

Table des matières

1	Préliminaires : Optimisation sans contraintes	7
1.1	Généralités sur les matrices	7
1.2	Calcul différentiel	8
1.3	Notion de convexité	9
1.3.1	Continuité des fonctions convexes	10
1.3.2	Différentiabilité des fonctions convexes	10
1.4	Formulation mathématique	11
1.4.1	Le vocabulaire de l'optimisation	11
1.4.2	Formulation abstraite	11
1.4.3	Extremums d'une fonction	12
1.5	Les problèmes de minimisation sans contraintes et leurs algorithmes	13
1.5.1	Algorithmes (déterministes)	13
1.5.2	Aspect général des algorithmes	14
1.5.3	Fonctions multivoques et algorithmes	14
1.6	Convergence globale des algorithmes	15
1.7	Modes et vitesse de convergence	16
1.8	Conditions d'optimalité	16
1.8.1	Pourquoi conditions d'optimalité ?	16
1.8.2	Conditions nécessaires d'optimalité	16
1.8.3	Conditions suffisantes d'optimalité	18
1.8.4	Conditions nécessaires et suffisantes d'optimalité dans le cas convexe	20
2	Les méthodes d'optimisation sans contraintes	21
2.1	Résultats d'existence et d'unicité	21
2.2	Méthodes à directions de descente	22

2.2.1	Principe général	22
2.2.2	Direction de descente	23
2.2.3	Description des méthodes à directions de descentes	23
2.3	La Recherche linéaire	24
2.3.1	Objectifs à atteindre	25
2.3.2	Types de recherche linéaire	26
2.4	Recherches linéaires exactes	26
2.4.1	Les inconvénients des recherches linéaires exactes pour une fonction non linéaire arbitraire	26
2.4.2	Intervalle de sécurité	27
2.5	Recherches linéaires inexactes	27
2.5.1	Schéma des recherches linéaires inexactes	27
2.5.2	La règle (le critère) d'Armijo	28
2.5.3	La règle (le critère) de Goldstein&Price.	29
2.5.4	La règle (le critère) de Wolfe	29
2.6	Convergence des méthodes à directions de descente	30
2.6.1	Condition de Zoutendijk	30
3	Résultats de convergence des méthodes du gradient conjugué non linéaire avec la recherche linéaire inexacte	33
3.1	Méthodes du gradient conjugué non linéaire	33
3.2	Méthode de Fletcher-Reeves	35
3.2.1	Algorithme de la méthode de Fletcher-Reeves	36
3.2.2	La propriété de descente de la méthode de FR	36
3.2.3	Convergence de la méthode de Fletcher-Reeves	38
3.3	Méthode de Polak-Ribière-Polyak	41
3.3.1	Algorithme de la méthode de Polak-Ribière-Polyak	41
3.3.2	Convergence de la méthode de PRP	41
3.4	Méthode de descente conjuguée	43
3.4.1	Algorithme de la méthode de descente conjuguée	43
3.4.2	La propriété descente de la méthode de descente conjuguée	44
3.4.3	Convergence de la méthode de descente conjuguée	45
3.5	Méthode de Dai-Yuan	46
3.5.1	Algorithme de la méthode de Dai-Yuan	46
3.5.2	La propriété de descente de la méthode de Dai-Yuan	47

3.5.3	Convergence de la méthode de Dai-Yuan	48
4	Convergence global de la méthode CGBB à direction de descente pour l'optimisation sans contrainte	49
4.1	Algorithme de la méthode CGBB	50
4.2	La propriété de descente de la méthode CGBB	50
4.3	Convergence de la méthode BB	52
4.4	Résultats et discussion numérique	54
4.4.1	Les profils de performance numérique	54

Introduction

L'optimisation est une méthode (ensemble des méthodes) qui consiste à rechercher le minimum ou le maximum d'une fonction linéaire ou non linéaire avec ou sans contraintes.

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

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

Parmi les plus anciennes méthodes utilisées pour résoudre les problèmes du type (\mathcal{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 ([30, 1952]), 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 ([22, 1964]) (*méthode de Fletcher-Reeves*) puis en 1969 par Polak, Ribière ([38, 1969]) et Ployak ([39, 1969]) (*méthode de Polak-Ribière-Ployak*). Une autre variante a été étudiée en 1987 par Fletcher ([21, 1987]) (*méthode de la descente conjuguée*). Une dernière version a été proposée par Dai et Yuan ([14, 1999]) (*méthode de Dai-Yuan*). Toutes ces méthodes génèrent une suite $\{x_k\}_{k \in \mathbb{N}}$ de la façon suivante

$$x_{k+1} = x_k + \alpha_k d_k \quad (0.1)$$

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

$$d_k = \begin{cases} -g_1 & \text{si } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 2 \end{cases} \quad (0.2)$$

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

Les différentes valeurs attribuées à β_k définissent les différentes formes du gradient conjugué.

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

1- Gradient conjugué variante Hestenes - Stiefel (HS) [30]

$$\beta_k^{HS} = \frac{g_{k+1}^\top y_k}{d_k^\top y_k}$$

2- Gradient conjugué variante Fletcher Reeves (FR) [22]

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

3- Gradient conjugué variante Daniel (D)[16]

$$\beta_k^D = \frac{g_{k+1}^\top \nabla^2 f(x_k) d_k}{d_k^\top \nabla^2 f(x_k) d_k}$$

4- Gradient conjugué variante Polak-Ribière-Polyak (PRP)[38,39]

$$\beta_k^{PRP} = \frac{g_k^\top y_k}{\|g_{k-1}\|^2}$$

5- Gradient conjugué variante descente – Fletcher (CD)[21]

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

6- Gradient conjugué variante Liu - Storey (LS)[33]

$$\beta_k^{LS} = -\frac{g_{k+1}^\top y_k}{d_k^\top g_k}$$

7- Gradient conjugué variante de Dai-Yuan (DY)[14]

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^\top y_{k-1}}$$

8- Gradient conjugué variante de Dai-Liao (DL)[13]

$$\beta_k^{DL} = \frac{g_{k+1}^\top (y_k - t s_k)}{y_k^\top s_k}$$

9- Gradient conjugué variante Hager-Zhang (HZ)[29]

$$\beta_k^{HZ} = \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^\top y_k} \right)^\top \frac{g_{k+1}}{d_k^\top y_k}$$

10- Gradient conjugué variante de Z. Wei [51]

$$\beta_k^* = \frac{g_k^\top \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{\|g_{k-1}\|^2}$$

11- Gradient conjugué variante de Hao Fan, Zhibin Zhu et Anwa Zhou [23]

$$\beta_k^{MN} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^\top g_{k-1}}{\mu |g_k^\top d_{k-1}| + \|g_{k-1}\|^2}$$

12- Gradient conjugué variante Rivaie-Mustafa-Ismail-Leong (RMIL)[46]

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

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

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

Malgré que toutes ces méthodes ont été très importantes pour la résolution du problème (\mathcal{P}), leur étude s'est faite de façon individuelle et chacune à part par Flecher, Powell, Wolf, El Baali, Polak-Ribière, Polyak, Y. H. Dai and Y. Yuan,.... Il est alors légitime de se poser le problème suivant

Peut on définir une nouvelle méthodes dans une même classe et étudier leurs propriétés de descente et de convergence ?

M. Belloufi et R. Benzine ont répondu dans ([7, 2014]) positivement à cette question. Le but principal de ce mémoire est d'étudier de façon approfondie le travail de M. Belloufi et R. Benzine ([7, 2014]).

La nouvelle forme du gradient conjugué en question, est de la forme (0.1) où d_k est la forme générale suivante

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

Le mémoire comporte quatre chapitres

Chapitre 1

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

Chapitre 2

On expose dans ce chapitre les grandes lignes des méthodes d'optimisation sans contraintes basées sur les directions de descente et les recherches linéaires.

On insistera dans ce chapitre sur les recherches linéaires inexactes dites de Wolfe, Armijo et Goldstein-Price.

Chapitre 3

On établit dans ce chapitre une synthèse des résultats de convergence concernant les différentes variantes de la méthode du gradient conjugué avec la recherche linéaire inexacte.

Chapitre 4

On expose dans le chapitre 4 la nouvelle méthode du Gradient conjugué et qui a été introduite par M. Belloufi et R. Benzine ([7,2014]). Les auteurs montrent que cette nouvelle méthode est globalement convergente avec **les conditions de Wolfe fortes**.

Chapitre 1

Préliminaires : Optimisation sans contraintes

On introduit dans ce chapitre les notions préliminaires de base concernant l'optimisation sans contraintes. On commence par quelques éléments de Topologie et de calcul différentiel.

On termine ce chapitre par quelques notions sur les problèmes de minimisation sans contraintes et leurs algorithmes

1.1 Généralités sur les matrices

Dans tout ce qui suit on considère l'ensemble $M_n(\mathbb{k})$ des matrices carrées ($n \times n$) à coefficients dans un corps \mathbb{k} (en pratique \mathbb{R} ou \mathbb{C}).

Définition 1.1 [9] (*Matrice symétrique*)

Soit A une matrice carrée de $M_n(\mathbb{k})$ tel que $A = [a_{ij}]_{1 \leq i, j \leq n}$.

On dit que A est symétrique si $A = A^t$

Ou A^t est la matrice transposée $A^t = [a_{ji}]_{1 \leq i, j \leq n}$

Définition 1.2 [9] (*Matrice (semi-définie) positive*)

Soit A une matrice carrée d'ordre n à coefficients réels.

On dit que A est semi-définie positive si

$$\forall x \in \mathbb{R}^n \quad (Ax, x) = x^t Ax \geq 0$$

Attention cela ne signifie pas automatiquement que tous les coefficients de A sont positifs.

Définition 1.3 [9] (*Matrice définie positive*)

Soit A une matrice carrée d'ordre n .

On dit que A est définie positive si A est semi-définie positive et

$$(Ax, x) = 0 \Leftrightarrow x = 0$$

1.2 Calcul différentiel

Définition 1.4 (*Dérivées partielles*)

Soit f une fonction définie de \mathbb{R}^n dans \mathbb{R} , on suppose que f admet des dérivées partielles par rapport à chaque variable au point a

On appelle différentielle de f en a et on note df_a l'application linéaire de \mathbb{R}^n dans \mathbb{R} définie par

$$df_a = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) h_i \quad (1,1)$$

Définition 1.5 [47] (*Gradient*)

On appelle gradient de la fonction f en un point $x \in \mathbb{R}^n$ et on note $\nabla f(x)$ tel que

$$(\nabla f(x))^\top = \frac{\partial f}{\partial x}(x) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) (x) \quad (1,2)$$

Exemple 1.1 1) fonction $f(x) = -x^4 + 12x^3 - 47x^2 + 60x$

$$\text{Gradient } \nabla f(x) = -4x^3 + 36x^2 - 94x + 60$$

$$2) \text{ fonction } f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

$$\text{Gradient } \nabla f(x) = \begin{pmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}$$

Définition 1.6 [10] (*point stationnaire*)

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

Définition 1.7 [10] (*Hessienne*)

Soit M_n l'anneau des matrices carrées ($n \times n$) dans \mathbb{R}

On appelle Hessienne de f en un point $x \in \mathbb{R}^n$ la matrice symétrique de M_n notée $H(x)$ définie par

$$H(x) = \nabla(\nabla^\top f)(x) = \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right) (x) \quad i = \overline{1, n} \quad j = \overline{1, n} \quad (1,3)$$

Théorème 1.1 [12] (Schwarz)

Soit U un ouvert de \mathbb{R}^n et $a \in U$, Si $f : U \rightarrow \mathbb{R}$ est deux fois différentiable en a alors

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a) \quad \forall i, j \in \{1, \dots, n\}$$

Autrement dit, la matrice $\nabla^2 f(a)$ est symétrique, ou bien la forme bilinéaire $D^2 f(a)$ est symétrique.

Définition 1.8 [5] (Dérivée directionnelle)

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

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

Autrement dit

$$\delta f(x, d) = \lim_{h \rightarrow 0} \frac{f(x + hd) - f(x)}{h} = \nabla^\top f(x) d \quad (1,4)$$

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

1.3 Notion de convexité

Le cas où les données sont convexes est un cas très important car les problèmes quadratiques sont à la base de nombreux algorithmes non linéaires. Nous rappelons quelques définitions et propriétés.

