

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Université 8 Mai 1945 Guelma

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



17/510.245



Mémoire

Présenté en vue de l'obtention du diplôme de

Master en Mathématiques

Option : EDP Et Analyse numérique

Par :

Melle. Guernine Sarra

**Intitulé**

**Sur la convergence de quelques nouvelles modifications des  
recherches linéaires inexactes**

Dirigé par : Mme. MELLAL Remaissa

Devant le jury

PRESIDENT	Dr. BELHIRECHE Hanane	MCB	Univ-Guelma
RAPPORTEUR	Dr. MELLAL Remaissa	MCB	Univ-Guelma
EXAMINATEUR	Dr. BOUAFIA Mousaab	MCB	Univ-Guelma

Session Juin 2018

---

# Table des matières

<b>1</b>	<b>Notions Préliminaires</b>	<b>3</b>
1.1	Rappels et définitions . . . . .	3
1.2	Problèmes d'optimisation sans contraintes . . . . .	5
1.2.1	Conditions d'optimalité des problèmes d'optimisation sans contraintes	5
1.3	Méthodes à direction de descente avec recherche linéaire inexacte . . . . .	6
1.3.1	Direction de descente . . . . .	6
1.3.2	Schéma général des méthodes à direction de descente avec recherche linéaire inexacte . . . . .	8
1.3.3	Méthode du gradient . . . . .	8
1.4	Mode de convergence et Critère d'arrêt . . . . .	9
1.4.1	Mode de convergence . . . . .	9
1.4.2	Critère d'arrêt . . . . .	10
<b>2</b>	<b>Optimisation unidimensionnelle recherches linéaires inexactes</b>	<b>11</b>
2.1	Recherches linéaires . . . . .	11
2.1.1	Principe des recherches linéaires . . . . .	11
2.1.2	But de la recherche linéaire . . . . .	13
2.1.3	Intervalle de sécurité . . . . .	13
2.2	Algorithme générale des recherches linéaires inexactes . . . . .	14
2.2.1	Algorithme des recherches linéaires inexactes . . . . .	14
2.3	La règle d'Armijo . . . . .	15
2.3.1	Principe . . . . .	15

---

2.3.2	Représentation géométrique . . . . .	16
2.3.3	Algorithme de la Règle d'Armijo . . . . .	17
2.3.4	Théorème d'existence du pas d'Armijo . . . . .	17
2.3.5	Exemple . . . . .	18
2.4	La règle de Goldstein&Price . . . . .	19
2.4.1	Principe . . . . .	20
2.4.2	Représentation géométrique . . . . .	21
2.4.3	Algorithme de la Règle de Goldstein&Price . . . . .	22
2.4.4	Théorème d'existence du pas de Goldstein&Price . . . . .	22
2.4.5	Exemple . . . . .	24
2.5	La règle de Wolfe . . . . .	25
2.5.1	Principe . . . . .	26
2.5.2	Représentation géométrique . . . . .	27
2.5.3	Algorithme de la règle Wolfe . . . . .	28
2.5.4	Règle de Wolfe forte . . . . .	28
2.5.5	Théorème d'existence du pas de Wolfe . . . . .	29
2.5.6	Exemple . . . . .	30
<b>3</b>	<b>Modification de La recherche linéaire inexacte d'Armijo</b>	<b>32</b>
3.1	Définition et hypothèses . . . . .	32
3.2	Principe de la recherche linéaire inexacte d'Armijo modifiée . . . . .	33
3.3	L'analyse de la convergence globale . . . . .	34
3.4	Taux de la convergence linéaire . . . . .	39
<b>4</b>	<b>Simulations numériques</b>	<b>42</b>
4.1	Méthode du gradient avec la règle d'Armijo modifiée . . . . .	43
4.2	Méthode du gradient avec la règle d'Armijo . . . . .	43

---

# *Remerciements*

*En préambule à ce mémoire, je tiens à remercier en premier lieu **ALLAH** qui m'a donné le pouvoir d'effectuer ce modeste travail.*

*Mes profonds remerciements, ma reconnaissance et ma gratitude vont à mon directeur de mémoire madame **Mellal Romaiissa**. Je la remercie sincèrement pour la confiance qu'elle m'a accordé, ses encouragements, ses conseils précieux ainsi que pour le temps qu'elle m'a accordé malgré ses obligations et ses responsabilités. Je la remercie également pour sa rigueur, sa bonne humeur et sa modestie.*

*J'adresse aussi mes sincères remerciements aux membres de jury qui ont accepté la lourde tâche de lire, commenter et juger ce mémoire.*

*Mes vifs remerciements vont également à tous les professeurs qui nous ont enseigné et qui par leurs compétences nous ont soutenu dans la poursuite de nos études.*

*Mes derniers remerciements, et pas les moindres, vont à **ma mère** et à **mon père** qui m'ont tant apporté d'amour, d'encouragement et sans eux je n'aurais pas pu*

*Aller au bout de ce travail, à mes chères frères **Mehdi** et **Abdelkarim** et **Mouhamed Yassine**, qui m'ont toujours apporté leur support moral, ainsi que toute ma famille*

*Je termine avec un grand remerciement bien particulier à toute personne qui de près ou de loin a contribué à ma réussite.*

---

# *Dédicaces*

*Je dédie ce modeste travail:*

*A la mémoire de mes grands pères, la mémoire de ma grand-mère paternelle  
Qui ont été toujours dans mon esprit et dans mon cœur, je vous dédie aujourd'hui  
ma réussite. Que Dieu, le miséricordieux, vous accueille  
dans son éternel paradis.*

*À ma grand-mère maternelle « Nous lui souhaitons la longévité »*

*À mes parents*

*À ma très chère Mère; et à mon très cher Père; pour ses sacrifices*

*De tous les instants.*

*À mes beaux frères*

*À Toute ma famille Guernine et Bazine*

*À tous Mes enseignants sans Exception*

*À mes tendres amies*

*Rawnek, Marwa, Nour el houda, sarra, nessrine.*

*Et à tous ceux qui ont toujours cru en moi, m'ont accompagné et soutenu . . .*

*Trouvez ici ma profonde reconnaissance. . .*

*Guernine sarra*

# Sur la convergence de quelques nouvelles modifications des recherches linéaires inexactes

## Résumé

**Résumé :** L'objectif de ce mémoire consiste à étudier quelques recherches linéaires inexactes. Il s'agit des règles d'Armijo, Goldstein-Price et de Wolfe. Récemment une généralisation de la règle d'Armijo à été proposée. Cette nouvelle règle assure la convergence de la classe des méthodes à directions de descente.

Dans ce travail on va étudier ces règles tout en donnant leurs algorithmes, citant les théorèmes qui assurent l'existence de leurs pas et on termine chaque partie par un exemple. Par la suite on va prouver numériquement moyennant le logiciel Scilab, la convergence de la méthode du gradient avec les règles d'Armijo et d'Armijo modifiée. On illustre que la règle d'Armijo modifiée accélère la convergence de la méthode du gradient mieux que celle d'Armijo.

**Mots-clés :** Recherche linéaire inexacte, Règle d'Armijo, Règle d'Armijo modifiée, Règle de Goldstein-Price, Règle de Wolfe, Méthode du gradient.

---

# Introduction

Considérons le problème d'optimisation sans contrainte suivant:

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

Les méthodes itératives sont en général les plus utilisées pour la résolution des problèmes du type  $(P)$ .

Ces méthodes génèrent une suite  $\{x_k\}_{k \in \mathbb{N}}$  de la façon suivante:

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

Où  $x_k$  est l'itération donnée à l'étape  $k$ .

Le pas  $\alpha_k \in \mathbb{R}_+$  est déterminé par une optimisation unidimensionnelle, autrement dit une recherche linéaire.

Les vecteurs  $d_k$  sont appelés directions et leurs choix donnent un nom à la méthode.

Dans ce travail on s'intéresse au calcul du pas  $\alpha$  par les recherches linéaires inexactes d'Armijo[1], de Goldstein-Price [11], de Wolfe [22] et une généralisation de la règle d'Armijo [1] nommée règle d'Armijo modifiée proposée par Z. J. Shi et J. Shen en 2005 [20].

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.

Faire de la recherche linéaire inexacte revient à déterminer

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

Autrement dit, résoudre le problème unidimensionnel

$$(P') : \min_{\alpha > 0} \varphi_k(\alpha)$$

Cette nouvelle règle donne des résultats plus fiables que celle d'Armijo [1].

Nos simulations numériques (Scilab) montrent que la recherche linéaire inexacte d'Armijo modifiée [20] est considérée plus performante que celle d'Armijo [1].

Notre mémoire est divisée en quatre chapitres :

Dans le premier chapitre on introduit les outils de base. Tout d'abord nous rappelons quelques généralités sur le problème d'optimisation sans contrainte. Ensuite nous décrivons la direction de descente, Mode de convergence et Critère d'arrêt. Dans le deuxième chapitre on va décrire les recherches linéaires inexactes d'Armijo [1], Goldstein-Price [11] et de Wolfe [22], on va donner leurs algorithmes et citer les théorèmes qui assurent l'existence du pas de ces recherches avec des exemples. Dans le troisième chapitre on commence par présenter la recherche linéaire inexacte d'Armijo modifiée [20]. Ensuite nous illustrons numériquement dans le dernier chapitre que la règle d'Armijo modifiée [20] est plus fiable que celle d'Armijo[1]. On termine ce mémoire par une conclusion.

Dans ce chapitre on va rappeler quelques notions de base qui seront utilisées par la suite.

## 1.1 Rappels et définitions

**Définition 1.1.1** Soit  $E$  un espace vectoriel, on appelle *norme* toute application de  $E$  dans  $\mathbb{R}_+$ , on dit que  $\|\cdot\|$  est une norme si elle vérifie:

$$\langle x, x \rangle = \|x\|^2;$$

**Définition 1.1.2** Une application  $\|\cdot\| : V \rightarrow \mathbb{R}$  est dit *norme vectorielle* si:

- 1)  $\|v\| \geq 0 \forall v \in V$  et  $\|v\| = 0 \iff v = 0$
- 2)  $\|\lambda v\| = |\lambda| \|v\| \quad \forall \lambda \in \mathbb{R}; v \in V$
- 3)  $\|v + u\| \leq \|v\| + \|u\| \quad \forall v, u \in V$

**Définition 1.1.3** Soit  $E$  un espace vectoriel, on appelle *produit scalaire* toute application de  $E$  dans  $\mathbb{R}$ , vérifiant les conditions suivantes :

Pour tout  $x, y$  et  $z \in E$ , et  $\alpha \in \mathbb{R}$  on a.

$$\left\{ \begin{array}{l} a) \quad \langle x, x \rangle \geq 0 \text{ et } \langle x, x \rangle = 0 \iff x = 0, \\ b) \quad \langle r, y \rangle = \langle y, r \rangle, \\ c) \quad \langle \alpha x, y \rangle = \alpha \langle x, y \rangle, \\ d) \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle; \end{array} \right.$$

Dans tout ce qui suit on prend  $E = \mathbb{R}^n$  et pour tout  $x$  et  $y$  dans  $\mathbb{R}^n$ , on note par  $x^T y$  le produit scalaire  $(x, y)$ , autrement dit :

$$x^T y = (x, y);$$

**Définition 1.1.4** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , le vecteur de  $\mathbb{R}^n$ ,  $\nabla f$  défini ci-dessus, est appelé **gradient** de la fonction  $f$  au point  $x_0 \in \mathbb{R}^n$ , et est noté :

$$\nabla f(x_0) = \left( \frac{\partial f}{\partial x_1}(x_0), \frac{\partial f}{\partial x_2}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0) \right); \quad (1.1.1)$$

On trouve aussi comme notation

$$\overrightarrow{\text{Grad}}(f)(x_0);$$

**Définition 1.1.5** a)  $f$  est dite **convexe** sur  $\mathbb{R}^n$  si :

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2); \quad (1.1.2)$$

pour tous point  $x_1$  et  $x_2 \in \mathbb{R}^n$  et pour tout  $t \in [0, 1]$ .

b) Si l'inégalité précédente est stricte pour tous points  $x_1$  et  $x_2$  distincts et pour tout  $t \in ]0, 1[$ , alors  $f$  est dite **strictement convexe**.

