

17/15/10.058

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

Université 8 Mai 1945 – Guelma

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



Mémoire de Fin d'Etude  
Master Académique en Mathématiques  
Option : EDP

**THEME**

**Méthodes à régions de confiance**

Présenté par :

Bousslah Imen

Dirigé par :

*Dr R. Mellal*

Jury :

Examineur 1 : Dr N.Azzouza ..... Univ-Guelma

Examineur 2 : Dr D. Bellaouar ..... Univ-Guelma

## Remerciements

Je tiens à exprimer toutes mes reconnaissances au Docteur **R.Meilal**, de m'avoir proposé ce sujet de recherche et d'avoir dirigé mon travail

Je lui témoigne aussi ma gratitude pour son soutien sa grande disponibilité et surtout ses conseils et ses encouragements tout au long mes recherches

J'exprime également mes chaleureux remerciements au :

Docteur **N . Azzouza** et docteur **D.Bellaouar**, pour l'honneur qu'ils m'ont fait d'avoir accepté de faire partie de ce jury.

## DEDICACES

### **Vouloir c'est pouvoir**

Avec beaucoup de respect et d'amour je dédie ce modeste travail :

**\*\*Ma mère Halima, pour toute la tendresse. Et le souci qu'elle a fait pour moi.**

**\*\* Mon père Ahmed, vous qui avez fait preuve d'amour et sacrifices.**

Je sais bien que vous méritez beaucoup plus, mais pour le moment Je n'ai que mon amour et mémoire à vous offrir, que dieu vous garde.

**\*\*A mes très chers frères Soufiane, Loukman.**

**\*\* A mes très chères sœurs Amel.**

**\*\*A tous mes amis qui m'ont tant aimé, encouragé et tous ceux qui m'ont supporté, réconforté et inspiré : Hanane, Raffika et Nadia .**

**\*\*A tous mes collègues et le group de mathématique.**

**\*\*A toute la famille : Bousslah.**

**\*\*A tous ceux dont j'ai oublié de mentionner le nom.**

# Table des matières

<b>1</b>	<b>Notions préliminaires</b>	<b>7</b>
1.1	Aspect général des algorithmes . . . . .	7
1.2	La notion de convergence globale . . . . .	8
1.3	Conditions d'optimalité des problèmes d'optimisation sans contraintes . . . . .	8
1.3.1	Conditions nécessaires d'optimalité . . . . .	9
1.3.2	Conditions suffisantes d'optimalité . . . . .	10
1.3.3	Cas convexe . . . . .	11
<b>2</b>	<b>Méthode à direction de descente et Recherche linéaire in- exacte</b>	<b>12</b>
2.1	Méthode à direction de descente . . . . .	13
2.1.1	Algorithme de la méthode à direction de descente . . . . .	14
2.2	Recherche linéaire . . . . .	14
2.2.1	Objectifs à atteindre . . . . .	15
2.2.2	Type de recherche linéaire . . . . .	16
2.2.3	Schéma général de résolution . . . . .	17
2.2.4	Intervalle de sécurité . . . . .	17
2.2.5	Algorithme de base . . . . .	18
2.3	Recherches linéaires inexactes . . . . .	18
2.3.1	Schéma des recherches linéaires inexactes : . . . . .	19
2.3.2	La règle d'Armijo . . . . .	20
2.3.3	La Règle de Goldstein and Price. . . . .	22
2.3.4	La règle de Wolfe . . . . .	25
2.4	Contribution des recherches linéaires inexacte dans la conver- gence des méthodes à direction de descente . . . . .	28



<b>3</b>	<b>Méthode à région de confiance</b>	<b>32</b>
3.1	Principe des méthodes à région de confiance . . . . .	33
3.2	Algorithme de la méthode à région de confiance . . . . .	34
3.3	Analyse du sous- problème quadratique . . . . .	36
	3.3.1 Condition d'optimalité . . . . .	36
	3.3.2 Résolution numérique de $(RC_k)$ . . . . .	37
3.4	Résultats de convergence . . . . .	38
	3.4.1 Convergence avec modèle du premier ordre . . . . .	38
	3.4.2 Convergence avec modèle du second ordre . . . . .	43

## Résumé

On s'intéresse dans ce mémoire aux méthodes à direction de descente avec des recherches linéaires inexactes et les méthodes à région de confiance. On va présenter ces méthodes, donner leurs principes, algorithmes et les théorèmes assurant leurs convergence.

**Mots clés :** Région de confiance, Méthodes à direction de descente, Recherche linéaires inexactes, Optimisation sans contraintes, Algorithmes, Convergence.

## Introduction

Les méthodes à région de confiance et les méthodes à direction de descente servent à résoudre des problèmes d'optimisation sans contrainte, c'est-à-dire :

$$\min_{x \in \mathbb{R}^n} f(x) \quad (0.1)$$

La fonction  $f$  est supposée différentiable.

Les méthodes à direction de descente utilisent en général le gradient ou le hessien de la fonction objectif  $f$  afin de construire une direction de descente  $d$ , or on ne peut pas garantir la convergence vers un minimum, même local. C'est pourquoi, ces méthodes sont couplées à une recherche linéaire effectuée le long de la direction  $d$  afin de déterminer un pas  $\alpha$  qui réduise la fonction objectif  $f$  de telle sorte que la décroissance soit assurée, c'est-à-dire  $f(x + \alpha d) < f(x)$ . Ainsi, on retrouve dans l'algorithme la propriété de convergence globale vers un minimum local.

La technique de région de confiance peut être vue comme une méthode de globalisation d'algorithme pour la résolution du problème (0.1), ce qui veut dire qu'elle permet de forcer la convergence des itérés lorsque le démarrage des algorithmes se fait en un point éloigné d'une solution.

Le principe des méthodes à région de confiance consiste à remplacer le problème d'optimisation initial (0.1) par une suite de sous-problèmes d'optimisation, plus simples à résoudre. Dans chaque sous-problème, la fonction objectif  $f$  est remplacée par une fonction modèle  $\psi$ , à un itéré courant  $x$ .

Dans ce mémoire on aborde la méthode à région de confiance appliquée aux problèmes quadratiques parce que dans le cas non quadratique aucun résultat de convergence n'a été démontré.

La région de confiance est définie comme la région à l'intérieur de laquelle la confiance est donnée à la fonction modèle qui fournit une bonne approximation de la fonction objectif. Ainsi, une contrainte supplémentaire existe sur les pas  $s$ .

Si la décroissance de la fonction  $f$ , évaluée entre le point courant  $x$  et le point  $\hat{x}$  qui minimise la fonction modèle  $\psi_k$ , n'est pas jugée suffisante, le rayon de la région de confiance  $\Delta$  est diminué, et en général, la direction de recherche est modifiée. Une bonne détermination du rayon de la région de confiance est alors un point essentiel dans la performance de telles méthodes.

De ce point de vue, cette technique concurrence les méthodes à direction de descente. Un peu plus difficile à mettre n'en ouvre que cette dernières, elles ont divers avantages, notamment en ce qui concerne les résultats de convergence, qui sont plus forts sous des hypothèses plus faibles. Elles ont aussi la réputation d'être plus robustes, c'est-à-dire de pouvoir résoudre des problème plus difficile, moins bien conditionnés.

Le mémoire est divisé en trois chapitres :

**Chapitre 1 :** Dans ce chapitre on va introduire les outils de base.

**Chapitre 2 :** L'objectif de ce chapitre consiste à décrire les méthodes à direction de descente et les principales règles de recherche linéaire inexacte (Wolfe, Armijo et Goldstein and Price ), et on termine par étudier la contribution de la recherche linéaire inexacte dans la convergence des algorithmes à directions de descente.

**Chapitre 3 :** Dans ce chapitre on étudie la méthode à région de confiance appliquée à des modèles quadratiques. On commencera par donner son principe ainsi que son algorithme. On passera ensuite à l'analyse du sous problème quadratique dont on citera une condition nécessaire et suffisante pour la convergence et on terminera par donner les théorèmes de convergence pour des modèles du premier et second ordre.



# Chapitre 1

## Notions préliminaires

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . On appelle problème de minimisation sans contraintes le problème (1.1) suivant :

$$(p) \begin{cases} \min f(x) \\ x \in \mathbb{R}^n \end{cases} \quad (1.1)$$

Où  $x = (x_1, x_2, \dots, x_n)^t$  est un vecteur de  $\mathbb{R}^n$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est la fonction que l'on désire minimiser appelée (fonction objectif, fonction-coût, ou critère du problème) .

Dans ce chapitre on va présenter quelques généralités sur les problèmes de minimisation sans contraintes et leurs algorithmes.

### 1.1 Aspect général des algorithmes

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

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

où  $d_k$  détermine la direction de déplacement à partir du point  $x_k$  et  $\alpha_k$  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  $\alpha_k$  à chaque itération.