Définition 1.9 [9] (Ensemble convexe)

On dit que l'ensemble $S \subset \mathbb{R}^n$ est **convexe** si

$$\forall (x, y) \in S \times S, \forall \lambda \in [0, 1] \quad \lambda x + (1 - \lambda)y \in S \quad (1,5)$$

Définition 1.10 [9] (Fonction convexe)

On dit que la fonction $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ est **convexe** si S est convexe et si

$$\forall (x, y) \in S \times S, \forall \lambda \in [0, 1] \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad (1,6)$$

Définition 1.11 [9] (Fonction strictement convexe)

On dit que la fonction $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ est **strictement convexe** si S est convexe et si

$$\forall (x, y) \in S \times S \text{ avec } x \neq y \forall \lambda \in]0, 1[\quad f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y) \quad (1,7)$$

Définition 1.12 [31] (Fonction fortement convexe)

Supposons que f soit différentiable, f est dit **fortement convexe** de module $\varepsilon > 0$ si

$$\forall (x, y) \in \mathbb{R}^n, \forall \lambda \in [0, 1] \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{\varepsilon}{2} \lambda(1 - \lambda) \|x - y\|^2 \quad (1.8)$$

Définition 1.13 [31] (Fonction pseudo convexe)

Supposons que f soit différentiable, f est dit **pseudo convexe** si

$$\forall x_1 \text{ et } x_2 \text{ vérifiant } \nabla f(x_1)^t(x_2 - x_1) \geq 0 \text{ on a } f(x_2) \geq f(x_1) \quad (1.9)$$

1.3.1 Continuité des fonctions convexes

Donnons maintenant quelques propriétés (topologiques) importantes des fonctions convexes.

Lemme 1.1 [9] Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction convexe et majorée par une constante et si au voisinage d'un point $x_0 \in H$, f est continue en x_0

1.3.2 Différentiabilité des fonctions convexes

Donnons maintenant quelques propriétés de différentiabilité.

Théorème 1.2 Soit $S \subset \mathbb{R}^n$ convexe non vide ouvert et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable dans S alors

1) f est convexe dans S si et seulement si pour tout $\hat{x} \in S$

$$f(x) \geq f(\hat{x}) + \nabla^T(x - \hat{x}) \quad \forall x \in S$$

2) f est strictement convexe dans S si et seulement si pour tout $\hat{x} \in S$

$$f(x) > f(\hat{x}) + \nabla^T(x - \hat{x}) \quad \forall x \in S$$

Théorème 1.3 Soit $S \subset \mathbb{R}^n$ non vide ouvert convexe et $f : S \rightarrow \mathbb{R}$ une fonction deux fois différentiable, si $H(x)$ est définie positive on tout x de S alors f est strictement convexe dans S .

1.4 Formulation mathématique

1.4.1 Le vocabulaire de l'optimisation

Soit V est un espace vectoriel normé muni de la norme $\|\cdot\|$

On s'intéresse au problème suivant

$$\inf_{x \in K} f(x) \quad (1,10)$$

Où $K \subset V$ et $f : K \rightarrow \mathbb{R}$ est une fonction appelle fonction coût ou critère.

* Si $K = V$ on dit que (1, 10) est un problème d'optimisation sans contrainte.

* Si $K \subsetneq V$ on dit que (1, 10) est un problème d'optimisation sous contrainte.

* Si $\dim K < +\infty$ (resp $\dim K = +\infty$) on dit que (1, 10) est un problème d'optimisation en dimension finie (resp infinie).[44]

1.4.2 Formulation abstraite

Les exemples d'optimisation peuvent tous s'écrire sous la forme générale suivante

$$(P') \begin{cases} \min f(x) \\ g(x) \leq 0 \\ h(x) = 0 \\ x \in \mathbb{R}^n \end{cases} \quad (1,11)$$

Où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction de plusieurs variables ($x = (x_1, x_2, \dots, x_n)$) à valeurs réelles cette fonction (que l'on minimise) est appelée indifféremment fonction coût, objectif ou critère.

$g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est une fonction de plusieurs variables $x \in \mathbb{R}^n$ à valeurs dans \mathbb{R}^p : elle à p composantes et on peut écrire $g(x) = (g_1(x), \dots, g_p(x))$.

Chaque fonction g_i étant définie sur \mathbb{R}^n et à valeurs dans \mathbb{R} , La fonction g représente les contraintes en inégalité.

La notation $g(x) \leq 0$ signifie qu'on considère les inégalités composante par composante

$$g(x) \leq 0 \stackrel{\text{def}}{\Leftrightarrow} \forall i = \overline{1, p} \quad g_i(x) \leq 0$$

$h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ est une fonction de plusieurs variables $x \in \mathbb{R}^n$ à valeurs dans \mathbb{R}^q : elle a q composantes et on peut écrire $h(x) = (h_1(x), \dots, h_q(x))$.

Chaque fonction h_i étant définie sur \mathbb{R}^n et à valeurs dans \mathbb{R} , La fonction h représente les contraintes en égalité [9].

1.4.3 Extremums d'une fonction

Définition 1.14 [6] (*Problème Minimalici*)

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

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

Considérons le problème de minimisation sans contraintes (P)

a) $x^* \in \mathbb{R}^n$ s'appelle minimum global de problème (P) si

$$f(x^*) \leq f(x) \quad \forall x \in \mathbb{R}^n \quad (1,12)$$

b) $x^* \in \mathbb{R}^n$ est un minimum local s'il existe un voisinage $V_\varepsilon(x^*)$ de x^* tel que

$$f(x^*) \leq f(x) \quad \forall x \in V_\varepsilon(x^*) \quad (1,13)$$

c) x^* est minimum local strict s'il existe un voisinage $V_\varepsilon(x^*)$ de x^* tel que

$$f(x^*) < f(x) \quad \forall x \in V_\varepsilon(x^*) \quad \text{et } x \neq x^* \quad (1,14)$$

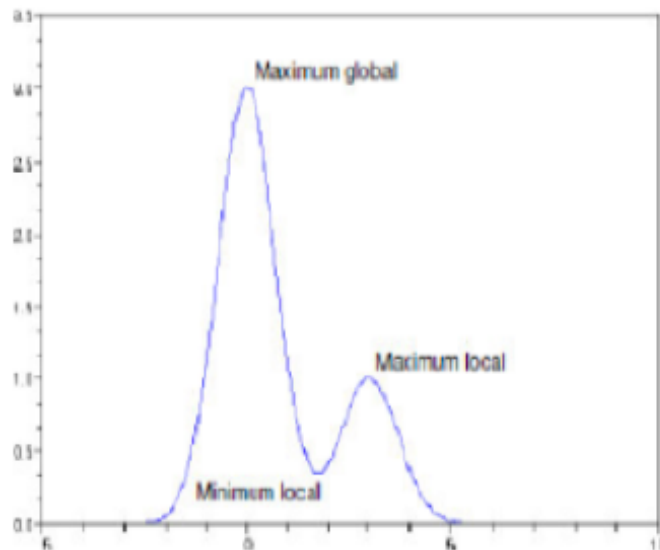


Fig1.1 Le Maximum global, local et Le Minimum local

Exemple 1.2 La fonction $f(x) = 3x^4 - 16x^3 + 18x^2$ présente sur $I = [-1, 4]$

Trois extrema locaux, $x = 0$ et $x = 3$ (minima), $x = 1$ (maximum).

Seul $m = f(3) = -27$ est un minimum global sur I

Le maximum global étant atteint à l'extrémité gauche de l'intervalle I , $M = f(-1) = 37$.

1.5 Les problèmes de minimisation sans contraintes et leurs algorithmes

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ On appelle problème de minimisation sans contraintes le problème (P) cité précédemment. L'étude de ces problèmes est importante pour des raisons diverses. Beaucoup de problèmes d'optimisation avec contraintes sont transformés en des suites de problèmes d'optimisation sans contraintes (multiplicateur de Lagrange, méthodes des pénalités, ...). L'étude des problèmes d'optimisation sans contraintes trouve aussi des applications dans la résolution des systèmes différentiels non linéaires.

Les méthodes numériques de résolution de divers problèmes de minimisation sans contraintes ont pris ces dernières années un bel essor si bien que la bibliographie correspondante contient des centaines d'ouvrages et d'articles. Cet intérêt n'est nullement fortuit, il reflète le rôle de premier plan que les problèmes d'optimisation jouent dans les applications.

La construction d'algorithmes consacrés à la recherche efficace de minimum d'une fonction sans contraintes est un problème complexe, car il ne suffit pas d'élaborer un algorithme mais il faut montrer de plus qu'il est plus performant que ceux déjà connus.

On compare des algorithmes en se basant sur plusieurs critères, par exemple, la précision du résultat, le nombre d'évaluations fonctionnelles et celui d'évaluations du gradient, la vitesse de la convergence ainsi que le temps de calcul.

1.5.1 Algorithmes (déterministes)

La définition des classes de problèmes fait appel aux notions d'algorithme déterministe [2]

Un algorithme déterministe est un algorithme classique qui fait des choix déterministes à chaque étape.

Définition 1.15 [9] (*Algorithme*)

Un algorithme est défini par une application A de \mathbb{R}^n dans \mathbb{R}^n permettant la génération d'une suite d'éléments de \mathbb{R}^n par la formule :

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné} & k = 0 \text{ Etape d'initialisation} \\ x_{k+1} = A(x_k) & k = k + 1 \text{ Itération } k \end{cases}$$

Ecrire un algorithme n'est ni plus ni moins que se donner une suite $(x_k)_{k \in \mathbb{N}}$ de \mathbb{R}^n étudier la convergence de l'algorithme, c'est étudier la convergence de la suite $(x_k)_{k \in \mathbb{N}}$.

1.5.2 Aspect général des algorithmes

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

$$x_{k+1} = x_k + \alpha_k d_k \quad (1,15)$$

Où d_k détermine la direction de déplacement à partir du point x_k et α_k est un facteur numérique dont le grandeur donne la longueur du pas dans la direction d_k .

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

La façon avec laquelle on construit les vecteurs d_k et les scalaires α_k détermine directement les propriétés du processus et spécialement en ce qui concerne la convergence de la suite $\{x_k\}$ ainsi que la vitesse de convergence de la méthode.

Pour s'approcher de la solution optimale du problème (P) (dans le cas général, c'est un point en lequel ont lieu peut être avec une certaine précision les conditions nécessaires d'optimalité de f), on se déplace naturellement à partir du point x_k dans la direction de la décroissance de la fonction f [10].

1.5.3 Fonctions multivoques et algorithmes

Une application multivoque est une application A qui à $x \in \mathbb{R}^n$ fait correspondre un sous ensemble $A(x)$ de \mathbb{R}^n .

Etant donné un point x_k , En appliquant les instructions d'un certain algorithme, on obtient un nouveau point x_{k+1} . Cette procédure peut être décrite par une application multivoque A appelée application algorithmique.

Donc étant donné un point $x_1 \in \mathbb{R}^n$, l'application algorithmique génère une suite $x_1, x_2, \dots, x_{k+1} \in \mathbb{R}^n$ pour tout k .

La notion d'application algorithmique fermé est directement liée à la convergence des algorithmes [10].

Définition 1.16 [10] Soient X et Y deux sous ensembles fermés non vides de \mathbb{R}^n et \mathbb{R}^m respectivement et $A : X \rightarrow Y$ une application multivoque. A est dite fermée au point $x \in \mathbb{R}^n$ si

$$\left. \begin{array}{l} x_k \in X, x_k \rightarrow x \\ y_k \in Y, y_k \rightarrow y \end{array} \right\} \text{ implique que } y \in A(x) \quad (1,16)$$

Donnons un théorème de convergence qui utilise les fonctions multivoques et qui nous sera utile par la suite. Avant cela notons qu'à cause de la non convexité, de la taille du problème et d'autres

difficultés on peut arrêter le processus itératif si on trouve un point appartenant à un ensemble spécifique qu'on appelle ensemble des solutions Ω .

Voici ci-dessous quelques exemples typiques de cet ensemble.

$$\Omega = \{x^* : \nabla f(x^*) = 0\}$$

$$\Omega = \{x^* : x^* \text{ est une solution optimale locale du problème } (P)\}$$

$$\Omega = \{x^* : f(x^*) < \nu^* + \varepsilon\} \text{ où } \varepsilon > 0 \text{ est une tolérance définie à l'avance et } \nu^* \text{ est la valeur minimale de la fonction objectif.}$$

Nous dirons que l'application algorithmique $A : X \rightarrow X$ converge dans $Y \subseteq X$ si commençant par n'importe quel point initial $x_1 \in Y$, la limite de toute sous suite convergente, extraite de la suite x_1, x_2, \dots , générée par l'algorithme, appartient à Ω .

