$

L'ensemble des directions de descente

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

de f en x , forme un demi-espace ouvert de \mathbb{R}^n .

1.3.1 Exemple de choix de directions de descente

Direction du gradient

Si on choisit l'opposé du gradient comme directions de descente; c'est à dire si on prend :

$$d = -\nabla f(x) \quad \text{avec} \quad \nabla f(x) \neq 0. \quad (1.3.8)$$

On obtient alors la direction du **gradient**

Bien sur $-\nabla f(x)$ est une direction de descente, puisque

$$f(x)^T d = -\nabla f(x)^T \nabla f(x) = \langle \nabla f(x), \nabla f(x) \rangle = -\|\nabla f(x)\|^2 < 0. \quad (1.3.9)$$

Direction de Newton

On va supposer que la fonction f est deux fois différentiable en x

La direction de **newton** en un point x en lequel le hessien de f est inversible est définie par :

$$d = -(\nabla^2 f(x))^{-1} \nabla f(x). \quad (1.3.10)$$

Il s'agit bien d'une direction de descente si :

- $\nabla f(x) \neq 0$
- $\nabla^2 f(x)$ est définie positive.

1.4 Condition nécessaire d'optimalité du premier ordre :

Théorème 1.4.1 . [17] Soit $f: \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $\tilde{x} \in \mathbb{R}^n$. Si \tilde{x} est un minimum local de (1.2.1) alors

$$\nabla f(\tilde{x}) = 0. \quad (1.4.1)$$

Preuve 1.4.1 En effet, supposons que

$$\nabla f(\tilde{x}) \neq 0.$$

Puisque la direction

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

est une direction de descente, il existe $\delta > 0$ tel que :

$$f(\tilde{x} + \alpha d) < f(\tilde{x}), \forall \alpha \in]0, \delta[. \quad (1.4.2)$$

Ceci est contradiction avec le fait que \tilde{x} est une solution optimal local de (1.2.1) d'où :

$$\nabla f(\tilde{x}) = 0. \quad (1.4.3)$$

1.5 Condition nécessaire d'optimalité du second ordre :

Théorème 1.5.1 [18] Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable au point $\tilde{x} \in \mathbb{R}^n$. Si \tilde{x} est un minimum locale de (1.2.1) alors

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

et la matrice hessienne de f au point \tilde{x} , qu'on note $H(\tilde{x})$, est semi définie positive.

Preuve 1.5.1 Soit $x \in \mathbb{R}^n$ quelconque, f étant deux fois différentiable au point \tilde{x} , on aura pour tout $\alpha \neq 0$

$$f(\tilde{x} + \alpha d) = f(\tilde{x}) + \frac{1}{2} \alpha^2 d^t H(\tilde{x}) d + \alpha^2 \|d\|^2 o(\tilde{x}, \alpha d). \quad (1.5.1)$$

avec

$$o(\tilde{x}, \alpha d) \rightarrow 0_{\alpha \rightarrow 0}$$

Ceci implique

$$\frac{f(\tilde{x} + \alpha d) - f(\tilde{x})}{\alpha^2} = \frac{1}{2} d^t H(\tilde{x}) d + \|d\|^2 o(\tilde{x}, \alpha d). \quad (1.5.2)$$

Comme \tilde{x} est un optimum local alors

$$f(\tilde{x} + \alpha d) \geq f(\tilde{x}).$$

pour α suffisamment petit, d'où

$$d^t H(\tilde{x}) d + \|d\|^2 o(\tilde{x}, \alpha d) \leq 0. \quad (1.5.3)$$

pour α petit. On passe à la limite quand $\alpha \rightarrow 0$, On obtient

$$d^t H(\tilde{x}) d \leq 0.$$

d'où $H(\tilde{x})$ est semi définie positive.

1.5.1 Condition suffisantes d'optimalité

Théorème 1.5.2 . [17] Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable au point $\tilde{x} \in \mathbb{R}^n$. Si

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

et $H(\tilde{x})$ est définie positive, alors \tilde{x} est un minimum local strict de (1.2.1).

Preuve 1.5.2 f est deux fois différentiable au point \tilde{x} . Alors $\forall x \in \mathbb{R}^n$, on obtient :

$$f(x) = f(\tilde{x}) + \frac{1}{2} (x - \tilde{x})^t H(\tilde{x}) (x - \tilde{x}) + \|x - \tilde{x}\|^2 o(\tilde{x}; x - \tilde{x}). \quad (1.5.4)$$

avec

$$o(\tilde{x}; x - \tilde{x}) \rightarrow_{x \rightarrow \tilde{x}} 0.$$

Supposons que \tilde{x} n'est pas un minimum local strict. Donc il existe une suite x_k convergente vers \tilde{x} telle que

$$f(x_k) \leq f(\tilde{x}). \quad (1.5.5)$$

avec $x_k \neq \tilde{x}, \forall k$

Posons

$$d_k = \frac{x_k - \tilde{x}}{\|x_k - \tilde{x}\|},$$

Donc

$$\|d_k\| = 1,$$

et on obtient :

$$\frac{f(x_k) - f(\tilde{x})}{\|x_k - \tilde{x}\|} = \frac{1}{2} d_k^t H(\tilde{x}) d_k + o(x; x_k - \tilde{x}) \leq 0, \forall k \quad (1.5.6)$$

et comme

$$\|d_k\| = 1, \forall k$$

alors il existe $\{d_k\}_{k \in N_1 \subset \mathbb{N}}$ telle que $d_k \rightarrow d$ pour $k \rightarrow \infty$ et $k \in N_1$.

On a bien sûr

$$d_k = \frac{x_k - \tilde{x}}{\|x_k - \tilde{x}\|}.$$

Considérons donc $\{d_k\}_{k \in N_1}$ et le fait que $o(\tilde{x}; \tilde{x} - x) \rightarrow 0$ pour $k \rightarrow \infty$ et $k \in N_1$. Alors :

$$d^t H(\tilde{x}) d \leq 0. \quad (1.5.7)$$

Ce qui contredit le fait que $H(\tilde{x})$ est définie positive car $\|d_k\| = 1$ (donc $d \neq 0$). Donc \tilde{x} est un minimum local strict.

1.5.2 Cas convexe

Théorème 1.5.3 . [21] Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que f est convexe et différentiable. Alors \tilde{x} est un minimum globale de f si seulement si

$$\nabla f(\tilde{x}) = 0. \quad (1.5.8)$$

Remarque 1.5.1 Dans le cas où la fonction f est convexe, il n'y a pas de différence entre un minimum local et un minimum global, c'est à dire que tout minimum local est également global. De plus, si f est strictement convexe, tous les minima locaux deviennent non seulement globaux, mais également uniques.

Recherches linéaires inexactes

Le but de ce chapitre consiste à décrire les principales règles de recherche linéaire inexacte (Armijo, Wolfe), ainsi que l'étude de la contribution de cette dernière à la convergence des algorithmes à directions de descente.

2.1 Recherche linéaire

2.1.1 Principe des recherches linéaires

Si la recherche linéaire nous intéresse, ce n'est pas seulement parce que dans les applications, on rencontre naturellement des problèmes unidimensionnels, mais plutôt parce que la recherche linéaire est une pièce fondamentale de toutes les méthodes traditionnelles d'optimisation multidimensionnelle.

Effectuer une recherche linéaire consiste à déterminer un pas α_k le long d'une direction de descente d_k , en d'autres termes résoudre le problème unidimensionnel suivant :

$$\min f(x_k + \alpha d_k), \quad \alpha \in \mathbb{R}_+ \quad (2.1.1)$$

Habituellement, nous avons le schéma suivant d'une méthode de minimisation multidimensionnelle sans contrainte :

En observant le comportement local de la fonction-coût f sur l'itération courante x_k , la méthode sélectionne "la direction du mouvement" d_k (qui est souvent une direction de descente de l'objectif f) :

$$\nabla^T f(x) d < 0.$$

et exécute un pas dans cette direction de la forme suivante :

$$x_k \rightarrow x_{k+1} = x_k + \alpha_k d_k. \quad (2.1.2)$$

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

$$\min_{\alpha > 0} f(x_k + \alpha d_k). \quad (2.1.3)$$

en vue de réaliser une certaine progression de la valeur de l'objectif f , c'est-à-dire de faire en sorte que

$$f(x_k + \alpha_k d_k) \leq f(x_k + \alpha d_k), \quad \forall \alpha > 0. \quad (2.1.4)$$

La plupart des méthodes considèrent que x_k, d_k sont des valeurs fixes et que le pas α_k est déterminé par le minimisation unidimensionnelle de la fonction $h_k : \mathbb{R}_+ \rightarrow \mathbb{R}$ tel que :

$$\alpha \rightarrow h_k(\alpha) = f(x_k + \alpha d_k). \quad (2.1.5)$$

Il convient de noter que dans les problèmes d'optimisation sans contraintes, nous devons résoudre un problème d'optimisation dans \mathbb{R} pour chaque itération x_k .