**Définition 1.1.6** Soit  $A$  une matrice carrée  $(n \times n)$ ,  $A \in \mathcal{M}_{n \times n}$ .

a) On dit que  $A$  est **symétrique** si et seulement si :  $A = A^t$ ;

b) On dit que  $A$  est **définie positive** si et seulement si :

$$\forall x \in \mathbb{R}^n, x^t A x > 0; \quad (1.1.3)$$

**Théorème 1.1.1 (de la valeur moyenne)** Pour toute fonction  $f$  à valeurs réelles, définie et continue sur un segment  $[a, b]$ , avec  $a < b$ , il existe un réel  $c \in [a, b]$  vérifiant :

$$f(b) - f(a) = (b - a)f'(c) \quad (1.1.4)$$

**Définition 1.1.7** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , on dit que  $f$  est une **fonction quadratique** si elle s'écrit sous la forme

$$f(x) = \frac{1}{2}(A, x) + (b, x) + c. \quad (1.1.5)$$

Tel que  $A$  est une matrice carré positive symétrique et  $b \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$ .

**Théorème 1.1.2 (Inégalité de Cauchy Schwarz)** Soit  $x$  et  $y$  dans  $\mathbb{R}^n$  On a :

$$|\langle x, y \rangle| \leq \|x\| \|y\| \quad (1.1.6)$$

## 1.2 Problèmes d'optimisation sans contraintes

Le problème d'optimisation sans contrainte opte a minimiser une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  sur tout l'espace  $\mathbb{R}^n$ .

Ces problèmes sont plus simples à analyser et à résoudre que les problèmes d'optimisation avec contrainte.

**Définition 1.2.1** soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , On Appelle *problème d'optimisation sans contrainte*, le problème  $(P)$  suivant :

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

C'est un problème d'optimisation sans conditions sur les variables.

### 1.2.1 Conditions d'optimalité des problèmes d'optimisation sans contraintes

**Définition 1.2.2** Considérons le problème de minimisation sans contraintes  $(P)$ .

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

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

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

$$f(x^*) \leq f(x), \forall x \in V_\varepsilon(x^*). \quad (1.2.3)$$

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

$$f(x^*) < f(x), \forall x \in V_\varepsilon(x^*) \quad \text{et} \quad x \neq x^*. \quad (1.2.4)$$

### Conditions nécessaires d'optimalité

**Théorème 1.2.1** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  deux fois différentiable en  $x^*$ . Supposons que  $x^*$  soit un *minimum local*. Alors  $\nabla f(x^*) = 0$  et  $H(x^*)$  est semi définie positive.

**Corollaire 1.2.1** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  différentiable en  $x^*$ , si  $x^*$  est un *minimum local* alors  $\nabla f(x^*) = 0$ .

### Conditions suffisantes d'optimalité

**Théorème 1.2.2** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . (i) Supposons que  $f$  est deux fois différentiable en  $x^*$ . Si  $\nabla f(x^*) = 0$  et  $H(x^*)$  est définie positive, alors  $x^*$  est *minimum local strict*.

(ii) Supposons que  $f$  est convexe et différentiable. Alors  $x^*$  est un *minimum global* de  $f$  si et seulement si  $\nabla f(x^*) = 0$ .

(iii) Les résultats (i) et (ii) demeurent vrais si on remplace  $\mathbb{R}^n$  par un ouvert  $S$  de  $\mathbb{R}^n$ .

**Corollaire 1.2.2** Dans le cas où  $f$  est convexe, alors tout minimum local est aussi global. De plus si  $f$  est strictement convexe, alors tout minimum local devient non seulement global mais aussi unique.

## 1.3 Méthodes à direction de descente avec recherche linéaire inexacte

Considérons le problème d'optimisation sans contrainte :

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1.3.1}$$

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

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

### 1.3.1 Direction de descente

**Théorème 1.3.1** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  telle que  $f$  soit différentiable au point  $x \in \mathbb{R}^n$ . Soit  $d \in \mathbb{R}^n$  telle que  $\nabla f(x)^t d < 0$ . Alors il existe  $\delta > 0$  tel que  $f(x + \alpha d) < f(x)$  pour tout  $\alpha \in ]0, \delta[$ . La direction  $d$  s'appelle dans ce cas *direction de descente*.

**Remarque 1.3.1**  $d$  fait avec l'opposé du gradient  $-\nabla f(x)$  un angle  $\theta$  strictement plus petit que  $90^\circ$  :

$$\theta = \arccos \frac{-\nabla^T f(x) \cdot d}{\|\nabla^T f(x)\| \|d\|} \in [0, \frac{\pi}{2}[. \quad (1.3.2)$$

L'ensemble des directions de descente de  $f$  en  $x$ ,  $\{d \in \mathbb{R}^n : \nabla^T f(x_k) \cdot d < 0\}$  forme un demi-espace ouvert de  $\mathbb{R}^n$

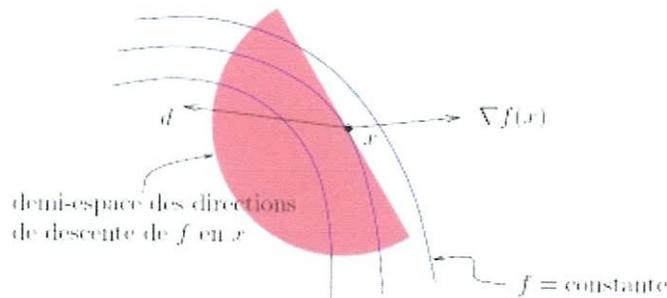


Figure 1.1 direction de descente

De telles directions sont intéressantes en optimisation car, pour faire décroître  $f$ , il suffit de faire un déplacement le long de  $d$ .

Les méthodes à directions de descentes utilisent cette idée pour minimiser une fonction. Elles construisent la suite des itérés  $\{x_k\}_{k \geq 1}$ , approchant une solution  $x_k$  de (1.2.1) par la récurrence :

$$x_{k+1} = x_k + \alpha_k d_k, \text{ pour } k \geq 1 \quad (1.3.3)$$

où  $\alpha_k$  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  $\alpha_k$ ; c'est ce que l'on appelle : la recherche linéaire.

Décrivons cette classe d'algorithmes de manière précise.

### 1.3.2 Schéma général des méthodes à direction de descente avec recherche linéaire inexacte

Présentons maintenant l'algorithme des méthodes à direction de descente avec recherche linéaire inexacte.

**Algorithme 1.1 (méthode à directions de descente- une itération)**

**Etape 0: (initialisation)**

On suppose qu'au début de l'itération  $k$ , on dispose d'un itéré

$$x_k \in \mathbb{R}^n$$

**Etape 1:**

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

**Etape 2:**

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

**Etape 3:**

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

**Etape 4:**

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.

Dans le chapitre suivant on va discuter comment choisir le pas à chaque itération, en utilisant quelque règles de recherche linéaire (Armijo [1], Goldstien-Price [11] ou Wolfe [22]).

### 1.3.3 Méthode du gradient

Le principe de cette méthode remonte au moins à Cauchy [5].

Dans l'algorithme (1.1), le choix de la direction  $d_{i_0} = -\nabla f(\tau_{i_0})$  nous donne l'algorithme de la méthode du gradient.

Donc on peut dire que la méthode du gradient est une méthode à directions de descente. En effet, on peut facilement vérifier que la direction  $d_k = -\nabla f(x_k)$  est une direction de descente.

L'algorithme du gradient est également connu sous le nom d'algorithme de la plus forte pente ou de la plus profonde descente (steepest descent, en anglais) parce que le gradient est la pente de la fonction au point courant et est donc, localement, sa plus forte pente.

Algorithme de la méthode du gradient

**Algorithme 1.2 (Algorithme du gradient)**

**Etape 0: (initialisation)**

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

poser  $k = 0$  et aller à l'étape 1.

**Etape 1:**

si  $\|\nabla f(x_k)\| \simeq 0$  : 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 :

$$\alpha_k \text{ est déterminé par une recherche linéaire} \tag{1.3.4}$$

$$d_k = -\nabla f(x_k) \tag{1.3.5}$$

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

## 1.4 Mode de convergence et Critère d'arrêt

### 1.4.1 Mode de convergence

**Définition 1.4.1** Une méthode itérative est dite *convergente* si la suite de points générés  $(x_n)$  tend vers la racine exacte quand  $n$  tend vers l'infini, autrement dit :

$$\lim_{n \rightarrow +\infty} x_n = \hat{x}. \tag{1.4.1}$$

**Définition 1.4.2** On dit qu'une suite  $(x_n)$  construite par une méthode numérique converge vers  $\hat{x}$  avec un ordre (taux)  $p \geq 1$  s'il existe un indice  $k_0$  tel que pour tout  $k \geq k_0$  on a :

$$\lim_{k \rightarrow +\infty} \frac{|x_{k+1} - \hat{x}|}{|x_k - \hat{x}|^p} = \beta < \infty. \quad (1.4.2)$$

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

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

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

### 1.4.2 Critère d'arrêt

Typiquement une procédure itérative démarre avec un point initial  $x_0$  et on conclut que la méthode converge si le rapport  $|x_k - x_{k-1}|$  ou  $|f(x_k)|$  décroît vers 0.

Dans la pratique, on doit déterminer la solution de problème  $(P)$  en un nombre fini d'étapes. Il est par conséquent impossible de faire tendre  $n$  vers l'infini. Il faut donc arrêter le processus itératif au bout d'un certain nombre d'itérations ce qui a pour conséquence de n'obtenir qu'une approximation de la solution  $\hat{x}$ .

Dans cet ordre d'idée plusieurs formules de test d'arrêt sont utilisées pour terminer le processus de calcul après un certain nombre d'opérations :

$$|x_n - x_{n-1}| < a \times \text{Tolérance};$$

$$|x_n - x_{n-1}| < \text{Tolérance} \times (1 + |x_n|);$$

$$|x_n - x_{n-1}| < r \times \text{Tolérance} \times |x_n|;$$

$$|f(x_n)| < \text{Tolérance}.$$

Avec *Tolérance* : l'erreur maximale tolérée,

$a$  et  $r$  sont des constantes définies par l'utilisateur.

## 2.1 Recherches linéaires

Notre problème consiste à minimiser une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

La recherche linéaire consiste à trouver  $\alpha_k$  de façon à diminuer la fonction  $f$  *suffisamment* le long de cette direction.

Ce "*suffisamment*" sera quantifié dans cette section dans la description des conditions dites d'Armijo [1], de Goldstein-Price [11] et de Wolfe [22].

### 2.1.1 Principe des recherches linéaires

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

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

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 :  $\nabla^T f(x).d < 0$ ) et exécute un pas dans cette direction :

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

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 :

$$\varphi_k(\alpha) = f(x_k + \alpha d_k). \tag{2.1.2}$$

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

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

Dans la recherche linéaire il existe deux grandes classes de méthodes qui s'intéressent à l'optimisation unidimensionnelle:

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

### 2.1.2 But de la recherche linéaire

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

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

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

$$\varphi'(\alpha^*) = 0$$

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

Autrement dit,  $\alpha^*$  est un minimum local de  $\varphi'$  qui assure de plus la décroissance de  $f$ .

On a habitué de faire des recherches linéaire inexact dans les algorithmes d'optimisation en générale mais actuellement on ne fait jamais ça, car le fait de trouver  $\alpha^*$  exige le calcul d'un grand nombre de fois la fonction  $\varphi'_k$  et cela conduit a la notion d'intervalle de sécurité.

### 2.1.3 Intervalle de sécurité

On est habitué de faire des recherches linéaires inexactes dans les Algorithmes d'optimisation en générale mais actuellement on ne fait jamais ça, car le fait de trouver  $\alpha_k$  exige le calcul d'un grand nombre de fois la fonction  $\varphi_k$  et cela conduit a la notion d'intervalle de sécurité.

**Définition 2.1.1** *On dit que  $]\alpha_g, \alpha_d[$  est un intervalle de sécurité s'il permet de classer les valeurs de  $\alpha$  de façon suivante:*

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

Dans notre travail on s'intéresse avec les recherches linéaires inexactes.

## 2.2 Algorithme générale des recherches linéaires inexactes

Le principe de cette recherche est de faire une réduction « de façon importante » pour la valeur de l'objectif par un pas  $x_{k+1} = x_k + \alpha_k d_k$  de  $x_k$  dans la direction de descente  $d_k$ . Pour ce raison a la plusieurs mathématiciens (Armijo [1], Goldstein&Price [11], Wolfe [22]) ont proposé des règles suivant cette recherche.

Cette section bute de présenter les principaux tests.