Théorème 1.4 [6] Soient X un ensemble non vide fermé dans \mathbb{R}^n et Ω un ensemble des solutions non vide. Soit $\xi : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue, on considère que l'application algorithmique $C : X \rightarrow X$ vérifie la propriété suivante : "étant donné $x \in X$ alors $\xi(y) \leq \xi(x)$ pour $y \in C(x)$ " : Soit $B : X \rightarrow X$ une application algorithmique fermée sur le complémentaire de Ω qui vérifie : $\xi(y) < \xi(x)$ pour $y \in B(x)$ et $x \notin \Omega$ Maintenant considérons l'application algorithmique composée $A = CB$: Soit $x_1 \in X$ la suite $\{x_k\}$ est générée comme suit : Si $x_k \in \Omega$ stop, sinon poser $x_{k+1} \in A(x_k)$ remplacer k par $(k + 1)$ et répéter.

Supposons que $\Lambda = \{x \in X, \xi(x) \leq \xi(x_1)\}$ est compact. Alors, ou bien l'algorithme s'arrête à un nombre fini d'itérations pour un point de Ω , ou tous les points d'accumulations de $\{x_k\}$ appartiennent à Ω .

1.6 Convergence globale des algorithmes

Définition 1.17 [28] Soit un algorithme itératif qui génère une suite $(x_k)_{k \in \mathbb{N}}$ dans \mathbb{R}^n afin de résoudre le problème

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

Où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une application de classe C^1 , l'algorithme est dit globalement convergent si quelque soit le point initial $x_0 \in \mathbb{R}^n$

$$\lim_{k \rightarrow +\infty} \inf \|\nabla f(x_k)\| = 0$$

1.7 Modes et vitesse de convergence

Définition 1.18 [9] Soit $\{x_k\}$ une suite de vecteurs convergente vers x telle que $x_k \neq \hat{x}$ pour tout k , l'ordre de la convergence de la suite $\{x_k\}$ est le supremum des nombres p non négative tel que

$$\overline{\lim}_{k \rightarrow +\infty} \frac{\|x_{k+1} - \hat{x}\|}{\|x_k - \hat{x}\|^p} = \tau < \infty \quad (1,17)$$

Si $p = 1$ et $\tau < 1$ alors la convergence est dite linéaire.

Si $p = 1$ et $\tau = 0$ alors la convergence est dite super linéaire.

Si $p > 1$ et $\tau < \infty$ alors la convergence est super linéaire d'ordre p .

En particulier, si $p = 2$ et $\tau < 1$ alors la convergence est dite quadratique.

1.8 Conditions d'optimalité

1.8.1 Pourquoi conditions d'optimalité ?

Afin d'analyser ou de résoudre de manière efficace un problème d'optimisation, il est fondamental de pouvoir disposer de conditions d'optimalité.

En effet, celles-ci nous servent non seulement à vérifier la validité des solutions obtenues, mais souvent l'étude de ces conditions aboutit au développement des algorithmes de résolution eux-mêmes.

Des conditions équivalentes peuvent être obtenues de diverses manières, en procédant à des analyses suivants différentes "lignes directrices".

L'approche considérée ici pour l'obtention de conditions est basée sur les notions de descente et de direction admissible.[10]

1.8.2 Conditions nécessaires d'optimalité

Théorème 1.5 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que f soit différentiable au point $\hat{x} \in \mathbb{R}^n$. Soit $d \in \mathbb{R}^n$ telle que $\nabla f(\hat{x})^t d < 0$ Alors

$$\exists \delta > 0 \text{ tel que } f(\hat{x} + \alpha d) < f(\hat{x}) \forall \alpha \in]0, \delta[$$

La direction d s'appelle dans ce cas direction de descente.

Preuve. [6] Comme f est différentiable en \hat{x} alors

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

Où $\lambda(\hat{x}, \alpha d) \rightarrow 0$ pour $\alpha \rightarrow 0$, Ceci implique

$$\frac{f(\hat{x} + \alpha d) - f(\hat{x})}{\alpha} = \nabla f(\hat{x})^t d + \|d\| \lambda(\hat{x}, \alpha d) \quad \alpha \neq 0$$

Et comme $\nabla f(\hat{x})^t d < 0$ et $\lambda(\hat{x}, \alpha d) \rightarrow 0$ pour $\alpha \rightarrow 0$, $\exists \delta > 0$ tel que

$$\nabla f(\hat{x})^t d + \|d\| \lambda(\hat{x}, \alpha d) < 0 \quad \forall \alpha \in]0, \delta[$$

Et par conséquent on obtient

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

■

Conditions nécessaires du premier ordre

Théorème 1.6 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable en \hat{x} , Si \hat{x} est un minimum local alors

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

Preuve. [35] Par contre, on suppose que $\nabla f(\hat{x}) \neq 0$

Si on pose $d = -\nabla f(\hat{x})$, on obtient

$$\nabla f(\hat{x})^t d = -\|\nabla f(\hat{x})\|^2 < 0$$

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

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

Mais ceci est contradiction avec le fait que \hat{x} est un minimum local, D'où $\nabla f(\hat{x}) = 0$. ■

Remarque 1.1 Le théorème dernier vrai si on remplace \mathbb{R}^n par un ouvert V de \mathbb{R}^n

Conditions nécessaires du deuxième ordre

Théorème 1.7 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable en \hat{x} Supposons que \hat{x} soit minimum local
Alors

$$\nabla f(\hat{x}) = 0 \text{ et } H(\hat{x}) \text{ est semi définie positive}$$

Preuve. [18] Le théorème ci-dessus montre la première proposition

Pour la deuxième proposition on a

$$f(\hat{x} + \alpha d) = f(\hat{x}) + \alpha \nabla f(\hat{x})^t d + \frac{1}{2} \alpha^2 d^t H(\hat{x}) d + \alpha^2 \|d\|^2 \lambda(\hat{x}, \alpha d)$$

Où $\lambda(\hat{x}, \alpha d) \rightarrow 0$ pour $\alpha \rightarrow 0$, Ceci implique

$$\frac{f(\hat{x} + \alpha d) - f(\hat{x})}{\alpha^2} = \frac{1}{2} d^t H(\hat{x}) d + \|d\|^2 \lambda(\hat{x}, \alpha d) \quad \alpha \neq 0$$

Comme \hat{x} est un minimum local alors $f(\hat{x} + \alpha d) \geq f(\hat{x})$

Pour α suffisamment petit, d'où

$$\frac{1}{2} d^t H(\hat{x}) d + \|d\|^2 \lambda(\hat{x}, \alpha d) \geq 0 \quad \text{pour } \alpha \text{ petit}$$

En passant à la limite quand $\alpha \rightarrow 0$, on obtient que $d^t H(\hat{x}) d \geq 0$

D'où $H(\hat{x})$ est semi définie positive. ■

Exemple 1.3 Soit $f(x) = x_1^4 + x_2^4 - 4x_1x_2$

$$\text{Gradient } \nabla f(x) = \begin{pmatrix} 4x_1^3 - 4x_2 \\ -4x_1 + 4x_2^3 \end{pmatrix}$$

Points stationnaires vérifient $\nabla f(x) = 0$.

$$\text{Trois solutions } x^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, x^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x^{(3)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

1) $x^{(2)}$ et $x^{(3)}$ sont des minimum alors que $x^{(1)}$ ne l'est pas

2) Tout point stationnaire n'est pas forcément un minimum Nécessité d'une autre condition

1.8.3 Conditions suffisantes d'optimalité

Théorème 1.8 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiable en \hat{x}

Si $\nabla f(\hat{x}) = 0$ et $H(\hat{x})$ est définie positive, alors \hat{x} est minimum local strict.

Preuve. [36] La fonction f est deux fois différentiable au point \hat{x}

Alors $\forall x \in \mathbb{R}^n$, on obtient

$$f(x) = f(\hat{x}) + \nabla f(\hat{x})^t (x - \hat{x}) + \frac{1}{2} (x - \hat{x})^t H(\hat{x}) (x - \hat{x}) + \|x - \hat{x}\|^2 \lambda(\hat{x}, x - \hat{x}) \quad (1,18)$$

Où $\lambda(\hat{x}, x - \hat{x}) \rightarrow 0$ pour $x \rightarrow \hat{x}$

Supposons que \hat{x} n'est pas un minimum local strict.

Donc il existe une suite $\{x_k\}$ convergente vers \hat{x} telle que

$$f(x_k) \leq f(\hat{x}) \quad x_k \neq \hat{x} \quad \forall k$$

Posons

$$d_k = \frac{(x_k - \hat{x})}{\|x_k - \hat{x}\|}$$

Donc $\|d_k\| = 1$ et on obtient à partir de (1.18)

$$\frac{f(x_k) - f(\hat{x})}{\|x_k - \hat{x}\|^2} = \frac{1}{2} d_k^t H(\hat{x}) d_k + \lambda(\hat{x}, x_k - \hat{x}) \leq 0 \quad \forall k \quad (1,19)$$

Et comme $\|d_k\| = 1 \quad \forall k$ alors $\exists \{d_k\}_{k \in N_1 \subset \mathbb{N}}$ telle que

$$d_k \rightarrow d \text{ pour } k \rightarrow \infty \text{ et } k \in N_1$$

On a bien sûr $\|d\| = 1$.

Considérons donc $\{d_k\}_{k \in N_1}$ et le fait que

$$\lambda(\hat{x}, x - \hat{x}) \rightarrow 0 \text{ pour } k \rightarrow \infty \text{ et } k \in N_1$$

Alors (1.19) donne $d^t H(\hat{x}) d \leq 0$,

Ce qui contredit le fait que $H(\hat{x})$ est définie positive car $\|d\| = 1$ (donc $d \neq 0$).

Donc \hat{x} est un minimum local strict. ■

Exemple 1.4 Illustration de la condition nécessaire du 2e ordre

$$f(x) = x_1^4 + x_2^4 - 4x_1x_2$$

$$\text{Gradient : } \nabla f(x) = \begin{pmatrix} 4x_1^3 - 4x_2 \\ -4x_1 + 4x_2^3 \end{pmatrix}$$

$$\text{Points stationnaires : } x^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, x^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x^{(3)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\text{Hessienne } H(x) = \begin{pmatrix} 12x_1^2 & -4 \\ -4 & 12x_2^2 \end{pmatrix}$$

	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$
Hessienne	$\begin{pmatrix} 0 & -4 \\ -4 & 12 \end{pmatrix}$	$\begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}$	$\begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}$
Val. propres	4 et -4	8 et 16	8 et 16
Type de solution	Point selle	Minimum	Minimum

1.8.4 Conditions nécessaires et suffisantes d'optimalité dans le cas convexe

Théorème 1.9 [6] Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que f est convexe et différentiable Alors

\hat{x} est un minimum global de f si et seulement si $\nabla f(\hat{x}) = 0$

Remarque 1.2 Si 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 (voir [6])

Théorème 1.10 [6] Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que f est pseudo convexe

Soit $\hat{x} \in \mathbb{R}^n$ tel que $\nabla f(\hat{x}) = 0$ alors \hat{x} est un minimum global de f

Chapitre 2

Les méthodes d'optimisation sans contraintes

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

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

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

On insistera dans ce chapitre sur les recherches linéaires inexactes dites de Wolfe, Armijo et Goldstein-Price.

2.1 Résultats d'existence et d'unicité

Avant d'étudier les propriétés de la solution (ou des solutions) de (P) il faut s'assurer de leur existence. Nous donnerons ensuite des résultats d'unicité.

Définition 2.1 [19] (*Fonction coercive*)

On dit que $f : H \rightarrow \mathbb{R}$ est coercive si

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty \quad (2,1)$$

Exemple 2.1 1. La fonction $f_1(x) = \|x\|_2$ est coercive

2. $f_2(x) = x_1^2 - x_2^2$ n'est pas coercive.

En effet, la suite de terme général $x_n = (0, n)$ $n \in \mathbb{N}$ est telle que

$$\lim_{n \rightarrow +\infty} \|x_n\| = \lim_{n \rightarrow +\infty} n = +\infty \text{ mais } \lim_{n \rightarrow +\infty} f_2(x_n) = \lim_{n \rightarrow +\infty} -n^2 = -\infty$$

Théorème 2.1 (Existence)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ propre, continue et coercive

Alors (P) admet au moins une solution.

