

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



N° Réf :.....

Centre Universitaire
Abd elhafid boussouf Mila

Institut des sciences et de la technologie

Département de Mathématiques et Informatique

**Mémoire préparé En vue de l'obtention du diplôme de
Licence
En : Filière mathématiques**

Méthodes de recherche linéaire

Préparé par :

- Benchaoui warda
- Chergui sara
- Derghal saad
- Ben khelifa fateh

Encadrer par : Fadel wahida

Année universitaire : 2014 /2015

Remerciements :

Je tiens adresser mes remerciements les plus chaleureux et ma profonde gratitude mon encadreur madame FADEL WAHIDA, pour m'avoir proposé le sujet de ce mémoire. C'est grâce à sa grande disponibilité, ses conseils, ses orientations, et ses encouragements que j'ai pu mener bien ce travail.

Nous remercions nos parents, pour tous les sacrifices qu'ils ont consentis Pour nous permettre de suivre nos études dans les meilleures conditions et de nous avoir en courage tout au long de ces années.

Un remerciement particulier à nos très chers frères, sœurs, collègues et amies et à tout la famille respectives qui nous ont encouragés, soutenu durant tout notre parcours.

Résumé :

On propose dans ce mémoire quelques méthodes de recherche linéaire on optimisation.

La première approche tend vers ramener les problèmes d'optimisation avec contrainte et sans contraintes.

La deuxième approche utilise la méthode numérique pour minimiser une fonction.

Dans le troisième approche on pressante quelques méthodes de recherche linéaire.

Mots clés :

Optimisation sans contrainte, optimisation avec contrainte, convexe point critique, recherche linéaires.

:

نقترح في هذ

:

: هي عرض المشكل (Le Pb d'optimisation)

الثانية: فهي عرض الطريقة العددية لحل

: تتضمن موضوع الدراسة وهو عرض بعض طرق البحث الخطية.

مفتاحية:

,

,

,

.

Table des matières

Introduction	2
1 Optimisation sans contraintes et avec contraintes	4
1.1 Élément d'analyse convexe	4
1.2 Optimisation sans contraintes	5
1.2.1 Résultats d'existence et d'unicité	5
1.2.2 Conditions d'optimalité	6
1.3 Optimisation avec contraintes	7
1.3.1 Résultats d'existence et d'unicité	7
1.3.2 Condition d'optimalité	8
1.3.3 Classification des problèmes d'optimisation	11
2 Méthode numérique pour la minimisation d'une fonction	12
2.1 Introduction	12
2.2 Fonction convexe deux fois défférentiable	12
2.3 Conditions d'optimalité des problèmes de minimisation sans contraintes . .	13
2.3.1 Minima locaux et globaux	14
2.4 Conditions nécessaires d'optimalité	14
2.4.1 Condition nécessaire d'optimalité du premier ordre	15
2.4.2 Conditions nécessaires d'optimalité du second ordre	15
2.5 Conditions suffisantes d'optimalité	16
2.5.1 Condition suffisante d'optimalité du second ordre	16
3 Les méthodes de recherches linéaires	19
3.1 Objectifs de la recherche linéaire	19
3.1.1 Le premier objectif	19
3.1.2 Le second objectif	20
3.2 But de la recherche linéaire	20
3.3 Intervalle de sécurité	21

3.4	Recherche linéaire d'ordre zéro	21
3.5	Méthode de gradient conjugué	22
3.5.1	La méthode du gradient conjugué vue comme une méthode directe .	23
3.5.2	La méthode du gradient conjugué vue comme une méthode itérative	23
3.6	Méthode de Dai-Yuan	25
3.7	Direction de descente	25
3.7.1	Exemples de direction de descente	26
3.7.2	Direction du gradient	26
3.7.3	Direction de Newton	26
3.7.4	Direction de quasi-Newton	27
3.7.5	Direction de Gauss-Newton	27
	Bibliographie	28
	Annexe	29

Introduction

L'optimisation (c'est-à-dire les techniques permettant de chercher les minima ou les maxima de fonctions ou de fonctionnelles) intervient dans pratiquement tous les processus de modélisation actuels.

Qu'il s'agisse de problèmes directs (ajustement de données, contrôle optimal, résolution des systèmes linéaires par moindres carrés, etc ..) ou inverses (identification des paramètres, contrôle de frontières libres etc..), il est rare qu'un problème d'optimisation plus ou moins complexe n'intervienne pas à un stade donné de la modélisation et/ou de la simulation.

Avant de donner les définitions et principes de base de la théorie de l'optimisation, nous allons présenter quelques exemples simples permettant d'introduire et d'illustrer par anticipation notre propos.

Ce mémoire commence par une introduction générale et réparti en trois chapitres actuels.

Dans le premier nous rappelons des notions sur l'optimisation sans contraintes et avec contraintes.

Le deuxième chapitre contient la présentation de quelques méthodes numériques pour minimiser une fonction et dans le dernier chapitre, nous présentons quelques méthodes de recherche linéaire (méthode fibonacci , newton...) et leurs buts.

Chapitre 1

Optimisation sans contraintes et avec contraintes

1.1 Élément d'analyse convexe

Un sous ensemble C de \mathbb{R}^n est dit convexe si :

$$(1 - \lambda)x + \lambda y \in C, \forall x, y \in C, \forall \lambda \in [0, 1]$$

Autrement dit, si le segment de droite joient deux points quelconques :

$$x, y \in C : [x, y] = \{(1 - \lambda)x + \lambda y : 0 \leq \lambda \leq 1\}$$

est entièrement inclu dans C .