La méthode de recherche linéaire est ainsi une brique de base de toute méthode multidimensionnelle.

Il existe deux grandes catégories principales de méthodes qui concernent l'optimisation unidimensionnelle :

- ◇ Les recherches linéaires exactes.
- ◇ Les recherches linéaires inexactes.

2.1.2 Objectifs à atteindre

Il s'agit d'atteindre deux objectifs [16].

Le premier objectif

Consiste à diminuer f suffisamment. Cela se fait le plus souvent en réalisant une égalité sous la forme :

$$f(x_k + \alpha_k d_k) = f(x_k) + \text{"un terme négatif"}. \quad (2.1.6)$$

Le terme négatif, dit k , joue un rôle essentiel pour la convergence de l'algorithme qui utilise cette recherche linéaire. L'argument va comme suit.

Dans le cas où $f(x_k)$ est minorée il existe une constante C telle que

$$f(x_k) \geq C, \forall k$$

alors ce terme négatif doit nécessairement tendre vers zéro. C'est bien souvent à partir de la convergence vers zéro de cette suite que nous arrivons à montrer que le gradient lui-même doit tendre vers zéro. Le terme négatif doit donc prendre une forme tout à fait particulière si l'on veut parvenir à en tirer des informations.

Notamment, il ne suffit pas d'imposer

$$f(x_k + \alpha_k d_k) < f(x_k).$$

Le second objectif

Cela consiste à éviter que le pas $\alpha_k > 0$ soit trop petit, trop proche de zéro. En fait, le premier objectif est insuffisant puisque l'inégalité (2.1.6) est généralement satisfaite par des pas $\alpha_k > 0$ arbitrairement petit.

Mais cela peut conduire à une "fausse convergence", c'est-à-dire à la convergence des itérés vers un point non stationnaire.

2.2 Recherche linéaire inexacte d'Armijio

2.2.1 Principe de la règle d'Armijio

On vise à réduire d'une manière importante la valeur de l'objectif par un pas

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

de x_k dans la direction d_k , tel que

$$f(x_k + \alpha_k d_k) < f(x_k).$$

La règle d'Armijio [5] impose une contrainte sur le choix de α_k suffisante pour minimiser localement h .

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

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla^T f(x_k) d_k. \quad (2.2.1)$$

Elle est de la forme (2.2.1) car ρ devra être choisi dans $]0, 1[$.

Il faut qu'en α_k , la fonction h_k prenne une valeur plus petite que celle prise par la fonction

$$\psi_\rho(\alpha) : \alpha \rightarrow h_k(0) + \rho h'_k(0)\alpha,$$

autrement dit

$$\alpha \rightarrow f(x_k) + \rho \alpha_k \nabla^T f(x_k) d_k.$$

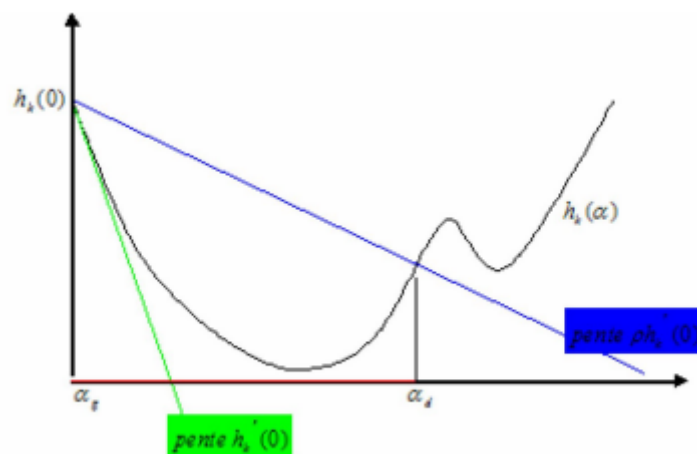


FIGURE 2.1 – Règle d'Armijio

On voit bien à la figure (2.1) ce que signifie cette condition.

2.3 Recherche linéaire inexacte de Wolfe

En 1969, WOLFE [25] a proposé une méthode basée sur le choix de deux paramètres.

2.3.1 Conditions de Wolfe faibles

Étant donnés deux réels ρ et σ tel que $0 < \rho < \sigma < 1$, ces conditions sont :

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha \nabla^T f(x_k) d_k. \quad (2.3.1)$$

$$\nabla^T f(x_k + \alpha d_k) d_k \geq \sigma \nabla^T f(x_k) d_k. \quad (2.3.2)$$

Autrement dit :

$$h_k(\alpha) \leq h_k(0) + \rho h'_k(0) \alpha. \quad (2.3.3)$$

$$h'_k(\alpha) \geq \sigma h'_k(0). \quad (2.3.4)$$

On voit bien à la figure (2.2) ce que signifie cette condition .

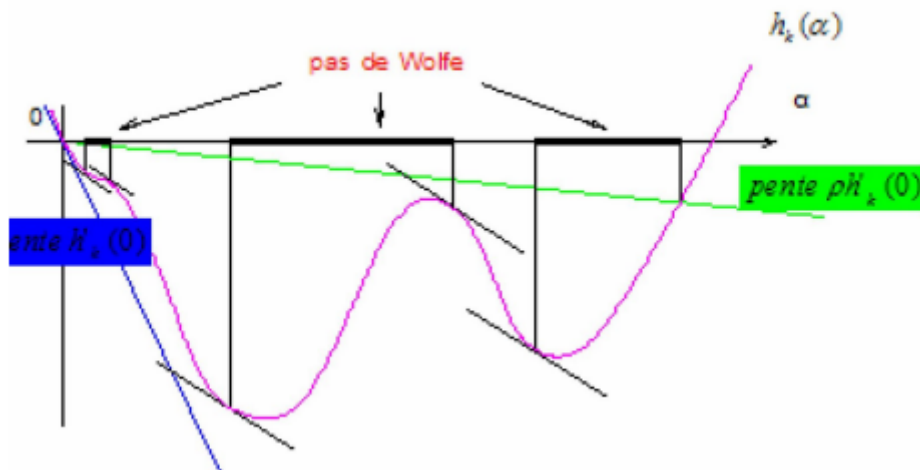


FIGURE 2.2 – Règle de Wolfe

2.3.2 Conditions de Wolfe fortes

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha \nabla^T f(x_k) d_k. \quad (2.3.5)$$

$$|\nabla^T f(x_k + \alpha d_k) d_k| \leq -\sigma \nabla^T f(x_k) d_k. \quad (2.3.6)$$

autrement dit :

$$h_k(\alpha) \leq h_k(0) + \rho h'_k(0) \alpha. \quad (2.3.7)$$

$$|h'_k(\alpha)| \leq -\sigma h'_k(0). \quad (2.3.8)$$

où

$$0 < \rho < \sigma < 1. \quad (2.3.9)$$

2.4 Condition de Zoutendijk

Dans cette partie , nous allons étudier la contribution de la recherche linéaire inexacte à la convergence des algorithmes de classe des méthodes à directions de descente.

Tout d'abord on va définir une condition importante sur les méthodes à direction de descente a fin d'assurer leurs convergence .

Définition 2.4.1 *On dit qu'une règle de recherche linéaire vérifie la condition de Zoutendijk s'il existe une constante $C > 0$ de telle sorte que pour chaque indice $K \geq 1$ on ait :*

$$f(x_{k+1}) \leq f(x_k) - C \|\nabla f(x_k)\|^2 \cos^2 \Theta_k. \quad (2.4.1)$$

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

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

Voici comment on se sert de la condition de Zoutendijk[Voir [26]].

Proposition 2.4.1.