Preuve. [Voir 45] ■

L'unicité résulte en général de propriétés de convexité (de f et de K).

Théorème 2.2 (Unicité)

Soit $f : K \subset \mathbb{R}^n \rightarrow \mathbb{R}$ strictement convexe sur K convexe, Le minimum de f sur K s'il existe est unique.

Preuve. [45] Soit $\hat{x} \in K$ tel que

$$f(\hat{x}) \leq f(x) \quad \forall x \in K$$

Supposons qu'il existe $\hat{y} \neq \hat{x}$ tel que

$$f(\hat{y}) \leq f(x) \quad \forall x \in K$$

Formons pour $\lambda \in]0, 1[$ le vecteur

$$u = \lambda \hat{y} + (1 - \lambda) \hat{x}$$

D'après la stricte convexité de f et puisque nécessairement $f(\hat{y}) = f(\hat{x})$ on a

$$f(u) < \lambda f(\hat{y}) + (1 - \lambda) f(\hat{x}) = f(\hat{x})$$

Ce qui contredit le fait que \hat{x} soit un minimum on a donc $\hat{x} = \hat{y}$. ■

2.2 Méthodes à directions de descente

2.2.1 Principe général

Considérons le problème d'optimisation sans contrainte [10]

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

Où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est supposée régulière.

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

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

2.2.2 Direction de descente

Définition 2.2 [10] (direction de descente)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$, d un vecteur non nul de \mathbb{R}^n est dit une direction de descente s'il existe un réel $\delta > 0$ tel que pour tout α dans l'intervalle $]0, \delta[$ on a

$$f(x + \alpha d) < f(x) \quad (2,2)$$

Théorème 2.3 [10] Si f est différentiable en un point $x \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$ telle que

$$\nabla^\top f(x) \cdot d < 0 \quad (2,3)$$

Alors d est une direction de descente en x

Remarque 2.1 [10] La direction d fait avec l'opposé du gradient $-\nabla f(x)$ un angle θ strictement plus petit que 90°

$$\theta = \arccos \frac{-\nabla^\top f(x) \cdot d}{\|\nabla^\top f(x)\| \|d\|} \in [0, \frac{\pi}{2}[\quad (2,4)$$

L'ensemble des directions de descente de f en x , $\{d \in \mathbb{R}^n : \nabla^\top f(x) \cdot d < 0\}$ forme un demi-espace ouvert de \mathbb{R}^n

On voit que si d est une direction de descente $f(x + \alpha d) < f(x)$ pour tout $\alpha > 0$ 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 .

2.2.3 Description des méthodes à directions de descentes

Les méthodes à directions de descentes utilisent cette idée pour minimiser une fonction [11, 1847].

Elles construisent la suite des itérés $\{x_k\}_{k \geq 1}$ approchant une solution x_k du problème (\mathcal{P}) par la récurrence

$$x_{k+1} = x_k + \alpha_k d_k \quad \text{pour } k \geq 1 \quad (2,5)$$

Où α_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 α_k c'est ce que l'on appelle la recherche linéaire.

Algorithme 2.1(méthode à directions de descente- une itération)[10]**Etape 0 : (initialisation)**

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

Etape1 :

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

Etape2 :

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

Etape3 :

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

Etape4 :

Si la recherche linéaire réussie $x_{k+1} = x_k + \alpha_k d_k$;

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

2.3 La Recherche linéaire

Faire de la recherche linéaire veut dire déterminer un pas α_k le long d'une direction de descente d_k , autrement dit résoudre le problème unidimensionnel

$$\min f(x_k + \alpha d_k) \quad \alpha \in \mathbb{R} \quad (2,6)$$

Notre intérêt pour la recherche linéaire ne vient pas seulement du fait que dans les applications on rencontre, naturellement, des problèmes unidimensionnels, mais plutôt du fait que la recherche linéaire est un composant fondamental de toutes les méthodes traditionnelles d'optimisation multidimensionnelle. D'habitude, nous avons le schéma suivant d'une méthode de minimisation sans contraintes multidimensionnelle : en regardant le comportement local de l'objectif f sur l'itération courante x_k , la méthode choisit la "direction du mouvement" d_k (qui normalement, est une direction de descente de l'objectif : $\nabla^\top f(x).d < 0$) et exécute un pas dans cette direction

$$x_k \longmapsto x_{k+1} = x_k + \alpha_k d_k \quad (2,7)$$

Afin de réaliser un certain progrès en valeur de l'objective, c'est-à-dire pour assurer que : $f(x_k + \alpha_k d_k) \leq f(x_k)$. Et dans la majorité des méthodes le pas dans la direction d_k est choisi par la minimisation unidimensionnelle de la fonction

$$h_k(\alpha) = f(x_k + \alpha d_k) \quad (2,8)$$

Ainsi, la technique de recherche linéaire est une brick de base fondamentale de toute méthode multidimensionnelle.

2.3.1 Objectifs à atteindre

Il s'agit de réaliser deux objectifs

Le premier objectif

Consiste à faire décroître f suffisamment. Cela se traduit le plus souvent par la réalisation d'une inégalité de la forme

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \text{"un terme négatif"} \quad (2,9)$$

Le terme négatif disons ν_k joue un rôle-clé dans la convergence de l'algorithme utilisant cette recherche linéaire. L'argument est le suivant. Si $f(x_k)$ est minorée (il existe une constante C telle que $f(x_k) \geq C$ pour tout k), alors ce terme négatif tend nécessairement vers zéro : $\nu_k \rightarrow 0$. C'est souvent à partir de la convergence vers zéro de cette suite que l'on parvient à montrer que le gradient lui-même doit tendre vers zéro. Le terme négatif devra prendre une forme bien particulière si on veut pouvoir en tirer de l'information.

En particulier, il ne suffit pas d'imposer ([31])

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

Le second objectif

Consiste d'empêcher le pas $\alpha_k > 0$ d'être trop petit, trop proche de zéro.

Le premier objectif n'est en effet pas suffisant car l'inégalité (2,9) est en général satisfaite par des pas $\alpha_k > 0$ 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 ([20]).

Si on prend

$$0 < \alpha_k \leq \frac{\varepsilon}{2^k \|d_k\|} \quad (2,10)$$

La suite $\{x_k\}$ générée par (2,7) est de Cauchy, puisque pour $1 \leq l < k$ on a

$$\|x_k - x_l\| = \left\| \sum_{i=1}^{i=k-1} \alpha_i d_i \right\| \leq \sum_{i=1}^{i=k-1} \frac{\varepsilon}{2^i} \rightarrow 0 \quad \text{lorsque } l \rightarrow \infty \quad (2,11)$$

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 $\bar{x} \in \bar{B}(x_1, \varepsilon)$ et donc \bar{x} ne saurait être solution s'il n'y a pas de solution dans $\bar{B}(x_1, \varepsilon)$. 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 $\{x_k + \alpha d_k / \alpha \in \mathbb{R}\} \subset \mathbb{R}^n$ comme la fonction

$$h_k : \mathbb{R}^+ \rightarrow \mathbb{R} \quad \alpha \mapsto h_k(\alpha) = f(x_k + \alpha d_k) \quad (2,12)$$

2.3.2 Types de recherche linéaire

Il existe deux grandes classes de méthodes qui s'intéressent à l'optimisation unidimensionnelle :

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

2.4 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 α_k comme solution du problème

$$\min h_k(\alpha) \quad \alpha \geq 0 \quad (2,13)$$

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 h_k qui fait décroître cette fonction

$$\alpha_k = \inf \{ \alpha \geq 0 : h_k(\alpha) < h_k(0) \} \quad (2,14)$$

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. ([10])

Remarque 2.2 [10] *Ces deux règles ne sont utilisées que dans des cas particuliers, par exemple lorsque h_k est quadratique.*

Le mot exact prend sa signification dans le fait que si f est quadratique la solution de la recherche linéaire s'obtient de façon exacte et dans un nombre fini d'itérations.

2.4.1 Les inconvénients des recherches linéaires exactes 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 (recherche linéaire inexacte), moins gourmandes en temps de calcul. ([10])

2.4.2 Intervalle de sécurité

Dans la plupart des algorithmes d'optimisation modernes, on ne fait jamais de recherche linéaire exacte, car trouver α_k signifie qu'il va falloir calculer un grand nombre de fois la fonction h_k et cela peut être dissuasif du point de vue du temps de calcul.

En pratique, on recherche plutôt une valeur de α^* qui assure une décroissance suffisante de f . Cela conduit à la notion d'intervalle de sécurité ([31]).

Définition 2.3 [24] *On dit que $[\alpha_g, \alpha_d]$ est un intervalle de sécurité s'il permet de classer les valeurs de α de la façon suivante :*

- Si $\alpha < \alpha_g$ alors est considéré trop petit,
- Si $\alpha_d \geq \alpha \geq \alpha_g$ alors est satisfaisant,
- Si $\alpha > \alpha_d$ alors est considéré trop grand.

Le problème est de traduire de façon numérique sur h_k les trois conditions précédentes, ainsi que de trouver un algorithme permettant de déterminer α_g et α_d .

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

$$\nabla^\top f(x_k) \cdot d_k < 0 \quad (2,15)$$

Le but est de réduire “de façon importante” la valeur de l'objectif par un pas

$$x_k \longmapsto x_{k+1} = x_k + \alpha_k d_k \quad (2,16)$$

De x_k dans la direction d_k pour cela de nombreux mathématiciens (Armijo, Goldstein, Wolfe, Al-baali, Fletcher...) ont élaboré plusieurs règles (tests). L'objectif de cette section consiste à présenter les principaux tests. D'abord présentons le schéma d'une recherche linéaire inexacte. ([31])

2.5.1 Schéma des recherches linéaires inexactes

Elles reviennent à déterminer, par tâtonnement un intervalle $[\alpha_g, \alpha_d]$, où $\alpha^* \in [\alpha_g, \alpha_d]$, dans lequel

$$h_k(\alpha_k) < h_k(0) \quad (f(x_k + \alpha_k d_k) < f(x_k)) \quad (2,17)$$

Le schéma de l'algorithme est donc

Algorithme 1.1 (Schéma général des recherches linéaires inexactes)[31]

Etape 0 : (initialisation)

$\alpha_{g,1} = \alpha_{d,1} = 0$, choisir $\alpha_1 > 0$, poser $k = 1$ et aller à l'étape 1.

Etape 1 :

Si α_k est satisfaisant (suivant un certain critère) : STOP($\alpha^* = \alpha_k$).

Si α_k est trop petit (suivant un certain critère) : nouvel intervalle :

$[\alpha_{g,k+1} = \alpha_k, \alpha_{d,k+1} = \alpha_d]$ et aller à l'étape 2.

Si α_k est trop grand (suivant un certain critère) : nouvel intervalle :

$[\alpha_{g,k+1} = \alpha_k, \alpha_{d,k+1} = \alpha_d]$ et aller à l'étape 2.

Etape 2 :

Si $\alpha_{d,k+1} = 0$ déterminer $\alpha_{k+1} \in]\alpha_{g,k+1}, +\infty[$

Si $\alpha_{d,k+1} \neq 0$ déterminer $\alpha_{k+1} \in]\alpha_{g,k+1}, \alpha_{d,k+1}[$

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

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

2.5.2 La règle (le critère) d'Armijo

On bute à réduire "de façon importante" la valeur de l'objectif par un pas $x_k \mapsto x_{k+1} = x_k + \alpha_k d_k$ de x_k dans la direction d_k , tel que $f(x_k + \alpha_k d_k) < f(x_k)$.

Or cette condition de décroissance stricte n'est pas suffisante pour minimiser h_k au moins localement. Par exemple, avec la fonction $h : \mathbb{R} \rightarrow \mathbb{R}; x \mapsto h(x) = x^2$ et $x_1 = 2$, les choix $d_k = (-1)^{k+1}$; $\alpha_k = 2 + 3 \times 2^{-(k+1)}$, donnent : $x_k = (-1)^k (1 + 2^{-k})$.

$h(x_k)$ est bien strictement décroissante mais $\{x_k\}_{k \geq 0}$ ne converge pas vers le minimum zéro mais vers 1 (dans cet exemple le pas est trop grand).