D'abord on présente le schéma général d'une recherche linéaire inexacte Elles reviennent à déterminer, par tâtonnement un intervalle  $[\alpha_g, \alpha_d]$  où  $\alpha^* \in [\alpha_g, \alpha_d]$  dans laquelle :

$$\varphi_k(\alpha_k) \leq \varphi_k(0) \quad \text{et} \quad f(x_k + \alpha_k d_k) < f(x_k) \quad (2.2.1)$$

### 2.2.1 Algorithme des recherches linéaires inexactes

Présentons l'algorithme des recherches linéaires inexactes :

#### Algorithme 2.1 Schéma général des la recherches linéaires inexactes

Etape 0 : Initialisation

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

Etape 1:

- Si  $\alpha_k$  satisfaisant (suivant un certain critère): STOP( $\alpha^* = \alpha_k$ ).
- Si  $\alpha_k$  est trop petit (suivant certain critère): nouvel intervalle:  $[\alpha_{g,k+1} = \alpha_k, \alpha_{d,k+1} = \alpha_d]$  et aller a l'étape 2.
- Si  $\alpha_k$  est très grand (suivant certain critère): nouvel intervalle:  $[\alpha_{g,k+1} = \alpha_g, \alpha_{d,k+1} = \alpha_k]$  et aller a 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) notre pas est considéré trop petit ou trop grand ou satisfaisant.

## 2.3 La règle d'Armijo

On bute à réduire « de façon importante » la valeur de l'objectif par un pas

$x_k \longrightarrow x_{k+1} = x_k + \alpha_k d_k$  de  $x_k$  dans la direction  $d_k$ , tel que  $f(x_k + \alpha_k d_k) \leq f(x_k)$ .

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

$(x_k)$  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 plus grand)

La règle d'Armijo [1] impose une contrainte sur le choix de  $\alpha_k$  suffisante pour minimiser localement  $\varphi$

### 2.3.1 Principe

Soit  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ ,  $x_k \in \mathbb{R}^n$ ,  $d_k \in \mathbb{R}^n$  une direction de descente.

On dit que le pas  $\alpha_k$  vérifie la règle d'Armijo [1] si on a:

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha_k \beta \varphi'_k(0) \quad \forall \beta \in ]0, 1[ \quad (2.3.1)$$

$$\text{autrement dit } f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k \quad \forall \beta \in ]0, 1[ \quad (2.3.2)$$

Teste d'Armijo :

◆ Si

$$\varphi_k(\alpha_k) < \varphi_k(0) + \alpha_k \beta \varphi'_k(0)$$

$$\text{autrement dit } f(x_k + \alpha_k d_k) < f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k$$

alors  $\alpha_k$  est convenant .

◆ Si

$$\varphi_k(\alpha_k) > \varphi_k(0) + \alpha_k \beta \varphi'_k(0)$$

$$\text{autrement dit } f(x_k + \alpha_k d_k) > f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k$$

alors  $\alpha_k$  est trop grand.

### 2.3.2 Représentation géométrique

$T$  : la tangente de graphe de  $\varphi_k(\alpha)$  qui passe par le point  $(0, \varphi_k(0))$

$$T := \{(\alpha, y) \mid y = \varphi_k(0) + \alpha \varphi'_k(0)\}$$

$\Delta_1$ : la droite qui passe par le point  $(0, \varphi_k(0))$  et qui à  $\beta \varphi'_k(0)$  coefficient de direction

$$\Delta_1 := \{(\alpha, y) \mid y = \varphi_k(0) + \alpha \beta \varphi'_k(0)\}$$

On a :

$$\beta \varphi'_k(0) \geq \varphi'_k(0) \implies \alpha \beta \varphi'_k(0) \geq \alpha \varphi'_k(0)$$

$$\implies \varphi_k(0) + \alpha \beta \varphi'_k(0) \geq \varphi_k(0) + \alpha \varphi'_k(0)$$

On voit bien  $\Delta$  est au dessus de  $T$ .

$\alpha$  vérifier la règle d'Armijo [1] si  $\alpha \in [0, \tau_1]$

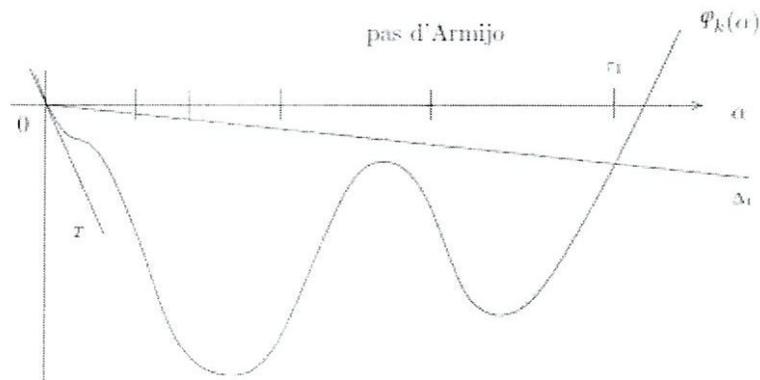


Figure 2.1 la recherche linéaire inexacte d'Armijo

### 2.3.3 Algorithme de la Règle d'Armijo

Dans cette partie nous allons présenter l'Algorithme de la règle d'Armijo [1] qui suit le schéma suivant :

#### Algorithme 2.2 La règle d'Armijo

**Etape 1:** (initialisation)

$\alpha_{g,k} = \alpha_{d,k} = 0$ , choisir  $\alpha_k > 0$ ,  $\beta \in ]0, 1[$  poser  $k = 1$  et aller à l'étape 2

**Etape 2:** (l'étape principale)

- Si  $\varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha_k \beta \varphi'_k(0)$  :STOP( $\alpha^* = \alpha_k$ ).

- Si  $\varphi_k(\alpha_k) > \varphi_k(0) + \alpha_k \beta \varphi'_k(0)$  alors

$\alpha_{g,k+1} = \alpha_g$ ,  $\alpha_{d,k+1} = \alpha_k$  et aller a l'étape 3.

**Etape 3:**

- 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 2

### 2.3.4 Théorème d'existence du pas d'Armijo

**Théorème 2.3.1** Si  $\varphi_k : \mathbb{R}_+ \mapsto \mathbb{R}$ ; définie par (2.1.2)  $f(x_k + \alpha d_k) = \varphi_k(\alpha)$  est continue et bornée inférieurement, si  $d_k$  est une direction de descente en  $x_k$  ( $\varphi'_k(0) < 0$ ) et si  $\beta \in ]0, 1[$  alors l'ensemble des pas vérifiant la règle d'Armijo (2.3.2) est non vide.

**Démonstration:**

On a:

$$\varphi_k(\alpha) = f(x_k + \alpha d_k)$$

$$\psi_\beta(\alpha) = f(x_k) + \alpha \beta \nabla^T f(x_k) d_k$$

Le développement de Taylor-Yong en  $\alpha = 0$  de  $\varphi_k$  est :

$$\varphi_k(\alpha) = f(x_k + \alpha d_k) = f(x_k) + \alpha \nabla^T f(x_k) d_k + \xi(\alpha) \text{ Ou } \xi(\alpha) \longrightarrow 0, \alpha \rightarrow 0$$

Et comme  $\beta \in ]0, 1[$  et  $\varphi'_k(0) = \nabla^T f(x_k) d_k < 0$  on déduit:

$$f(x_k) + \alpha \nabla^T f(x_k) d_k < f(x_k) + \alpha \beta \nabla^T f(x_k) d_k \quad \text{pour } \alpha > 0$$

On voit que pour  $\alpha > 0$  assez petit on a:

$$\varphi_k(\alpha) < \psi_\beta(\alpha)$$

De ce qui précède du fait que  $\varphi_k$  est bornée inférieurement et

$$\psi_\beta(\alpha) \longmapsto -\infty; \alpha \longmapsto +\infty$$

On déduit que la fonction  $\psi_\beta(\alpha) - \varphi_k(\alpha)$  possède la propriété :

$$\begin{cases} \psi_\beta(\alpha) - \varphi_k(\alpha) > 0 \text{ pour } \alpha \text{ assez petit} \\ \psi_\beta(\alpha) - \varphi_k(\alpha) < 0 \text{ pour } \alpha \text{ assez grand} \end{cases}$$

Donc s'annule au moins une fois pour  $\alpha > 0$ .

En choisissant le plus petit de ces zéros on voit qu'il existe  $\bar{\alpha} > 0$  tel que:

$$\varphi_k(\bar{\alpha}) = \psi_\beta(\bar{\alpha}) \text{ et } \varphi_k(\alpha) < \psi_\beta(\alpha) \text{ pour } 0 < \alpha < \bar{\alpha}. \blacksquare$$

### 2.3.5 Exemple

Prenons

$$f(x, y) = 3x^2 + y^2$$

Soit  $X_0 = (x_0, y_0) = (1, 1)$  et  $d_0 = -\nabla f(x_0, y_0) = (-2, -3)$ .

1) Calculons  $\varphi_0(\alpha)$ ,  $\varphi_0(0)$ ,  $\varphi'_0(\alpha)$ ,  $\varphi'_0(0)$

$$\varphi_0(\alpha) = f(x_0 + \alpha d_0) = f[(1, 1) + \alpha(-2, -3)]$$

$$\varphi_0(\alpha) = 21\alpha^2 - 18\alpha + 4; \varphi_0(0) = 4$$

$$\varphi'_0(\alpha) = 42\alpha - 18; \varphi'_0(0) = -18$$

Cherchons un point stationnaire de  $\varphi_0$ , autrement dit un pas pour lequel  $\varphi'_0(\alpha) = 0$

$$\varphi'_0(\alpha) = 0 \implies 42\alpha - 18 = 0 \implies \alpha^* = \frac{9}{21}$$

Comme  $\varphi_0''(\alpha) = 42 > 0 \implies \alpha^*$  est un solution minimale locale stricte

2) Trouvons le pas d'Armijo :

-)  $\alpha_{g,1} = \alpha_{d,1} = 0$ . On pose que:  $\alpha_1 = 1$  et  $\beta = 0.9$ ;  $g_0(\alpha) = \varphi_0(0) + \alpha\beta\varphi_0'(0) = 4 - 16.2\alpha$

On va calculer les  $\varphi_0$  et  $g_0$  puis les comparé jusqu'à ce que (2.3.2) soit vérifié

$$\varphi_0(1) = 7; \quad g_0(1) = -8,6 \implies \varphi_0(1) > g_0(1)$$

-)  $\alpha_{g,1} = 0, \alpha_{d,1} = 1$  on prend  $\alpha_2 = \frac{1}{2}$

$$\varphi_0\left(\frac{1}{2}\right) = 0.25; \quad g_0\left(\frac{1}{2}\right) = -2.3 \implies \varphi_0\left(\frac{1}{2}\right) > g_0\left(\frac{1}{2}\right)$$

-)  $\alpha_{g,1} = 0, \alpha_{d,1} = \frac{1}{2}$  on prend  $\alpha_3 = \frac{1}{4}$

$$\varphi_0\left(\frac{1}{4}\right) = 0.8125; \quad g_0\left(\frac{1}{4}\right) = 0.85 \implies \varphi_0\left(\frac{1}{4}\right) < g_0\left(\frac{1}{4}\right)$$

STOP  $\alpha^* = \alpha_3 = \frac{1}{4}$ .

Ainsi on obtient la nouvelle itérée  $X_1 = (x_1, y_1) = x_0 + \alpha_0 d_0$

$$f(x_0, y_0) = 3(1)^2 + (1)^2 = 4; \quad f(x_1, y_1) = f[(x_0, y_0) + \alpha^* d_0] = 0.8125$$

$$f(x_1, y_1) < f(x_0, y_0)$$

On voit bien que le pas d'Armijo [1] vérifié la descente de la fonction  $f$ .

## 2.4 La règle de Goldstein&Price

Dans la règle d'Armijo [1] nécessite de décroissance de la fonction objective à chaque pas mais c'est insuffisant; pour cela on prend l'exemple suivant :

$$(P) : \min_{x \in \mathbb{R}^n} x^2$$

La solution optimale globale minimal exacte de ce problème est  $x^* = 0$ .

On prend  $x_1 = 2, d_1 = -1$  le choix  $\alpha_k = \frac{1}{2^{k+1}}$  donnent  $x_k = 1 + \frac{1}{2^k}$ .