Si la suite $\{x_k\}$ générée par un algorithme d'optimisation vérifie la condition de Zoutendijk (2.4.1) et si la suite $\{f(x_k)\}$ est minorée, alors

$$\sum_{k \geq 1} \|\nabla f(x_k)\|^2 \cos^2 \Theta_k < \infty. \quad (2.4.2)$$

Preuve 2.4.1 *En sommant les inégalités (2.4.1), 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 constant C' telle que pour tout k ,

$$f(x_k) \geq C'.$$

La proposition suivante précise les circonstances dans les quelles la condition de Zoutendijk (2.4.1) est vérifiée avec la règle de Wolfe.

Proposition 2.4.2.

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continument différentiable dans un voisinage de

$$\Lambda = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}.$$

on considère un algorithme à directions de descente d_k , qui génère une suite $\{x_k\}_{k \in \mathbb{N}}$ en utilisant la recherche linéaire de Wolfe (2.3.1)-(2.3.2).

Alors il existe une constante $C > 0$ telle que , pour tout $k \geq 1$ la condition de Zoutendijk (2.4.1) est vérifiée.

Preuve 2.4.2 D'après (2.3.2)

$$\begin{aligned} \nabla^T f(x_k + \alpha_k d_k) d_k &\geq \sigma \nabla^T f(x_k) d_k \\ (\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k))^T d_k &\geq (\sigma - 1) \nabla^T f(x_k) d_k \\ &= -(1 - \sigma) \nabla^T f(x_k) d_k = (1 - \sigma) \left| \nabla^T f(x_k) d_k \right| \\ (1 - \sigma) \left| \nabla^T f(x_k) d_k \right| &\leq (\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k))^T d_k \end{aligned}$$

Et du fait que f est continument différentiable :

$$\begin{aligned} (1 - \sigma) \left| \nabla^T f(x_k) d_k \right| &= (1 - \sigma) \left\| \nabla^T f(x_k) \right\| \|d_k\| \cos \Theta_k \\ &\leq \left\| (\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k)) \right\| \|d_k\| \\ (1 - \sigma) \left\| \nabla^T f(x_k) \right\| \cos \Theta_k &\leq L \alpha_k \|d_k\| \\ \alpha_k \|d_k\| &\leq \frac{(1 - \sigma)}{L} \left\| \nabla^T f(x_k) \right\| \cos \Theta_k \end{aligned}$$

En utilisant (2.3.1), on aura :

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \rho \alpha_k \nabla f(x_k) d_k \\ &\leq f(x_k) + \rho \alpha \nabla f(x_k) d_k \\ &\leq f(x_k) + \left| \rho \alpha \nabla f(x_k) d_k \right| \\ &\leq f(x_k) - \rho \alpha \nabla f(x_k) d_k \\ &\leq f(x_k) - \rho \alpha \left\| \nabla f(x_k) \right\| \|d_k\| \cos \Theta_k \\ &\leq f(x_k) - \frac{\rho(1 - \sigma)}{L} \left\| \nabla f(x_k) \right\|^2 \cos^2 \Theta_k \end{aligned}$$

On en déduit (2.4.1) □

Méthode du gradient conjugué

3.1 Méthode du gradient

Il s'agit d'une méthode fondamentale proposée par Cauchy [7] en 1847 afin de minimiser des problèmes d'optimisation sans contrainte. En effet, cette méthode prend comme direction : $d_k = -g_k$. Dans ce cas donc on peut dire que la méthode du gradient est une méthode à directions de descente.

Au point courant x_k , la direction du gradient négatif représente la meilleure direction pour la recherche d'un minimum de l'objectif f , c'est pour ce là que l'algorithme du gradient est également connu sous le nom d'algorithme de la plus profonde descente ou la plus forte pente (steepest descente).

3.1.1 Algorithme de la méthode du gradient

Algorithme 3 : Méthode du gradient

Etape 0(initialisation)

Soit x_0 le point de départ .

Poser $d_0 = -\nabla f(x_0)$, Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

"Test d'arrêt" : Si $\|\nabla f(x_k)\| \simeq 0$: STOP ($\tilde{x} = x_k$)

Si non aller a l'étape 2.

Etape 2 :

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

$$d_k = -\nabla f(x_k)$$

α_k est déterminé par une recherche linéaire .

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

3.2 La méthode du gradient conjugué linéaire

3.2.1 Principe général

La méthode du gradient conjugué est une méthode très puissante et très utilisée pour la résolution des problèmes d'optimisation sans contraintes surtout de grand taille .

Elle se base sur le concept des directions conjuguées, car les gradients successifs sont à la fois orthogonaux entre eux et aux directions précédentes.

L'idée principale de cette méthode est de construire une suite de direction de descente que sont mutuellement conjuguées par rapport à la fonction quadratique afin de résoudre le problème d'optimisation sans contraintes (P) suivant :

$$(P) : \{ \min f(x) : x \in \mathbb{R}^n \}. \quad (3.2.1)$$

Où f est une fonction quadratique strictement convexe de n variables .
ça veut dire que la minimisation de $f(x)$ revient à la minimisation de $q(x)$ où :

$$q(x) = \frac{1}{2} x^T A x - b^T x + c, \text{ avec } x \in \mathbb{R}^n \quad (3.2.2)$$

- $A \in M_{n \times n}$ est une matrice symétrique et définie positive.
- $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ un vecteur de \mathbb{R}^n .
- $c \in \mathbb{R}$ est une constante .

La méthode du gradient conjugué linéaire est une méthode itérative, elle consiste à générer une suite $\{x_k\}$ de la manière suivante :

✓ On commence par un vecteur x_0 donné comme point de départ.

✓ A chaque étape k si on a $x_k \in \mathbb{R}^n$ alors le successeur x_{k+1} de x_k est déterminé par la relation suivante :

$$x_{k+1} = x_k + \alpha_k d_k. \quad (3.2.3)$$

Où $\{d_0, d_1, \dots, d_n\}$ constitue un ensemble de directions de descente ayant la propriété d'être mutuellement conjuguées par rapport à la fonction quadratique $q(x)$.

Les directions de descente d_k cherchées sont définies avec les formule de récurrence qui suit :

$$d_k = \begin{cases} -g_0 = \nabla q(x_0) & \text{Si } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{Si } k \geq 1 \end{cases} \quad (3.2.4)$$

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

3.2.2 Détermination du pas α_k

En effet, une méthode pour déterminer le pas α_k est de résoudre le problème d'optimisation unidimensionnel suivant :

$$\alpha_k = \min f(x_k + \alpha d_k), \alpha > 0. \quad (3.2.5)$$

On note

$$g_k = \nabla f(x_k) = \nabla q(x_k) = A x_k - b. \quad (3.2.6)$$

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

$$\frac{dq}{d\alpha}(\alpha_k) = q'(\alpha_k) = d_k^T \nabla q(x_{k+1}) = 0, \quad (3.2.7)$$

$$d_k^T \nabla q(x_{k+1}) = d_k^T (Ax_{k+1} - b) = 0, \quad (3.2.8)$$

D'où :

$$d_k^T A(x_k + \alpha_k d_k) - d_k^T b = 0, \quad (3.2.9)$$

Alors

$$d_k^T A \alpha_k d_k = -d_k^T A x_k + d_k^T b, \quad (3.2.10)$$

$$d_k^T A \alpha_k d_k = -d_k^T (A x_k + b), \quad (3.2.11)$$

D'où l'on tire

$$\alpha_k = \frac{-d_k^T (A x_k + b)}{d_k^T A d_k}, \quad (3.2.12)$$

Donc

$$\alpha_k = \frac{-d_k^T g_k}{d_k^T A d_k}. \quad (3.2.13)$$

Le pas α_k obtenu ainsi s'appelle le pas optimal.