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

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  $f$ .

## 1.2 La notion de convergence globale

La question importante est donc celle de la convergence de la suite  $\{x_k\}_k$  vers « ce qu'il faut » quelque soit l'itéré initiale (convergence globale)

**Définition 1.2.1** *On dit qu'un algorithme est globalement convergent (ou encore : possède la propriété de convergence globale) si, quel que soit le point de départ  $x_0$  choisi, la suite  $\{x_k\}_k$  générée par cet algorithme (ou une sous-suite) converge vers un point satisfaisant une condition nécessaire d'optimalité.*

La notion de convergence globale concerne le fait qu'on aura limite même si l'itéré initial est très éloigné de la limite  $x^*$ . Au contraire, on aura seulement une convergence locale si une suite  $\{x_k\}_k$  converge vers  $x^*$ .

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

**Définition 1.3.1** *Considérons le problème de minimisation sans contraintes  $(P)$ .*

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

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

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

$$f(x^*) \leq f(x), \forall x \in V_\epsilon(x^*).$$

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

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

**Définition 1.3.2** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . La fonction  $f$  est dite différentiable en  $\tilde{x}$  s'il existe un vecteur  $\nabla f(\tilde{x}) \in \mathbb{R}^n$  appelé gradient de  $f$  en  $\tilde{x}$  et une fonction  $B(\tilde{x}, x - \tilde{x}) \rightarrow 0$  quand  $x \rightarrow \tilde{x}$  tel que :

$$f(x) = f(\tilde{x}) + [\nabla f(\tilde{x})]^T(x - \tilde{x}) + \|x - \tilde{x}\| B(\tilde{x}, x - \tilde{x}), \forall x \in \mathbb{R}^n.$$

### 1.3.1 Conditions nécessaires d'optimalité

**Théorème 1.3.1** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  différentiable au point  $\tilde{x} \in \mathbb{R}^n$ . Soit  $d \in \mathbb{R}^n$  telle que  $\nabla f(\tilde{x})^t d < 0$ . Alors il existe  $\delta > 0$  tel que

$$f(\tilde{x} + \alpha d) < f(\tilde{x}) \text{ pour tout } \alpha \in ]0, \delta[.$$

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

**Démonstration 1** Comme  $f$  est différentiable en  $\tilde{x}$  alors

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

où  $\lambda(\tilde{x}; \alpha d) \rightarrow 0$  pour  $\alpha \rightarrow 0$ . Ceci implique :

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

et comme  $\nabla f(\tilde{x})^t d < 0$  et  $\lambda(\tilde{x}; \alpha d) \rightarrow 0$  pour  $\alpha \rightarrow 0$ , il existe  $\delta > 0$  tel que

$$\nabla f(\tilde{x})^t d + \|d\| \lambda(\tilde{x}; \alpha d) < 0 \text{ pour tout } \alpha \in ]0, \delta[$$

et par conséquent on obtient :

$$f(\tilde{x} + \alpha d) < f(\tilde{x}) \text{ pour tout } \alpha \in ]0, \delta[.$$

**Lemme 1.3.1** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  différentiable au point  $\tilde{x} \in \mathbb{R}^n$ . Si  $\tilde{x}$  est un minimum local de  $(P)$  alors  $\nabla f(\tilde{x}) = 0$ .

**Démonstration 2** Par contre, on suppose que  $\nabla f(\tilde{x}) \neq 0$ . Si on pose  $d = -\nabla f(\tilde{x})$ , on obtient :

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

et par le théorème précédent (1.3.1), il existe  $\delta > 0$  tel que :

$$f(\tilde{x} + \alpha d) < f(\tilde{x}) \text{ pour tout } \alpha \in ]0, \delta[$$

mais ceci est contradictoire avec le fait que  $\tilde{x}$  est un minimum local, d'où :

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

**Théorème 1.3.2** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  deux fois différentiable au point  $\tilde{x} \in \mathbb{R}^n$ . Si  $\tilde{x}$  est un minimum local de (P) alors  $\nabla f(\tilde{x}) = 0$  et la matrice hessienne de  $f$  au point  $\tilde{x}$ , qu'on note  $H(\tilde{x})$ , est semi définie positive.

**Démonstration 3** Le corollaire ci-dessus montre la première proposition, pour la deuxième proposition on a :

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

où  $\lambda(\tilde{x}; \alpha d) \rightarrow 0$  pour  $\alpha \rightarrow 0$ . Ceci implique :

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

Comme  $\tilde{x}$  est un minimum local alors  $f(\tilde{x} + \alpha d) \geq f(\tilde{x})$  pour  $\alpha$  suffisamment petit, d'où

$$d^t H(\tilde{x}) d + \|d\|^2 \lambda(\tilde{x}; \alpha d) \geq 0 \text{ pour } \alpha \text{ petit.}$$

En passant à la limite quand  $\alpha \rightarrow 0$ , on obtient que  $d^t H(\tilde{x}) d \geq 0$ , d'où  $H(\tilde{x})$  est semi définie positive.

### 1.3.2 Conditions suffisantes d'optimalité

**Théorème 1.3.3** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  deux fois différentiable au point  $\tilde{x} \in \mathbb{R}^n$ . Si  $\nabla f(\tilde{x}) = 0$  et  $H(\tilde{x})$ , est définie positive alors  $\tilde{x}$  est un minimum local strict de (P).

**Démonstration 4**  $f$  est deux fois différentiable au point  $\tilde{x}$ . Alors  $\forall x \in \mathbb{R}^n$ , on obtient :

$$f(x) = f(\tilde{x}) + \nabla f(\tilde{x})^t (x - \tilde{x}) + \frac{1}{2} (x - \tilde{x})^t H(\tilde{x}) (x - \tilde{x}) + \|x - \tilde{x}\|^2 \lambda(\tilde{x}; x - \tilde{x})$$

où  $\lambda(\tilde{x}; x - \tilde{x}) \rightarrow 0$  pour  $x \rightarrow \tilde{x}$

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

$$f(x_k) \leq f(\tilde{x}), x_k \neq \tilde{x}, \forall k.$$

Posons  $d_k = \frac{x_k - \tilde{x}}{\|x_k - \tilde{x}\|}$ , Donc  $\|d_k\| = 1$  et on obtient :

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



et comme  $\|d_k\| = 1, \forall k$  alors il existe  $\{d_k\}_{k \in \mathbb{N}_1 \subset \mathbb{N}}$  telle que  $d_k \rightarrow d$  pour  $k \rightarrow \infty$  et  $k \in \mathbb{N}_1$ . On a bien sûr  $\|d\| = 1$ . Considérons donc  $\{d_k\}_{k \in \mathbb{N}_1}$  et le fait que  $\lambda(x; x - x) \rightarrow 0$  pour  $k \rightarrow \infty$  et  $k \in \mathbb{N}_1$ . Alors :

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

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

### 1.3.3 Cas convexe

**Définition 1.3.3** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

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

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

Pour tous point  $x_1$  et  $x_2$  de  $\mathbb{R}^n$  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

**Théorème 1.3.4** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  telle que  $f$  est convexe et différentiable. Alors  $x_*$  est un minimum global de  $f$  si et seulement si  $\nabla f(x_*) = 0$ .

**Démonstration 5** Voir[6]

**Remarque 1.3.1** Dans le cas où  $f$  est convexe, alors tout minimum local est aussi global. De plus si  $f$  est strictement convexe, alors tout minimum local devient non seulement global mais aussi unique.

## Chapitre 2

# Méthode à direction de descente et Recherche linéaire inexacte

Notre problème consiste à minimiser une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . On s'intéresse à la classe des méthodes à direction de descente, or on ne peut pas garantir la convergence vers un minimum, même local. C'est pourquoi, ces algorithmes sont couplés à une recherche linéaire. Ainsi, on retrouve dans l'algorithme la propriété de convergence globale vers un minimum local.

On suppose connaître la direction de descente  $d_k$  au point  $x_k$ . La recherche linéaire consiste à trouver  $\alpha_k$  de façon à diminuer la fonction  $f$  suffisamment le long de cette direction. L'objectif de ce chapitre consiste à décrire les principales règles de recherche linéaire inexacte (Wolfe, Armijo, Goldstein and Price) et d'étudier la contribution de la recherche linéaire inexacte à la convergence des algorithmes à directions de descente.

On commencera notre chapitre par un petit rappel de l'algorithme de descente ensuite on décrira la recherche linéaire inexacte, où on va présenter des théorèmes garantissant la convergence des méthodes à directions de descente qui les utilisent.