$\varphi_k$  est bien strictement décroissante mais  $\{x_k\}_{k \in \mathbb{N}}$  ne converge pas vers le minimum zero mais  $x_k \rightarrow 1$  quand  $k \rightarrow +\infty$

Alors la condition d'Armijo [1] est satisfaite.

### 2.4.1 Principe

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x_k \in \mathbb{R}^n$ ,  $d_k \in \mathbb{R}^n$  une direction de descente

On dit que le pas  $\alpha_k$  vérifie la règle de Goldstein&Price [11] si on a :

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha_k \beta \varphi'_k(0) \quad ; \quad \forall \beta \in ]0, 1[ \quad (2.4.1)$$

$$\varphi_k(\alpha_k) \geq \varphi_k(0) + \alpha_k \rho \varphi'_k(0) \quad ; \quad \forall \rho \in ]\beta, 1[ \quad (2.4.2)$$

Autrement dit; pour  $0 < \beta < \sigma < 1$  on a :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k \quad ; \quad \forall \beta \in ]0, 1[ \quad (2.4.3)$$

$$f(x_k + \alpha_k d_k) \geq f(x_k) + \alpha_k \rho \nabla^T f(x_k) d_k \quad ; \quad \forall \rho \in ]\beta, 1[ \quad (2.4.4)$$

**Teste de Goldstein&Price :**

◆ Si

$$\varphi_k(0) + \alpha_k \rho \varphi'_k(0) \leq \varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha_k \beta \varphi'_k(0)$$

$$\text{autrement dit } f(x_k) + \alpha_k \rho \nabla^T f(x_k) d_k \leq f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k$$

alors  $\alpha_k$  est convient.

◆ Si

$$\varphi_k(\alpha_k) > \varphi_k(0) + \alpha_k \beta \varphi'_k(0)$$

$$\text{autrement dit } f(x_k + \alpha_k d_k) > f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k$$

alors  $\alpha_k$  est trop grand.

◆ Si

$$\varphi_k(\alpha_k) < \varphi_k(0) + \alpha_k \rho \varphi'_k(0)$$

$$\text{autrement dit } f(x_k + \alpha_k d_k) < f(x_k) + \alpha_k \rho \nabla^T f(x_k) d_k$$

alors  $\alpha_k$  est trop petit

### 2.4.2 Représentation géométrique

$T$  : la tangente de graphe de  $\varphi_k(\alpha)$  qui passe par le point  $(0, \varphi_k(0))$

$$T := \{(\alpha, y) \mid y = \varphi_k(0) + \alpha\varphi'_k(0)\}$$

$\Delta_1$ : la droite qui passe par le point  $(0, \varphi_k(0))$  et qui a  $\beta\varphi'_k(0)$  coefficient de direction

$$\Delta_1 := \{(\alpha, y) \mid y = \varphi_k(0) + \alpha\beta\varphi'_k(0)\}$$

$\Delta_2$ : la droite qui passe par le point  $(0, \varphi_k(0))$  et qui a  $\rho\varphi'_k(0)$  coefficient de direction

$$\Delta_2 := \{(\alpha, y) \mid y = \varphi_k(0) + \alpha\rho\varphi'_k(0)\}$$

On a:  $0 < \beta < \rho < 1 \implies$

$$\rho\varphi'_k(0) < \beta\varphi'_k(0) \implies \alpha\rho\varphi'_k(0) < \alpha\beta\varphi'_k(0) \implies \varphi_k(0) + \alpha\rho\varphi'_k(0) < \varphi_k(0) + \alpha\beta\varphi'_k(0)$$

On voit bien que  $\Delta_2$  est au dessous de  $\Delta_1$ .

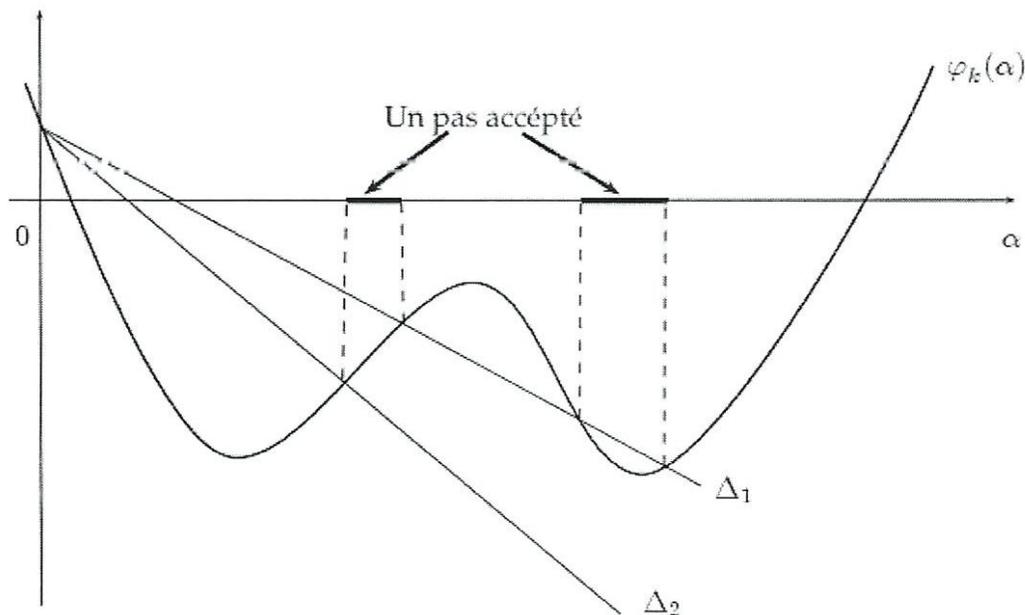


Figure 2.2 la recherche linéaire inexacte de Goldstein&Price

### 2.4.3 Algorithme de la Règle de Goldstein&Price

Dans cette partie nous allons présenter l'Algorithme de la règle de Goldstein&Price [11] qui suit le schéma suivant :

#### Algorithme 2.3 La règle de Goldstein&Price

**Etape 1:** (initialisation)

$\alpha_{g,1} = \alpha_{d,1} = 0$  choisir  $\alpha_k, \beta \in ]0, 1[, \rho \in ]\beta, 1[$ , poser  $k=1$  et allez à l'étape 2.

**Etape 2:**

◆ Si  $\varphi_k(0) + \alpha\rho\varphi'_k(0) \leq \varphi_k(\alpha) \leq \varphi_k(0) + \alpha\beta\varphi'_k(0)$  : STOP  $\alpha^* = \alpha_k$ .

◆ Si  $\varphi_k(\alpha) \geq \varphi_k(0) + \alpha\beta\varphi'_k(0)$  alors

$\alpha_{d,k+1} = \alpha_k, \alpha_{g,k+1} = \alpha_{g,k}$  et allez à l'étape 3.

◆ Si  $\varphi_k(\alpha) \leq \varphi_k(0) + \alpha\rho\varphi'_k(0)$  alors

$\alpha_{g,k+1} = \alpha_k, \alpha_{d,k+1} = \alpha_{d,k}$  et allez à l'étape 3.

**Etape 3:**

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

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

(par exemple on prend  $\alpha_{k+1} = \frac{\alpha_{g,k+1} + \alpha_{d,k+1}}{2}$ )

Remplacer  $k$  par  $k+1$  et allez à l'étape 2.

### 2.4.4 Théorème d'existence du pas de Goldstein&Price

**Théorème 2.4.1** Si  $\varphi_k : \mathbb{R}_+ \rightarrow \mathbb{R}$ ; définie par (2.1.2)  $f(x_k + \alpha d_k) = \varphi_k(\alpha)$  est continue et bornée inférieurement, si  $d_k$  est une direction de descente en  $x_k$  ( $\varphi'_k(0) < 0$ ) et si  $\beta \in ]0, 1[$  et  $\rho \in ]\beta, 1[$  alors l'ensemble des pas vérifiant la règle de Goldstein-Price (2.4.3)-(2.4.4) est non vide.

**Démonstration:**

On a

$$\begin{aligned} f(x_k + \alpha d_k) &= \varphi_k(\alpha) \\ \psi_\beta(\alpha) &= f(x_k) + \alpha \beta \nabla^T f(x_k) d_k \\ \psi_\rho(\alpha) &= f(x_k) + \alpha \rho \nabla^T f(x_k) d_k \end{aligned}$$

Le développement de Taylor-Yong en  $\alpha = 0$  de  $\varphi_k$  est :

$$\varphi_k(\alpha) = f(x_k + \alpha d_k) = f(x_k) + \alpha \nabla^T f(x_k) d_k + \xi(\alpha) \text{ Ou } \xi(\alpha) \longrightarrow 0, \alpha \rightarrow 0$$

Et comme  $\beta \in ]0, 1[$ ,  $\rho \in ]\beta, 1[$  et  $\varphi'_k(0) = \nabla^T f(x_k) d_k < 0$  on déduit:

$$f(x_k) + \alpha \nabla^T f(x_k) d_k < f(x_k) + \alpha \rho \nabla^T f(x_k) d_k < f(x_k) + \alpha \beta \nabla^T f(x_k) d_k \quad \text{pour } \alpha > 0$$

On voit que pour  $\alpha > 0$  assez petit on a:

$$\varphi_k(\alpha) < \psi_\rho(\alpha) < \psi_\beta(\alpha)$$

De ce qui précède du fait que  $\varphi_k$  est bornée inférieurement et

$$\psi_\beta(\alpha) \longmapsto -\infty; \alpha \longmapsto +\infty$$

On déduit que la fonction  $\psi_\beta(\alpha) - \varphi_k(\alpha)$  possède la propriété :

$$\begin{cases} \psi_\beta(\alpha) - \varphi_k(\alpha) > 0 \text{ pour } \alpha \text{ assez petit} \\ \psi_\beta(\alpha) - \varphi_k(\alpha) < 0 \text{ pour } \alpha \text{ assez grand} \end{cases}$$

Donc s'annule au moins une fois pour  $\alpha > 0$ .

En choisissant le plus petit de ces zéros on voit qu'il existe  $\bar{\alpha} > 0$  tel que:

$$\varphi_k(\bar{\alpha}) = \psi_\beta(\bar{\alpha}) \text{ et } \varphi_k(\alpha) < \psi_\beta(\alpha) \text{ pour } 0 < \alpha < \bar{\alpha}$$

De la même manière:

De ce qui précède du fait que  $\varphi_k(\alpha)$  est bornée inférieurement et

$$\psi_\rho(\alpha) \longmapsto -\infty; \alpha \longmapsto +\infty$$

On déduit que la fonction  $\psi_\rho(\alpha) - \varphi_k(\alpha)$  possède la propriété :

$$\begin{cases} \psi_\rho(\alpha) - \varphi_k(\alpha) > 0 \text{ pour } \alpha \text{ assez petit} \\ \psi_\rho(\alpha) - \varphi_k(\alpha) < 0 \text{ pour } \alpha \text{ assez grand} \end{cases}$$

Donc s'annule au moins une fois pour  $\alpha > 0$ .

En choisissant le plus petit de ces zéros on voit qu'il existe  $\tilde{\alpha} > 0$  tel que

$$\varphi_k(\tilde{\alpha}) = \psi_\rho(\tilde{\alpha}) \text{ et } \varphi_k(\alpha) < \psi_\rho(\alpha) \text{ pour } 0 < \alpha < \tilde{\alpha}$$

Et comme  $\psi_\rho(\alpha) < \psi_\beta(\alpha)$  pour  $\alpha > 0$ , forcément  $\tilde{\alpha} < \bar{\alpha}$  et  $\alpha = \bar{\alpha}$  satisfait (2.4.3)-(2.4.4)

$$\left( \begin{array}{l} \psi_\rho(\tilde{\alpha}) = \varphi_k(\tilde{\alpha}) < \psi_\beta(\alpha) \text{ n'est autre que} \\ f(x_k) + \tilde{\alpha}\rho\nabla^T f(x_k)d_k = f(x_k + \tilde{\alpha}d_k) < f(x_k) + \alpha\beta\nabla^T f(x_k)d_k \blacksquare \end{array} \right)$$

### 2.4.5 Exemple

Prenons

$$f(x, y) = 3x^2 + y^2$$

Soit  $X_0 = (x_0, y_0) = (1, 1)$  et  $d_0 = -\nabla f(x_0, y_0) = (-2, -3)$ .