Définition 1.1.1 Une fonction $f : C \rightarrow \mathbb{R}$ est dit convexe si :

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

Définition 1.1.2 une fonction $f : C \rightarrow \mathbb{R}$ est dit strictement convexe si :

$$\forall t \in [0, 1], \forall x, y \in C, f((1 - \lambda)x + \lambda y) < (1 - \lambda)f(x) + \lambda f(y).$$

Définition 1.1.3 Un ensemble convexe de la forme :

$$X = \{x / Ax = b, x \geq 0\}$$

est appelé un polytope convexe ,un polytope borné sera appelé un polydère convexe .

Définition 1.1.4 Soit X un ensemble convexe non vide de \mathbb{R}^n un point $x \in X$ dit extrémal si l'on a :

$$\forall t \in [0, 1], \forall y, z \in X : x = (1 - t)y + tz \Rightarrow x = y = z$$

1.2 Optimisation sans contraintes

Le problème d'optimisation sans contraintes s'écrit sous la forme suivante :

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

où f est une fonction de \mathbb{R}^n vers $\mathbb{R} \cup \{+\infty\}$

1.2.1 Résultats d'existence et d'unicité

Avant d'étudier les propriétés de la solution (où des solutions) éventuelle(s) de (Q) il faut s'assurer de son/leurs existence.

Nous donnerons ensuite des résultats d'unicité.

Définition 1.2.1 On dit que $f : C \rightarrow \mathbb{R}$ est coercive si :

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$$

Dans le cas où $C = \mathbb{R}^n$ les normes sont toutes équivalentes et $\| \cdot \|$ désigne une norme quelconque de \mathbb{R}^n .

Théorème 1.2.2 : (existence)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ propre, continue et coercive alors (Q) admet au moins une solution.

Toutes les fois, on n'a pas forcément unicité. Nous donnons ci-dessous un critère pour l'unicité.

Théorème 1.2.3 (unicité)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ strictement convexe alors le problème (Q) admet au plus une solution.

Donnons pour terminer un critère qu'une fonction soit strictement convexe et coercive.

Théorème 1.2.4 Soit f une fonction C^1 de \mathbb{R}^n dans \mathbb{R} . On suppose qu'il existe $\alpha > 0$ tels que :

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2$$

alors f est strictement convexe et coercive, en particulier (Q) admet une solution unique. On donne dans ce qui suit les conditions pour pouvoir calculer la (où les) solution(s) optimale(s) de (Q).

1.2.2 Conditions d'optimalité

Conditions nécessaires et suffisantes du premier ordre

Les conditions que nous allons donner sont des conditions différentielles qui dépendent de la dérivée de la fonction à minimiser.

Théorème 1.2.5 (condition nécessaire d'optimalité du premier ordre)

Soit C une partie de \mathbb{R}^n et $f : C \rightarrow \mathbb{R}$ une fonction différentiable sur C , si x^* réalise un minimum (global ou local) de f sur C alors $\nabla f(x^*) = 0$.

Définition 1.2.6 Un point x^* de C vérifiant $\nabla f(x^*) = 0$ est appelé "point critique" ou "point stationnaire".

Théorème 1.2.7 (CNS du premier ordre dans le cas convexe)

Soit $f : C \rightarrow \mathbb{R}$ fonction différentiable et convexe sur C , un point x^* réalise un minimum global de f sur C si et seulement si $\nabla f(x^*) = 0$.

Remarque 1.2.8 On peut raffiner le résultat précédent en ne supposant que la locale convexité de f au voisinage de x^* , c'est-à-dire en supposant que f est convexe sur une boule centré en x^* .

A ce moment là, nous pouvons affirmer que x^* est un minimum "local" de f .

Le cas où f est convexe est fréquent dans la pratique mais pas systématique.

Nous allons donc donner maintenant des conditions suffisantes pour qu'un point critique réalise un minimum (ou un maximum).

Ces conditions vont faire intervenir la dérivée seconde de f , ce sont des conditions du "second ordre".

Conditions du second ordre

Nous commençons par une condition nécessaire permettant de préciser encore les éventuels minimax.

Théorème 1.2.9 (*Condition nécessaire du second ordre*)

On suppose que x^* est un minimum (local) de f et que f est deux fois différentiables sur C , alors :

a) $\nabla f(x^*) = 0$

b) $\langle \nabla^2 f(x^*)x, x \rangle \geq 0, \forall x \in C$

Remarque 1.2.10 Dans le cas où $C = \mathbb{R}^n$, b) est équivalent à dire que la matrice Hessienne de f en x^* : $\nabla^2 f(x)$ est semi-définie positive.

Théorème 1.2.11 (*Condition suffisante du seconde ordre*)

Soit f deux fois dérivable sur f vérifiant :

$$\nabla f(x^*) = 0 \text{ et } \exists \alpha > 0$$

tels que :

$$\forall x \in C \langle \nabla^2 f(x^*)x, x \rangle \geq \alpha \|x\|^2 \quad (*)$$

alors la fonction f admet un minimum "local strict" en x^* .

Remarque 1.2.12 Si $C = \mathbb{R}^n$ la condition (*) revient à dire que la matrice Hessienne $\nabla^2 f(x^*)$ est définie positive, un choix possible pour α étant alors la plus petite valeur propre, c'est une condition deité (locale) stricte au voisinage de x^* .

1.3 Optimisation avec contraintes

La forme de problème avec contraintes est :

$$\begin{cases} \min f(x) \\ x \in C \end{cases} \quad (P)$$

où C est l'ensemble des contraintes.

1.3.1 Résultats d'existence et d'unicité

Commençons par donner un résultat d'existence.

Théorème 1.3.1 *Supposons que f est continue, que C est un sous-ensemble fermé non vide de \mathbb{R}^n et que l'une des conditions suivantes est réalisée :*

a) *Soit C est borné.*

b) *Soit f est coercive*

Alors le problème (P) admet au moins une solution.