## 2.1 Méthode à direction de descente

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

**Définition 2.1.1** On dit que  $d$  est une direction de descente de  $f$  en  $x \in \mathbb{R}^n$  si  $\nabla^T f(x)d < 0$ . Ou encore que  $d$  fait avec l'opposé du gradient  $-\nabla f(x)$  un angle  $\theta$  strictement plus petit que  $90^\circ$  :

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

Par la suite l'angle  $\theta_k$  jouera un rôle important dans le processus de la convergence.

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

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

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

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 (2.1) par la récurrence :

$$x_{k+1} = x_k + \alpha_k \cdot d_k, k \geq 1 \tag{2.1}$$

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.

### 2.1.1 Algorithme de la méthode à direction de descente

Algorithme 2.1 (Algorithme de la méthode à direction de descente)

**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.

## 2.2 Recherche linéaire

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 :

$$\min f(x_k + \alpha \cdot d_k), \alpha \in \mathbb{R} \quad (2.2)$$

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 est une direction de descente de l'objectif :

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

et exécute un pas dans cette direction :

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



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 :  $h_k(\alpha) = f(x_k + \alpha d_k)$ .

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

### 2.2.1 Objectifs à atteindre

L'idée générale consiste à proposer une méthode de sélection du pas  $\alpha_k$  qui permette de prouver la convergence

$$\lim_{x \rightarrow \infty} \|\nabla f(x)\| \rightarrow 0.$$

d'avoir une bonne vitesse de convergence, et qui soit simple à implémenter.

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 :

$$f(x_k + \alpha_k d_k) = f(x_k) + \text{"un terme négatif"} \quad (2.3)$$

Le terme négatif, disons  $\nu_k$ , 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  $f(x_k + \alpha_k d_k) < f(x_k)$

#### 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 (2.3) 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 :

Si on prend

$$0 < \alpha_k \leq \frac{\epsilon}{2^k \|d_k\|}$$

la suite  $x_k$  générée par (2.1) est de Cauchy, puisque pour  $1 \leq l < k$  on a :

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

Donc  $x_k$  converge, disons vers un point  $x$ . En prenant  $l = 1$  et  $k \rightarrow \infty$  dans l'estimation ci-dessus, on voit que  $x \in \bar{B}(x_1, \epsilon)$  et donc  $x$  ne saurait être solution s'il n'y a pas de solution dans  $\bar{B}(x_1, \epsilon)$ .

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 :

$$h_k : \mathbb{R}_+ \rightarrow \mathbb{R}; \alpha_k \rightarrow h_k(\alpha) = f(x_k + \alpha d_k) \quad (2.4)$$

## 2.2.2 Type de recherche linéaire

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

- a) Les recherches linéaires exactes.
- b) Les recherches linéaires inexactes.

Les recherches linéaires exactes, malgré qu'elles n'aboutissent qu'à une solution optimale approchée, nécessitent beaucoup d'observations à chaque itération de l'algorithme principale.

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.

On aurait aimé restrictif notre étude sur ce domaine sympathique, malheureusement, c'est très rare de rencontrer des problèmes quadratiques, et donc par exemple pour une fonction non linéaire arbitraire,

- 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.

Au lieu de demander que  $\alpha_k$  minimise, on préfère imposer des conditions moins restrictives, plus facilement vérifiées, qui permettent toute fois de contribuer à la convergence des algorithmes. En particulier, il n'y aura plus un unique pas (ou quelques pas) vérifiant ces conditions mais tout un intervalle de pas (ou plusieurs intervalles), ce qui rendra d'ailleurs leur recherche plus aisée. C'est ce que l'on fait avec les règles d'Armijo, de Goldstein et de Wolfe décrites dans la prochaine section.

### 2.2.3 Schéma général de résolution

Un algorithme de recherche linéaire résolvant le problème (2.2) génère une suite de valeurs  $\{\alpha_1, \alpha_2, \dots, \alpha_k, \dots\}_{k \in \mathbb{N}^*}$  est un test à trois sorties étant donné un  $\alpha_1 > 0$  (une valeur initial) répond si :

- $\alpha_k$  est satisfaisant donc c'est une solution
- $\alpha_k$  est trop grand
- $\alpha_k$  est trop petit

On appellera  $\alpha^*$  un pas optimal,  $\alpha_g$  un pas trop petit et  $\alpha_d$  un pas trop grand

### 2.2.4 Intervalle de sécurité

Dans la plupart des algorithmes d'optimisation modernes, on ne fait jamais de recherche linéaire exacte, car trouver  $\alpha_k$  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  $\alpha^*$  qui assure une décroissance suffisante de  $f$ . Cela conduit à la notion d'intervalle de sécurité.

**Définition 2.2.1** *On dit que  $[\alpha_g, \alpha_d]$  est un intervalle de sécurité s'il permet de classer les valeurs de  $\alpha$  de la façon suivante :*

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

### 2.2.5 Algorithme de base

Algorithme 2.2 (Algorithme de base)

#### Etape 0 : (initialisation)

$\alpha_g = \alpha_d = 0$ , choisir  $\alpha_1 > 0$ , poser  $k = 1$  et aller à l'étape 1 :

#### Etape 1 :

- Si  $\alpha_k$  convient, poser  $\alpha^* = \alpha_k$  et on s'arrête.
- Si  $\alpha_k$  est trop petit on prend  $\alpha_{g,k+1} = \alpha_k, \alpha_d = \alpha_d$  et on va à l'étape 2.
- Si  $\alpha_k$  est trop grand on prend  $\alpha_{d,k+1} = \alpha_k, \alpha_g = \alpha_g$  et on va 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 faut maintenant préciser quelles sont les relations sur  $h_k$  qui vont nous permettre de caractériser les valeurs de  $\alpha$  convenables, ainsi que les techniques utilisées pour réduire l'intervalle.

## 2.3 Recherches linéaires inexactes

On considère la situation qui est typique pour l'application de la technique de recherche linéaire à l'intérieur de la méthode principale multidimensionnelle.

Sur une itération  $k$  de la dernière méthode nous avons l'itération courante  $x_k \in \mathbb{R}^n$  et la direction de recherche  $d_k \in \mathbb{R}^n$  qui la direction de descente pour notre objectif :  $f : \mathbb{R}^n \rightarrow \mathbb{R}$

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



Le but est de réduire "de façon importante" la valeur de l'objectif par un pas  $x_k \rightarrow x_{k+1} = x_k + \alpha_k d_k$  de  $x_k$  dans la direction  $d_k$ . Pour cela de nombreux mathématiciens (Armijo, Goldstein, Wolfe, Albaali, Lemaréchal, Fletcher...) ont élaboré plusieurs règles. L'objectif de cette section consiste à présenter les principales tests. D'abord présentons le schéma d'une recherche linéaire inexacte.

### 2.3.1 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 le quel :  $h_k(\alpha_k) < h_k(0)$ ; ( $f(x_k + \alpha_k d_k) < f(x_k)$ ) Le schéma de l'algorithme est donc :

Algorithme 2.3 (Schéma général des recherches linéaires inexactes)

#### 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  $\alpha_k$  est satisfaisant (suivant un certain critère) :  
STOP ( $\alpha^* = \alpha_k$ ).
- si  $\alpha_k$  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  $\alpha_k$  est trop grand (suivant un certain critère), nouvel intervalle :  
 $[\alpha_{g,k+1} = \alpha_g, \alpha_{d,k+1} = \alpha_k]$  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 critère (s)  $\alpha$  est trop petit ou trop grand ou satisfaisant.

### 2.3.2 La règle d'Armijo

On bute à réduire "de façon importante" la valeur de l'objectif par un pas  $x_k \rightarrow x_{k+1} = x_k + \alpha_k d_k$  de  $x_k$  dans la direction  $d_k$ , tel que  $f(x_k + \alpha_k d_k) < f(x_k)$ . Or cette condition de décroissance stricte n'est pas suffisante pour minimiser  $h_k$  au moins localement. Par exemple, avec la fonction

$$h : \mathbb{R} \rightarrow \mathbb{R} : x \rightarrow h(x) = x^2$$

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

La règle d'Armijo [1] impose une contrainte sur le choix de  $\alpha_k$  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 :

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k \nabla^T f(x_k) d_k \quad (2.5)$$

Elle est de la forme (2.3), car  $\rho$  devra être choisi dans  $]0, 1[$ .

Il faut qu'en  $\alpha_k$ , la fonction  $h_k$  prenne une valeur plus petite que celle prise par la fonction  $\psi_\rho(\alpha) : \alpha \rightarrow h_k(0) + \rho h'_k(0)\alpha$  autrement dit

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

Test d'Armijo :

◇ Si

$h_k(\alpha) \leq h_k(0) + \rho h'_k(0)\alpha$  autrement dit  $f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha \nabla^T f(x_k) d_k$  alors  $\alpha$  convient.

◇ Si

$h_k(\alpha) > h_k(0) + \rho h'_k(0)\alpha$  autrement dit  $f(x_k + \alpha d_k) > f(x_k) + \rho \alpha \nabla^T f(x_k) d_k$  alors  $\alpha$  est trop grand.