1) Calculons  $\varphi_0(\alpha)$ ,  $\varphi_0(0)$ ,  $\varphi'_0(\alpha)$ ,  $\varphi'_0(0)$

$$\varphi_0(\alpha) = f(x_0 + \alpha d_0) = f[(1, 1) + \alpha(-2, -3)]$$

$$\varphi_0(\alpha) = 21\alpha^2 - 18\alpha + 4; \varphi_0(0) = 4$$

$$\varphi'_0(\alpha) = 42\alpha - 18; \varphi'_0(0) = -18$$

Cherchons un point stationnaire de  $\varphi_0$ , autrement dit un pas pour lequel  $\varphi'_0(\alpha) = 0$

$$\varphi'_0(\alpha) = 0 \longrightarrow 42\alpha - 18 = 0 \longrightarrow \alpha^* = \frac{9}{21}$$

On a :  $\varphi_k''(\alpha) = 42 > 0 \implies \alpha^*$  est un solution minimale locale stricte

2) Trouvons le pas de Goldstein&Price :

-)  $\alpha_{g,1} = \alpha_{d,1} = 0$  On pose que:  $\alpha_1 = 1$  et  $\beta = 0.1, \rho = 0.8$ ;  $g_0(\alpha) = \varphi_0(0) + \alpha\beta\varphi_0'(0) = 4 - 1.8\alpha$ ,

$$h_0(\alpha) = \varphi_0(0) + \alpha\rho\varphi_0'(0) = 4 - 14.4\alpha$$

On va calculer les  $\varphi_0$  et  $g_0$  puis les comparé jusqu'à ce que (2.4.3) soit vérifié

$$\varphi_0(1) = 7; \quad g_0(1) = 2.2 \implies \varphi_0(1) > g_0(1)$$

-)  $\alpha_{g,1} = 0, \alpha_{d,1} = 1$  on prend  $\alpha_2 = \frac{1}{2}$

$$\varphi_0\left(\frac{1}{2}\right) = 0.25; \quad g_0\left(\frac{1}{2}\right) = 3.1 \implies \varphi_0\left(\frac{1}{2}\right) < g_0\left(\frac{1}{2}\right)$$

On va calculer les  $\varphi_0$  et  $h_0$  puis les comparé à chaque fois jusqu'à ce que (2.4.4) soit vérifié

-)  $\alpha_2 = \frac{1}{2}$

$$\varphi_0\left(\frac{1}{2}\right) = 0.25; \quad h_0\left(\frac{1}{2}\right) = -3,2 \implies \varphi_0\left(\frac{1}{2}\right) > h_0\left(\frac{1}{2}\right)$$

STOP  $\alpha^* = \alpha_2 = \frac{1}{2}$ .

Ainsi on obtient la nouvelle itérée  $X_1 = (x_1, y_1) = x_0 + \alpha_0 d_0$

$$f(x_0, y_0) = 3(1)^2 + (1)^2 = 4; \quad f(x_1, y_1) = f[(x_0, y_0) + \alpha^* d_0] = 0.25$$

$$f(x_1, y_1) < f(x_0, y_0)$$

On voit bien que le pas de Goldstein&Price [11] vérifié la descente de la fonction  $f$ .

## 2.5 La règle de Wolfe

La condition (2.4.3)-(2.4.4) « règle de Goldstein&Price [11] » peuvent exclure un minimum ce qui est peut être un inconvénient.

Les conditions de Wolfe [22] n'ont pas cet inconvénient.

### 2.5.1 Principe

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x_k \in \mathbb{R}^n$ ,  $d_k \in \mathbb{R}^n$  une direction de descente

On dit que le pas  $\alpha_k$  vérifie la règle de Wolfe faible [22] si on a:

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha_k \beta \varphi_k'(0) \quad \forall \beta \in ]0, 1[ \quad (2.5.1)$$

$$\varphi_k'(\alpha_k) \geq \sigma \varphi_k'(0) \quad \forall \sigma \in ]\beta, 1[ \quad (2.5.2)$$

Autrement dit; pour  $0 < \beta < \sigma < 1$  on a:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k \quad (2.5.3)$$

$$\nabla^T f(x_k + \alpha_k d_k) \geq \sigma \nabla^T f(x_k) d_k \quad (2.5.4)$$

Teste de Wolfe:

◆ Si

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha_k \beta \varphi_k'(0) \text{ et } \varphi_k'(\alpha_k) \geq \sigma \varphi_k'(0)$$

alors  $\alpha_k$  est convenit.

◆ Si

$$\varphi_k(\alpha_k) \geq \varphi_k(0) + \alpha_k \beta \varphi_k'(0)$$

alors  $\alpha_k$  est trop grand.

◆ Si

$$\varphi_k'(\alpha_k) \leq \sigma \varphi_k'(0)$$

alors  $\alpha_k$  est trop petit.

## 2.5.2 Représentation géométrique

$T$ : la tangente de graphe de  $\varphi_k(\alpha)$  qui passe par le point  $(0, \varphi_k(0))$

$$T := \{(\alpha, y) \mid y = \varphi_k(0) + \alpha\varphi_k'(0)\}$$

$\Delta_1$ : la droite qui passe par le point  $(0, \varphi_k(0))$  et qui a  $\beta\varphi_k'(0)$  coefficient de direction

$$\Delta_1 := \{(\alpha, y) \mid y = \varphi_k(0) + \alpha\beta\varphi_k'(0)\}$$

$\Delta_3$ : la droite qui a  $\sigma\varphi_k'(0)$  coefficient de direction

$$\Delta_3 := \{y = \sigma\varphi_k'(0)\}$$

On a:  $0 < \beta < \sigma < 1 \implies$

$$\sigma\varphi_k'(0) < \beta\varphi_k'(0) \implies \sigma\varphi_k'(0) < \alpha\beta\varphi_k'(0) \implies \sigma\varphi_k'(0) < \varphi_k(0) + \alpha\beta\varphi_k'(0)$$

On voit bien que  $\Delta_3$  est au dessous de  $\Delta_1$ .

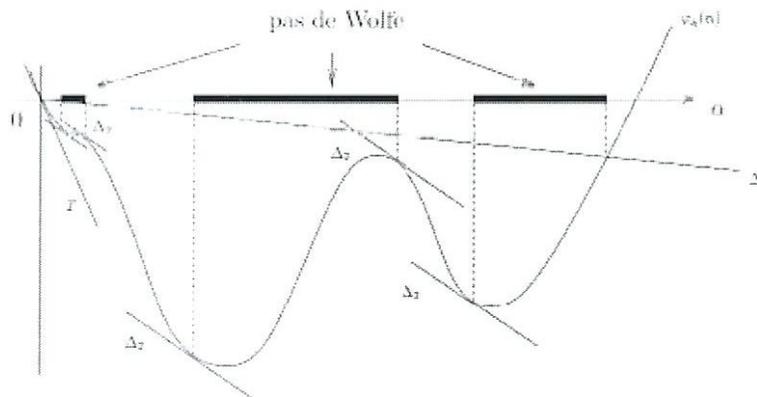


Figure 2.3 la recherche linéaire inexacte de Wolfe

### 2.5.3 Algorithme de la règle Wolfe

Dans cette partie nous allons présenter l'Algorithme de la règle de Wolfe [22] qui suit le schéma suivant :

#### Algorithme 2.4 La règle Wolfe

##### Etape 1: (initialisation)

$\alpha_{g,1} = \alpha_{d,1} = 0$  choisir  $\alpha_k > 0$ ,  $\beta \in ]0, 1[$ ,  $\sigma \in ]\beta, 1[$ , poser  $k=1$  et allez à l'étape 2.

##### Etape 2:

◆ Si  $\varphi_k(\alpha) \leq \varphi_k(0) + \alpha\beta\varphi'_k(0)$  et  $\varphi'_k(\alpha_k) \geq \sigma\varphi'_k(0)$  : STOP  $\alpha^* = \alpha_k$ .

◆ Si  $\varphi_k(\alpha) \geq \varphi_k(0) + \alpha\beta\varphi'_k(0)$  alors

$\alpha_{d,k+1} = \alpha_k$ ,  $\alpha_{g,k+1} = \alpha_{g,k}$  et allez à l'étape 3.

◆ Si  $\varphi'_k(\alpha_k) \leq \sigma\varphi'_k(0)$  alors

$\alpha_{g,k+1} = \alpha_k$ ,  $\alpha_{d,k+1} = \alpha_{d,k}$  et allez à l'étape 3.

##### Etape 3:

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

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

(par exemple on prend  $\alpha_{k+1} = \frac{\alpha_{g,k+1} + \alpha_{d,k+1}}{2}$ )

Remplacer  $k$  par  $k+1$  et allez à l'étape 2.

### 2.5.4 Règle de Wolfe forte

Pour certains Algorithmes il est parfois nécessaire d'avoir une condition plus restrictive que (2.5.4) pour cela la deuxième condition (2.5.4) est remplacée par

$$|\nabla^T f(x_k + \alpha_k d_k)| \leq \sigma |\nabla^T f(x_k) d_k| = -\sigma \nabla^T f(x_k) d_k$$

On aura donc les conditions de Wolfe forte [25]

$$\varphi_k(\alpha_k) \leq \varphi_k(0) + \alpha_k \beta \varphi'_k(0) \quad \forall \beta \in ]0, 1[ \quad (2.5.5)$$

$$|\varphi'_k(\alpha_k)| \leq -\sigma \varphi'_k(0) \quad \forall \sigma \in ]\beta, 1[ \quad (2.5.6)$$

Autrement dit; pour  $0 < \beta < \sigma < 1$  on a:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \beta \nabla^T f(x_k) d_k \quad (2.5.7)$$

$$|\nabla^T f(x_k + \alpha_k d_k)| \leq -\sigma \nabla^T f(x_k) d_k \quad (2.5.8)$$

**Remarque 2.5.1** On voit bien que les conditions de Wolfe forte [25] impliquent les conditions de Wolfe faible [22]

$$\begin{aligned} |\varphi'_k(\alpha_k)| \leq -\sigma \varphi'_k(0) &\Leftrightarrow \sigma \varphi'_k(\alpha_k) \leq \varphi'_k(\alpha_k) \leq -\sigma \varphi'_k(0) \\ \implies \varphi'_k(\alpha_k) \geq \sigma \varphi'_k(0) &\implies \text{la condition 2 de Wolf faible} \end{aligned}$$

### 2.5.5 Théorème d'existence du pas de Wolfe

**Théorème 2.5.1** Si  $\varphi_k : \mathbb{R}_+ \mapsto \mathbb{R}$ ; définie par (2.1.2)  $f(x_k + \alpha d_k) = \varphi_k(\alpha)$  est continue et bornée inférieurement, si  $d_k$  est une direction de descente en  $x_k$  ( $\varphi'_k(0) < 0$ )

et si  $\beta \in ]0, 1[$  et  $\sigma \in ]\beta, 1[$  alors l'ensemble des pas vérifiant la règle de Wolfe (3.53) et (3.5.4) est non vide.

#### Démonstration

On a

$$\begin{aligned} f(x_k + \alpha d_k) &= \varphi_k(\alpha) \\ \psi_\beta(\alpha) &= f(x_k) + \alpha \beta \nabla^T f(x_k) d_k \end{aligned}$$

Le développement de Taylor-Yong en  $\alpha = 0$  de  $\varphi_k$  est :

$$\varphi_k(\alpha) = f(x_k + \alpha d_k) = f(x_k) + \alpha \nabla^T f(x_k) d_k + \xi(\alpha) \text{ Ou } \xi(\alpha) \longrightarrow 0, \alpha \rightarrow 0$$

Et comme  $\beta \in ]0, 1[$  et  $\varphi'_k(0) = \nabla^T f(x_k) d_k < 0$  on déduit:

$$f(x_k) + \alpha \nabla^T f(x_k) d_k < f(x_k) + \alpha \beta \nabla^T f(x_k) d_k \quad \text{pour } \alpha > 0$$

On voit que pour  $\alpha > 0$  assez petit on a:

$$\varphi_k(\alpha) < \psi_\beta(\alpha)$$

De ce qui précède du fait que  $\varphi_k$  est bornée inférieurement et

$$\psi_\beta(\alpha) \longmapsto -\infty; \alpha \longmapsto +\infty$$

On déduit que la fonction  $\psi_\beta(\alpha) - \varphi_k(\alpha)$  possède la propriété :

$$\begin{cases} \psi_\beta(\alpha) - \varphi_k(\alpha) > 0 \text{ pour } \alpha \text{ assez petit} \\ \psi_\beta(\alpha) - \varphi_k(\alpha) < 0 \text{ pour } \alpha \text{ assez grand} \end{cases}$$

Donc s'annule au moins une fois pour  $\alpha > 0$ .