3.2.3 Détermination du coefficient β_k

Les coefficients β_k étant choisis de telle manière que d_k soit conjuguée avec toutes les directions précédentes. Autrement dit il faut avoir :

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

on en déduit :

$$d_{k+1}^T A d_k = 0 \Rightarrow (-\nabla q(x_{k+1}) + \beta_{k+1} d_k)^T A d_k = 0,$$

$$-\nabla^T q(x_{k+1}) A d_k + \beta_{k+1} d_k^T A d_k = 0,$$

$$\beta_{k+1} = \frac{\nabla^T q(x_{k+1}) A d_k}{d_k^T A d_k},$$

$$\beta_{k+1} = \frac{g_{k+1}^T A d_k}{d_k^T A d_k}.$$

3.2.4 Algorithme du gradient conjugué linéaire

Nous allons utiliser l'algorithme suivant pour résumer tout ce qui est dit ci-dessus :

Algorithme 4 : Méthode du gradient conjugué linéaire**Etape 0(initialisation)**

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

Poser $d_0 = -g_0$.

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

Si $\|g_k\| = 0$: STOP ($\tilde{x} = x_k$). "Test d'arrêt"

Si non aller à l'étape 2.

Etape 2 :

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

$$\alpha_k = \frac{-d_k^T g_k}{d_k^T A d_k}. \quad (3.2.14)$$

$$d_{k+1} = -g_{k+1} + \beta_{k+1} d_k. \quad (3.2.15)$$

$$\beta_{k+1} = \frac{g_{k+1}^T A d_k}{d_k^T A d_k}. \quad (3.2.16)$$

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

3.2.5 Les avantages de la méthode du gradient conjugué linéaire

1- L'algorithme du gradient conjugué linéaire est très utile dans le but de résoudre des grands systèmes creux, en effet il suffit de savoir comment appliquer la matrice A à un vecteur.

2-La convergence peut être très rapide si A admet seulement r ($r < n$) valeurs propres distinctes la convergence a lieu en au plus r itérations.

3.3 Exemple

Exercice 3.3.1 . Dans cet exemple on va minimiser la fonction quadratique suivante en utilisant la méthode du gradient conjuguée linéaire :

Soit

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto f(x, y) = x^2 + 4y^2 - 4x - 8y.$$

On peut l'écrire sous la forme :

$$f(x) = \frac{1}{2} x^T \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} x - \begin{pmatrix} 4 \\ 8 \end{pmatrix} x, \text{ avec } x_0 = (0, 0)^T$$

Remarquons tout d'abord que $A = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}$ est définie positive car :

$$\Delta_1 = 2 > 0 \text{ et } \Delta_2 = \begin{vmatrix} 2 & 0 \\ 0 & 8 \end{vmatrix} = 16 > 0 \Rightarrow f \text{ est une fonction quadratique strictement convexe.}$$

De plus on a :

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

et on prend comme direction initiale $d_0 = -g_0$, $\varepsilon = 10^{-6}$

$$g_0 = \nabla f(x_0) = Ax_0 - b = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 4 \\ 8 \end{pmatrix} = -\begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

$$g_0 = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

$$d_0 = -g_0 = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

on calcule

$$\|g_0\| = \sqrt{80} > 10^{-6} = \varepsilon.$$

Avant toute chose, rappelons que les itérations de la méthode du gradient conjugué sont définies comme suit : $x_{k+1} = x_k + \alpha_k d_k$, avec :

$$\alpha_k = \frac{-d_k^T g_k}{d_k^T A d_k},$$

Donc, on obtient :

$$\alpha_0 = \frac{-d_0^T g_0}{d_0^T A d_0} = \frac{5}{34}$$

$k = 1$:

$$x_1 = x_0 + \alpha_0 d_0$$

$$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{5}{34} \begin{pmatrix} 4 \\ 8 \end{pmatrix} = \begin{pmatrix} \frac{20}{34} \\ \frac{40}{34} \end{pmatrix}$$

$$g_1 = \nabla f(x_1) = Ax_1 - b$$

$$g_1 = \begin{pmatrix} -\frac{48}{17} \\ \frac{24}{17} \end{pmatrix}$$

$$\|g_1\| = \sqrt{9.9653979239} = 3.1568018506 > 10^{-6} = \varepsilon$$

Comme

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

$$\beta_{k+1} = \frac{g_{k+1}^T A d_k}{d_k^T A d_k} \Rightarrow \beta_1 = \frac{g_1^T A d_0}{d_0^T A d_0} = 0,124567474$$

$$d_1 = -g_1 + \beta_1 d_0 = \begin{pmatrix} 3,3218 \\ -0,4152 \end{pmatrix}$$

$$\alpha_1 = \frac{-d_1^T g_1}{d_1^T A d_1} = 0,42500$$

$k = 2$:

$$x_2 = x_1 + \alpha_1 d_1$$

$$x_2 = \begin{pmatrix} \frac{20}{34} \\ \frac{40}{34} \end{pmatrix} + 0,43 \begin{pmatrix} 3,3218 \\ -0,4152 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$g_2 = \nabla f(x_2) = Ax_2 - b$$

$$g_2 = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 4 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

$$\|g_2\| = 0 < 10^{-6} = \varepsilon$$

En conclusion, la méthode converge en 2 itérations, c'est à dire que le min de f est atteint en $\tilde{x} = x_2$ et on a :

$$\tilde{x} = x_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$f(\tilde{x}) = \frac{1}{2} (2 \ 1) \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 4 \\ 8 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

En effet, $f(\tilde{x}) = -8$ est le minimum globale de la fonction f avec la précision donnée.

3.4 La méthode du gradient conjugué non linéaire

3.4.1 Principe général

La méthode du gradient conjugué linéaire a été proposée afin de minimiser des fonctions quadratique strictement convexes. La question qui se pose maintenant est peut-on utiliser cette méthode pour minimiser des fonctions non convexes ou même des fonction non linéaires ?

Afin de minimiser une fonction non linéaire f , la méthodes du gradient conjugué doit être adaptée selon les étapes ci-dessous :

→ D'abord, on effectue une recherche linéaire pour déterminer le pas α_k qui minimise la fonction non linéaire f le long de d_k .

→ Ensuite, le résidu ∇_q devra être remplacé par le gradient de la fonction f .

On obtient alors un algorithme très efficace pour l'optimisation non-linéaire.

Pour garantir que la direction d_{k+1} est une direction descente, le paramètre α_k devra vérifier certaines conditions. En effet, si la recherche linéaire est exacte, on voit que :

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

Par contre, si la recherche linéaire n'est pas exacte (puisque la recherche exacte est trop coûteuse), d'autres conditions doivent être imposées.

L'idée de cette méthode est de construire des directions d_0, \dots, d_k mutuellement conjuguées d'une manière itérative.

La direction d_k étant obtenue à chaque étape k comme une combinaison linéaire du gradient au point x_k et de la direction précédente d_{k-1} , les coefficients seront choisis de telle manière que d_k soit conjugué avec toutes les directions précédentes.

Nous nous intéressons ici à la minimisation d'une fonction f de n variables à valeurs réelles

$$\min f(x) : x \in \mathbb{R}^n. \quad (3.4.1)$$

pas nécessairement quadratique.

Comme indiqué précédemment, la direction d_k est définie par la formule de récurrence suivante :

$$d_k = \begin{cases} -g_0 & \text{si } k = 0, \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 1 \end{cases} \quad (3.4.2)$$

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

$$x_{k+1} = x_k + \alpha_k d_k. \quad (3.4.3)$$

et le pas α_k étant calculé par une recherche linéaire.

3.4.2 Algorithme du gradient conjugué non linéaire avec la recherche linéaire inexacte de Wolfe forte

Nous allons présenter ci-dessous le schéma général de l'algorithme du gradient conjugué non linéaire pour des fonctions quelconques.

Algorithme 5 : La méthode du gradient conjugué non-linéaire

Etape 0 : (initialisation)

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

Etape 1 :

Si $\|g_k\| = 0$: STOP ($\tilde{x} = x_k$). "Test d'arrêt" Si non aller à l'étape 2.

Etape 2 :

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

α_k : déterminée par la recherche linéaire inexacte de Wolfe forte.

$$d_{k+1} = -g_{k+1} + \beta_{k+1} d_k, \quad (3.4.4)$$

où

β_k : défini selon la méthode.

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

3.4.3 Convergence des méthodes du gradient conjugué non linéaires

Le premier résultat de convergence de la méthode du gradient conjugué non linéaire avec des recherches linéaires inexacte était démontré par Al-Baali [1].

Le résultat qui suit garantit la convergence de toutes les variantes de la méthode du gradient conjugué non linéaire avec la recherche linéaire inexacte de Wolfe pour des fonctions quelconque.