## Algorithme de la Règle d'Armijo

### Algorithme 2.4 (Règle d'Armijo)

#### Etape 0 : (initialisation)

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

#### Etape 1

- si  $h_k(\alpha_k) \leq h_k(0) + \rho h'_k(0)\alpha_k$  : STOP ( $\alpha^* = \alpha_k$ ).
- si  $h_k(\alpha_k) > h_k(0) + \rho h'_k(0)\alpha_k$ , alors  $\alpha_{d,k+1} = \alpha_k$ ,  $\alpha_{g,k+1} = \alpha_g$  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.

**Remarque 2.3.1** En pratique, la constante  $\rho$  est prise très petite, de manière à satisfaire (2.5) le plus facilement possible. Typiquement,  $\rho = 10^{-4}$ . Notons que cette constante ne doit pas être adaptée aux données du problème et donc que l'on ne se trouve pas devant un choix de valeur délicat.

**Remarque 2.3.2** L'inégalité (2.5) est toujours vérifiée si  $\alpha_k > 0$  est suffisamment petit. Cela se démontre aussi facilement ([1]).

En effet dans le cas contraire, on aurait une suite de pas strictement positifs  $\{\alpha_{k,i}\}_{i=1}$  convergeant vers 0 lorsque  $i \rightarrow \infty$  et tels que (2.5) n'ait pas lieu pour  $\alpha_k = \alpha_{k,i}$ . En retranchant  $f(x_k)$  dans les deux membres, en divisant par  $\alpha_{k,i}$  et en passant à la limite quand  $i \rightarrow \infty$ , on trouverait

$$\nabla^T f(x_k) d_k \geq \rho \nabla^T f(x_k) d_k$$

ce qui contredirait le fait que  $d_k$  est une direction de descente ( $\rho < 1$ ).

### Existence du pas d'Armijo

Pour assurer l'existence du pas d'Armijo en posant quelques conditions sur la fonction  $h_k$ .

**Théorème 2.3.1** Si  $h_k : \mathbb{R}_+ \rightarrow \mathbb{R}$ ; définie par (2.4)  $h_k(\alpha) = f(x_k + \alpha d_k)$  est continue et bornée inférieurement, si  $d_k$  est une direction de descente en  $x_k$  ( $h'_k(0) < 0$ ) et si  $\rho \in ]0, 1[$ , alors l'ensemble des pas vérifiant la règle d'Armijo est non vide.

**Démonstration 6** On a

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

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

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

$$h_k(\alpha) = f(x_k + \alpha d_k) = f(x_k) + \alpha \nabla^T f(x_k) d_k + \varepsilon(\alpha) \text{ ou } \varepsilon(\alpha) \rightarrow 0, \alpha \rightarrow 0.$$

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

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

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

$$h_k(\alpha) < \psi_\rho(\alpha)$$

De ce qui précède et du fait que  $h_k$  est bornée inférieurement, et

$$\psi_\rho(\alpha) \rightarrow -\infty; \alpha \rightarrow +\infty,$$

on déduit que la fonction  $\psi_\rho(\alpha) - h_k(\alpha)$  possède la propriété :

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

$$\psi_\rho(\alpha) - h_k(\alpha) < 0 \text{ un grand } \alpha$$

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

$$h_k(\bar{\alpha}) = \psi_\rho(\bar{\alpha}) \text{ et } h_k(\alpha) < \psi_\rho(\alpha) \text{ pour } 0 < \alpha < \bar{\alpha}$$

Ce qui achève la démonstration.

Décrivons maintenant la règle de Goldstein and Price.

### 2.3.3 La Règle de Goldstein and 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

$$h : \mathbb{R} \rightarrow \mathbb{R}; x \rightarrow h(x) = x^2$$



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  $\{x_k\}_{k=0}$  ne converge pas vers le  
 minimum zéro mais vers 1.

Les conditions de Goldstein and Price ([4, 1969]) 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  $\rho$  et  $\delta$  tels que  $0 < \rho < \delta < 1$ ; ces conditions sont :

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

$$f(x_k + \alpha d_k) \geq f(x_k) + \delta \alpha \nabla^T f(x_k) d_k$$

Autrement dit :

$$h_k(\alpha) \leq h_k(0) + \rho h'_k(0) \alpha \quad (2.6)$$

$$h_k(\alpha) \geq h_k(0) + \delta h'_k(0) \alpha \quad (2.7)$$

**Test de Goldstein and Price :**

◊ Si

$$h_k(0) + \delta h'_k(0) \alpha \leq h_k(\alpha) \leq h_k(0) + \rho h'_k(0) \alpha$$

Autrement dit

$$f(x_k) + \delta \alpha \nabla^T f(x_k) d_k \leq f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha \nabla^T f(x_k) d_k$$

Alors  $\alpha$  convient.

◊ Si

$$h_k(\alpha) > h_k(0) + \rho h'_k(0) \alpha$$

Autrement dit

$$f(x_k + \alpha d_k) > f(x_k) + \rho \alpha \nabla^T f(x_k) d_k$$

Alors  $\alpha$  est trop grand.

◊ Si

$$h_k(\alpha) < h_k(0) + \delta h'_k(0) \alpha$$

Autrement dit

$$f(x_k + \alpha d_k) < f(x_k) + \delta \alpha \nabla^T f(x_k) d_k$$

Alors  $\alpha$  est trop petit.

**Démonstration 10** Si le pas  $\alpha_k = \alpha_1$  est accepté, on a (2.11), car  $\alpha_1$  est uniformément positif. Dans le cas contraire, (2.5) n'est pas vérifiée avec un pas  $\alpha'_k \leq \frac{\alpha_k}{\tau}$ , c'est-à-dire  $f(x_k + \alpha'_k d_k) > f(x_k) + \rho \alpha'_k \nabla^T f(x_k) d_k$ . Comme  $f$  est continument différentiable, on a pour tout  $\alpha_k > 0$  :

$$f(x_k + \alpha_k d_k) = f(x_k) + \alpha_k \nabla^T f(x_k) d_k + \int_0^1 [\nabla f(x_k + \tau \alpha_k d_k) - \nabla f(x_k)]^T \alpha_k d_k dt$$

$$\Rightarrow f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k \nabla^T f(x_k) d_k + C \alpha_k^2 \|d_k\|^2$$

où  $C > 0$  est une constante. Avec l'inégalité précédente, et le fait que, on obtient :

$$\begin{cases} f(x_k + \alpha'_k d_k) - f(x_k) > \rho \alpha'_k \nabla^T f(x_k) d_k \\ f(x_k + \alpha'_k d_k) - f(x_k) \leq \alpha'_k \nabla^T f(x_k) d_k + C \alpha_k'^2 \|d_k\|^2 \end{cases}$$

$$\Rightarrow \rho \alpha'_k \nabla^T f(x_k) d_k \leq \alpha'_k \nabla^T f(x_k) d_k + C \alpha_k'^2 \|d_k\|^2$$

$$\Rightarrow -C \alpha_k'^2 \|d_k\|^2 \leq (1 - \rho) \alpha'_k \nabla^T f(x_k) d_k$$

or

$$\rho < 1 \Rightarrow 0 < 1 - \rho < 1 \Rightarrow \frac{1}{1 - \rho} > 1$$

d'où

$$\nabla^T f(x_k) d_k \geq \frac{-C}{1 - \rho} \alpha_k'^2 \|d_k\|^2 \Rightarrow -\nabla^T f(x_k) d_k \leq \frac{C}{1 - \rho} \alpha_k'^2 \|d_k\|^2$$

$$\Rightarrow |\nabla^T f(x_k) d_k| = \|\nabla^T f(x_k)\| \|d_k\| \cos \theta_k \leq \frac{C}{1 - \rho} \alpha_k'^2 \|d_k\|^2$$

ce qui permet de minorer  $\alpha' \|d_k\|$  et donc aussi  $\|d_k\|$  par une constante fois  $\|\nabla^T f(x_k)\| \|d_k\|$ . Cette minoration et l'expression suivante de (2.5)

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \underbrace{\rho \alpha_k \|\nabla^T f(x_k)\| \|d_k\|}_{\text{minoré}} \cos \theta_k$$

conduit à (2.11).

**Proposition 2.4.3** Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continument différentiable dans un voisinage de  $\bar{\Gamma} = \{x \in \mathbb{R}^n \rightarrow \mathbb{R} : f(x) \leq f(x_1)\}$ . On considère un algorithme à directions de descente  $d_k$ , qui génère une suite  $x_k$  en utilisant la recherche linéaire de Wolfe (2.8)-(2.9). Alors il existe une constante  $C > 0$  telle que, pour tout  $k = 1$ , la condition de Zoutendijk (2.11) est vérifiée.