En choisissant le plus petit de ces zéros on voit qu'il existe  $\bar{\alpha} > 0$  tel que

$$\varphi_k(\bar{\alpha}) = \psi_\beta(\bar{\alpha}) \text{ et } \varphi_k(\alpha) < \psi_\beta(\alpha) \text{ pour } 0 < \alpha < \bar{\alpha}.$$

La formule des accroissements finis fournit alors un nombre  $\hat{\alpha}$ ,  $0 < \hat{\alpha} < \bar{\alpha}$  tel que:

$$\begin{aligned} \varphi_k(\bar{\alpha}) - \varphi_k(0) &= \bar{\alpha}\varphi_k'(\hat{\alpha}) = \bar{\alpha}\nabla^T f(x_k + \hat{\alpha}d_k)d_k \\ &\implies \bar{\alpha}\beta\nabla^T f(x_k)d_k = \bar{\alpha}\nabla^T f(x_k + \hat{\alpha}d_k)d_k \\ &\implies \nabla^T f(x_k + \hat{\alpha}d_k)d_k = \beta\nabla^T f(x_k)d_k \geq \sigma\nabla^T f(x_k)d_k \end{aligned}$$

Car  $0 < \beta < \sigma < 1$  et  $\nabla^T f(x_k)d_k < 0$ .

Donc  $\hat{\alpha}$  satisfait (2.5.4) D'autre part  $\alpha = \hat{\alpha}$  satisfait (2.5.3) en effet,  $\hat{\alpha}$  satisfait  $\varphi_k(\hat{\alpha}) < \psi_\beta(\hat{\alpha})$  n'est autre que:

$$f(x_k + \hat{\alpha}d_k) \leq f(x_k) + \hat{\alpha}\beta\nabla^T f(x_k)d_k \blacksquare$$

En pratique, on utilise des algorithmes spécifique pour trouver un pas de Wolfe [22].

### 2.5.6 Exemple

Prenons

$$f(x, y) = 3x^2 + y^2$$

Soit  $X_0 = (x_0, y_0) = (1, 1)$  et  $d_0 = -\nabla f(x_0, y_0) = (-2, -3)$ .

1) Calculons  $\varphi_0(\alpha)$ ,  $\varphi_0(0)$ ,  $\varphi'_0(\alpha)$ ,  $\varphi'_0(0)$

$$\varphi_0(\alpha) = f(x_0 + \alpha d_0) = f[(1, 1) + \alpha(-2, -3)]$$

$$\varphi_0(\alpha) = 21\alpha^2 - 18\alpha + 4; \varphi_0(0) = 4$$

$$\varphi'_0(\alpha) = 42\alpha - 18; \varphi'_0(0) = -18$$

Cherchons un point stationnaire de  $\varphi_0$ , autrement dit un pas pour lequel  $\varphi'_0(\alpha) = 0$

$$\varphi'_0(\alpha) = 0 \implies 42\alpha - 18 = 0 \implies \alpha^* = \frac{9}{21}$$

On a :  $\varphi''_0(\alpha) = 42 > 0 \implies \alpha^*$  est une solution minimale locale stricte

2) Trouvons le pas de Wolfe :

-)  $\alpha_{g,1} = \alpha_{d,1} = 0$  On pose que:  $\alpha_1 = 1$  et  $\beta = 0.2, \sigma = 0.6$ ;

$$g_0(\alpha) = \varphi_0(0) + \alpha\beta\varphi'_0(0) = 4 - 3.6\alpha.$$

On va Calculer les  $\varphi_0$  et  $g_0$  puis les comparé jusqu'à ce que la condition (2.5.3) soit vérifié

$$\varphi_0(1) = 7; \quad g_0(1) = 0.4 \implies \varphi_0(1) > g_0(1)$$

-)  $\alpha_{g,1} = 0, \alpha_{d,1} = 1$  on prend  $\alpha_2 = 0.4$

$$\varphi_0(0.4) = 0.16; \quad g_0(0.4) = 2.56 \implies \varphi_0(0.4) < g_0(0.4)$$

On va Calculer les  $\varphi'_0$  et  $\sigma\varphi'_0$  puis les comparé jusqu'à ce que (2.5.4) soit vérifié

-)  $\alpha_2 = 0.4$

$$\varphi'_0(0.4) = -1.2; \quad \sigma\varphi'_0(0) = -18 * 0.6 = -10.8 \implies \varphi'_0(0.4) > \sigma\varphi'_0(0)$$

STOP  $\alpha^* = \alpha_2 = 0.4$ .

Ainsi on obtient la nouvelle itérée  $X_1 = (x_1, y_1) = x_0 + \alpha_0 d_0$

$$f(x_0, y_0) = 3(1)^2 + (1)^2 = 4; \quad f(x_1, y_1) = f[(x_0, y_0) + \alpha^* d_0] = 0.16$$

$$f(x_1, y_1) < f(x_0, y_0)$$

On voit bien que le pas de Wolfe [22] vérifié la descente de la fonction  $f$ .

# CHAPITRE 3

---

## Modification de La recherche linéaire inexacte d'Armijo

Dans ce chapitre on s'intéresse à une généralisation de la recherche linéaire inexacte d'Armijo [1] dite règle d'Armijo modifiée [20]. Cette règle a été proposée en 2005 par Z. J. Shi et J. Shen

Avant de donner le principe de cette nouvelle recherche linéaire inexacte on va donner la définition d'une direction de descente suffisante ainsi que deux hypothèses qui seront utilisées par la suite

### 3.1 Définition et hypothèses

**Définition 3.1.1** *On dit que le vecteur  $d_k$  vérifie la condition de descente suffisante si et seulement si*

$$\nabla^T f(x) d_k \leq -c \|\nabla f(x)\|^2 \quad (3.1.1)$$

Où  $c > 0$ .

**Remarque 3.1.1** Choisir une direction vérifiant la condition de descente suffisante revient à dire :

$$\cos \langle -\nabla f_k, d_k \rangle = \frac{-\nabla^T f_k d_k}{\|\nabla f_k\| \|d_k\|} \geq \eta_0 \quad (3.1.2)$$

Où  $\eta_0 \in [0,1]$  et  $\langle -\nabla f_k, d_k \rangle$  est l'angle entre  $\nabla f_k$  et  $d_k$

**Hypothèse 1 :** Supposons que la fonction possède une limite inférieure sur l'ensemble  $L_0 = \{x \in \mathbb{R}^n : f(x) < f(x_0)\}$ , où  $x_0$  est donnée

**Hypothèse 2 :** Supposons que le gradient  $\nabla f(x)$  de  $f(x)$  est Lipschitzien continue dans un ensemble convexe ouvert  $B$  cela contient  $L_0$ ; C'est à dire il existe  $L$  tel que

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \quad \forall x, y \in B \quad (3.1.3)$$

## 3.2 Principe de la recherche linéaire inexacte d'Armijo modifiée

On va décrire une nouvelle recherche linéaire inexacte pour laquelle la règle d'Armijo [1] n'est qu'un cas particulier.

On pose  $s_k = \frac{-\nabla^T f(x_k) d_k}{L_k \|d_k\|^2}$  on voit bien que  $s_k > 0$ ; soit  $\lambda \in ]0, 1[$  et  $\beta \in ]0, \frac{1}{2}[$ ,  $\mu \in [0, 2[$  et soit  $\alpha_k$  le plus grand  $\alpha$  dans l'ensemble  $\{s_k, \lambda s_k, \lambda^2 s_k, \dots\}$  vérifiant:

$$f(x_k + \alpha d_k) \leq f(x_k) + \alpha \beta [\nabla^T f(x_k) d_k + \frac{1}{2} \alpha \mu L_k \|d_k\|^2] \quad (3.2.1)$$

**Remarque 3.2.1** Supposons que  $\alpha_k$  est définie par la règle d'Armijo [1] et que  $\alpha'_k$  définie par la nouvelle règle d'Armijo modifiée [20] on voit bien que  $\alpha'_k > \alpha_k$ .

Soit  $A^c$  désigne l'ensemble des pas d'Armijo modifiée [20] et  $A^e$  l'ensemble des pas d'Armijo alors  $A^c \subset A^e$ , en fait si  $L_k = L$  et il existe  $\alpha_k$  vérifier la condition d'Armijo  $\square$  alors  $\alpha_k$  est satisfait (3.2.1)

D'ailleurs si  $\mu = 0$  alors la nouvelle règle de recherche linéaire inexacte (Armijo modifiée [20]) se réduit à la règle d'Armijo [1].

Dans l'Algorithme modèle, les Algorithmes correspondants à la règle de recherche linéaire (3.2.1) sont noté par (C')

### 3.3 L'analyse de la convergence globale

**Théorème 3.3.1** *Supposons que Hypothèse 1 et Hypothèse 2 sont vérifiées la direction  $d_k$  satisfait  $\nabla^T f(x_k)d_k < 0$  et  $\alpha_k$  déterminé par la règle de recherche linéaire d'Armijo modifiée l'Algorithme (C') génère une suite infinie  $\{x_k\}$  et*

$$0 < L_k < m_k L \quad (3.3.1)$$

Où  $m_k$  un entier positive et  $m_k < M_0 < +\infty$  avec  $M_0$  étant un grand constante positive alors

$$\sum_{k=1}^{+\infty} (\nabla^T f(x_k)d_k / \|d_k\|)^2 < +\infty \quad (3.3.2)$$

#### Démonstration

Pour démontrer l'inégalité (3.3.2) on va étudier les deux cas suivant :

◆ Si  $s_k$  satisfait la condition de la recherche linéaire d'Armijo modifiée [20]:

Dans ce cas on prend  $k \in K_1$  où  $K_1 = \{k : \alpha_k = s_k\}$  alors

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq \alpha_k \beta [\nabla^T f(x_k)d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2] \\ &= -\beta [\nabla^T f(x_k)d_k / L_k \|d_k\|^2] [\nabla^T f(x_k)d_k - \frac{1}{2} \mu \nabla^T f(x_k)d_k] \\ &= -\left[ \beta(1 - \frac{1}{2} \mu) / L_k \right] (\nabla^T f(x_k)d_k)^2 / \|d_k\|^2; \end{aligned}$$

Ainsi

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\left[ \beta(1 - \frac{1}{2} \mu) / L_k \right] (\nabla^T f(x_k)d_k)^2 / \|d_k\|^2 \quad k \in K_1 \quad (3.3.3)$$

Soit

$$\eta_k = -\beta(1 - \frac{1}{2} \mu) / L_k, \quad k \in K_1$$

D'après (3.3.1) nous avons

$$\begin{aligned} \eta_k &= -\beta(1 - \frac{1}{2} \mu) / L_k \\ &\leq -\beta(1 - \frac{1}{2} \mu) / m_k L \\ &\leq -\beta(1 - \frac{1}{2} \mu) / M_0 L < 0. \end{aligned}$$

Soit

$$\eta' = \beta(1 - \frac{1}{2}\mu)/M_0L.$$

Ceci et (3.3.3) implique que  $\eta_k \leq -\eta'$  et

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\eta'((\nabla^T f(x_k) d_k) / \|d_k\|)^2, \quad k \in K_1 \quad (3.3.4)$$

◆ Si  $s_k$  ne vérifie pas la condition de la recherche linéaire d'Armijo modifiée [20]:

On choisit  $k \in K_2$  où  $K_2 = \{k : \alpha_k < s_k\}$ , dans ce cas il existe  $\lambda \in ]0, 1[$  telle que  $\alpha_k < \lambda s_k$  ne vérifie pas aussi la condition de la recherche linéaire d'Armijo modifiée [20].