La règle d'Armijo ([4]) 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'Armijo* ou *condition de décroissance linéaire*

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla^\top f(x_k) d_k \quad (2,18)$$

Elle est de la forme (2,18), 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 $\psi_\rho(\alpha)$

$$\alpha \mapsto h_k(0) + \rho h'_k(0) \alpha \quad \text{autrement dit } \alpha \mapsto f(x_k) + \rho \alpha_k \nabla^\top f(x_k) d_k$$

2.5.3 La règle (le critère) de Goldstein&Price.

Dans la règle d'Armijo on assure la décroissance de la fonction objectif à chaque pas, mais c'est ne pas suffisant, reprenant l'exemple de la fonction $h : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto h(x) = x^2$ et $x_1 = 2$, et cette fois $d_k = -1$ le choix $\alpha_k = 2^{-(k+1)}$ donnent : $x_k = (1 + 2^{-k})$.

h_k est bien strictement décroissante mais $\{x_k\}_{k \geq 0}$ ne converge pas vers le minimum zéro mais vers 1. Alors que la condition d'Armijo est satisfaite.

Les conditions de Goldstein & Price ([27]) suivant sont, comme on va le prouver, suffisante pour assurer la convergence sous certaines conditions et indépendamment de l'algorithme qui calcule le paramètre.

Etant données deux réels ρ et δ tels que $0 < \rho < \delta < 1$ ces conditions sont

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha \nabla^\top f(x_k) d_k \quad (2,19)$$

$$f(x_k + \alpha d_k) \geq f(x_k) + \delta \alpha \nabla^\top f(x_k) d_k \quad (2,20)$$

2.5.4 La règle (le critère) de Wolfe

La conditions (2,19)-(2,20) "règle de Goldstein&Price" peuvent exclure un minimum ce qui est peut être un inconvénient. Les conditions de wolfe ([49]) n'ont pas cet inconvénient.

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

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha \nabla^\top f(x_k) d_k \quad (2,21)$$

$$\nabla^\top f(x_k + \alpha d_k) d_k \geq \sigma \nabla^\top f(x_k) d_k \quad (2,22)$$

Conditions de Wolfe fortes

Pour certains algorithmes (par exemple le gradient conjugué non linéaire) il est parfois nécessaire d'avoir une condition plus restrictive que (2.22).

Pour cela la deuxième condition (2.22) est remplacée par

$$|\nabla^\top f(x_k + \alpha d_k) d_k| \leq \sigma |\nabla^\top f(x_k) d_k| = -\nabla^\top f(x_k) d_k$$

On aura donc les conditions de Wolfe fortes ([50])

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha \nabla^\top f(x_k) d_k \quad (2,23)$$

$$|\nabla^\top f(x_k + \alpha d_k) d_k| \leq -\sigma \nabla^\top f(x_k) d_k \quad (2,24)$$

Remarque 2.3 [10] La seconde condition (2.22) ou condition de courbure interdit le choix de pas trop petit pouvant entraîner une convergence lente ou prématurée.

Remarque 2.4 [10] On voit bien que les conditions de Wolfe fortes impliquent les conditions de Wolfe faibles. Effectivement (2.23) est équivalente à (2.21), tandis que (2.24)

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

2.6 Convergence des méthodes à directions de descente

2.6.1 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. Ce n'est qu'une contribution, parce que la recherche linéaire ne peut à elle seule assurer la convergence des itérés. On comprend bien que le choix de la direction de descente joue aussi un rôle. Cela se traduit par une condition, dite de Zoutendijk, dont on peut tirer quelques informations qualitatives intéressantes. 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

$$f(x_{k+1}) \leq f(x_k) - C \|\nabla f(x_k)\|^2 \cos^2 \theta_k \quad (2,25)$$

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

$$\cos \theta_k = \frac{-\nabla^\top f(x_k) d_k}{\|\nabla f(x_k)\| \|d_k\|} \quad (2,26)$$

Voici comment on se sert de la condition de Zoutendijk ([10])

Proposition 2.1 Si la suite $\{x_k\}$ générée par un algorithme d'optimisation vérifie la condition de Zoutendijk (2,25) et si la suite $\{f(x_k)\}$ est minorée, alors

$$\sum_{k \geq 1} \|\nabla f(x_k)\|^2 \cos^2 \theta_k < \infty \quad (2,27)$$

Preuve. [25] En sommant les inégalités (2,25), 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 $f(x_k) \geq C'$ ■
 Les deux propositions suivantes précisent les circonstances dans lesquelles la condition de Zoutendijk (2,25) est vérifiée avec les règles d'Armijo et de Wolfe.

Proposition 2.2 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continuellement différentiable dans un voisinage de $\Gamma = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$

On considère un algorithme à directions de descente d_k qui génère une suite $\{x_k\}$ en utilisant la recherche linéaire d'Armijo(2,18) avec $\alpha_1 > 0$

Alors il existe une constante $C > 0$ telle que pour tout $k \geq 1$ l'une des conditions

$$f(x_{k+1}) \leq f(x_k) - C \nabla^\top f(x_k) d_k \quad (2,28)$$

Ou

$$f(x_{k+1}) \leq f(x_k) - C \|\nabla f(x_k)\|^2 \cos^2 \theta_k \quad (2,29)$$

Est vérifiée.

Preuve. [25] Si le pas $\alpha_k = \alpha_1$ est accepté, on a (2,25) car α_1 est uniformément positif.

Dans le cas contraire, (2,18) n'est pas vérifiée avec un pas $\alpha'_k \leq \frac{\alpha_k}{\tau}$ c'est-à-dire

$$f(x_k + \alpha'_k d_k) > f(x_k) + \rho \alpha'_k \nabla^\top f(x_k) d_k$$

Comme f est continuellement différentiable, on a pour tout $\alpha_k > 0$

$$\begin{aligned} f(x_k + \alpha_k d_k) &= f(x_k) + \alpha_k \nabla^\top f(x_k) d_k + \int_0^1 [\nabla f(x_k + t \alpha_k d_k) - \nabla f(x_k)]^\top \alpha_k d_k dt \quad (2,30) \\ &\Rightarrow f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \nabla^\top f(x_k) d_k + C \alpha_k^2 \|d_k\|^2 \end{aligned}$$

Où $C > 0$ est une constante. Avec l'inégalité précédente et le fait que On obtient

$$\begin{cases} f(x_k + \alpha'_k d_k) - f(x_k) > \rho \alpha'_k \nabla^\top f(x_k) d_k \\ f(x_k + \alpha'_k d_k) - f(x_k) \leq \alpha'_k \nabla^\top f(x_k) d_k + C \alpha_k'^2 \|d_k\|^2 \end{cases} \quad (2,31)$$

$$\Rightarrow \rho \alpha'_k \nabla^\top f(x_k) d_k \leq \alpha_k \nabla^\top f(x_k) d_k + C \alpha_k'^2 \|d_k\|^2 \quad (2,32)$$

$$\Rightarrow -C \alpha_k'^2 \|d_k\|^2 \leq (1 - \rho) \alpha'_k \nabla^\top f(x_k) d_k$$

Or

$$\rho < 1 \Rightarrow 0 < 1 - \rho < 1 \Rightarrow \frac{1}{1 - \rho} > 1 \quad (2,33)$$

D'où

$$\nabla^\top f(x_k) d_k \geq \frac{-C}{1 - \rho} \alpha_k' \|d_k\|^2 \Rightarrow -\nabla^\top f(x_k) d_k \leq \frac{C}{1 - \rho} \alpha_k' \|d_k\|^2 \quad (2,34)$$

$$\Rightarrow |\nabla^\top f(x_k) d_k| = \|\nabla^\top f(x_k)\| \|d_k\| \cos \theta_k \leq \frac{C}{1 - \rho} \alpha_k' \|d_k\|^2$$

Ce qui permet de minorer $\alpha' \|d_k\|$ et donc aussi $\alpha \|d_k\|$ par une constante fois $\|\nabla^\top f(x_k)\| \|d_k\|$. Cette minoration et l'expression suivante de (2,18)

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \rho \alpha_k \|\nabla^\top f(x_k)\| \|d_k\| \cos \theta_k \quad (2,35)$$

Conduit à (2,25). ■

Proposition 2.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continuellement différentiable dans un voisinage de

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

On considère un algorithme à directions de descente d_k qui génère une suite $\{x_k\}$ en utilisant la recherche linéaire de Wolfe (2,23)-(2,24)

Alors il existe une constante $C > 0$ telle que, pour tout $k \geq 1$ la condition de Zoutendijk (2,25) est vérifiée.

Preuve. [25] D'après (2,22)

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

Et du fait que f est continuellement différentiable

$$\begin{aligned} (1 - \sigma) |\nabla^\top f(x_k) d_k| &= (1 - \sigma) \|\nabla^\top f(x_k)\| \|d_k\| \cos \theta_k \\ &\leq \|\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k)\| \|d_k\| \\ \Rightarrow (1 - \sigma) \|\nabla^\top f(x_k)\| \cos \theta_k &\leq L \alpha_k \|d_k\| \\ \Rightarrow \alpha_k \|d_k\| &\leq \frac{(1 - \sigma)}{L} \|\nabla^\top f(x_k)\| \cos \theta_k \end{aligned}$$

En utilisant (2,23) on aura

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \rho \alpha_k \nabla^\top f(x_k) d_k \\ \Rightarrow f(x_k + \alpha_k d_k) &\leq f(x_k) + \rho \alpha_k \nabla^\top f(x_k) d_k \leq f(x_k) + |\rho \alpha_k \nabla^\top f(x_k) d_k| \\ \Rightarrow f(x_k + \alpha_k d_k) &\leq f(x_k) + |\rho \alpha_k \nabla^\top f(x_k) d_k| \leq f(x_k) - \rho \alpha_k \nabla^\top f(x_k) d_k \\ \Rightarrow f(x_k + \alpha_k d_k) &\leq f(x_k) - \rho \alpha_k \|\nabla^\top f(x_k)\| \|d_k\| \cos \theta_k \\ \Rightarrow f(x_k + \alpha_k d_k) &\leq f(x_k) - \frac{\rho(1 - \sigma)}{L} \|\nabla^\top f(x_k)\|^2 \cos^2 \theta_k \end{aligned}$$

On en déduit (2,25). ■

Chapitre 3

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

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

3.1 Méthodes du gradient conjugué non linéaire

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

$$\min \{f(x) : x \in \mathbb{R}^n\} \tag{3,1}$$

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

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

$$x_{k+1} = x_k + \alpha_k d_k \tag{3,2}$$

Où la direction recherchée est définie par la formule de récurrence suivante

$$d_k = \begin{cases} -g_0 & \text{si } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 1 \end{cases} \quad (3,3)$$

Le coefficient β_k détermine la méthode du gradient conjugué en question (Fletcher Reeves, Polak Ribière Polyak,...). Le pas $\alpha_k \in \mathbb{R}^+$ étant déterminé par une recherche linéaire. ([31])

L'hypothèse 3.1.1 [10]

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

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

$$\exists l > 0 \text{ tel que } \|g(x) - g(\tilde{x})\| \leq l \|x - \tilde{x}\| \quad \forall x, \tilde{x} \in \mathcal{N} \quad (3,4)$$

Remarque 3.1 [10] Notons que ces suppositions impliquent qu'il existe $\lambda > 0$ tel que

$$\|g(x)\| \leq \lambda \quad \forall x \in L \quad (3,5)$$

Définition 3.1 [31] On dit que d_k est une direction de descente suffisante si

$$d_k^\top g_k \leq -C \|g_k\|^2 \quad \text{où } C > 0 \quad (3,6)$$

Rappelons les conditions de Wolfe faibles

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha_k d_k^\top g_k \quad (3,7)$$

$$d_k^\top g_{k+1} \geq \sigma d_k^\top g_k \quad (3,8)$$

Où $0 < \rho < \sigma < 1$

Les conditions de Wolfe fortes

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \rho \alpha_k d_k^\top g_k \\ |d_k^\top g_{k+1}| &\leq -\sigma d_k^\top g_k \end{aligned} \quad (3,9)$$

Où $0 < \rho < \sigma < 1$

Les conditions de Wolfe relaxée

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \rho \alpha_k d_k^\top g_k \\ \sigma_1 d_k^\top g_k &\leq d_k^\top g_{k+1} \leq -\sigma_2 d_k^\top g_k \end{aligned} \quad (3,10)$$

Où $0 < \rho < \sigma_1 < 1$ et $\sigma_2 > 0$

Présentons maintenant un théorème fondamental qui assure la satisfaction de la condition de Zoutendijk, dans laquelle le pas α_k est déterminé par la règle de Wolfe faible. Ce théorème était démontré par Zoutendijk ([53, 1970]) et Powell ([40, 1971]).