Théorème 1.3.2 *Sous les hypothèses du théorème (1-2-5), si f est strictement convexe et si C est convexe, alors le problème (P) admet une solution unique.*

1.3.2 Condition d'optimalité

Tout comme dans le cas sans contraintes, nous allons établir des conditions d'optimalité du premier et du second ordre permettant de calculer les éventuels minima de f .

Conditions d'optimalité du premier ordre

La condition nécessaire suivante est l'analogue de celle que nous avons dans le cas sans contraintes. Elle fait bien sûr intervenir l'ensemble des contraintes.

Théorème 1.3.3 *(Condition nécessaire du premier ordre)*

Si f est une fonction différentiable et si C est un convexe fermé, alors toute solution x^ de (P) vérifie la condition nécessaire d'optimalité du premier ordre.*

$$\forall x \in C \quad \langle \nabla f(x^*), x - x^* \rangle \geq 0 \quad (1).$$

Théorème 1.3.4 *(CNS du premier ordre dans le cas convexe)*

Supposons f convexe différentiable et C convexe fermé de \mathbb{R}^n , soit x un élément de C .

La condition (1) est nécessaire et suffisant pour que x^ soit solution de (P). Elle caractérise donc les minima de f sur C .*

On suppose maintenant que C est défini comme suit :

$$C = \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, \dots, p, h_j(x) = 0, j = 1, \dots, q\}$$

Les contraintes $g_i(x) \leq 0$ sont appelées "contraintes d'inégalités" et les contraintes $h_j(x) = 0$ "contraintes d'égalités".

L'obtention de conditions nécessaires d'optimalité nécessite que certaines conditions appelées conditions de qualification des contraintes soient vérifiées.

Contraintes en égalités

problème (P) se réduit à

$$\min f(x), x \in \mathbb{R}^n, h(x) = 0, \text{ avec } h(x) = (h_1(x), \dots, h_p(x))$$

et h est continue de \mathbb{R}^n dans \mathbb{R}^p .

Théorème 1.3.5 (CN du premier ordre-contraintes en égalité)

f et h sont de classe C^1 sur \mathbb{R}^n .

Le problème (P) à une solution x^* .

Les p vecteurs de $\mathbb{R}^n : \nabla h_1(x^*), \dots, \nabla h_p(x^*)$ sont linéairement (et donc $p \leq n$) alors il existe p réels $(\lambda_1^*, \dots, \lambda_p^*)$ tels que :

$$\nabla f(x^*) + \sum_{j=1}^p \lambda_j^* \nabla h_j(x^*) = 0.$$

Contrainte d'inégalité

L'ensemble de contraintes est donné par :

$$C = \{x \in \mathbb{R}^n / g_i(x) \leq 0, \forall i = 1, \dots, m\}.$$

Contraintes d'égalité et d'inégalité

Le problème P est de la forme :

$$\min f(x), x \in C.$$

$$C = \{x \in \mathbb{R}^n / h_j(x) = 0, g_i(x) \leq 0\}.$$

Ici nous nous limiterons aux trois conditions de qualification suivantes :

★ (CQ1) C est un polyèdre convexe :

c'est le cas lorsque les fonctions g_i et h_j sont affines.

★ (CQ2) condition de Slater :

la condition de qualification des contraintes de Slater est comme suit :

1- Les fonctions g_i sont convexes et les fonctions h_j sont affines.

2- Il existe x^* tels que $g_i(x^*) < 0$ et $h_j(x^*) = 0$ pour tout i, j .

★ (CQ3) condition de Mangasarian-Fromowitz :

la condition de qualification des contraintes de Mangasarian-Fromowitz en un point $x \in C$ est comme suit :

1) Les q vecteurs $\nabla h_j(x^*)$ sont linéairement indépendants.

2) Il existe \bar{d} tels que $\langle \nabla h_j(x^*), \bar{d} \rangle = 0$ pour tout j et $\langle \nabla g_i(x^*), \bar{d} \rangle < 0$ pour tout i .

On associe au problème d'optimisation suivant :

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0, i = 1, \dots, p \\ h_j(x) = 0, j = 1, \dots, q \end{cases}$$

la fonction $L : \mathbb{R}^n \times [0, \infty[\times \mathbb{R}^n \rightarrow \mathbb{R}$ définie par :

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^p \lambda_i g_i(x) + \sum_{j=1}^q \mu_j h_j(x)$$

est appelée le " Lagrangien ".

Le théorème suivant de Karush-Kuhn-Tucker-Lagrange donne une condition nécessaire d'optimalité.

Supposons que les fonctions f, g_i, h_j sont C^1 dans un voisinage de $x^ \in C$ et que les contraintes vérifient une condition de qualification. Si f a un minimum local en x^* sur C alors il existe $\lambda \in \mathbb{R}^p$ et $\mu \in \mathbb{R}^q$ tels que :*

$$\nabla_x L(x^*, \lambda, \mu) = 0, \nabla_{\lambda} L(x^*, \lambda, \mu) \leq 0, \nabla_{\mu} L(x^*, \lambda, \mu) = 0, \lambda \geq 0, \langle \lambda, \nabla_{\lambda} L(x^*, \lambda, \mu) \rangle = 0.$$

Les quantités λ_i et λ_j sont appelées "multiplicateurs de Karush-Kuhn-Tucker". Lorsque le problème est convexe (f et g_i convexes et h_j affine) on a la condition suffisante d'optimalité.