Dans ce cas puisque le pas  $\alpha_k$  ne satisfait pas à la règle d'Armijo modifiée [20], alors toute valeur  $\alpha$  supérieure à  $\alpha_k$  ne peut satisfaire à cette condition, ainsi pour  $\alpha = \frac{\alpha_k}{\lambda}$  la règle d'Armijo modifiée [20] n'est pas satisfaite c'est-à-dire

$$f(x_k + \alpha_k d_k / \lambda) - f(x_k) > \beta \alpha_k / \lambda [\nabla^T f(x_k) d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2 / \lambda]$$

En utilisant le théorème de la valeur moyenne sur la côté gauche de l'égalité ci dessus nous voyons qu'il existe  $\theta_k \in [0, 1]$  tel que:

$$\alpha_k \nabla^T f(x_k + \theta_k \alpha_k d_k / \lambda) d_k / \lambda > \beta \alpha_k / \lambda [\nabla^T f(x_k) d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2 / \lambda];$$

Donc

$$\nabla^T f(x_k + \theta_k \alpha_k d_k / \lambda) d_k > \beta [\nabla^T f(x_k) d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2 / \lambda]. \quad (3.3.5)$$

D'après Hypothèse 2, l'inégalité de Cauchy Schwarz, et (3.3.5), on obtient

$$\begin{aligned} \alpha_k L \|d_k\|^2 / \lambda &\geq \|\nabla^T f(x_k + \theta_k \alpha_k d_k / \lambda) - \nabla^T f(x_k)\| \cdot \|d_k\| \\ &\geq [\nabla^T f(x_k + \theta_k \alpha_k d_k / \lambda) - \nabla^T f(x_k)]^T d_k \\ &> -(1 - \beta) \nabla^T f(x_k) d_k + \frac{1}{2} \alpha_k \beta \mu L_k \|d_k\|^2 / \lambda. \end{aligned}$$

Donc

$$\alpha_k L \|d_k\|^2 / \lambda > -(1 - \beta) \nabla^T f(x_k) d_k,$$

Ce qui implique que

$$\alpha_k > -[\lambda(1 - \beta)]/L [\nabla^T f(x_k) d_k / \|d_k\|^2], \quad k \in K_2 \quad (3.3.6)$$

Posons

$$s'_k = -[\lambda(1 - \beta))/L][\nabla^T f(x_k)d_k/\|d_k\|^2], \quad k \in K_2$$

Nous avons

$$s_k > \alpha_k > s'_k, \quad k \in K_2 \quad (3.3.7)$$

D'après (3.3.1) et (3.3.7) , nous avons

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq \beta \alpha_k [\nabla^T f(x_k) d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2] \\ &\leq \beta \max_{s_k \geq \alpha \geq s'_k} \{ \alpha [\nabla^T f(x_k) d_k + \frac{1}{2} \alpha \mu L_k \|d_k\|^2] \} \\ &\leq \beta \max_{s_k \geq \alpha \geq s'_k} \{ \alpha [\nabla^T f(x_k) d_k + \frac{1}{2} s_k \mu L_k \|d_k\|^2] \} \\ &= \beta s'_k (1 - \frac{1}{2} \mu) \nabla^T f(x_k) d_k \\ &= -[\lambda \beta (1 - \lambda) (1 - \frac{1}{2} \mu) / L] (\nabla^T f(x_k) d_k / \|d_k\|)^2. \end{aligned}$$

On pose

$$\eta'' = \lambda \beta (1 - \lambda) (1 - \frac{1}{2} \mu) / L, \quad (3.3.8)$$

Nous avons

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\eta'' (\nabla^T f(x_k) d_k / \|d_k\|)^2, \quad k \in K_2 \quad (3.3.9)$$

Poser

$$\eta'_0 = \min(\eta', \eta'');$$

D'après (3.3.4) et (3.3.9), nous affirmons que si  $s_k$  vérifie la recherche linéaire d'Armijo modifiée [20] ou non nous avons :

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\eta'_0 (\nabla^T f(x_k) d_k / \|d_k\|)^2, \quad \forall k. \quad (3.3.10)$$

Comme on a supposé que la fonction possède une limite inférieure sur l'ensemble  $L_0 = \{x \in \mathbb{R}^n : f(x) < f(x_0)\}$ , où  $x_0$  est donnée et D'après (3.3.10) on voit bien que (3.3.2) est vérifié, ce qui achève la démonstration.

**Corollaire 3.3.1** *Si les conditions du théorème 3.3.1 sont vérifiées, alors*

$$\lim_{k \rightarrow \infty} (\nabla^T f(x_k) d_k / \|d_k\|)^2 = 0 \quad (3.3.11)$$

En fait, l'hypothèse 3 peut être remplacée par l'hypothèse plus faible suivante :

**Hypothèse 3** Le gradient  $\nabla f(x_k)$  de  $f(x_k)$  est uniformément continue sur un ensemble convexe ouvert  $B$  qui contient  $L_0$ .

**Théorème 3.3.2** *Supposons que Hypothèse 1 et Hypothèse 2 sont vérifiées, la direction  $d_k$  satisfait  $\nabla^T f(x_k)d_k < 0$  et  $\alpha_k$  est déterminé par la règle d'Armijo modifiée, l'algorithme (C') génère une suite infinie  $\{x_k\}$  et*

$$0 \leq L_k \leq M'_0, \quad (3.3.12)$$

Où  $M'_0$  est une grande constante positive alors

$$\lim_{k \rightarrow +\infty} (-\nabla^T f(x_k)d_k / \|d_k\|) = 0. \quad (3.3.13)$$

### Démonstration

Pour  $\alpha_k = s_k$  ( $k \in K_1$ ), en suivant les mêmes étapes dans la démonstration de théorème 3.3.1, on trouve (3.3.4) en passant à la limite quand  $k \rightarrow +\infty$  on obtient :

$$\begin{aligned} \lim_{k \rightarrow +\infty} f(x_k + \alpha_k d_k) - f(x_k) &= 0 \\ \implies \lim_{k \rightarrow +\infty} -\eta'(\nabla^T f(x_k)d_k) / \|d_k\|^2 &= 0 \end{aligned}$$

Comme  $\eta' \neq 0$  alors

$$\lim_{k \in K_1, k \rightarrow \infty} (\nabla^T f(x_k)d_k / \|d_k\|) = 0. \quad (3.3.14)$$

Dans le cas où  $k \in K_2$  :

On suppose que  $\alpha_k$  vérifie la règle d'Armijo modifiée [20] c'est-à-dire

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \beta \alpha_k [\nabla^T f(x_k)d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2]$$

Comme  $k \in K_2$  donc  $\alpha_k < s_k$  ainsi

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq \beta \alpha_k [\nabla^T f(x_k)d_k + \frac{1}{2} s_k \mu L_k \|d_k\|^2] \\ &\leq \beta \alpha_k (1 - \frac{1}{2} \mu) \nabla^T f(x_k)d_k. \end{aligned}$$

D'après Hypothèse 1, nous avons

$$\lim_{k \in K_2, k \rightarrow \infty} (\nabla^T f(x_k)d_k) = 0 \quad (3.3.15)$$

S'il existe  $\xi_0 > 0$  et un sous ensemble infinie  $K_3 \subset K_2$  tel que

$$-\nabla^T f(x_k) / \|d_k\| \geq \xi_0, \quad \forall k \in K_3. \quad (3.3.16)$$

Puis par (3.3.15) et (3.3.16) , nous avons

$$\lim_{k \in K_3, k \rightarrow \infty} \alpha_k \|d_k\| = 0. \quad (3.3.17)$$

Comme  $x_{k+1} = x_k + \alpha_k d_k$  on a :

$$\alpha_k \|d_k\| = \|x_{k+1} - x_k\| = 0$$

Donc  $x_{k+1} = x_k$  est un point stationnaire

$$\implies \nabla f(x_k) = 0. \quad (\bullet)$$

D'après (3.3.5), nous avons

$$\nabla^T f(x_k + \theta_k \alpha_k d_k / \lambda) d_k \geq \beta \nabla^T f(x_k) d_k. \quad k \in K_3 \quad (3.3.18)$$

Où  $\theta_k \in [0, 1]$  et définie dans la démonstration de théorème 3.3.1 .par l'inégalité de Cauchy Schwarz et (3.3.18), nous avons

$$\begin{aligned} \|\nabla f(x_k + \theta_k \alpha_k d_k / \lambda) - \nabla f(x_k)\| &= \|\nabla f(x_k + \theta_k \alpha_k d_k / \lambda) - \nabla f(x_k)\| \|d_k\| / \|d_k\| \\ &\geq [\nabla f(x_k + \theta_k \alpha_k d_k / \lambda) - \nabla f(x_k)]^T d_k / \|d_k\| \\ &\geq -(1 - \beta) \nabla^T f(x_k) d_k / \|d_k\|, \quad k \in K_3. \end{aligned}$$

D'après Hypothèse 3 et (3.3.17), on obtient

$$\lim_{k \in K_3, k \rightarrow \infty} (-\nabla^T f(x_k) / \|d_k\|) = 0.$$

Ce qui contredit (3.3.16). Cela montre que

$$\lim_{k \in K_2, k \rightarrow \infty} (-\nabla^T f(x_k) / \|d_k\|) = 0. \quad (3.3.19)$$

D'après (3.3.14), (3.3.19), et notant que  $K_1 \cup K_2 = \{1, 2, 3, \dots\}$ , nous affirmons que (3.3.13) est vérifié.

Depuis Hypothèse 2 implique Hypothèse 3, théorème 3.3.1 et essentiellement un corollaire de théorème 3.3.2

### 3.4 Taux de la convergence linéaire

Pour analyser le taux de convergence, nous supposons que la suite  $\{x_k\}$  génère par l'Algorithme converge vers  $x^*$  nous faisons l'hypothèse suivante :

**Hypothèse 4 :**  $\nabla^2 f(x^*)$  est une matrice symétrique définie positive et  $f(x)$  est deux fois différentiable sur un voisinage  $V_0(x^*, \xi_0)$  de  $x^*$ .

**Lemme 3.4.1** *Suppose que Hypothèse 4 est vérifié. alors il existe  $\xi > 0$  et  $0 < m' < M'$  tel que Hypothèse 1 et Hypothèse 2 sont vérifiés pour  $x_0 \in V(x^*, \xi) \subseteq V_0(x^*, \xi_0)$  et*

$$m' \|y\|^2 \leq y^T \nabla^2 f(x) y \leq M' \|y\|^2 \quad \forall x, y \in V(x^*, \xi) \quad (3.4.1)$$

$$(1/2)m' \|x - x^*\|^2 \leq f(x) - f(x^*) \leq (1/2)M' \|x - x^*\|^2 \quad \forall x \in V(x^*, \xi) \quad (3.4.2)$$

$$m' \|x - y\|^2 \leq (\nabla f(x) - \nabla f(y))^T (x - y) \leq M' \|x - y\|^2 \quad \forall x, y \in V(x^*, \xi) \quad (3.4.3)$$

Ainsi

$$m' \|x - x^*\|^2 \leq \nabla^T f(x) (x - x^*) \leq M' \|x - x^*\|^2 \quad \forall x \in V(x^*, \xi) \quad (3.4.4)$$

D'après (3.4.4) et (3.4.3), on obtient aussi de l'inégalité de Cauchy Schwarz

$$m' \|x - x^*\| \leq \|\nabla^T f(x)\| \|x - x^*\| \leq M' \|x - x^*\| \quad \forall x \in V(x^*, \xi) \quad (3.4.5)$$

Et

$$\|\nabla f(x) - \nabla f(y)\| \leq M' \|x - y\| \quad \forall x, y \in V(x^*, \xi) \quad (3.4.6)$$

#### Démonstration

D'après références [6] et [27].

**Lemme 3.4.2** *Si Hypothèse 1 et Hypothèse 2 sont vérifiés, la direction  $d_k$  satisfait la propriété (3.1.2) pour chaque itération, Algorithme (C') génère une suite infinie  $\{x_k\}$ ; puis il existe  $\eta > 0$  tel que :*

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \eta \|\nabla^T f(x_k)\|^2, \quad \forall k. \quad (3.4.7)$$

**Démonstration**

Par théorème 3.3.1, (3.1.2), et (3.10), nous avons

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq -\eta'_0 (\nabla^T f(x_k) d_k / \|d_k\|)^2 \\ &\leq -\eta'_0 (\nabla^T f(x_k) d_k / \|d_k\| \|\nabla^T f(x_k)\|)^2 \cdot (\|\nabla^T f(x_k)\|)^2 \\ &\leq -\eta'_0 \eta_0^2 \|\nabla^T f(x_k)\|^2 \end{aligned}$$

Laisser

$$\eta = \eta'_0 \eta_0^2;$$

On obtient cela (3.4.7) qui détient.