**Démonstration 11** *D'après (2.8)*

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

et du fait que  $f$  est continument différentiable :

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

On en déduit (2.11).

## Chapitre 3

### Méthode à région de confiance

On bute à résoudre numériquement le problème d'optimisation sans contrainte :

$$\min_{x \in \mathbb{R}^n} f(x) \quad (3.1)$$

La fonction  $f$  est supposée différentiable.

La technique de région de confiance peut être vue comme une méthode de globalisation d'algorithme pour la résolution du problème (3.1), ce qui veut dire qu'elle permet de forcer la convergence des itérés lorsque le démarrage des algorithmes se fait en un point éloigné d'une solution.

De ce point de vue, cette technique concurrence les méthodes à direction de descente. Un peu plus difficile à mettre en œuvre que ces dernières, elles ont divers avantages, notamment en ce qui concerne les résultats de convergence, qui sont plus forts sous des hypothèses plus faibles. Elles ont aussi la réputation d'être plus robustes, c'est-à-dire de pouvoir résoudre des problèmes plus difficiles, moins bien conditionnés. Dans ce mémoire on aborde la méthode à région de confiance appliquée aux problèmes quadratiques parce que dans le cas non quadratique aucun résultat de convergence n'a été démontré.

Dans ce chapitre on étudie la méthode à région de confiance appliquée à des modèles quadratiques. On commencera par donner son principe ainsi que son algorithme. On passera ensuite à l'analyse du sous-problème quadratique dont on citera une condition nécessaire et suffisante pour la convergence et on terminera par donner les théorèmes de convergence pour des modèles du premier et second ordre.



### 3.1 Principe des méthodes à région de confiance

Les méthodes à région de confiance sont fondées sur le principe suivant, au point courant  $x_k \in \mathbb{R}^n$ , on suppose donner un modèle de la variation de  $f$  pour un incrément  $s$  de  $x^k$ , on s'intéresse dans ce chapitre aux modèles quadratiques. Le modèle quadratique est sous la forme :

$$\psi_k(s) = g_k^T s + \frac{1}{2} s^T H_k s;$$

Où  $g_k = \nabla f(x_k)$  est le gradient de  $f$  en  $x_k$  et  $H_k = \nabla^2 f(x_k)$  est le hessien de  $f$  en  $x_k$ . Le gradient et le hessien sont supposés être calculés pour le produit scalaire euclidien  $\langle u, v \rangle = \sum_i u_i v_i$ .

Soit le développement limité à l'ordre 2 :

$$f(x_k + s) \simeq f(x_k) + \psi_k(s).$$

Dans ce cas, le modèle quadratique est adéquat pour  $s$  voisin de 0.

Dans les méthodes à région de confiance, on considère que  $\psi_k$  est un modèle de la variation de  $f$  qui est acceptable dans un voisinage de la forme :

$$\{x \in \mathbb{R}^n : \|x\| \leq \Delta_k\},$$

Où  $\Delta_k > 0$  et  $\|\cdot\|$  est la norme euclidienne. Ce domaine est appelé la région de confiance du modèle  $\psi_k$  et  $\Delta_k$  est appelé le rayon de confiance.

Pour trouver l'accroissement  $s_k$  à donner à  $x_k$ , on minimise le modèle quadratique  $\psi_k$  sur la région de confiance.

On doit donc résoudre le sous-problème quadratique

$$(RC_k) \begin{cases} \min \psi_k(s) \\ \|s\| \leq \Delta_k \end{cases}$$

Soit  $s_k$  une solution (en générale approchée) de sous-problème quadratique.

Si  $f(x_k + s_k)$  est suffisamment plus petit que  $f(x_k)$ , on accepte le pas  $s_k$  et on passe à l'itération suivante avec  $x_{k+1} = x_k + s_k$ .

Dans le cas contraire, on résout à nouveau  $(RC_k)$  avec un rayon de confiance plus petit, car on considère que le modèle courant n'est pas fiable sur la boule  $\overline{B}(0, \Delta_k)$ .

Comme  $f$  et  $\psi_k$  coïncident au premier ordre, l'algorithme finit par trouver un rayon de confiance  $\Delta_k > 0$  pour lequel la solution de  $(RC_k)$  est acceptée.

On ajuste ensuite le modèle à l'itéré suivant  $x_{k+1}$  : On calcule  $g_{k+1}$ ,  $H_{k+1}$  et de  $\Delta_{k+1}$ . L'ajustement du rayon de confiance est basé sur la comparaison entre la décroissance réelle de  $f$  en passant de  $x_k$  à  $x_{k+1}$  et la décroissance prédite par le modèle, c'est-à-dire  $-\psi_k(s_k)$ . On peut alors passer à l'itération suivante.

### 3.2 Algorithme de la méthode à région de confiance

Décrivons de manière précise une itération de la méthode à région de confiance, celle passant de  $x_k$  à  $x_{k+1}$ .

On se donne des constantes indépendantes de l'indice  $k$  des itérations, qui sont :

- Des seuils de succès :  $0 < \omega_1 < \omega_2 < 1$ .

- Des facteurs de mise à jour de  $\Delta_k$  :  $0 < \tau_1 \leq \tau_2 < 1 < \tau_3$

Des valeurs typiques sont :  $\omega_1 = 10^{-2}$ ,  $\omega_2 = 0.8$ ,  $\tau_1 = \tau_2 = 0.5$ , et  $\tau_3 \in [2, 4]$ .

Étant donnés  $x_k, g_k, H_k$  et  $\Delta_k$ , on s'y prend alors de la manière suivante.

Algorithme 3.1 : Méthode à région de confiance (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 :** On calcule une solution approchée  $s_k$  de  $(RC'_k)$  telle que

$$\psi_k(s_k) < 0.$$

Si ce n'est pas possible on s'arrête.

**Etape 3 :** calcul du ratio  $\rho_k$  (critère de performance)

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{-\psi_k(s_k)}$$

**Etape 4 :** Actualisation du rayon de confiance :

**4.1** Si  $\rho_k \geq \omega_1$  L'étape est un succès et le pas est accepté.  $x_{k+1} = x_k + s_k$  et on choisit  $\Delta_{k+1} \in [\tau_2 \Delta_k, \Delta_k]$

- 4.2 Si  $\omega_1 \geq \rho_k \leq \omega_2$  L'étape est un succès et le pas est accepté.  $x_{k+1} = x_k + s_k$  et on choisit  $\Delta_{k+1} \in [\Delta_k, \tau_3 \Delta_k]$
- 4.3 Si  $\rho_k < \omega_1$  L'étape est un échec et le pas est rejeté  $x_{k+1} = x_k$  et on réduit le rayon de la région de confiance  $\Delta_k$  en choisissant  $\Delta_{k+1} \in [\tau_1 \|s_k\|, \tau_2 \|s_k\|]$

**Etape 5 :** mise à jour du modèle : calcul de  $g_{k+1}$  et  $H_{k+1}$  Poser  $k \leftarrow k+1$  et aller à l'étape 1.

**Remarque 3.2.1** On note dorénavant  $(s_k, \Delta_k)$  la solution approchée et le rayon de confiance obtenus à la sortie des étapes 2 et 4 respectivement. On ne tient donc pas compte des valeurs intermédiaires prises par ce couple lorsque la condition  $\rho_k > \omega_k$  n'est pas vérifiée de premier coup.

**Remarque 3.2.2** Si à l'étape 2, il n'est pas possible de réaliser  $\psi(s_k) < 0$ , c'est que  $g_k = 0$  et  $H_k$  est semi-définie positive. Ceci devrait vouloir dire qu'on ne peut pas faire décroître  $f$  non plus.

Donc si on s'impose un modèle  $\psi_k$  convexe ( $H_k$  semi-définie positive), on ne pourra trouver qu'un point stationnaires de  $f$ . Si  $f$  est convexe, c'est aussi un minimum.

Si  $f$  n'est pas convexe, le fait que  $H_k$  est semi-définie positive n'aide pas à éviter les points stationnaires qui ne sont pas des minimas, il est donc préférable d'y incorporer d'autres informations sur le hessien.

Cette observation met en évidence une différence importante avec les méthodes à directions de descente. Dans ces dernière, la direction de descente s'obtient également en minimisant un modèle quadratique, mais sur  $\mathbb{R}^n$  tout entier. Dans ce cas, il est essentiel d'avoir  $H_k$  (semi-) définie positive, si non le sous-problème quadratique n'a pas de solution. De plus, en prenant  $H_k$  définie positive, on est sûr que la direction est de descente pour  $f$ . Les conditions assurant le bon fonctionnement des région de confiance sont donc moins restrictives. C'est exactement sur ce point, grâce au fait qu'elles ne demandent pas d'avoir un modèle fortement convexe, que les régions de confiance gagnent en robustesse. Cela se traduira par un affaiblissement des hypothèses garantissant la convergence de cette approche.