Théorème 1.3.6 *Supposons que les fonctions f, g_i sont C^1 dans un voisinage de $x^* \in C$, convexes et que les fonctions h_j sont affines. S'il existe $\lambda \in \mathbb{R}^p$ et $\mu \in \mathbb{R}^q$ tels que :*

$$\nabla f(x^*) + \sum_{i=1}^p \lambda_i \nabla g_i(x^*) + \sum_{j=1}^q \mu_j \nabla h_j(x^*) = 0, \lambda_i \geq 0, g_i(x^*) \leq 0, \lambda_i g_i(x^*) = 0, h_j(x^*) = 0$$

$\forall i, j$ alors f a un minimum global en x^ sur C .*

Conditions d'optimalité nécessaires du deuxième ordre

Théorème 1.3.7 *on suppose que f, h et g sont de classe C^2 , que x^* est un minimum(local) de f sur C .*

1.3.3 Classification des problèmes d'optimisation

Les problèmes d'optimisation sont classifiés selon les caractéristiques des fonctions f, g_i, h_j .

Parmi les cas particuliers les plus étudiés, on note :

- La programmation linéaire (f linéaire, g_i, h_j affine, C orthant positif).
- La programmation convexe (f, g_i convexe, h_j affine, C convexe).
- La programmation en nombres entiers (C discret).

Chapitre 2

Méthode numérique pour la minimisation d'une fonction

2.1 Introduction

Nous allons exposer dans ce deuxième chapitre quelques concepts clés concernant l'optimisation, l'effort définitionnel portera surtout sur la méthode du gradient conjugué, où les questions suivantes seront prises en considération les problèmes proposés comportent-ils des contraintes ?

- les fonctions en jeu sont-elles quadratiques ? sont-elles convexes ?
- les domaines de définition des fonctions sont-ils continus ou discrets ?

Chaque problème possède sa structure spécifique et demande donc d'être traité de manière particulière.

Le présent projet ne se focalisera que sur les problèmes d'optimisation sans contraintes et non linéaires.

2.2 Fonction convexe deux fois différentiable

Définition 2.2.1 soit $C \subseteq \mathbb{R}^n$ non vide et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite deux fois différentiable S au point $x' \in \text{int}(C)$ s'il existe un vecteur $\nabla f(x')$ et une matrice symétrique $H(x')$ d'ordre (n, n) appelée matrice hessienne, et une fonction $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ tels que :

$$\forall x \in C : f(x) = f(x') + \nabla f(x')^T (x - x') + \frac{1}{2} (x - x')^T H(x') (x - x') + \|x - x'\|^2 \alpha(x', x - x')$$

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

$$H(x') = \begin{bmatrix} \frac{\partial^2 f(x')}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x')}{\partial x_1 \partial x_2} & \cdots & \cdots & \frac{\partial^2 f(x')}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x')}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x')}{\partial x_2 \partial x_2} & \cdots & \cdots & \frac{\partial^2 f(x')}{\partial x_2 \partial x_n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial^2 f(x')}{\partial x_n \partial x_1} & \frac{\partial^2 f(x')}{\partial x_n \partial x_2} & \cdots & \cdots & \frac{\partial^2 f(x')}{\partial x_n \partial x_n} \end{bmatrix}$$

★ La matrice $H(x')$ est dite semi-définie positive si :

$$\forall x \in \mathbb{R}^n : x^T H(x') x \geq 0$$

★ La matrice $H(x')$ est dite définie positive si :

$$\forall x \in \mathbb{R}^n, x \neq 0 : x^T H(x') x > 0$$

Théorème 2.2.2 soit $H(x')$ la matrice hessienne, on note par $\{t_i\}_{i=1 \dots n}$ ses valeurs propres (réelles) on a les équivalences suivantes :

$$\star H(x') \geq 0 \iff t_i \geq 0$$

$$\star H(x') > 0 \iff t_i > 0$$

2.3 Conditions d'optimalité des problèmes de minimisation sans contraintes

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

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

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

2.3.1 Minima locaux et globaux

Définition 2.3.1 1) $x' \in \mathbb{R}^n$ est un minimum local de (P) si et seulement si il existe un voisinage $v_\varepsilon(x')$ tels que :

$$f(x') \leq f(x) : \forall x \in V_\varepsilon$$

2) $x' \in \mathbb{R}^n$ est un minimum local strict de (P) si et seulement si il existe un voisinage $V_\varepsilon(x')$ tels que :

$$f(x') < f(x) : \forall x \in V_\varepsilon(x'), x \neq x'$$

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

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

Remarque 2.3.2 Dans le cas d'une fonction objectif convexe, il n'y a pas de distinction entre minimum local et global : tout minimum local est également global, comme l'établit le théorème suivant.

Théorème 2.3.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction convexe définie sur l'ensemble convexe X . Alors tout minimum local de f est également un minimum global.

Si f est strictement convexe, alors il existe au plus un minimum global de f .

2.4 Conditions nécessaires d'optimalité

Etant donné un vecteur x' , nous souhaiterions être capables de déterminer si ce vecteur est un minimum local ou global de (P) .

La propriété de différentiabilité de f fournit une première manière de caractériser une solution optimale.

Énonçons tout d'abord un théorème, pour établir une première condition nécessaire d'optimalité.

2.4.1 Condition nécessaire d'optimalité du premier ordre

Théorème 2.4.1 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable au point $x' \in \mathbb{R}^n$. si x' est un minimum local de (P) , alors $\nabla f(x') = 0$.

Preuve. C'est une conséquence directe du théorème (2-3-3). En effet, supposons que $\nabla f(x') = 0$. Puisque la direction $d = -\nabla f(x')$ est une direction de descente, alors il existe $\delta > 0$ tels que :

$$f(x' + td) < f(x') : \quad \forall t \in]0, \delta[$$

Ceci est contradiction avec le fait que x' est une solution optimale locale de (p) . ■

Remarque 2.4.2 si f est convexe, la condition nécessaire du premier ordre est également suffisante pour que x' soit un minimum global.