**Théorème 3.4.1** *Si Hypothèse 4 est vérifié, la direction  $d_k$  satisfait la propriété (3.1.2) pour chaque itération, Algorithme (C') génère une suite infinie  $\{x_k\}$  et  $x_k \in V(x^*, \xi)$  pour suffisamment grand  $k$ . puis  $\{x_k\} \rightarrow x^*$  au moins linéairement.*

**Démonstration**

D'après Hypothèse 4, lemme 3.4.1, lemme 3.4.2, et (3.4.5), il s'ensuit que

$$\lim_{k \rightarrow \infty} x_k = x^*.$$

D'après (3.4.5) et lemme 3.4.2, on obtient

$$\begin{aligned} f(x_k) - f(x_k + \alpha_k d_k) &> \eta \|\nabla^T f(x_k)\|^2 \\ &\geq \eta m'^2 \|x_k - x^*\|^2 \end{aligned}$$

D'après (3.4.2)

$$f(x_k) - f(x_k + \alpha_k d_k) \geq (\eta m'^2 / M') (f(x_k) - f(x^*)).$$

Posons

$$\theta = m' \sqrt{2\eta / M'}.$$

Nous pouvons prouver que  $\theta < 1$ . En fait par la définition de  $\eta$  et notant que  $M' \leq L$ , on obtient

$$\begin{aligned} \theta^2 &= 2m'^2 \eta / M' \leq 2m'^2 \eta'_0 \eta_0^2 / M' \leq 2m'^2 \eta'_0 / M' \\ &\leq 2m'^2 \eta / M' \leq [\beta(1 - \frac{1}{2}\mu) / M_0 L] (2m'^2 / M') \\ &\leq [2\beta(1 - \frac{1}{2}\mu) / M_0] (m'^2 / M'^2) \end{aligned}$$

Comme  $0 < m' < M'$  alors  $\frac{m'}{M'} < 1$

$$\theta^2 \leq (2\beta/M_0) < 1.$$

Soit

$$w = \sqrt{1 - \theta^2};$$

Évidemment  $w < 1$  on obtient de l'inégalité ci-dessus d'après

$$\begin{aligned} f(x_k) - f(x_k + \alpha_k d_k) &\geq \theta^2(f(x_k) - f(x^*)) \\ f(x_k + \alpha_k d_k) - f(x_k) &\leq \theta^2(f(x^*) - f(x_k)) \\ f(x_k + \alpha_k d_k) - f(x_k) - f(x^*) &\leq \theta^2(f(x^*) - f(x_k)) - f(x^*) \\ f(x_k + \alpha_k d_k) - f(x^*) &\leq (1 - \theta^2)(f(x_k) - f(x^*)) \\ &= w^2(f(x_k) - f(x^*)) \\ &\leq \dots \\ &\leq w^{2k}(f(x_1) - f(x^*)). \end{aligned}$$

D'après lemme 3.1 et l'inégalité ci-dessus nous avons

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq (2/m')(f(x_k + \alpha_k d_k) - f(x^*)) \\ &< w^{2k}[2(f(x_1) - f(x^*))/m']; \end{aligned}$$

Ainsi

$$\begin{aligned} \|x_{k+1} - x^*\| &\leq w^k \sqrt{2(f(x_1) - f(x^*))/m'} \\ \|x_{k+1} - x^*\|^{1/k} &\leq w \sqrt{2(f(x_1) - f(x^*))/m'}^{1/k} \end{aligned}$$

Donc

$$\begin{aligned} \lim_{k \rightarrow +\infty} \|x_{k+1} - x^*\|^{1/k} &= \lim_{k \rightarrow +\infty} [w^k \sqrt{2(f(x_1) - f(x^*))/m'}]^{1/k} \\ &= \lim_{k \rightarrow +\infty} w \sqrt{2(f(x_1) - f(x^*))/m'}^{1/k} \\ &= w \lim_{k \rightarrow +\infty} \sqrt{2(f(x_1) - f(x^*))/m'}^{1/k} \\ &= w < 1. \end{aligned}$$

Ce qui montre que  $\{x_k\}$  converge vers  $x^*$  au moins linéairement. ■

# 4 Simulations numériques

---

Dans cette section, on va présenter deux choix différents du calcul du pas via la méthode du gradient, il s'agit des recherches linéaires inexactes d'Armijo [1] et d'Armijo modifiée [20].

Les algorithmes des recherches linéaires inexactes d'Armijo [1] et d'Armijo modifiée [20] assurent la décroissance de la fonction objective et ceci à chaque itération de l'algorithme. Dans nos simulations numériques on n'a pas limité le nombre des itérations nécessaires dans l'algorithme de la recherche linéaire, et le pas initial est donnée par  $\alpha_0 = 100$ .

Les paramètres sont donnés :  $\rho = 0.1$ ,  $\delta = 0.01$  et  $\beta = 0.95$ .

Les programmes sont écrits sous scilab 5.4.0 et exécutés dans un ordinateur Intel PC i5 avec une mémoire 3.10 GHz et un système Windows sept.

Le critère d'arrêt est  $\|g_k\| \leq \varepsilon$ , ou  $\varepsilon = 10^{-3}$ .

Dans nos tests numériques on va reprendre le même exemple étudié.

Prenons :

$$f(x, y) = 3x^2 + y^2$$

Avec un point initial  $x_0 = (10, 10)$ .

## 4.1 Méthode du gradient avec la règle d'Armijo modifiée

Le tableau ci dessous donne la valeur de la fonction objective et la norme du gradient des cinq premières itérations, pour la méthode du gradient avec la règle d'Armijo modifiée [20]

L'itération	$x_k$	$f_k$	$\ g_k\ $
0	(10,10)	400	63.245553
1	(- 10.205585,3.2648049)	323.12087	61.580669
2	(9.3843482,1.1758406)	265.58057	56.355178
3	(- 8.6291955,0.4234866)	223.56838	51.7821
4	(7.9348094,0.1525214)	188.90686	47.609834
5	(- 7.2963001, 0.0549316)	159.711	43.777939

## 4.2 Méthode du gradient avec la règle d'Armijo

Le tableau ci dessous donne la valeur de la fonction objective et la norme du gradient des cinq premières itérations, pour la méthode du gradient avec la règle d'Armijo [1]

L'itération	$x_k$	$f_k$	$\ g_k\ $
0	(10,10)	400	63.245553
1	(- 74.959215,- 18.319738)	17192.264	451.24524
2	(68.927293, - 6.5979721)	14296.448	413.77423
3	(- 63.380756,- 2.3763023)	12057.007	380.31423
4	(58.280545,- 0.8558406)	10190.598	349.68746
5	(- 53.590745,- 0.3082365)	8615.999	321.54506

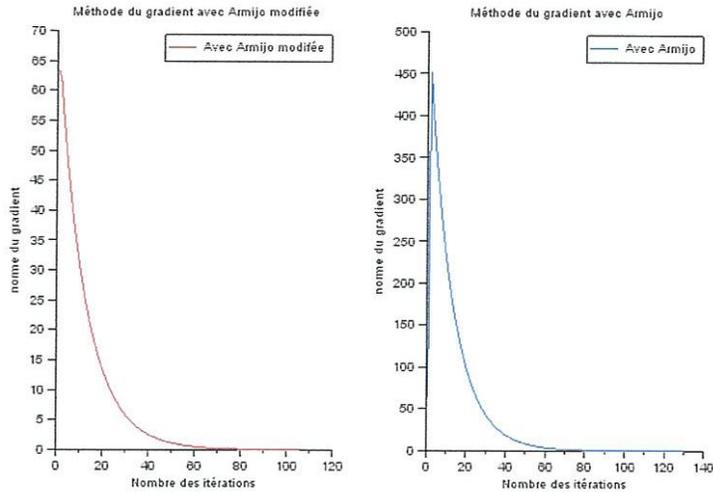


Fig 4.1: Méthode du gradient avec Armijo et Armijo modifiée

D'après ces deux tableaux on voit bien que la règle d'Armijo modifiée [20] accélère la convergence de la méthode du gradient mieux que celle d'Armijo [1].

### Conclusion

Dans ce mémoire on a étudié brièvement les recherches linéaires inexacts d'Armijo [1], Armijo modifiée [20], Goldstein-Price [11] et de Wolfe [22].

En imposant moins de condition sur le choix du pas, la règle d'Armijo [1] est considérée la meilleur numériquement. La règle d'Armijo modifiée [20] est une généralisation de celle d'Armijo [1].

D'après nos simulations numériques on a illustré que la règle d'Armijo modifiée [20] accélère la convergence de la méthode du gradient mieux que celle d'Armijo [1].

---

# Bibliographie

- [1] L. Armijo (1966), Minimization of function having lipschitz continous first partial derivatives, Pacific Journal of Mathematics, Vol. 16(1), pp.1-3.
- [2] M. S. Bazaraa, H. D. Sherali, et C. M. Shetty (1993), Nonlinear Programming, Theory and Algorithms, Wiley-Interscience.
- [3] 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.
- [4] M. Bergounioux (2001), Optimisation et Contrôle des Systèmes Linéaires, Dunod.
- [5] A. Cauchy (1847), Analyse mathématique, Méthode générale pour la résolution des systèmes d'équations simultanées, Comptes Rendus de l'Académie des Sciences de Paris, t-25, pp. 536-538.
- [6] A. I. Cohen(1981), Stepsize Analysis for Descent Methods, Journal for Optimizaton Theory and Applications, Vol. 33, pp. 187-205.
- [7] T. Diallo (2006), Travail pratique de master, Sujet : Etude et Illustration De Méthodes Itératives D'optimisation Non Linéaire, Responsable : Benjamin Leroy-Beaubeir, Ecole polytechnique Fédérale de Lausanne.
- [8] A. V. Fiacco and G. P. McCormick (1968), Nonlinear Programming, John Wiley, New York.

- 
- [9] J. C. Gilbert (2007), *Eléments d'Optimisation Différentiable : Théorie et Algorithmes*, Notes de cours, École Nationale Supérieure de Techniques Avancées, Paris.
- [10] A.A. Goldstein (1967), *Constructive Real Analysis*, A Harper International Edition.
- [11] A. A. Goldstein (1965), On steepest descent, *SIAM J. on Control A*, Vol. 3, No. 1, pp. 147-151.
- [12] A.A. Goldstein and J.F. Price (1969), An effective algorithm for minimization, *Num.Math.*, 10, pp. 184-189.
- [13] D. G. Luenberger (1969), *Optimization by Vector Space Methods*, John Wiley and Sons, Inc. New York.
- [14] M. Minoux (1983), *Programmation Mathématique, Théorie et Algorithmes*, tome 1, Dunod.
- [15] J. Nocedal (1991), Theory of algorithm for unconstrained optimization, *Acta Numerica*, pp. 199-242.
- [16] M.J.D. Powell (1984), Non convex minimization calculation and the conjugate gradient method, in : *Lecture Notes in Mathematics 1066* (Springer, Berlin), pp. 122-141.
- [17] M.J.D. Powell (1986), Convergence properties of algorithms for nonlinear optimization, *SIAM rev.*, 28, pp. 487-500.
- [18] M.J.D. Powell (1971), On the convergence of the variable metric algorithm; *J. Inst. Math. Appl*, 7, pp. 21-36.
- [19] M.J.D. Powell (1977), Restart procedures for the conjugate gradient method, *Math Programming*, 2, pp 241-254.
- [20] Z. J. Shi, J. Shen(2005), New Inexact Line Search Method for Unconstrained Optimization, *Journal of Optimization and Application* : Vol. 127, No. 2, pp. 425-446.
- [21] D. Touati-Ahmed and C. Storey (1990), Efficient hybrid conjugate gradient techniques, *JOTA*, 64, pp. 379-397.

- [22] P. Wolfe (1961), A duality theorem for nonlinear programming, *Quart. Appl. Math.*, 19, pp. 239-244.
- [23] P. Wolfe (1969), Convergence conditions for ascent methods, *SIAM Review*, 11, pp. 226-235.
- [24] P. Wolfe (1971), Conditions for ascent methods some corrections 2, *SIAM Review*, 13, pp.185-188.
- [25] P. Wolfe (1963), *Methods of Nonlinear Programming*, in *Recent Advances in Mathematical Programming* (Eds R. L. Graves and P. Wolfe), McGraw-Hill, New York.
- [26] G. Zoutendijk (1970), *Nonlinear Programming Computational Methods*, *Integer and Nonlinear Programming*, North Holland, Amsterdam.
- [27] Y. Yuan(1993), *Numerical Methods for Nonlinear Programming*, Shanghai Scientific and Technical Publishers.