**Remarque 3.2.3** Le ratio  $\rho_k$  calculé à l'étape 3 est appelé concordance (critère de performance). Celle-ci apprécie l'adéquation entre la fonction  $f$  et



son modèle  $\psi_k$  courant. Il s'agit en effet du rapport entre la décroissance réelle de  $f$  c'est-à-dire  $(f(x_k) - f(x_k + s_k))$  et la décroissance prédite par le modèle  $(-\psi_k(s_k) = \psi_k(0) - \psi_k(s_k) > 0)$ . Si ce rapport n'est pas assez grand ( $\rho_k \leq \omega_1$ ,  $\rho_k$  sera négatif si le pas  $s_k$  ne fait pas décroître  $f$ ), l'algorithme considère que le modèle ne permet pas de prédire la variation réelle de  $f$  et ne convient donc pas. En diminuant  $\Delta_k$  on accroît l'importance de la partie linéaire de  $\psi_k$  qui coïncide avec l'approximation de premier ordre de  $f$  :  $\nabla\psi_k(0) = \nabla f(x_k)$ .

**Remarque 3.2.4** Lorsqu'on passe à l'étape 4.1, on a  $\rho_k \geq \omega_1$ , ce qui s'écrit encore

$$f(x_k + s_k) \leq f(x_k) + \omega_1 \psi_k(s_k)$$

Cette condition rappelle la condition d'Armijo pour les méthode à direction de descente et recherche linéaire, mais le terme contrôlant la décroissance de  $f$ , ici  $\omega_k \psi_k(s_k) \leq 0$ , n'a pas le caractère linéaire qu'il a dans (3.5). Comme dans la règle d'Armijo), on force la réalisation de cette condition en faisant décroître  $\Delta_k$ . La structure du schéma (3.1) met clairement en évidence l'analogie entre les deux approches. L'étape 4 qui détermine  $\Delta_k$  dans le schéma (3.1) correspond à l'étape de calcul du pas  $\alpha_k$  en recherche linéaire. La différence essentielle est que la valeur de  $\Delta_k$  a une incidence sur la direction  $s_k \in \mathbb{R}^n$ , alors que  $\alpha_k$  n'affecte pas la direction  $d_k \in \mathbb{R}^n$  suivant la quelle on se déplace en recherche linéaire.

### 3.3 Analyse du sous- problème quadratique

Le sous problème quadratique à résoudre à chaque itération est de la forme suivante

$$\begin{cases} \min g_k^T(s) + \frac{1}{2} s^T H_k s \\ \|s\| \leq \Delta_k \end{cases}$$

où  $g_k \in \mathbb{R}^n$ ,  $H_k \in \mathbb{R}^{n \times n}$  est symétrique (sans être nécessairement définie positive) et  $\Delta > 0$ .

#### 3.3.1 Condition d'optimalité

Le résultat suivant est fondamentale pour l'étude des méthodes à région de confiance. Il donne des conditions nécessaires et suffisantes d'optimalité



pour un problème non convexe (on ne suppose pas que  $H$  est semi-définie positive), ce qui est assez rare. Cela vient de la structure particulière du problème : une fonction quadratique une seule contrainte quadratique convexe ( $\|s\|^2 \leq \Delta^2$ ).

**Théorème 3.3.1** *Supposons que  $\Delta > 0$ . Alors le point  $\hat{s} \in \mathbb{R}^n$  est solution de  $(RC_k)$  si et seulement s'il existe  $\hat{\lambda} \in \mathbb{R}$  tel que l'on ait :*

$$\left\{ \begin{array}{l} \text{a) } (H + \hat{\lambda}I)\hat{s} = -g \\ \text{b) } \|\hat{s}\| \leq \Delta \\ \text{c) } \hat{\lambda} \geq 0 \\ \text{d) } \hat{\lambda}(\Delta - \|\hat{s}\|) = 0 \\ \text{e) } (H + \hat{\lambda}I) \text{ est semi-définie positive} \end{array} \right.$$

**Démonstration 12** Voir [2]

### 3.3.2 Résolution numérique de $(RC_k)$

Dans cette partie on va résoudre le sous-problème par une méthode fondée sur le point de Cauchy (Définition 3.3.1). En général on ne peut pas résoudre le sous-problème quadratique  $(RC_k)$ . Le seul cas simple se présente lorsque  $H_k$  est définie positive et  $\Delta_k$  et suffisamment grand pour le minimum globale de  $\psi_k$ . La convergence des méthodes à région de confiance dépend aussi de l'algorithme utilisé pour résoudre le sous-problème quadratique.

Les algorithmes à région de confiance ont une propriété remarquable que n'ont pas les algorithmes à directions de descente :

Ils peuvent forcer les itérés pour converger vers un point stationnaire qui n'est pas un minimum local.

**Définition 3.3.1** *On appelle point de Cauchy du sous-problème quadratique  $(RC_k)$ , le point noté  $s_k^c$ , solution de :*

$$\left\{ \begin{array}{l} \min \psi_k(s) \\ \|s\| \leq \Delta_k \\ s = -\alpha g_k, \alpha \in \mathbb{R} \end{array} \right.$$

C'est donc le point minimisant  $\psi_k$  dans la région de confiance, le long de la droite de plus forte pente de  $\psi_k$  (ou de  $f$ , puisque l'on a supposé que  $\nabla \psi_k(0) = \nabla f(x_k)$ ).

**Proposition 3.3.1** *Le point de Cauchy  $s_k^c$  est unique et donné par*

$$s_k^c = \begin{cases} 0, & \text{si } g_k = 0; \\ -\frac{\Delta_k}{\|g_k\|} g_k, & \text{si } g_k \neq 0 \text{ et } g_k^T H_k g_k \leq 0; \\ -\min\left(\frac{\Delta_k}{\|g_k\|}, \frac{\|g_k\|^2}{g_k^T H_k g_k}\right) g_k. & \text{Sinon.} \end{cases} \quad (3.2)$$

**Démonstration 13** *Voir [2]*

Le point de Cauchy joue un rôle important dans la convergence des méthodes à région de confiance, alors on va résoudre le sous-problème quadratique de manière de trouver un point autant le point de Cauchy.

On pourrait évidemment prendre le point de Cauchy lui-même comme solution. En pratique, ce point n'est cependant pas très bon en ce qui concerne la vitesse de convergence, c'est pour ça on traduit cette condition pour assurer la convergence :

**Définition 3.3.2** *On dit que  $(RC_k)$  est résolu en  $s_k$  avec une condition de décroissance suffisante si il existe des constante strictement positives  $\beta_1$  et  $\beta_2$  telle que :*

$$\begin{cases} \psi_k(s_k) \leq \beta_1 \psi_k(s_k^c) \\ \|s_k\| \leq \beta_2 \Delta_k \end{cases} \quad (3.3)$$

## 3.4 Résultats de convergence

Dans cette section on va donner des résultats de convergence des méthodes à région de confiance. On considère le cas où le modèle est de premier ordre, on veut dire par là que la résultat de convergence ne s'intéresse pas à la matrice  $H_k$  (la seule chose demandée est que la suite  $\{H_k\}$  soit bornée), mais plutôt au point de Cauchy. C'est la condition de descente suffisante (3.2) qui joue un rôle-clé.

### 3.4.1 Convergence avec modèle du premier ordre

On suppose le sous-problème quadratique  $(RC_k)$  est résolu de façon approchée, de manière toutefois à satisfaire la condition de décroissance suffisante (3.2). On note :

$$N_1 = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$$

Pour les méthode à région de confiance, la proposition suivante joue le même rôle que la condition de Zoutendijk pour les méthodes à direction de descente : elle résume la contribution de la méthode de détermination du pas à la convergence de l'algorithme.