Théorème 3.1 *Considérons une méthode du type (3.2) et (3.3) dans laquelle d_k est une direction de descente et le pas α_k est déterminé par la règle de Wolfe faible (3.7)-(3.8) avec $\sigma \in]0, \frac{1}{2}[$. Considérons aussi que l'hypothèse 3.1.1 soit satisfaite. Alors pour une telle méthode la condition de Zoutendijk suivante*

$$\sum_{k \geq 1} \cos^2 \theta \|g_k\|^2 < \infty \quad (3,11)$$

Est vérifiée.

Preuve. ([53, 1970])

De (3.8) on a

$$d_k^\top (g_{k+1} - g_k) \geq (\sigma - 1) d_k^\top g_k$$

D'autre part

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

D'où

$$\alpha_k \geq \left(\frac{\sigma - 1}{L} \right) \frac{d_k^\top g_k}{\|d_k\|^2}$$

En remplaçant ceci dans (3.7) on aura

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

Où $\mu = \frac{\rho(\sigma-1)}{L}$

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

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

Ce qui achève la démonstration. ■

3.2 Méthode de Fletcher-Reeves

La méthode de Fletcher-Reeves ([22, 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.

Pour la méthode de Fletcher-Reeves la variante β_k est

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad (3,12)$$

3.2.1 Algorithme de la méthode de Fletcher-Reeves

Cette méthode est très intéressante, d'une part parce qu'elle nécessite le stockage de très peu d'informations (essentiellement trois vecteurs de dimension n) d'autre part, par sa vitesse de convergence très supérieure à celle des algorithmes du gradient classique([31])

Algorithme 4.1 Méthode de Fletcher-Reeves

Etape0 : (initialisation)

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

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

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

Sinon aller à l'étape 2.

Etape 2 :

Définir $x_{k+1} = x_k + \alpha_k d_k$ avec : $\alpha_k = \min_{\alpha > 0} f(x_k + \alpha d_k)$

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

Où

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

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

3.2.2 La propriété de descente de la méthode de FR

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

J.C. Gilbert et Nocedal ([24, 1992]) ont généralisé ce résultat pour toute méthode du type (3.2)-(3.3) dont

$$|\beta_k| \leq \beta_k^{FR} \tag{3,13}$$

Théorème 3.2 *Supposons que l'hypothèse 3.1.1 soit satisfaite.*

Considérons une méthode du type (3.2) et (3.3) dont β_k satisfait à (3.13) et le pas α_k satisfait à la règle de Wolfe forte (3.9) où $\sigma \in]0, \frac{1}{2}[$, Alors cette méthode génère des directions de descente.

De plus on a

$$\frac{-1}{1-\sigma} \leq \frac{d_k^\top g_k}{\|g_k\|^2} \leq \frac{2\sigma-1}{1-\sigma} \quad k = 1, \dots \quad (3,14)$$

Preuve. ([24, 1992])

La démonstration se fait par récurrence

1) Pour $k = 1$

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

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 (3.14) est satisfaite pour $k > 1$ et démontrons pour $k + 1$

Supposons que

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

On a

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

D'autre part de (3.12) on aura

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

D'où

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

En utilisant la condition de recherche linéaire (3.9) on aura

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

Remplaçant ceci dans (**)

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

De (*) on aura

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

Et de (3.13)

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

On aura

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

Ce qui achève la démonstration. ■

Corollaire 3.1 [31] D'après (3.14) on déduit que pour tout $k \geq 1$, d_k est une directions de descente suffisante avec $C = \frac{1-2\sigma}{1-\sigma} > 0$.

En effet

$$\begin{aligned} \frac{-1}{1-\sigma} &\leq \frac{d_k^\top g_k}{\|g_k\|^2} \leq \frac{2\sigma-1}{1-\sigma} \\ \Rightarrow d_k^\top g_k &\leq -C \|g_k\|^2 \text{ où } C = \frac{1-2\sigma}{1-\sigma} \end{aligned}$$

Remarque 3.2 [31] 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

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

Puisque

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

Donc α_k vérifie la condition nécessaire d'optimalité

$$h'_k(\alpha_k) = d_k^\top \nabla f(x_k + \alpha_k d_k) = d_k^\top g_{k+1} = 0 \quad \forall k \geq 1$$

3.2.3 Convergence de la méthode de Fletcher-Reeves

Le premier résultat de convergence de la méthode du gradient conjugué non linéaire (version Flecher Reeves) avec des recherches linéaires inexactes (recherche linéaire inexacte de Wolfe forte (3.9) où $\sigma < \frac{1}{2}$) était démontré par Al-Baali ([1, 1985]).

Touati Ahmed et Story ([48, 1990]) ont généralisé ce résultat pour $0 \leq \beta_k \leq \beta_k^{FR}$.

Gilbert et Nocedal ([24, 1992]) ont généralisé ce résultat pour $|\beta_k| \leq \beta_k^{FR}$.

Théorème 3.3 Supposons que l'hypothèse 3.1.1 soit satisfaite. Considérons une méthode du type (3.2) et (3.3) dont β_k satisfait à (3.13) et le pas α_k satisfait aux conditions de Wolfe fortes (3.9) où $\sigma \in]0, \frac{1}{2}[$. Alors cette méthode est globalement convergente, dans le sens suivant

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0 \tag{3,15}$$

Preuve. ([24, 1992]) Puisque les conditions du théorème 3.2 sont satisfaites alors on a

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

D'autre part de (3.9)

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

D'où

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

De (3.2), (3.13) et (***)

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

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

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

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

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

De (3.5) on a

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

D'où

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

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

D'autre part, puisque les conditions du Théorème 3.1. et du théorème 3.2 sont satisfaites on a

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

Et

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

D'où

$$\sum_{k \geq 1} c_1^2 \frac{\|g_k\|^2}{\|d_k\|^2} \|g_k\|^2 \leq \sum_{k \geq 1} \cos^2 \theta_k \|g_k\|^2 < \infty$$

$$\Rightarrow \sum_{k \geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty$$

$$\Rightarrow \sum_{k \geq 1} \frac{\omega^4}{\|d_k\|^2} < \infty$$

$$\Rightarrow \sum_{k \geq 1} \frac{1}{\|d_k\|^2} < \infty$$

Ce qui contredit (***) , d'où le résultat ([24, 1992])

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

■

Remarque 3.3 Lui, Han et Yuan ([32, 1995]) ont prouvé le résultat (3.15) pour cette méthode si $\rho = \sigma = \frac{1}{2}$. Dai et Yuan([17, 1996]) l'ont encore démontré mais par une approche plus simple.

Remarque 3.4 La méthode de Fletcher-Reeves possède de bon propriété théoriquement, mais en pratique, elle converge parfois lentement, et même prématurément. En effet, si les pas sont trop petits, il se peut que ce comportement s'élargit pour un grand nombre d'itérations, et c'est ce qu'il nous oblige à réinitialiser en posant $\beta_k^{FR} = 0$.

Powell ([41, 1977]) est le premier qui a observé ce comportement, ainsi il a donné un contre exemple avec une recherche linéaire exacte. On évite cet inconvénient en réinitialisant chaque n itération par exemple.

Nemirovsky et Yudin ([37, 1983]) ont démontré par un contre exemple que la méthode de Fletcher-Reeves converge plus lentement que la méthode de Steepest descent. (cas d'une fonction quadratique avec recherche linéaire exacte).

3.3 Méthode de Polak-Ribière-Polyak

Cette méthode fut découverte par Polak, Ribière ([38, 1969]) et Polyak ([39, 1969]).

Pour la méthode de Polak-Ribière-Polyak la variante β_k est :

$$\beta_k^{PRP} = \frac{g_k^\top y_{k-1}}{\|g_{k-1}\|^2} \quad (3,16)$$

3.3.1 Algorithme de la méthode de Polak-Ribière-Polyak

Algorithme 4.2 Méthode de Polak-Ribière-Polyak[31]

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

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

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

Où

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

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

3.3.2 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.

Pour remédier à cet inconvénient Powell ([43, 1986]) a modifié le choix de β_k^{PRP} afin d'assurer la convergence avec une recherche linéaire exacte.

Gilbert et Nocedal ([24, 1992]) ont proposé une nouvelle méthode qui converge dans le sens (3,15) avec ou bien une recherche linéaire exacte ou inexacte.

Le résultat suivant est du à Polak et Ribière ([38, 1969]).

Théorème 3.4 Si f est fortement convexe, continûment différentiable avec un gradient lipchitzien, 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 .

Preuve. [43] Montrons dans un premier temps que

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

Est uniformément positif. Grâce à la recherche linéaire exacte, on a

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

La forte convexité de f implique que

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

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

On en déduit, en utilisant la constante de lipchitz L de g

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

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

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

Ensuite

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

Ou encore

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

D'après la recherche linéaire exacte, la condition de Zoutendijk est vérifié.

Mais f et donc $\{f(x_k)\}$ est bornée inférieurement (car f est fortement convexe).

On en déduit que $g_k \rightarrow 0$. 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^* . ■

Remarque 3.5 [31] Si f n'est pas convexe, la méthode de Polak-Ribière-Polyak peut ne pas converger. Powell ([42, 1984]) a donné un exemple de fonction (de 3 variables, deux fois continûment différentiable) pour laquelle l'algorithme génère une suite $\{x_k\}$ dont aucun des points d'adhérence n'est stationnaire.

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

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

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

Remarque 3.6 Gilbert et Nocedal ([24, 1992]) ont assuré la convergence avec une recherche linéaire exacte ou inexacte, en hybridant les méthodes de FR et de PRP. Donc cette nouvelle méthode consiste à prendre la variante β_k comme suit

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

Remarque 3.7 Grippo et Lucidi ([26, 1997]) et P. Armand ([3, 2005]) ont suggéré des modifications dans le choix de α_k afin d'établir le résultat de la convergence.

3.4 Méthode de descente conjuguée

Cette méthode fut proposée en 1987 par Fletcher et Reeves([21, 1987]).

Pour la méthode de la descente conjuguée la variante β_k est

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}} \quad (3,17)$$

3.4.1 Algorithme de la méthode de descente conjuguée

Algorithme 4.3 Méthode de la descente conjuguée[10]

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

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

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

Où

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

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

3.4.2 La propriété descente de la méthode de descente conjuguée

Fletcher ([21, 1987]) a démontré que la méthode de la descente conjuguée est une méthode de descente si le pas α_k est déterminé par la règle forte de Wolfe (3.9) avec $\sigma \leq \frac{1}{2}$.

Dai et Yuan ([17, 1996]) ont démontré que cette méthode avec la règle de Wolfe relaxée (3.10) où $0 < \rho < \sigma_1 < 1$ et $0 \leq \sigma_2 \leq 1$, génère des directions de descente suffisante à chaque itération $k \geq 1$.

Théorème 3.5 Pour toute méthode du type (3.2) et (3.3) dont β_k satisfait à (3.17) et le pas α_k satisfait aux conditions de Wolfe relaxée (3.10)

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

Où $0 < \rho < \sigma_1 < 1$ et $0 \leq \sigma_2 \leq 1$.

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

Preuve. [17, 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 (3.10)

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

D'où

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

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

$$d_k^\top g_k \leq -C \|g_k\|^2 \quad \text{où } C = 1 - \sigma_2 > 0$$

Et donc d_k est une direction de descente suffisante. ■

3.4.3 Convergence de la méthode de descente conjuguée

Yuan ([52, 1993]) a démontré la convergence au sens (3.15) de cette méthode avec un pas satisfaisant aux conditions (3.7)-(3.10) si $\sigma_1 < \frac{1}{2}$ et $\sigma_2 = 0$.

Dai et Yuan ([17, 1996]) ont démontré ce résultat pour $\sigma_1 < 1$ et $\sigma_2 = 0$.