Dans le cas où f est deux fois différentiable, une autre condition nécessaire est donnée par le théorème (2-4-1). Elle est appelée condition nécessaire du second ordre car elle fait intervenir la matrice hessienne de f (que nous noterons $\nabla^2 f(x)$ dont les éléments sont ses secondes dérivées partielles).

2.4.2 Conditions nécessaires d'optimalité du second ordre

Théorème 2.4.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiables au point $x' \in \mathbb{R}^n$. Si x' est un minimum local de (P) , alors $\nabla f(x') = 0$ et $H(x')$ est semi définie positive.

Preuve. Soit $x \in \mathbb{R}^n$ quelconque, étant f deux fois différentiable au point x' on aura pour tout $\lambda \neq 0$:

$$f(x' + \lambda x) = f(x') + \frac{1}{2} \lambda^2 x^T H(x') x + \lambda^2 \|x\|^2 \alpha(x', \lambda x), \quad \alpha(x', \lambda x) \xrightarrow{\lambda \rightarrow 0} 0$$

Ceci implique :

$$\frac{f(x' + \lambda x) - f(x')}{\lambda^2} = \frac{1}{2} x^T H(x') x + \|x\|^2 \alpha(x', \lambda x) \dots (1.1)$$

x' est un optimum local, il existe alors $\delta > 0$ tels que :

$$\frac{f(x' + \lambda x) - f(x')}{\lambda^2} \geq 0 \quad \forall t \in]-\delta, +\delta[.$$

Si on prend en considération (1-1) et on passe à la limite quand $t \rightarrow 0$
 $t \neq 0$ on obtient :

$$x^T H(x') x \geq 0 \quad \forall x \in \mathbb{R}^n.$$

■

2.5 Conditions suffisantes d'optimalité

Les conditions données précédemment sont nécessaires (si f n'est pas convexe), c'est-à-dire qu'elles doivent être satisfaites pour tout minimum local .

Cependant, tout vecteur vérifiant ces conditions n'est pas nécessairement un minimum local.

Le théorème (2-4-3) établit une condition suffisante pour qu'un vecteur soit un minimum local, si f est deux fois différentiables.

2.5.1 Condition suffisante d'optimalité du second ordre

Théorème 2.5.1 Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois différentiables au point $x' \in \mathbb{R}^n$ Si $\nabla f(x') = 0$ et $H(x')$ est définie positive, alors x' est un minimum local strict de (P) .

Preuve. f étant deux fois différentiables au point x' , on aura pour tout $x \in \mathbb{R}^n$

$$f(x) = f(x') + \frac{1}{2} (x - x')^T H(x') (x - x') + \|x - x'\|^2 \alpha(x', x - x'), (\nabla f(x') = 0) \quad (1-2)$$

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

$$x_k \neq x' : \forall k \quad x_k \rightarrow x' \text{ et } f(x_k) \leq f(x') \quad (1-3)$$

Dans (1-3) prenons $x = x_k$ divisons le tout par $\|x_k - x'\|^2$ et notons $d_k = \frac{(x_k - x')}{\|x_k - x'\|^2}$ on obtient :

$$\frac{f(x_k) - f(x')}{\|x_k - x'\|^2} = \frac{1}{2} d_k^T H(x') d_k + \alpha(x', (x_k - x')) \xrightarrow{k \rightarrow \infty} 0 \quad (1-4)$$

(1-3) et (1.4) impliquent :

$$\frac{1}{2} d_k^T H(x') d_k + \alpha(x', (x_k - x')) \leq 0 \quad \forall k.$$

D'autre part la suite $\{d_k\}_{k \in \mathbb{N}^*}$ est bornée ($\|d_k\| = 1, \forall n$). Donc il existe une sous suite

$\{d_k\}_{k \in \mathbb{N}_1 \subset \mathbb{N}}$ tels que $d_k \xrightarrow[k \rightarrow \infty]{} d$. Finalement lorsque $k \rightarrow \infty$ on obtient :

$$\frac{1}{2} d_k^T H(x') d'_k \leq 0.$$

La dernière relation et le fait que $d' \neq 0$ ($\|d'\| = 1$) impliquent que la matrice hessienne $H(x')$ n'est pas définie positive. Ceci est en contradiction avec l'hypothèse. ■

Remarque 2.5.2 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.

Théorème 2.5.3 Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexe et différentiable au point $\bar{x} \in \mathbb{R}^n$ si $\nabla f(\bar{x}) = 0$

alors, \bar{x} est un minimum global de f .

f étant convexe on a :

$$f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x')$$

Ceci implique :

$$f(\lambda x + x' - \lambda x') \leq \lambda f(x) + (1 - \lambda)f(x')$$

$$f(x' + \lambda(x - x')) \leq \lambda f(x) + (1 - \lambda)f(x')$$

Ou encore :

$$f(x' + \lambda(x - x')) \leq \lambda f(x) + f(x') - \lambda f(x')$$

$$f(x' + \lambda(x - x')) \leq f(x') + \lambda(f(x) - f(x'))$$

$$f(x' + \lambda(x - x')) - f(x') \leq \lambda(f(x) - f(x'))$$

Divisons par λ on obtient :

$$\frac{f(x' + \lambda(x - x')) - f(x')}{\lambda} \leq f(x) - f(x')$$

Et comme f est dit defferentiable au point x' on a :

$$f(x' + \lambda d) = f(x') + \lambda \nabla^t f(x') d + \lambda \|d\| \alpha(x', \lambda d)$$

avec $\alpha(x', \lambda d) \rightarrow 0$ quand $t \rightarrow 0$, donc on obtient :

$$\frac{f(x' + \lambda d) - f(x')}{\lambda} = \nabla^t f(x') d$$

Si on pose $d = x - x'$ on obtient :

$$\nabla f(x') (x - x') = f(x) - f(x')$$

Et comme $\nabla f(x') = 0$ on a donc :

$$0 < f(x) - f(x') \Rightarrow f(x) \leq f(x')$$

D'ou x' est minimum global.