**Proposition 3.4.1** *Le point de Cauchy  $s_k^c$  vérifie la condition de Powell suivante avec  $s_k = s_k^c$  et  $C = 1/2$  :*

$$\psi_k(s_k) \leq -C \|g_k\| \min \left( \Delta_k \cdot \frac{\|g_k\|}{\|H_k\|} \right). \quad (3.4)$$

**Démonstration 14** *le cas où  $g_k = 0$  alors  $s_k^c = 0$  et  $\psi_k(s_k) = 0$ .*

*Supposons à présent que  $g_k \neq 0$  : si  $g_k^T H_k g_k \leq 0$ ,  $s_k^c = -\frac{\Delta_k}{\|g_k\|} g_k$  d'après la formule (3.2) donc*

$$\psi_k(s_k) = -\Delta_k \|g_k\| + \frac{\Delta_k^2}{2\|g_k\|^2} g_k^T H_k g_k \leq -\Delta_k \|g_k\|$$

*Alors (3.4) est vérifiée avec  $C = 1$ .*

*Il reste le cas où  $g_k \neq 0$  et  $g_k^T H_k g_k > 0$ . Si  $\Delta_k \leq \frac{\|g_k\|^3}{g_k^T H_k g_k}$ ,  $s_k = -\frac{\Delta_k}{\|g_k\|} g_k$  et on a*

$$\psi_k(s_k) = -\Delta_k \|g_k\| + \frac{\Delta_k^2}{2\|g_k\|^2} g_k^T H_k g_k \leq -\Delta_k 2\|g_k\|.$$

*Donc (3.4) est vérifiée avec  $C = \frac{1}{2}$ . Si nom*

$$s_k^c = \frac{\|g_k\|^2}{g_k^T H_k g_k} g_k$$

*et*

$$\psi_k(s_k) = \frac{\|g_k\|^4}{g_k^T H_k g_k} \leq \frac{\|g_k\|^2}{2H_k}$$

*Donc (3.4) est vérifiée avec  $C = \frac{1}{2}$ .*

La proposition suivante montre que si  $g_k \neq 0$ , si le rayon de confiance  $\Delta_k$  est assez petit, et si la solution approchée  $s_k$  de  $(RC_k)$  vérifie la condition de décroissance suffisante (4.2) alors la décroissance de  $f, f(x_k) - f(x_k + s_k)$ , est suffisante pour avoir

$$f(x_k + s_k) \leq f(x_k) + \omega_1 \psi_k(s_k)$$

C'est-à-dire  $\rho_k \geq \omega_1$  : Dans ces condition, il n'y a pas bouclage à l'étape 2 de l'algorithme.

**Proposition 3.4.2** *Supposons que  $f$  soit dérivable en  $x_k$  et que  $g_k \neq 0$ . Il existe  $\Delta_k^0 > 0$  tel que si  $0 < \Delta_k \leq \Delta_k^0$  et si la solution approchée de  $(RC_k)$  vérifie la condition de décroissance suffisante (4.2), alors le ratio  $\rho_k > \omega_1$ .*

**Démonstration 15** *Raisonnons par l'absurde, si ce n'est pas le cas, il existe une suite de rayons de confiance  $\Delta_{k,i} \rightarrow 0$  pour  $i \rightarrow \infty$  et des solutions approchées  $s_{k,i}$ , vérifiant*

$$f(x_k + s_{k,i}) \geq f(x_k) + \omega_1 \psi_k(s_{k,i}) \quad (3.5)$$

$$\psi_k(s_{k,i}) \leq \beta_1 \psi_k(s_{k,i}^c) \quad (3.6)$$

$$\|s_{k,i}\| \leq \beta_2 \Delta_{k,i} \quad (3.7)$$

où  $s_{k,i}^c$  est le point de Cauchy correspondant au rayon de confiance  $\Delta_{k,i}$ . Comme  $f$  est différentiable en  $x_k$ , on a en développant  $f(x_k + s_{k,i})$  autour de  $x_k$  :

$$\begin{aligned} f(x_k + s_{k,i}) &= f(x_k) + g_k^T s_{k,i} + o(\|s_{k,i}\|) \\ &= f(x_k) + \psi_k(s_{k,i}) + o(\|s_{k,i}\|) \end{aligned}$$

Alors (3.5) et  $0 < \omega_1 < 1$  donnent

$$\psi_k(s_{k,i}) = o(\|s_{k,i}\|)$$

D'autre part, comme  $g_k \neq 0$ , on peut supposer que  $\Delta_{k,i} \leq \frac{\|g_k\|}{\|H_k\|}$ , et d'après la proposition (3.4.1) et (3.7) montrent que

$$|\psi_k(s_{k,i})| = -\psi_k(s_{k,i}) \geq \frac{\beta_1}{2} \|g_k\| \Delta_{k,i} \geq \frac{\beta_1}{2\beta_2} \|g_k\| \|s_{k,i}\|$$

Ceci est contradiction avec l'estimation précédente ( $g_k \neq 0$ ) et cette contradiction prouve la proposition.

**Théorème 3.4.1** *Supposons que  $f$  soit bornée inférieurement et de classe  $C^1$  dans une voisinage de  $N_1$ . Si dans la méthode à régions de confiance,  $\{M_k\}$  est bornée et la solution approchée  $s_k$  du sous-problème quadratique  $(RC_k)$  vérifie les conditions de décroissance suffisante (4.2), alors l'une des situations suivantes a lieu :*

- (i) il y a échec d'une itération  $k_0$ , en un point  $x_{k_0}$  où  $\nabla f(x_{k_0}) = 0$  ;
- (ii)  $\liminf \|g_k\| = 0$



**Démonstration 16** On raisonne par l'absurde, en supposant que la situation (i) n'a pas lieu (donc une suite  $\{x_k\}$  est générée) et qu'il existe un indice  $k_1$  et une constante  $\gamma > 0$  tels que

$$\|g_k\| \geq \gamma, \forall k \geq k_1 \quad (3.8)$$

Montrons que

$$\Delta_k \rightarrow 0 \text{ et } \sum_{k \geq 1} \|s_k\| < +\infty \quad (3.9)$$

D'après la condition de Powell (3.4), la supposition de départ (3.8) et la bornitude de  $\{H_k\}$ , on a

$$\psi_k(s_k) \leq -C \|g_k\| \min(\Delta_k, \frac{\|g_k\|}{H_k}) \leq -C \min(\Delta_k, 1). \quad (3.10)$$

Où  $C$  est une constante absorbante strictement positive indépendante de  $k$ . En utilisant (3.5), on obtient alors

$$f(x_{k+1}) \leq f(x_k) - C \min(\Delta_k, 1)$$

$$C \min(\Delta_k, 1) \leq f(x_{k+1}) - f(x_k)$$

Comme  $f$  est bornée inférieurement, la suite  $f(x_k) - f(x_{k+1})$  est sommable et on déduit de l'inégalité ci-dessus que  $\sum_{k \geq 1} \|s_k\| < +\infty$ . Comme  $\|s_k\| \leq \beta_1 \Delta_k$ , les assertions (3.9) s'en déduisent.

Montrons que le ratio  $\rho_k \rightarrow 1$ . D'après (3.9),  $\{x_k\}$  est une suite de Cauchy, donc elle converge et on a

$$\frac{|f(x_k + s_k) - f(x_k) - g_k^T s_k|}{\|s_k\|} \rightarrow 0 \quad (3.11)$$

D'autre part, comme  $\{H_k\}$  est bornée, on a

$$|\psi_k(s_k) - g_k^T s_k| \leq C \|s_k\|^2 \quad (3.12)$$

Enfin, on peut écrire successivement

$$\rho_k = \frac{f(x_{k+1}) - f(x_k)}{c_k(s_k)} = \frac{f(x_{k+1}) - f(x_k) - g_k^T s_k}{\psi_k(s_k)} + \frac{g_k^T s_k - \psi_k(s_k)}{c_k(s_k)} + 1$$

$$|\rho_k - 1| = \frac{f(x_{k+1}) - f(x_k)}{\psi_k(s_k)} = \frac{f(x_{k+1}) - f(x_k) - g_k^T s_k}{\psi_k(s_k)} + \frac{g_k^T s_k - \psi_k(s_k)}{\psi_k(s_k)}$$

D'après (3.10),  $\Delta_k \rightarrow 0$  et  $\|s_k\| \leq \beta_2 \Delta_k$ , on a  $|\psi_k(s_k)| \geq C\|s_k\|$ . Alors, la dernière inégalité (3.11) et (3.12) montrent que  $\rho_k \rightarrow 1$ .

Dès lors  $\rho_k > \omega_2$  pour  $k$  assez grand. d'après la règle de mise à jour de  $\Delta_k$  (étape 4 de l'algorithme), cela implique  $\Delta_k \geq \Delta > 0$ , qui est en contradiction avec la première assertion de (3.9). Cette contradiction prouve le théorème.

Le théorème (4.4.2) montre qu'avec un peu plus de régularité sur  $f$  (l'uniforme continuité de  $\nabla f$ ), on a la convergence de toute la suite des gradients.