Théorème 3.6 *Supposons que la l'hypothèse 3.1.1 est satisfaite. Toute méthode du type (3.2) et (3.3) dans laquelle β_k vérifie (3.17) et le pas α_k est déterminé par la règle de Wolfe relaxée (3.10) où $0 < \rho < \sigma_1 < 1$ et $\sigma_2 = 0$, est de descente convergente, dans le sens suivante*

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

Preuve. [17,1996] Du théorème 3.2 on a

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

Donc β_{k+1}^{CD} vérifie l'inégalité (3.13), D'après le théorème 3.3 on a

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

■

3.5 Méthode de Dai-Yuan

Cette méthode était découverte par Dai et Yuan ([14, 1999]).

Pour la méthode de la descente conjugué la variante β_k est

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^\top [g_k - g_{k-1}]} = \frac{\|g_k\|^2}{d_{k-1}^\top y_{k-1}} \quad (3,18)$$

Remarquons que cette variante a les mêmes numérateur et dénominateur de les variantes de Fletcher-Reeves et Hestenes-Stiefel respectivement.

Cette méthode possède plusieurs propriétés, par exemple elle possède la propriété de descente à chaque itération.

3.5.1 Algorithme de la méthode de Dai-Yuan

Algorithme 4.4 Méthode de Dai-Yuan[10]

Etape 0 : (initialisation)

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

Poser $k = 0$ et aller à l'étape 1.

Etape 1 :

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

Si non aller à l'étape 2.

Etape 2 :

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

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

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

Où

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

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

Algorithme 4.5 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 $x_{k+1} = x_k + \alpha_k d_k$ avec : α_k vérifie les conditions (3.7)-(3.8)

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

Où

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

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

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

Dai et Yuan ([15, 1998]) ont démontré qu'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 ([14, 1999]) s'ont généralisé ce résultat pour toute fonction régulière.

Théorème 3.7 *Supposons que l'hypothèse 3.1.1 soit satisfaite.*

Pour toute méthode du type (3.2) et (3.3) dont β_k satisfait à (3.18) et le pas α_k satisfait aux conditions de Wolfe faible :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho d_k^\top g_k$$

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

Où $0 < \rho < \sigma < 1$ Alors toutes les directions générées sont de descente, autrement dit

$$d_k^\top g_k < 0 \quad \forall k \geq 1 \tag{3.1}$$

Preuve. [14, 1999] La démonstration se fait par récurrence.

1) Pour $k = 1$

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

2) Supposons que (3.19) est satisfaite pour $k > 1$ et démontrons qu'elle le sera pour $k + 1$
 1Supposons que

$$d_k^\top g_k < 0 \quad k > 1$$

En utilisant (3.7) on aura

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

D'autre part

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

Or puisque $d_k^\top g_k < 0$, $d_k^\top y_k > 0$ il en résulte

$$d_{k+1}^\top g_{k+1} < 0$$

Ce qui achève la démonstration. ■

3.5.3 Convergence de la méthode de Dai-Yuan

Dai-Yuan ([14, 1999]) ont démontré la convergence de la méthode de Dai-Yuan au sens (3.15) si le pas α_k est déterminé par la règle de wolfe faible.

Théorème 3.8 [14] *Supposons que l'hypothèse 3.1.1 est satisfaite. La suite $\{x_k\}$ générée par l'algorithme converge dans le sens*

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

Chapitre 4

Convergence global de la méthode CGBB à direction de descente pour l'optimisation sans contrainte

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

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

Comme on l'a vu avant les différentes méthodes du gradient conjugué génèrent des suites $\{x_k\}_{k \in \mathbb{N}}$ de la façon suivante

$$x_{k+1} = x_k + \alpha_k d_k \quad (4,1)$$

Le pas $\alpha_k \in \mathbb{R}$ est déterminé par une optimisation unidimensionnelle ou recherche linéaire exacte ou inexacte, tel que la longueur de l'étape $\alpha_k > 0$ est calculé en effectuant certaines recherches linéaire. La recherche de Wolfe forte est de trouver une longueur de pas positif α_k telle que

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^\top d_k \quad (4,2)$$

$$\left| g(x_k + \alpha_k d_k)^\top d_k \right| \leq -\sigma g_k^\top d_k \quad (4,3)$$

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

$$d_k = \begin{cases} -g_1 & \text{si } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 2 \end{cases} \quad (4,4)$$

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

Les différentes valeurs attribuées à β_k définissent les différentes formes du gradient conjugué.

Nous avons vu dans les paragraphes précédents comment les résultats de convergences des différentes méthodes de gradient conjugué dépendaient essentiellement des scalaires β_k et ces résultats ont été étudiés séparément suivant les différentes valeurs de β_k , voir El-Baali ([1, 1985]), Gilbert et Nocdall ([24, 1992]), Grippo et Lucidi ([26, 1997]), Powell ([43, 1986]), Touati-Ahmed et Storey ([48, 1990]), Powell ([42, 1984]).

La question naturelle qui se pose est la suivante :

Peut-on obtenir des résultats de convergence avec ces variantes du Gradient Conjugué ou d'autres qui sont non liées à β_k en utilisant des recherches linéaires inexactes de Wolfe fortes ?

La réponse a été donnée positivement dans le travail récent de M. Belloufi et R. Benzine ([7, 2014]), qui a proposé des modifications du gradient conjugué de la forme (0.3).

Ce paragraphe détaille cette étude.

4.1 Algorithme de la méthode CGBB

[7] Nous sommes maintenant en mesure d'introduire la nouvelle Algorithme.

Dans cette section, nous donnons le spécifique de la nouvelle méthode de gradient conjugué proposé. Comme l'a signalé devant nos directions de recherche d_k^{BB} sont définis comme suit

$$d_k^{BB} = \begin{cases} -\frac{g_k}{\|g_k\|^2} & \text{si } k = 1 \\ -\frac{1}{\|g_k\|^2}g_k + d_{k-1} & \text{si } k \geq 2 \end{cases} \quad (4,5)$$

L'algorithme est donné comme suit

Algorithme 1.[7]

Étape 0 : Donnée $x_1 \in \mathbb{R}^n$, posé $d_1^{BB} = -\frac{g_1}{\|g_1\|^2}$, $k := 1$.

Étape 1 : Si $\|g_k\| = 0$ puis arrêter sinon passer à l'Étape 2.

Étape 2 : posé $x_{k+1} = x_k + \alpha_k d_k^{BB}$ où d_k^{BB} est défini par (4.4), et α_k est défini par (4.2) et (4.3).

Étape 3 : posé $k := k + 1$ et passez à l'Étape 1.

4.2 La propriété de descente de la méthode CGBB

Le théorème suivant indique que, si α_k est calculée par la recherche de linéaire de Wolfe fortes (4.2) et (4.3), puis la direction de recherche d_k^{BB} satisfait la propriété de descente.

Théorème 4.1 Pour toute méthode du type (4.2) et (4.3) dont le pas α_k satisfaisant aux conditions de Wolfe fortes où $\delta < \sigma < \frac{1}{2}$.

Alors les directions générées par cete méthode sont des directions de descente autrement dit

$$d_k^\top g_k < 0 \quad \forall k \geq 1 \quad (4,6)$$

De plus on a

$$-\sum_{j=0}^{k-1} \sigma^j \leq g_k^\top d_k \leq -2 + \sum_{j=0}^{k-1} \sigma^j$$

Preuve. [7] La démonstration se fait par récurrence.

1) Pour $k = 1$

$$d_1^\top g_1 = -\|g_1\|^2$$

2) Supposons que (4,6) et (4.7) est satisfaite pour $k \geq 1$ et démontrons qu'elle le sera pour $k + 1$
Supposons que

$$-\sum_{j=0}^{k-1} \sigma^j \leq g_k^\top d_k \leq -2 + \sum_{j=0}^{k-1} \sigma^j$$

On a

$$g_{k+1}^\top d_{k+1} = -1 + g_{k+1}^\top d_k \quad (4,8)$$

De (4.3) et (4.6)

$$-1 + \sigma g_k^\top d_k \leq g_{k+1}^\top d_{k+1} \leq -1 - \sigma g_k^\top d_k \quad (4,9)$$

D'autre part de (4.7), on aura

$$\begin{aligned} -\sum_{j=0}^k \sigma^j &= -1 - \sigma \sum_{j=0}^{k-1} \sigma^j \leq g_{k+1}^\top d_{k+1} \\ &\leq -1 + \sigma \sum_{j=0}^{k-1} \sigma^j = -2 + \sum_{j=0}^k \sigma^j \end{aligned}$$

D'où

$$-\sum_{j=0}^k \sigma^j \leq g_{k+1}^\top d_{k+1} \leq -2 + \sum_{j=0}^k \sigma^j$$

De

$$g_{k+1}^\top d_{k+1} \leq -2 + \sum_{j=0}^k \sigma^j \quad (4,10)$$

Et

$$\sum_{j=0}^k \sigma^j < \sum_{j=0}^{\infty} \sigma^j = \frac{1}{1 - \sigma} \quad (4,11)$$

Où $\sigma \in]0, \frac{1}{2}]$

Donc $1 - \sigma > \frac{1}{2}$ d'où $-2 + \sum_{j=0}^k \sigma^j < 0$

Par conséquent, en utilisant (4.10) nous avons

$$g_{k+1}^\top d_{k+1} < 0$$

Ce qui achève la démonstration. ■

4.3 Convergence de la méthode BB

Afin d'établir la convergence globale de la méthode proposée, on suppose que l'hypothèse suivante est satisfaite

Hypothèse 4.3.1 [7] Soit f deux fois continûment différentiable et l'ensemble

$$L := \{x \in \mathbb{R}^n, f(x) \leq f(x_1)\} \text{ est borné.}$$

Théorème 4.2 Supposons que l'hypothèse 4.3.1 soit satisfaite.

Considérons une méthode du type (4,1) et (4,4) avec le pas α_k satisfaisant aux conditions de Wolfe fortes (4,2) et (4,3) où $\delta < \sigma < \frac{1}{2}$. Alors cette méthode est globalement convergente, dans le sens suivant

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0 \quad (4,12)$$

Preuve. [7] Il est indiqué dans le théorème (4.1) que la propriété de descente (4.6) est satisfaite pour $\sigma \in \left]0, \frac{1}{2}\right]$, d'autre part de (4,3), (4,7) et (4.11) on aura

$$|g_k^\top d_{k-1}| \leq -\sigma g_{k-1}^\top d_{k-1} \leq \sigma \sum_{j=0}^{k-2} \sigma^j = \sum_{j=0}^{k-1} \sigma^j \leq \frac{\sigma}{1-\sigma} \quad (4,13)$$

Ainsi, à partir de la définition de d_k et en utilisant (4.13) on en déduit que

$$\begin{aligned} \|d_k\|^2 &= \frac{1}{\|g_k\|^2} - \frac{2}{\|g_k\|^2} g_k^\top d_{k-1} + \|d_{k-1}\|^2 \\ &\leq \frac{1}{\|g_k\|^2} + \frac{2\sigma}{1-\sigma} \frac{1}{\|g_k\|^2} + \|d_{k-1}\|^2 \\ &= \left(\frac{1+\sigma}{1-\sigma}\right) \frac{1}{\|g_k\|^2} + \|d_{k-1}\|^2 \end{aligned} \quad (4,14)$$

Par conséquent on aura

$$\begin{aligned} \|d_k\|^2 &\leq \left(\frac{1+\sigma}{1-\sigma}\right) \sum_{j=2}^k \frac{1}{\|g_j\|^2} + \frac{1}{\|g_1\|^2} \\ &\leq \left(\frac{1+\sigma}{1-\sigma}\right) \sum_{j=1}^k \frac{1}{\|g_j\|^2} \end{aligned} \quad (4,15)$$

D'où

$$\frac{1}{\|g_1\|^2} \leq \left(\frac{1+\sigma}{1-\sigma} \right) \frac{1}{\|g_j\|^2}$$

Nous prouvons maintenant (4.12) par l'absurde, Supposons maintenant que (4.12) n'est pas satisfaite, autrement dit : il existe une constante $\varepsilon > 0$ telle que

$$\|g_k\| \geq \varepsilon > 0 \quad (4.16)$$

Comme g_k est borné ci-dessus sur l'ensemble de niveau L et appliquons (4.15) on aura :

$$\|d_k\|^2 \leq c_1 k \quad (4.17)$$

Où c_1 est une constante positive, De (4.7) et (4.11) on obtient

$$\begin{aligned} \cos \theta_k &= -\frac{g_k^\top d_k}{\|g_k\| \|d_k\|} \geq \left(2 - \sum_{j=0}^{k-1} \sigma^j \right) \frac{1}{\|g_k\| \|d_k\|} \\ &\geq \left(\frac{1-2\sigma}{1-\sigma} \right) \frac{1}{\|g_k\| \|d_k\|} \end{aligned} \quad (4.18)$$

En remplaçant (4.17) et (4.16) dans (4.18) on aura

$$\sum_k \cos^2 \theta_k \geq \left(\frac{1-2\sigma}{1-\sigma} \right) \sum_k \frac{1}{\|g_k\|^2 \|d_k\|^2} \geq c_2 \sum_k \frac{1}{k} \quad (4.19)$$

Où c_2 est une constante positive. Par conséquent, la série $\sum_k \cos^2 \theta_k$ est divergent.