Chapitre 3

Les méthodes de recherches linéaires

Dans ce chapitre nous allons d'écrire les différentes manières de déterminer un pas $\lambda_k > 0$ le long d'une direction de descente d_k . C'est ce que l'on appelle faire de la recherche linéaire. Il existe deux grandes classes de méthodes qui s'intéressent à l'optimisation unidimensionnelle sans contraintes c-à-d aux méthodes numériques pour résoudre le problème de type : $f(x) \rightarrow \min \{ \infty \in \mathbb{R} \}$ f étant une fonction au continue sur l'axe D'habitude. Il ya principalement deux méthodes de recherche linéaire savoir :

- Recherche linéaire d'ordre zéro
- méthode de gradient conjugué.
- Direction de descente.
- Méthode de Dai-Yuan.
- Méthode de newton.

3.1 Objectifs de la recherche linéaire

Il s'agit de réaliser deux objectifs :

3.1.1 Le premier objectif

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

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

Le terme négatif, disons v_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 $f(x_k) \geq c \forall k$) alors ce terme négatif tend nécessairement vers zéro ($v_k \rightarrow 0$). 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 tend vers zéro. Le terme négatif devra prendre une forme bien particulière si on veut pouvoir entirer de l'information. En particulier, il ne suit pas d'imposer $f(x_k + \lambda_k d_k) < f(x_k)$.

3.1.2 Le second objectif

Consiste d'empêcher le pas $\lambda_k > 0$ d'être trop petit, pour assurer la convergence d'algorithme au point stationnaire. Le premier objectif n'est en effet pas suffisant car l'inégalité (2.4) est en générale satisfaite par des pas $\lambda_k > 0$ 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 :

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

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

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

lorsque $l \rightarrow \infty$ donc $\{x_k\}$ converge, disons vers un point \bar{x} . En prenant $l = 1$ et $k \rightarrow \infty$

dans l'estimation ci-dessus, on voit que $\bar{x} \in \bar{B}(x_{1,\varepsilon})$ et donc \bar{x} ne saurait être solution s'il n'y a pas de solution dans $\bar{B}(x_{1,\varepsilon})$. On a donc arbitrairement forcer la convergence de $\{x_k\}$ en prenant des pas très petits.

Pour simplifier les notations, on définit la restriction de f a la droite $\{x_k + td_k : t \in \mathbb{R}\}$ comme la fonction $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}; t \rightarrow \varphi(t) = f(x_k + td_k)$.

3.2 But de la recherche linéaire

Dans le cas non-quadratique les méthodes de descente (2, 2), nécessitent la recherche d'une valeur de $\lambda_k > 0$ optimale ou non, vérifiant :

$$f(x_k + \lambda_k d_k) \leq f(x_k)$$

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

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

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

Autrement dit, λ^* est un minimum local de φ qui assure de plus la décroissance de f .

3.3 Intervalle de sécurité

Définition 3.3.1 On dit que $[a, b]$ est un intervalle de sécurité s'il permet de classer les valeurs de λ de la façon suivante :

- ★ Si $\lambda < a$ alors λ est considéré trop petit.
- ★ Si $b \geq \lambda \geq a$ alors λ est satisfaisant.
- ★ Si $\lambda > b$ alors λ est considéré trop grand.

Le problème est de traduire de façon numérique sur φ les trois conditions précédentes, ainsi que de trouver un algorithme permettant de déterminer a et b . L'idée est de partir d'un intervalle suffisamment grand pour contenir $[a, b]$ et d'appliquer une bonne stratégie pour itérativement réduire cet intervalle.

3.4 Recherche linéaire d'ordre zéro

Nous commençons par la recherche linéaire d'ordre zéro, c-à-d par des méthodes qui utilisent des valeurs de f seulement pas ces dérivées.

Les méthodes que nous sommes sur le point de développer résolvent pas le problème $f(x) \rightarrow \min \mid x \in \mathbb{R}$ tels qu'il est, mais le problème $f(x) \rightarrow \min \mid a \leq x \leq b$ de minimisation de l'objectif sur un segment fini donné $[a, b]$ ($-\infty < a < b < \infty$) pour assurer que le problème soit bien conditionné .nous faisons l'hypothèse suivante : f est unimodale sur $[a, b]$ c-à-d possède un minimum local unique x sur le segment.cette hypothèse comme on le voit facilement implique qui f strictement décroissante sur $[a, b]$ à gauche de x^* :

$$a \leq x' \leq x'' \leq x^* \rightarrow f(x') > f(x'')$$

est strictement croissante sur $[a, b]$ à droite de x^* :

$$x^* \leq x' \leq x'' \leq b \rightarrow f(x') < f(x'')$$

En effet, si (2) étaient faux, il existerait x' et x'' tels que :

$$a \leq x' < x'' \leq x^* \rightarrow f(x') \leq f(x'')$$

Il suit que l'ensemble de minimiseurs de f sur $[a, x'']$ contient un minimiseur, x qui est différent de $x^{(3)}$.

Comme x^* est un minimiseur de f sur $[a, x'']$ et x diffère de x'' , x est un minimiseur local de f sur $[a, b]$ alors qu'on a supposé que le minimiseur local unique de f sur $[a, b]$ est x^* . Ceci donne la contradiction désirée. On a (3) de façon analogue.

Notez que les relations (2) et (3), à leur tour, impliquent que f est unimodal sur $[a, b]$ et même sur chaque segment $[a', b'] \subset [a, b]$ plus petit.