Théorème 3.4.1 [4] Supposons que l'ensemble $\Lambda = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ est borné où $x_1 \in \mathbb{R}^n$ est le point initial et que sur un voisinage V et Λ , la fonction objectif f est continûment différentiable et son gradient est lipschitzien. Considérons une méthode du gradient conjugué non linéaire du type (3.4.2) et (3.4.3) où α_k est déterminée avec la recherche linéaire inexacte de Wolfe forte (2.3.5) et (2.3.6). Alors

$$\lim_{k \rightarrow \infty} \|g_k\| = 0, \quad (3.4.5)$$

où

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \quad (3.4.6)$$

Preuve 3.4.1 De (3.4.4) pour tout $k \geq 1$,

$$d_{k+1} + g_{k+1} = \beta_{k+1} d_k, \quad (3.4.7)$$

en appliquant le carré sur les deux côtés de (3.4.7) résulte :

$$\|d_{k+1}\|^2 = -\|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1} + \beta_{k+1}^2 \|d_k\|^2. \quad (3.4.8)$$

Puisque d_k satisfait la condition de descente $g_k^T d_k < 0$ d'après (3.4.8), il en déduit que

$$\|d_{k+1}\|^2 \geq -\|g_{k+1}\|^2 + \beta_{k+1}^2 \|d_k\|^2, \quad (3.4.9)$$

De (3.4.4)

$$g_{k+1}^T d_{k+1} - \beta_{k+1} g_{k+1}^T d_k = -\|g_{k+1}\|^2,$$

lequel, avec la condition de recherche linéaire de Wolfe forte (2.3.5) montre que

$$|g_{k+1}^T d_{k+1}| + \sigma |\beta_{k+1} g_{k+1}^T d_k| = -\|g_{k+1}\|^2. \quad (3.4.10)$$

En appliquant l'inégalité $(a + \sigma b)^2 \leq (1 + \sigma^2)(a^2 + b^2)$ pour tout $a, b, \sigma \geq 0$ avec,

$a = |g_{k+1}^T d_{k+1}|$ et $b = |\beta_{k+1} g_{k+1}^T d_k|$, de (3.4.10) on peut écrire

$$(g_{k+1}^T d_{k+1})^2 + \beta_{k+1}^2 (g_{k+1}^T d_k)^2 \geq c \|g_{k+1}\|^4. \quad (3.4.11)$$

où $c = (1 + \sigma^2)^{-1}$ Par conséquent, de (3.4.9) et (3.4.11), il résulte que

$$\begin{aligned} &= \frac{1}{\|d_{k+1}\|^2} [(g_{k+1}^T d_{k+1})^2 + \frac{\|d_{k+1}\|^2}{\|d_k\|^2} (g_{k+1}^T d_k)^2] \\ \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_{k+1}^T d_k)^2}{\|d_k\|^2} &\geq \frac{1}{\|d_{k+1}\|^2} [(g_{k+1}^T d_{k+1})^2 + \beta_{k+1}^2 (g_{k+1}^T d_k)^2 - \frac{(g_{k+1}^T d_k)^2}{\|d_k\|^2} \|g_{k+1}\|^2] \\ &\geq \frac{1}{\|d_{k+1}\|^2} [c \|g_{k+1}\|^4 - \frac{\|d_{k+1}\|^2}{\|d_k\|^2} \|g_{k+1}\|^2] \end{aligned} \quad (3.4.12)$$

Or, si (3.4.5) n'est pas vrai, alors (3.4.12) et la condition de Zoutendijk (2.4.2) impliquent que l'inégalité suivante

$$\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_{k+1}^T d_k)^2}{\|d_k\|^2} \leq \frac{c \|g_{k+1}\|^4}{2 \|d_k\|^2}. \quad (3.4.13)$$

est vérifiée pour k suffisamment grand. Avec cela, l'inégalité (3.4.6) suit (3.4.13) et (3.4.12).

La convergence de quelques méthodes du gradient conjugué non linéaire

4.1 Introduction

La méthode de gradient conjugué (*GC*) a joué un rôle particulier pour la résolution à grande problèmes d'optimisation non linéaire en raison de la simplicité de leur itération et leurs besoins en mémoire très faibles. En fait, la méthode de la *GC* est pas parmi les algorithmes d'optimisation les plus rapides ou plus robuste pour les problèmes non linéaire disponibles aujourd'hui, mais il reste très populaire pour les ingénieurs et les mathématiciens qui sont intéressés à résoudre les grands problèmes.

La méthodes non linéaire de gradient conjugué est conçu pour résoudre les problèmes d'optimisation sans contrainte suivant :

$$\min f(x) : x \in \mathbb{R}^n. \quad (4.1.1)$$

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

$$x_{k+1} = x_k + \alpha_k d_k. \quad (4.1.2)$$

où α_k est une longueur de pas obtenu par une recherche linéaire unidimensionnelle, et d_k est le sens de la recherche défini par :

$$d_k = \begin{cases} -g_1 & \text{si } k = 1 \\ -g_k + \beta_k d_{k-1} \dots & \text{si } k \geq 2 \end{cases} \quad (4.1.3)$$

où $g_k = \nabla f(x_k)$; et β_k est un scalaire. Il ya au moins six formules pour β_k , qui sont donnés ci-dessous

$$\beta_K^{HS} = -\frac{g_K^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \quad \text{Hestenes-Stiefel [22], 1952} \quad (4.1.4)$$

$$\beta_K^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \quad \text{FletcherReeves [10], 1964} \quad (4.1.5)$$

$$\beta_K^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \quad \text{Polak- Ribière-Polyak ([19], [12]), 1969} \quad (4.1.6)$$

$$\beta_K^{CD} = -\frac{\mathbf{g}_k^T \mathbf{g}_k}{d_{K-1}^T \mathbf{g}_{k-1}}, \quad \text{descente conjugué [9], 1987} \quad (4.1.7)$$

$$\beta_K^{LS} = -\frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})}{d_{k-1}^T \mathbf{g}_{k-1}}, \quad \text{Liu - Storey [23], 1991} \quad (4.1.8)$$

$$\beta_K^{DY} = \frac{\mathbf{g}_k^T \mathbf{g}_k}{(\mathbf{g}_k - \mathbf{g}_{k-1})^T d_{k-1}}, \quad \text{Dai - yuan [8], 1999} \quad (4.1.9)$$

$$\beta_K^{wyl} = \frac{\mathbf{g}_k^T (\mathbf{g}_k - \frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{k-1}\|} \mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T \|\mathbf{g}_{k-1}\|}, \quad [24 \frown] \quad (4.1.10)$$

Dans cette section on va présenter trois différentes variantes de la classe des méthodes du gradient conjugué non linéaire méthode de WYL [24], méthodes FR [10] et la méthode PRP ([19], [12]). Après avoir présenté ces variantes on va citer leurs algorithmes, prouver le théorème qui assure leurs convergence.

4.2 Partie théorique

4.2.1 Méthode de WYL

Dans cette section, nous allons Montrer que la formule WYL avec la recherche linéaire de Wolfe - Powell forte posséder la condition suffisante de descente. Les α_k vérifient les relations de wolfe forte suivantes :

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \mathbf{g}_K^T d_k, \quad (4.2.1)$$

$$|\mathbf{g}(x_k + \alpha_k d_k)^T d_k| \leq \sigma |\mathbf{g}_K^T d_k|. \quad (4.2.2)$$

Où

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

4.2.2 La preuve de la condition suffisantes de descente

Le théorème suivant montre que la formule WYL avec la recherche linéaire de SWP possèdent la condition suffisantes de descente.