**Théorème 3.4.2** *Supposons, en plus des hypothèses et conditions du théorème (4.4.1), que  $\nabla f$  soit uniformément continue sur  $N_1$ . Alors soit la solution (i) du théorème (4.4.1) a lieu, soit  $g_k \rightarrow 0$ .*

**Démonstration 17** *On raisonne par l'absurde en supposant que la situation (i) du théorème (3.4.1) n'a pas lieu (donc une suite  $\{x_k\}$  est générée) et qu'il existe une constante  $\gamma > 0$  et une sous-suite  $\{k_j\}_{j \geq 1} \subset \mathbb{N}$  telles que*

$$\|g_{k_j}\| \geq \gamma, \forall j \geq 1. \quad (3.13)$$

Comme  $\liminf \|g_k\| = 0$ , en extrayant une sous-suite au besoin, on peut supposer que pour tout  $j$  on peut trouver  $l_j$  tel que

$$\begin{cases} k_j < l_j < k_{j+1}, \\ \|g_k\| \geq \frac{\gamma}{2} \text{ pour } k = k_j, \dots, l_{j-1}, \\ \|g_{l_j}\| < \frac{\gamma}{2}. \end{cases} \quad (3.14)$$

En utilisant la condition de Powell (3.4) et la bornitude de  $\{H_k\}$ , on a comme la démonstration du théorème (3.4.1) :

$$f(x_{k+1}) - f(x_k) \leq C\|g_k\| \min(\Delta_k, \|g_k\|).$$

Donc pour  $k = k_j, \dots, l_{j-1}$  et  $j \geq 1$ , on a

$$f(x_{k+1}) - f(x_k) \leq C \min(\Delta_k, 1).$$

Comme  $f(x_{k+1}) - f(x_k) \rightarrow 0$ ,  $\min(\Delta_k, 1) = \Delta_k$  pour  $k$  assez grand. En utilisant le fait que  $\|s_k\| \leq \beta_2 \Delta_k$ , on a finalement

$$\|s_k\| \leq C(f(x_k) - f(x_{k+1})), k = k_j, \dots, l_{j-1},$$

$$\|x_{l_j} - x_{k_j}\| \leq \sum_{k=k_j}^{l_j-1} \|s_k\| \leq C(f(x_{k_j}) - f(x_{k_{j+1}}))$$

Donc  $\|x_{l_j} - x_{k_j}\| \rightarrow 0$ , quand  $j \rightarrow \infty$ . Alors par uniforme continuité de  $\nabla f$ , on a :

$$\|g_{l_j} - g_{k_j}\| \rightarrow 0, \text{ quand } j \rightarrow \infty.$$

Ceci est en contraction avec le fait que , par(3.13)et(3.14)

$$\|g_{l_j} - g_{k_j}\| \geq \|g_{l_j}\| - \|g_{k_j}\| \geq \frac{\gamma}{2}.$$

cette contraction prouve le théorème.

### 3.4.2 Convergence avec modèle du second ordre

Dans cette section, nous étudions la convergence et les propriétés des points d'adhérence des suite générées par les méthode à région de confiance lorsque le sous-problème quadratique  $(RC_k)$  est résolu de façon fine, de manière à satisfaire la condition de décroissance forte :

$$\begin{cases} \psi_k(s_k) \leq \beta_1 \widehat{\psi}_k \\ \|s\| \leq \beta_2 \Delta_k \end{cases} \quad (3.15)$$

Où  $\widehat{\psi}_k = \min\{\psi_k(s) : \|s\| \leq \Delta_k\}$ , et lorsque le modèle  $\psi_k$  est du second ordre, c'est-à-dire lorsque l'on prend

$$H_k = \nabla^2 f(x_k)$$

On suppose donc que la fonction  $f$  est deux fois continûment dérivable et qu'on calcule le hessien. Si ceci est coûteux en temps de calcul, cela peut en vouloir la peine. On montre en effet (voir Théorème pour les hypothèses précises) que la suite  $\{x_k\}$  a au moins un point d'adhérence où le hessien est semi-défini positif et qu'en tout point d'adhérence isolé, le hessien est semi-défini positif. On montre aussi (Théorème) que si  $\{x_k\}$  a un point d'adhérence où le hessien est défini positif alors toute suite converge vers ce point.

La condition de décroissance forte (3.15) est clairement plus forte que la condition de décroissance suffisante (3.3), elle implique également la condition de Powell et le résultat de proposition (3.4.2). Le résultat suivant est plus précis. Il montre qu'il n'y a pas bouclage dans l'algorithme à région

de confiance avec modèle du second ordre ( $H_k = \nabla^2 f(x_k)$ ) si  $x_k$  ne vérifie pas ces conditions, l'itération  $k$  est interrompue, car on ne peut pas réaliser  $\psi_k(s_k) < 0$ .

**Proposition 3.4.3** *Supposons que  $f$  soit de classe  $C^1$  dans un voisinage de  $x_k$  et deux fois dérivable en  $x_k$  et que soit  $\lambda_1(\nabla^2 f(x_k)) < 0$ . Supposons également que dans  $(RC_k)$ , prenne  $H_k = \nabla^2 f(x_k)$ . Dans ce cas, il existe  $\Delta_k^0$  tel que si  $0 < \Delta_k \leq \Delta_k^0$  et si la solution approchée de  $(RC_k)$  vérifie la condition de décroissance forte (3.15), alors le ratio  $\rho_k > \omega_1$ .*

**Démonstration 18** Voir [2]

**Lemme 3.4.1** *Soit  $\bar{x}$  un point d'adhérence isolé d'une suite  $\{x_k\}$  de  $\mathbb{R}^n$ . Si  $\{x_k\}$  ne converge pas vers  $\bar{x}$ , il existe  $\epsilon > 0$  et une sous-suite  $\{x_{k_i}\}_{i \geq 1}$  qui converge vers  $\bar{x}$  et vérifie*

$$\|x_{k_{i+1}} - x_{k_i}\| \geq \epsilon, \forall i \geq 1$$

**Démonstration 19** Voir [2]

**Théorème 3.4.3** *Supposons que  $f$  soit bornée inférieurement, de classe  $C^2$  dans un voisinage de  $N_1$  et que son hessien  $\nabla^2 f$  soit borné sur  $N_1$ . On suppose que l'on prend  $H_k = \nabla^2 f(x_k)$  dans la méthode à région de confiance et que la solution approchée  $s_k$  de  $RC_k$  vérifie la condition de décroissance forte (3.15). Alors, soit il y a échec d'une itération  $k_0$  en un point  $x_{k_0}$  vérifiant les conditions nécessaires d'optimalité du second ordre ( $\nabla f(x_{k_0}) = 0$  et  $\nabla^2 f(x_{k_0})$  est semi-définie positive), soit une suite  $\{x_k\}$  est générée et*

- (i)  $g_k \rightarrow 0$ ;
- (ii) Si  $\{x_k\}$  est bornée, alors  $\limsup \lambda_1(\nabla^2 f(x_k)) \geq 0$ , ce qui revient à dire que  $\{x_k\}$  a au moins un point d'adhérence  $\bar{x}$  tel que  $\nabla^2 f(\bar{x})$  soit semi-définie positive;
- (iii) Si  $\bar{x}$  est un point d'adhérence isolé de  $\{x_k\}$ , alors  $\nabla^2 f(\bar{x})$  est semi-définie positive.

**Démonstration 20** Voir [2]

Dans le théorème suivant, suppose seulement que le problème  $(RC_k)$  est résolu avec condition de décroissance suffisante (3.3).



**Théorème 3.4.4** *On suppose que  $f$  vérifie les mêmes propriétés dans le théorème (3.4.3), que  $H_k = \nabla^2 f(x_k)$  et que la solution approchée de  $(RC_k)$  vérifie la condition de décroissance suffisante (3.3). Alors, soit il y a échec d'une itération  $k_0$  en un point  $x_{k_0}$  tel que  $\nabla f(x_{k_0}) = 0$ , soit une suite  $\{x_k\}$  est générée et*

(i)  $g_k \rightarrow 0$  :

(ii) Si  $\bar{x}$  est un point d'adhérence de  $\{x_k\}$  tel que  $\nabla^2 f(\bar{x})$  soit définie positive, alors  $x_k \rightarrow \bar{x}$  et  $\Delta_k$  est uniformément  $> 0$ .

**Démonstration 21** *La démonstration se fait en suivant la même stratégie suivie dans la preuve du théorème 3.4.2.*

## 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] A.R. CONN AND N.I.M COULD AND P.L. TOINT (2000) : *Trust region methods*, *MPS/SIAM Series On Optimization*.
- [3] J.C.GILBERT (2005) : *Eléments d'Optimisation Différentiable : Théorie et Algorithmes*, *INRIA Rocquencourt*.
- [4] A.A. GOLDSTEIN AND J.F. PRICE (1969) : *An effective algorithm for minimzation*, *Num.Math.*, 10, pp. 184-189.
- [5] D.G.LUEBERGER (1969) : *Optimization by vector space methods*, *john Wiley and Sons, Inc. New york*.
- [6] M.MINOUX(1983) : *Programation Mathématique. Théorie et Algorithmes, tome 1*, *Dunod*.
- [7] M.J.D.POWELL(1986) : *convergence properties of algorithme for nonlinear optimisation*, *SIAM rev*, 28, pp.487-500.
- [8] G. ZOUTENDIJK (1970) : *Nonlinear Programming Computational Methods, nteger and Nonlinear Programming*, *North Holland, Amsterdam*.