Etant donné que f est unimodal sur $[a, b]$ nous pouvons préciser une stratégie pour approcher x^* :

choisissons deux points x^- et x^+ dans $[a, b]$.

$$a < x^- < x^+ < b$$

et calculons les valeurs $f(x^-)$ et $f(x^+)$ On observe que si :

- Cas A : $f(x^-) \leq f(x^+)$ alors x^* se trouve à gauche de x^+ [en effet, si x^* était à droite de x^+ on aurait $f(x^-) > f(x^+)$ d'après (2)].

- Cas B : $f(x^-) > f(x^+)$, x^* est alors à droite de x^- [raisonnement "symétrique"].

En conséquence dans le cas A nous pouvons remplacer le "segment d'incertitude" initial $\Delta_0 = [a, b]$ par le nouveau segment d'incertitude $\Delta_1 = [a, x^+]$ et dans le cas B par le segment $\Delta_1 = [x^-, b]$

Dans les deux cas le nouveau "segment d'incertitude" Δ_1 couvre x et est strictement plus petit que Δ_0 .

Puisque, l'objectif étant unimodal sur le segment initial $\Delta_0 = [a, b]$, est unimodal également sur le segment plus petit $\Delta_1 \subset \Delta_0$ nous pouvons réitérer ce procédé choisir deux points dans Δ_1 calculer les valeurs de l'objectif en ces points, comparez les résultats et remplacez Δ_0 par un plus petit segment Δ_2 contenant la solution désirée x et ainsi de suite.

3.5 Méthode de gradient conjugué

L'objectif est de minimiser la fonction $f : x \rightarrow \frac{1}{2}(Ax, x) - (b, x)$ ou A est une matrice carrée symétrique définie positive de taille n .

Le calcul montre qu'une solution du problème est la solution du système $Ax = b$. En effet, on a $\nabla f(x) = Ax - b$

3.5.1 La méthode du gradient conjugué vue comme une méthode directe

On rappelle que deux vecteurs non nuls u et v sont conjugués par rapport à A si $u^T A v = 0$.

Sachant que A est symétrique définie positive, on en déduit un produit scalaire $\langle u, v \rangle_A = \langle Au, v \rangle = \langle u, A^T v \rangle = \langle u, Av \rangle = u^T Av$

Deux vecteurs sont conjugués s'ils sont donc orthogonaux pour ce produit scalaire.

La conjugaison est une relation symétrique : si u est conjugué à v pour A , alors v est conjugué à u .

Supposons que $\{p_k\}$ est une suite de n directions conjuguées deux à deux. Alors les $\{p_k\}$ forment une base de R^n , ainsi la solution x^* de $Ax = b$ dans cette base :

$$x^* = \sum_{i=1}^n \alpha_i p_i.$$

Les coefficients sont donnés par :

$$b = Ax^* = \sum_{i=1}^n \alpha_i A p_i.$$

$$p_k^T b = p_k^T Ax^* = \sum_{i=1}^n \alpha_i p_k^T A p_i = \alpha_k p_k^T A p_k.$$

(car $\forall i \neq k, p_i, p_k$ sont conjugués deux à deux)

$$\alpha_k = \frac{p_k^T b}{p_k^T A p_k} = \frac{\langle p_k, b \rangle}{\langle p_k, p_k \rangle_A} = \frac{\langle p_k, b \rangle}{\|p_k\|_A^2}$$

On a ainsi l'idée directrice de la méthode pour résoudre le système $Ax = b$: trouver une suite de n directions conjuguées, et calculer les coefficients α_k .

3.5.2 La méthode du gradient conjugué vue comme une méthode itérative

En choisissant correctement les directions conjuguées p_k , il n'est pas nécessaire de toutes les déterminer pour obtenir une bonne approximation de la solution x^* . Il est ainsi possible de considérer la méthode du gradient conjugué comme une méthode itérative. Ce choix permet ainsi de considérer la résolution de systèmes de très grande taille, où le calcul de l'ensemble des directions aurait été très long.

On considère ainsi un premier vecteur x_0 , qu'on pourra supposer nul (sinon, il faut considérer le système $Az = b - Ax_0$). L'algorithme va consister, partant de x_0 , à se « rapprocher » de la solution x^* inconnue, ce qui suppose la définition d'une métrique. Cette métrique vient du fait que la solution x^* est l'unique minimiseur de la forme quadratique :

$$f(x) = \frac{1}{2}x^T Ax - x^T b, x \in \mathbb{R}^n$$

Ainsi, si $f(x)$ diminue après une itération, alors on s'approche de x^* .

Ceci suggère donc de prendre la première direction p_1 comme l'opposé du gradient de f à $x = x_0$. Le gradient vaut $Ax_0 - b = -b$, d'après notre première hypothèse. Les vecteurs suivants de la base seront ainsi conjugués au gradient, d'où le nom « méthode du gradient conjugué ».

Soit r_k le résidu à la k^e itération :

$$r_k = b - Ax_k$$

Notons que r_k est l'opposé du gradient de f en $x = x_k$, ainsi, l'algorithme du gradient indique d'évoluer dans la direction r_k . On rappelle que les directions p_k sont conjuguées deux à deux. On veut aussi que la direction suivante soit construite à partir du résidu courant et des directions précédemment construites, ce qui est une hypothèse raisonnable en pratique.

La contrainte de conjugaison est une contrainte d'orthonormalité, aussi le problème partage des similitudes avec le procédé de Gram-Schmidt.

On a ainsi :

$$p_{k+1} = r_k - \sum_{i \leq k} \frac{p_i^T Ar_k}{p_i^T Ap_i} p_i$$

Suivant cette direction, le point suivant est donné par :

$$x_{k+1} = x_k + \alpha_{k+1} p_{k+1}$$

avec :

$$\alpha_{k+1} = \frac{p_{k+1}^T b}{p_{k+1}^T Ap_{k+1}} = \frac{p_{k+1}^T (r_k + Ax_k)}{p_{k+1}^T Ap_{k+1}} = \frac{p_{k+1}^T r_k}{p_{k+1}^T Ap_{k+1}}$$

la dernière égalité venant du fait que p_{k+1} et x_k sont conjugués.