Théorème 4.2.1 *Supposer que les séquences \mathbf{g}_k et d_k sont générés par la méthode de la forme (4.1.2), (4.1.3) et (4.1.10) et le pas α_k est déterminée par la recherche linéaire de Wolf - Powell fort (4.2.1) et (4.2.2), si $\sigma < \frac{1}{4}$ donc la séquence d_k possèdent la condition suffisante de descente*

$$\mathbf{g}_K^T d_k \leq -c \|\mathbf{g}_k\|^2. \quad (4.2.3)$$

Preuve 4.2.1 *Nous avons d'abord prouver la propriété de descente de d_k : Par (4.1.3)*

Nous avons

$$\mathbf{g}_K^T d_k = -\|\mathbf{g}_k\|^2 + \beta_K^{wyl} \mathbf{g}_k^T d_{k-1}. \quad (4.2.4)$$

combinant cette équation avec (β_K^{wyl}) , nous avons

$$\frac{\mathbf{g}_k^T d_k}{\|\mathbf{g}_k\|^2} = -1 + \frac{\sigma \mathbf{g}_{k-1}^T d_{k-1}}{\|\mathbf{g}_{k-1}\|^2} \left(1 - \frac{\mathbf{g}_k^T \mathbf{g}_{k-1}}{\|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|} \right) \leq \frac{\mathbf{g}_k^T d_k}{\|\mathbf{g}_k\|^2} \leq -1 + \frac{-\sigma \mathbf{g}_{k-1}^T d_{k-1}}{\|\mathbf{g}_{k-1}\|^2} \left(1 - \frac{\mathbf{g}_k^T \mathbf{g}_{k-1}}{\|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|} \right). \quad (4.2.5)$$

puisque

$$\mathbf{g}_1^T \mathbf{d}_1 = -\|\mathbf{g}_1\|^2 < 0.$$

si $\mathbf{g}_1 \neq 0$, Supposons maintenant que $\mathbf{d}_i, i = 1, 2, 3, \dots, k-1$

Sont toutes les direction de descente : $\mathbf{g}_k^T \mathbf{d}_k < 0$.

De l'inégalité de Cauchy - Schwarz, nous obtenons

$$0 \leq 1 - \frac{\mathbf{g}_k^T \mathbf{g}_{k-1}}{\|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|} \leq 2, \quad (4.2.6)$$

(4.2.5) et (4.2.6) en déduire

$$-1 + 2\sigma \frac{\mathbf{g}_k^T \mathbf{g}_{k-1}}{\|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|} \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq -1 - 2\sigma \frac{\mathbf{g}_k^T \mathbf{g}_{k-1}}{\|\mathbf{g}_k\| \|\mathbf{g}_{k-1}\|}, \quad (4.2.7)$$

En répétant de ce processus et le fait $\mathbf{g}_1^T \mathbf{d}_1 = -\|\mathbf{g}_1\|^2$, nous avons :

$$-\sum_{j=0}^{k-1} (2\sigma)^j \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq -2 + \sum_{j=0}^{k-1} (2\sigma)^j, \quad (4.2.8)$$

puisque

$$\sum_{j=0}^{k-1} (2\sigma)^j < \sum_{j=0}^{k-1} (2\sigma)^j = \frac{1}{1-2\sigma},$$

(4.2.8) peut être écrit comme

$$-\frac{1}{1-2\sigma} \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq -2 + \frac{1}{1-2\sigma}. \quad (4.2.9)$$

En faisant la restriction $\sigma \in (0, \frac{1}{4})$ nous avons $\mathbf{g}_k^T \mathbf{d}_k < 0$, donc par induction

$\forall k \in \mathbb{N}, \mathbf{g}_k^T \mathbf{d}_k < 0$ est vérifiée.

Maintenant, nous prouvons la propriété suffisante descente de \mathbf{d}_k si

$$\sigma \in (0, \frac{1}{4}) \text{ et } c = 2 - \frac{1}{1-2\sigma} \text{ puis } 0 < c < 1.$$

et (4.2.9) s'avère être

$$c - 2 \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq -c. \quad (4.2.10)$$

$\Rightarrow \mathbf{g}_k^T \mathbf{d}_k \leq -c \|\mathbf{g}_k\|^2$ où $C = -2 + \frac{1}{1-2\sigma}$ ce qui signifie que (4.2.3) est vérifiée.

4.2.3 Propriétés convergentes

Dans ce paragraphe, nous allons réintroduire les propriétés de convergence de la formule (4.1.10), (4.2.1) et (4.2.2).

Les résultats réintroduits sont donnés dans ([27]), de sorte que les preuves de détail sont négligés.

Pour plus commodité, nous avons d'abord introduire certaines hypothèses et de lemmes qui sont souvent utilisés dans les cours de méthodes de gradient conjugué...

Maintenant nous donnons l'algorithme suivant d'abord .

Étape 1 : Point initiale

$x_1 \in \mathbb{R}^n$, $\epsilon \geq 0$, fixé $d_1 = -g_1$, si $\|g_1\| \leq \epsilon$ alors stop .

Étape 2 :

Calculer α_k par quelques recherches linéaire $\delta = 0,01$ et $\sigma = 0,1$.

Étape 3 :

définir $x_{k+1} = x_k + \alpha_k d_k$,Si $\|g_{k+1}\| \leq \epsilon$ alors Stop $\epsilon = 10^{-6}$.

Étape 4 :

Calculer β_k par quelques formules et générer d_{k+1} par (4.1.3).

Étape 5 :

poser $k = k + 1$.passez à l'étape 2.

Sur les études de méthodes de gradient conjugué la condition suffisante de descente (4.2.3) joue un rôle important. Malheureusement ,cette condition ne peut être satisfaite pour de nombreuses. méthode de gradient conjugué .

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

Hypothèses B : Dans certains voisinage N de Ω , $f(x)$ est continûment différentiable et son gradient est Lipschitz continue,c'est-à-dire,pour toutes $x, y \in \mathbb{N}$,il existe une constante $L \geq 0$ telle que :

$$\|g(x) - g(y)\| \leq L \|x - y\|. \quad (4.2.11)$$

Lemme 1 Supposons que les hypothèses A et B ont vérifiées .considérer la méthode sous la forme d'(1.1.2) et (1.1.3) ,où d_k satisfait $g_k^T d_k < 0$,pour tout k et α_k est obtenu par SWP (4.2.1) et(4.2.2) alors,

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \quad (4.2.12)$$

Preuve 4.2.2 Nous avons $g_k^T d_k < 0$ pour tout $K \geq 1$ Nous avons aussi de (4.2.2) et (4.2.11)

$$-(1 - \sigma)g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq L\alpha_k \|d_k\|^2. \quad (4.2.13)$$

$$\alpha_k \geq \frac{-(1 - \sigma)g_k^T d_k}{L \|d_k\|^2}. \quad (4.2.14)$$

laquelle associant (4.2.1),nous obtenons

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \geq \delta \frac{(1 - \sigma)(g_k^T d_k)^2}{L \|d_k\|^2}. \quad (4.2.15)$$

En outre,à partir de Hypothèse (A) nous avons $\{f(x_k)\}$ est une suite décroissante et a une limite ci-dessous dans Ω ,et montre

$$\lim_{k \rightarrow \infty} f(x_{k+1}) < +\infty.$$

et (4.2.16) présente

$$+\infty > f(x_1) - \lim_{k \rightarrow \infty} f(x_{k+1}) = \sum |f(x_k) - f(x_{k+1})| \geq \delta \frac{(1-\sigma)(g_k^T d_k)^2}{L \|d_k\|^2}. \quad (4.2.16)$$

Nous pouvons conclure que (4.2.12) est vérifiée.

alors

$$\sum \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty.$$

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

4.2.4 Méthode de Fletcher-Reeves

La méthode de Fletcher-Reeves ([10],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.

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

Gradient conjugué variante Fletcher-Reeves.

Algorithme de la méthode de Fletcher-Reeves

Étape 0 : (initialisation)

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

Poser $k = 0$ et aller à l'étape 1.

Étape 1 :

Si $g_0 = 0$:STOP ($x^* = x_k$). "Test d'arrêt".

Si non aller à l'étape 2.

Étape 2 :

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

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

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

4.2.5 La propriété de descente de la méthode de FR

Powell ([20],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 ([1],1985) a démontré la satisfaction de la propriété de descente de la fonction objective pour la méthode de Fletcher-Reeves avec la recherche linéaire inexacte de Wolfe forte.

Convergence de la méthode de Fletcher-Reeves

Le premier résultat de convergence de la méthode du gradient conjugué non linéaire (version Fletcher-Reeves) avec recherches linéaires inexactes (recherche linéaire inexacte de Wolfe forte (4.2.1)-(4.2.2). où $\sigma < \frac{1}{2}$) était démontré par Al-Baali ([1],1985).