Soit M la borne supérieure de $\|\nabla^2 f(x)\|$ sur l'ensemble de niveau L

$$g_{k+1}^\top d_k = (g_k + a_k \nabla^2 f(x))^\top d_k \leq g_k^\top d_k + M a_k \|d_k\|^2$$

En utilisant (4.3), on aura

$$a_k \geq -\frac{(1-\sigma)}{M \|d_k\|^2} g_k^\top d_k \quad (4.20)$$

D'autre part de (4.20) et (4.2), on a

$$\begin{aligned} f_{k+1} &\leq f_k - \frac{(1-\sigma)\delta}{M} \left(\frac{g_k^\top d_k}{\|d_k\|} \right)^2 \\ &= f_k - c_3 \|g_k\|^2 \cos^2 \theta_k \end{aligned}$$

Où $c_3 = \frac{(1-\sigma)\delta}{M} > 0$. D'autre part, puisque f est borné au-dessous alors on a

$$\sum_k \|g_k\|^2 \cos^2 \theta_k \text{ converge}$$

Ce qui veut dire que $\sum_k \cos^2 \theta_k$ est converge. Ce qui contredit (4.19), d'où le résultat. Donc

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

■

4.4 Résultats et discussion numérique

Dans cette partie nous allons effectuer des tests numériques pour montrer l'efficacité et la performance du nouvel algorithme CGBB muni de la recherche linéaire Wolfe fortes.

Une comparaison a été réalisée avec les autres méthodes de CG y compris BB, CG_DESCENT, PRP et FR. Nous sommes considérés $\varepsilon = 10^{-6}$ et le critère d'arrêt est $\|g_k\| \leq 10^{-6}$.

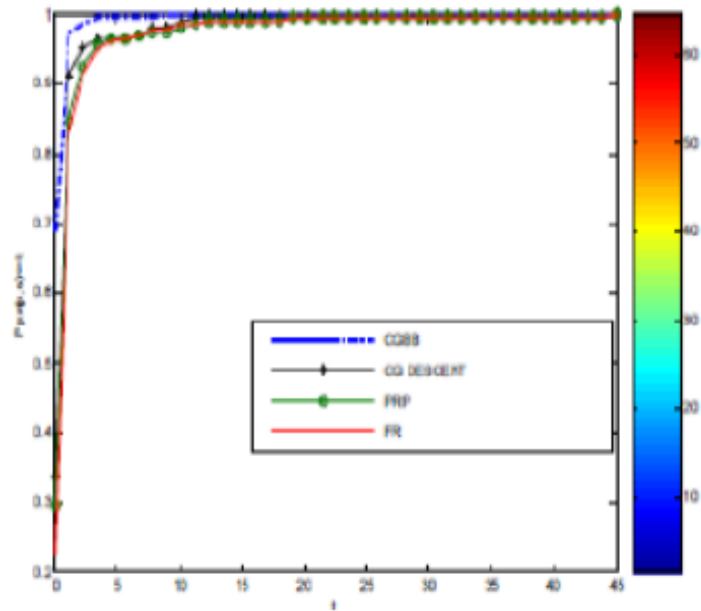


Fig.4.1-Performance basée sur le temps CPU.

4.4.1 Les profils de performance numérique

Les profils de performance numérique de la méthode du gradient conjugué BB, CG_DESCENT, PRP et FR basés sur le temps CPU.

Conclusion

Dans ce mémoire nous avons présenté une synthèse générale de la méthode de minimisation multi-dimensionnelle sans contrainte en particulier la méthode à direction de descente comme la méthode de la descente conjuguée avec les recherches linéaires exactes et inexacts, la méthode de Newton pour obtenir la convergence global (local) des méthodes précédentes.

Ensuite, nous avons étudiée une nouvelle famille des méthodes de gradient conjugué non linéaire, la propriété de descente est assurée à chaque itération et certains résultats généraux de convergence sont montrés pour cette famille avec la recherche linéaire de Wolfe fortes.

Il est clair d'après la figure. 4.1, que la méthode **CGBB** est numériquement plus performante que les autres méthodes du gradient conjugué **CG_DESCENT**, **PRP** et **FR**

Bibliographie

- [1] M. Al-Baali (1985), Descent property and global convergence of the Fletcher-Reeves method with inexact line search. *IMA J. Num. Anal.*, Vol. 5, pp.121-124.
- [2] Morelle Albin 2010 ,Morelle introduction à l'algorithmique , ESIEE Paris Mars 2010
- [3] 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.
- [4] L. Armijo, Minimization of function having Lipschitz continuous first partial derivatives, *Pacific Journal of Mathematics*, Vol. 16 (1966), pp.1-3.
- [5] M. Barrault et C.Le. Bris (1999), Optimisation Numérique et Différentiation Automatique pour un Problème Industriel, Notes de cours, Ecole Nationale de Ponts et Chaussées, Paris.
- [6] M.S.Bazaraa, H.D.Sherali et C.M.Shett, *Nonlinear Programming, Theory and Algorithms*, Wiley-Interscience,(1993).
- [7] M. Belloufi and Benzine Rachid, Descent property and global convergence of a new search direction method for unconstrained optimization. In *Numerical Functional Analysis and Optimization* , Vol 36, No(2), 169-180, Taylor & Francis Group, 2014, <http://www.tandfonline.com/doi/abs/10.1080/01630563.2014.976796>.
- [8] M. Belloufi, *Développements récents de la méthode du gradient conjugué*, Thèse Doctorat, Université Badji Mokhtar Annaba (2014).
- [9] M. Bergounioux, *Optimisation et contrôle des systèmes linéaires*, Dunod, Paris, 2001.
- [10] T.Bouli, *Convergence des méthodes du gradient conjugué avec la recherche linéaire non monotone*, Université Badji Mokhtar Annaba, 2014-2015
- [11] A. Cauchy, *Analyse mathématique, Méthode générale pour la résolution des systèmes d'équations simultanées*, *Comptes Rendus de l'Académie des Sciences de Paris*, (1847) t-25, pp. 536-538.

-
- [12] T.Cluzeau, Mathématiques pour l'Ingénieur école Nationale Supérieure d'Ingénieurs de Limoges 87068 Limoges CEDEX.
- [13] Y. H. Dai and L. Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.*, 43(2001), pp. 87–101.
- [14] 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.
- [15] 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.
- [16] J. W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, *SIAM J. Numer. Anal.*, 4 (1967), pp. 10–26.
- [17] Y. H. Dai and Y. Yuan (1996), Convergence properties of the conjugate descent method, *Advances in Mathematics*, 6, pp.552-562.Y.
- [18] Y. H. Dai and Y. Yuan, Nonlinear conjugate gradient methods, Shanghai Science and Technology Publisher, Shanghai, 2000.
- [19] Daniele Di Pietro optimisation numérique conditions d'existence et unicité A.A,2012-2013.
- [20] R. P. Fedorenko, A relaxation method for solving elliptic difference equation. U.S.S.R.Comput. Math. And Math. Phys., (1962).
- [21] R. Fletcher (1987), Practical methods of optimization, *John Wiley&Sons, Chichester*.
- [22] R. Fletcher and C. Reeves (1964), Function minimization by conjugate gradients. *Comput. J.*, 7, pp.149-154.
- [23] H. Fan, Z. Zhu et A. Zhou, A New Descent Nonlinear Conjugate Gradient Method for Unconstrained Optimization, *Applied Mathematics*, 2011, 2, 1119-1123.
- [24] 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.
- [25] J. C. Gilbert (2007), Eléments d'Optimisation Différentiable : Théorie et Algorithmes, *Notes de cours, École Nationale Supérieure de Techniques Avancées, Paris*.
- [26] L. Grippo and S. Lucidi (1997), A Globally Convergent Version of the Polak-Ribière Conjugate Gradient Method, *Mathematical Programming*, Vol. 78, pp. 375–391.
- [27] A. A. Goldstein, On steepest descent, *SIAM J. on Control A*, Vol. 3, No. 1 (1965), pp. 147-151.
- [28] Frédéric de Gournay & Aude Rondepierre , Introduction à l'Optimisation Numérique DEPARTEMENT STPI , 3EME ANNEE MIC

- [29] W. W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *November 17, (2003)*.
- [30] M. R. Hestenes and E.L. Stiefel (1952), Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards Sect.*, 5(49), pp. 409-436
- [31] JAHIDA ,Résultats de convergence des méthodes du gradient conjugué non linéaire avec la recherche linéaire inexacte,Université de Tebessa,2014-2015
- [32] 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.
- [33] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, Part 1 : Theory, *J. Optim. Theory Appl.*, 69 (1991), pp. 129–137.
- [34] O. L. Mangasarian, R. R. MEYER and S. M. Robinson, Convergence properties of minimization algorithms, in *Nonlinear Programming*, 2 eds. *Academic Press, New York, 1975*, pp. 1-27.
- [35] M. Minoux (1983), *Programmation Mathématiques, Théorie et Algorithmes*, tome 1 , Dunod.
- [36] J.J.Moré, B.S.Garbow and K.E. Hillstrom, Testing unconstrained optimization software, *ACM Transactions on Mathematical Software* 7(1981) 17–41.
- [37] A.S. Nemirovsky et D.B. Yudin (1983), *Problem Complexity and Method Efficiency*, *New York, Wiley*.
- [38] E. Polak and G. Ribière (1969), Note sur la convergence de directions conjuguées, *Rev. Française Informat. Recherche Operationelle*, 3e année, 16, pp. 35-43.
- [39] B.T. Polyak (1969), The conjugate gradient method in extremem problems, *Comput. Math. Math. Phys.*, 9, pp. 94-112.
- [40] M.J.D. Powell (1971), On the convergence of the variable metric algorithm ; *J. Inst. Math. Appl*, 7, pp. 21-36.
- [41] M.J.D. Powell (1977), Restart procedures for the conjugate gradient method, *Math Programming*, 2 , pp 241-254.
- [42] M.J.D. Powell (1984), Non convex minimzation calculation and the conjugate gradient method, in : *Lecture Notes in Mathematics 1066 (Springer, Berlin)*, pp. 122-141.
- [43] M.J.D. Powell (1986), Convergence properties of algorithms for nonlinear optimization, *SIAM rev.*, 28, pp. 487-500.

-
- [44] Yannick PRIVAT, Introduction à l'optimisation Aspects théoriques et numériques, Cours de première année - année universitaire 2014-2015
- [45] Rodolphe Le Riche, Stéphane Mottelet, Eric Touboul, Ecole des Mines de Saint-Etienne, Université de Technologie de Compiègne, CNRS 2010
- [46] M. Rivaie, M. Mamat, L. W. June, and I. Mohd, A new class of nonlinear conjugate gradient coefficients with global convergence properties. *Applied Mathematics and Computation* 218 (2012) 11323–11332.
- [47] François Rouvière. *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*. Cassini, 2003. Deuxième édition revue et augmentée.
- [48] D. Touati-Ahmed and C. Storey (1990), Efficient hybrid conjugate gradient techniques, *JOTA*, 64, pp. 379-397.
- [49] P. Wolfe, Convergence conditions for ascent methods, *SIAM Review*, 11 (1969) 226-235
- [50] P. Wolfe (1971), Conditions for ascent methods some corrections 2, *SIAM Review*, 13, pp. 185-188.
- [51] Z. Wei, S. Yao and L. Liu, The Convergence Properties of Some New Conjugate Gradient Methods, *Applied Mathematics and Computation*, Vol. 183, No. 2, 2006, pp. 1341-1350.
- [52] Y.X. Yuan (1993), Numerical Methods for Non linear Programming, *Shanghai scientific & Technical Publishers (in Chinese)*.
- [53] G. Zoutendijk (1970) Nonlinear Programming Computational Methods, Integer and Nonlinear Programming, North Holland, Amsterdam.