Théorème 4.2.2 ([14],1992]) *Supposons que l'hypothèse 4.1 soit satisfaite. Considérons un méthode du type (4.1.2) et (4.1.3) dont β_k (Fletcher-Reeves) et le pas α_k satisfait aux conditions de Wolfe forte (4.2.1)-(4.2.2) où $\sigma \in]0, \frac{1}{2}[$: Alors cette méthode est globalement convergente, dans le sens suivant :*

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (4.2.17)$$

Méthode de Polak-Ribière-Polyak

Cette méthode fut découverte par Polak,Ribière ([12],1969) et Polyak ([19],1969) .Rappelons que pour la méthode de Polak-Ribière-Polyak la variante β_k est :

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}. \quad (4.2.18)$$

Algorithme de la méthode de Polak-Ribière-Polyak

Étape 0 : (initialisation).

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

Poser $k = 0$ et aller à l'étape 1.

Étape 1 :

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

Si non aller à l'étape 2.

Étape 2 :

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

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

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

où

$$\beta_{k+1}^{PRP} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|^2} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}.$$

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

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 4.2.3 *Si f est fortement convexe ,continument différentiable avec un gradient lipschitzien, alors la méthode de Polak Ribière avec recherche linéaire exacte génère une suite $\{x_k\}$ converge vers l'unique point x^* réalisant le minimum de f .*

Preuve 4.2.3 ([8]) *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

$$d_{k-1}^T y_{k-1} = d_{k-1}^T (g_k - g_{k-1})$$

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

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 :

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

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

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

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^* .

5.1 Résultats numérique et discussions

Dans cette section nous présentons les résultats numériques de la Méthode *WYL* et on comparation avec les méthodes *GC* de *FR,PRP* pour analyse l'efficacité de ces méthode nous avons considère $\varepsilon = 10^{-6}$, $\sigma = 0.1$, $\delta = 0.01$ et la valeur du gradient comme critère d'arrêt. La tolérance $\varepsilon = 10^{-6}$ est choisie pour tous les algorithmes pour étudier la Rapidité d'itération de ces algorithmes vers la solution optimale.

Nous utilisons l'algorithme suivant :

Etape 0 : (initialisation)

Soit x_0 le point de départ, $g_0 = \nabla f(x_0)$.

Poser $d_0 = -g_0$.

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

Si $g_k = 0$: STOP ($x^* = x_k$). "Test d'arrêt" $\|g_k\| < 10^{-6}$.

Si non aller a l'étape 2.

Etape 2 :

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

α_k : calculer par la recherche linéaire (Wolf-fort (4.1.2)(4.1.3), avec $\sigma = 0.1$, $\delta = 0.01$)

$d_k = -g_k + \beta_k d_{k-1}$, β_k^{WYL} , β_k^{FR} , β_k^{PRP} .

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

Les résultats numérique ont été obtenus sur les fonctions suivantes :

Problème 1

$$f(x, y) = x^2 + 4y^2 - 4x - 8y \text{ avec } x_0 = (0, 0), x^* = (2, 1), f(x^*) = 0.$$

Problème 2

$$f(x, y) = (x - 2)^4 + (x - 2y)^2 \text{ avec } x_0 = (0, 0), x^* = (2, 1), f(x^*) = 0.$$

la fonction

$$f(x, y) = (x + 2y - 7)^2 + (2x + y - 5)^2 \text{ (Bouth function) avec } x^* = (1, 3), f(x^*) = 0.$$

application des méthodes de gradients sur fonction quadratique, Nous testons

La condition d'arrêt est $\|g_k\| < 10^{-6}$.

Dans cette partie ,on adopter les notations suivantes :

$CPU\ time$: le temps.

NI nombre d'itérations.

$\|g\|$ la norme du gradient.

5.1.1

Problème 1 : Considérons le problème quadratique suivante :

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

La solution optimal est (2 1).

La méthode du gradient à pas fixe ,à pas optimal,puis le gradient conjuguée pour comparer ces méthode on choisit de démarrer les algorithmes que désencadrer au point de coordonnée $x_0 = (0, 0)$.

Méthode du gradient à pas fixe

Pour tester cette méthode, on pourra choisir par exemple un pas initial $\alpha = 0.001, \alpha = 0.8, \alpha = 0.20$.

Tableau 1 : Résultats numérique de la méthode du gradient à pas fixe

α	NI	CPU time	la solution	$\ g\ $
0.001	999	0.159	(1.72, 0.99)	0.5413
0.8	467	0.123	échoué	échoué
0.20	34	0.003	(1.99, 1)	9×10^{-7}

La convergence n'est pas assuré dans le cas 2 ou $\alpha = 0.8$ (plus grand).

Dans le troisième cas la convergence vers la solution optimale (2 1) est mieux .

Méthode du gradient à pas optimale

$$\alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}. (exacte)$$

Tableau 2 : Résultats numérique de la méthode du gradient à pas optimale

alpha	NI	CPU time	la solution	$\ g\ $
0.14705	25	0.0087	(1.99,1)	3.7×10^{-7}

L'algorithme converge vers le $x^* = (1.99, 1)$ qui est très proche du point optimale détecte théoriquement $x^* = (2, 1)$.

Méthode du gradient conjuguée à pas optimale

$$\alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}. (exacte)$$

Tableau 3 : Comparaison la performance de la méthode de WYL avec d'autre

Méthode	β_k	le pas α_k	NI	CPU time	la solution
FR	0.1245	0.147	2	0.00147	(2 1)
PRP	0.1245	0.147	2	0.00120	(2 1)
WYL	0.1245	0.147	2	0.001	(2 1)

Remarque 5.1.1 Dans le cas quadratique avec recherche linéaire exacte on a vu que

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

Problème 2 : Le minimum de la fonction

$$f(x, y) = (x - 2)^4 + (x - 2y)^2.$$

Correspondant au point de coordonnées à (2 1)

Le gradient de cette fonction ce calcul de la manière suivante :

$$\frac{\delta f(x, y)}{\delta x} = 4(x - 2)^3 + 2(x - 2y).$$

$$\frac{\delta f(x, y)}{\delta y} = -4(x - 2y).$$

La norme du gradient de ce calcul de la manière suivante :

$$\|g\| = \sqrt{g_x^2 + g_y^2}.$$

g_x et g_y représentant la composante en x et en y du gradient g .

Méthode du gradient à pas fixe Tableau 4 : Résultat numérique pour la fonction (x,y)

Point initiale x_0	le pas	NI	CPU time	le min	$\ g\ $
(-1,50, -0,0175)	0,01	3674	9,0609	(1,934, 0,96)	$9,98 * 10^{-4}$
(-1,40, -4,45)	0,01	3686	9,0617	(1,946, 0,9673)	$9,99 * 10^{-4}$
(-1,65, -0,017)	0,07	échoué	0,01756	échoué	/

Remarque 5.1.2 Dans le tableau 4, nous notons que la méthode ne converge pas si le pas est trop grand comme dans le troisième cas. Généralement lorsque le pas est fixée, nous avons besoin d'un grand nombre d'itérations, c'est l'inconvénient de cette méthode.

Méthode du gradient conjugué

Le programme Matlab pour la méthode du gradient conjugué trouvé à [29].

Résultats numériques pour la fonction :

$$f(x, y) = (x - 2)^4 + (x - 2y)^2 \text{ et } x_0 = (-1, 5, -0, 5).$$

La condition d'arrêt est $\|g\| < 10^{-6}$.

Tableau 5 : Comparaison de la performance de la méthode de WYL avec d'autre méthode de gradient conjugué

	WYL	FR	PRP
NI	9	36	16
CPU	0,32	1,3864	0,58
X optimal	(1,99, 1)	(1,96, 0,98)	(1,9, 0,99)
$\ g\ $	$2,5 * 10^{-4}$	$9,54 * 10^{-4}$	$3,04 * 10^{-4}$

La méthode WYL fonctionne mieux que d'autre méthode et converge rapidement vers la solution optimale. Moins d'itération et moins de temps. Il est claire que la méthode de WYL est un peut mieux que les autres méthode.

Problème 3 : Fonction teste Bouth Fonction

Résultat numériques pour la fonction

$$f(x, y) = (x + 2y - 7)^2 + (2x + y - 5)^2.$$

(Bouth fonction) avec la méthode du gradient conjugué nous avons utilisé le code Matlab [29] $x_0 = (-6, 4)$ la condition d'arrêt et $\|g_k\| < 10^{-4}$.

Tableau 6 : Comparaison de la performance de la méthode de WYL avec d'autres méthodes du gradient conjugué

	WYL	FR	PRP
NI	41	99	49
CPU	0,1429	0,2448	0,3202
X optimal	(1, 3)	(1, 3)	(1, 3)
$\ g\ $	$1,1 * 10^{-7}$	$2,15 * 10^{-7}$	$8,042 * 10^{-8}$

Dans le tableau 6, on note que la méthode WYL est plus rapide que PRP et FR.

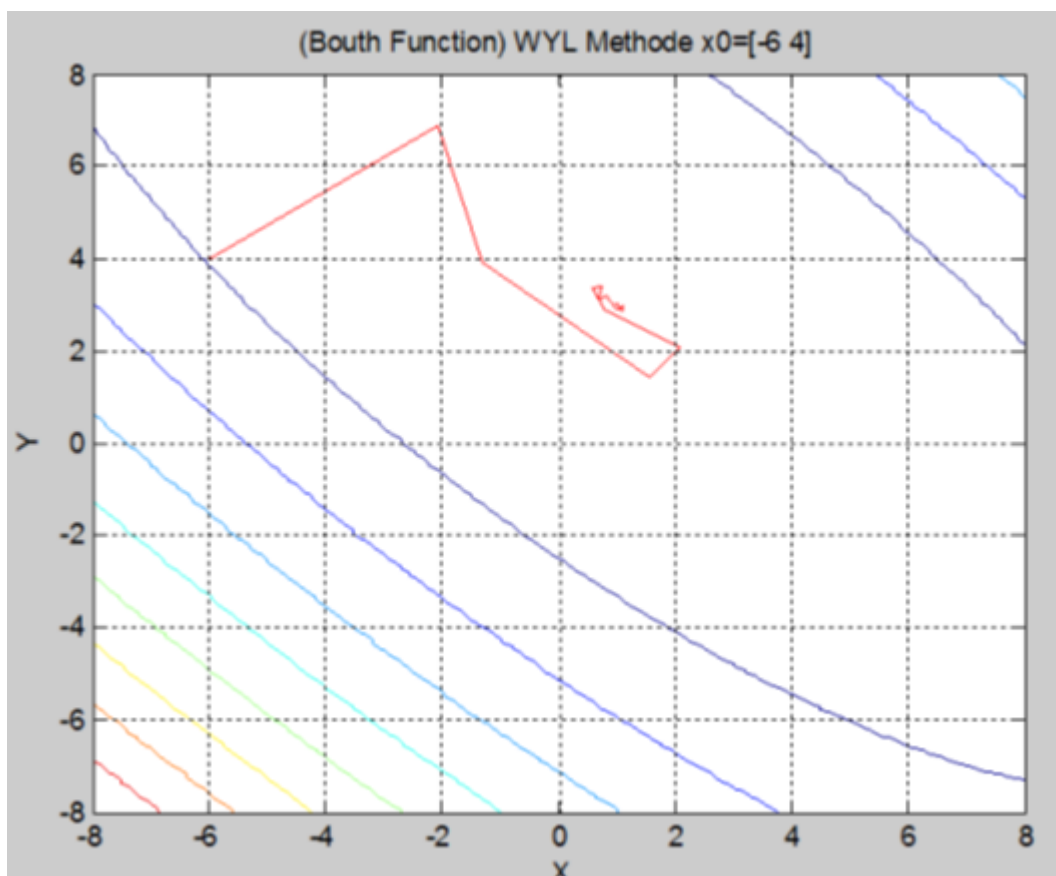


FIGURE 5.1 – Bouth Fonction

Conclusion

Dans ce travail, nous avons étudié la convergence de certaines méthodes du gradient conjugué non linéaire (WYL, fR, PRP) avec la recherche linéaire inexacte de Wolfe fort pour la résolution des problèmes d'optimisation sans contraintes .

Nous avons également mené des expériences numériques sur certaines fonctions de test pour déterminer l'efficacité de ces fonctions à travers le temps pris et le nombre d'itérations.

Afin de tester la performance des méthodes étudiées les expériences numériques, montre l'efficacité de la méthode du gradient conjugué de WYL.

Bibliographie

- [1] Mehiddin AL-BAALI.« Descent property and global convergence of the Fletcher-Reeves method with inexact line search ».In :IMA Journal of Numerical Analysis 5.1 (1985),p.121-124.
- [2] Neculai ANDREI.« An unconstrained optimization test functions collection ».In :Adv.Model.Optim10.1(2008),p.147-161.
- [3] Neculai ANDREI. Another accelerated conjugate gradient algorithm with guaranteed descent and conjugacy conditions for large-scale unconstrained optimization. Rapp. tech. ICI Technical Report, 2010.
- [4] Neculai ANDREI. Nonlinear Conjugate Gradient Methods for Unconstrained Optimization. Springer, 2020.
- [5] Larry ARMIJO. « Minimization of functions having Lipschitz continuous first partial derivatives ». In : Pacific Journal of mathematics 16.1 (1966), p. 1-3.
- [6] Stephen BOYD, Stephen P BOYD et Lieven VANDENBERGHE. Convex optimization. Cambridge university press, 2004.
- [7] Augustin CAUCHY et al. « Méthode générale pour la résolution des systèmes d'équations simultanées ». In : Comp. Rend. Sci. Paris 25.1847 (1847), p. 536-538.
- [8] 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.
- [9] R. Fletcher (1987),Practical methods of optimization, John WileySons,Chichester.
- [10] Reeves FLETCHER et Colin M REEVES. « Function minimization by conjugate gradients ». In : The Computer journal 7.2 (1964), p. 149-154.

- [11] JC GILBERT. « Eléments d'Optimisation Différentiable : Théorie et Algorithmes, notes de cours ». In : École Nationale Supérieure de Techniques Avancées, Paris (2007).
- [12] E. Polak and G.Ribière (1969), Note sur la convergence de directions conjuguées, Rev.Franaise Informat. Recherche Operationelle, 3e année 16, pp;35-43.
- [13] Robert L GRAVES et Philip WOLFE. « Recent advances in mathematical programming ». In : (1963).
- [14] 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.
- [15] Shengwei, Lin Suihua, A new conjugate gradient method combined HS and DY formulas , Department of Mathematics and Information Science, Guangxi University, 2.
- [16] Romaiissa MELLAL. « Méthode du gradient conjugué et convergence uniforme des méthodes multigrilles ». Thèse de doct. Université Badji Mokhtar, 2015.
- [17] Michel MINOUX. Programmation mathématique. Théorie et algorithmes. Dunod, 1983.
- [18] Jorge NOCEDAL et Stephen WRIGHT. Numerical optimization. Springer Science Business Media, 2006.
- [19] B.T. Polyak (1969), The conjugate gradient method in extremem problema, USSR Comput.Math.Math.Phys., 9, pp.94-112...
- [20] M.J.D. Powell (1984), Non convex minimzation calculation and the conjugate gradient method, in :Lecture Notes in Mathematics 1066 (Springer, Berlin), pp.122-14.
- [21] M Bazaraa HD Sherali CM SHETTY. Nonlinear programming. Wiley, 1993.
- [22] M.R. Hestenes, E.Stiefel, Method of conjugate gradient for solving linear equations, J.Res.Nat.Bur.Stand.49 (1952) 409-436
- [23] Y.Liu, C.Storey, Ecient generalized conjugate gradient algorithms, part 1 :theory, J Optim.TheoryAppl.69 (1992) 129-137.
- [24] Lin Suihua, Yao Shengwei, A new conjugate gradient method of modified LS formula, Department of Mathematics and Information Science, Guangxi University, 2006.
- [25] Philip WOLFE. « Convergence conditions for ascent methods ». In : SIAM review 11.2 (1969), p. 226-235.

- [26] G ZOUTENDIJK. « Nonlinear programming, computational methods ». In : Integer and nonlinear programming (1970), p. 37-86.
- [27] Wei et al, The convergence properties of some conjugate gradient methodes, Appl.Math.Comput.183 (2006) 1341-1350 .
- [28] The proof of the succient 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.
- [29] Programme matlab,Conjugate gradient (<https://www.mathworks.com/matlab> Central).