3.6 Méthode de Dai-Yuan

une motivation pour nos nouvelles méthodes est la propriété de descente de la méthode de descente conjuguée :

$$\beta_{k+1}^{cd} = \frac{\|g_{k+1}\|^2}{-d_{t_k} g_k}$$

La méthode de descente conjuguée produit toujours une direction de descente si les conditions de Wolfe fortes sont satisfaites. Nous essayons de trouver des méthodes de gradient conjugué qui produisent des directions de descente avec seulement des conditions de Wolfe faibles. C'est l'objet du paragraphe suivant : Cette méthode a été découverte par Dai -Yuan.

3.7 Direction de descente

En optimisation différentiables, qui est une discipline d'analyse numérique en mathématiques étudiant en particulier les algorithmes minimisant des fonctions différentiables sur des ensembles, une direction de descente est une direction le long de laquelle la fonction à minimiser a une dérivée directionnelle strictement négative. Ces directions sont utilisées par les méthodes à directions de descente. C'est le long de ces directions qu'un déplacement est effectué afin de trouver l'itéré suivant, en lequel la fonction à minimiser prend une valeur inférieure à celle qu'elle a en l'itéré courant. Des directions de descente peuvent être calculées par de nombreuses techniques dont les plus classiques sont présentées ci-dessous.

Soient E un espace vectoriel et $f : E \rightarrow \mathbb{R}$ une fonction réelle définie sur E , admettant des dérivées directionnelles au point $x \in E$ considéré. On note :

$$f'(x; d) = \lim_{t \rightarrow 0} \frac{f(x + \lambda d) - f(x)}{t} t \in \mathbb{R}$$

la dérivée directionnelle (au sens de Dini) de f en $x \in E$ dans la direction $d \in E$. La notation $t \rightarrow 0$ signifie que le réel t tend vers zéro dans \mathbb{R} par des valeurs strictement positives.

La notion de direction de descente est surtout utilisée en optimisation numérique. Une direction de descente de f en x est un vecteur $d \in E$ tels que : $f'(x, d) < 0$.

On en déduit que $\forall \lambda > 0$ petit :

$$f(x + \lambda d) < f(x)$$

Si bien que f décroît en x dans la direction d . Cette propriété justifie le nom donné

à cette direction. Certains auteurs utilisent cette dernière propriété comme définition d'une direction de descente. cependant, comme cette propriété n'implique pas que la dérivée directionnelle soit strictement négative, elle n'est pas suffisamment forte pour les algorithmes à directions de descente.

3.7.1 Exemples de direction de descente

Il existe de nombreuses méthodes permettant de calculer une direction de descente. Les principales sont présentées dans cette section, chacune avec ses propres caractéristiques. Un algorithme qui utilise une telle direction hérite d'elle son nom. Ainsi l'algorithme du gradient est l'algorithme à directions de descente qui utilise la direction du gradient, l'algorithme du gradient conjugué utilise la direction du gradient conjugué, etc.

On suppose dans cette section que E est un espace hilbertien, dont le produit scalaire est noté $\langle \cdot, \cdot \rangle$ et la norme associée $\|\cdot\|$ et que $f : E \rightarrow \mathbb{R}$ est au moins une fois différentiable au point x considéré. On note $f'(x)$ sa dérivée, qui est une application linéaire continue de E dans \mathbb{R} et $f'(x) \cdot d = f'(x; d)$ la valeur en $d \in E$ de cette dérivée. Par le théorème de Riesz-Fréchet, il existe alors un vecteur $\nabla f(x) \in E$, appelé le gradient de f en x défini par :

$$\forall d \in E \quad f'(x) \cdot d = \langle \nabla f(x), d \rangle$$

3.7.2 Direction du gradient

La direction du gradient d est réalité l'opposé du gradient :

$$d = -\nabla f(x)$$

Il s'agit bien d'une direction de descente si $f'(x) \neq 0$ puisqu'alors :

$$f'(x) \cdot (-\nabla f(x)) = \langle \nabla f(x), -\nabla f(x) \rangle = -\|\nabla f(x)\|^2 < 0.$$

3.7.3 Direction de Newton

On suppose ici que la fonction f à minimiser est deux fois différentiable en x et on désigne par $\nabla^2 f(x)$ son hessien en x lequel est l'unique opérateur linéaire auto-adjoint $\nabla^2 f(x) : E \rightarrow E$ vérifiant :

$$\forall (h, k) \in E^2 : f''(x) \cdot (h, k) = \langle \nabla^2 f(x)h, k \rangle$$

La direction de Newton est définie en un point x en lequel le hessien de f est inversible

par :

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

Cette direction est une direction de descente si :

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

En effet :

$$f'(x) \cdot d = \langle \nabla f(x), d \rangle = \langle -\nabla f(x), \nabla^2 f(x)^{-1} \nabla f(x) \rangle \leq -\lambda_{\max}^{-1} \|\nabla f(x)\|^2 < 0$$

où λ_{\max} désigne la plus grande valeur propre de $\nabla^2 f(x)$. La seconde condition assurant le caractère descendant de la direction de Newton sera vérifiée dans le voisinage d'une solution vérifiant les conditions suffisantes d'optimalité du deuxième ordre.

3.7.4 Direction de quasi-Newton

Les algorithmes de quasi-Newton en optimisation définissent une direction de descente en prenant une approximation convenable du hessien du critère au moyen d'un opérateur M auto-adjoint $M \sim \nabla^2 f(x)$.

Une direction de quasi-Newton est donc de la forme :

$$d = -M^{-1} \nabla f(x).$$

Comme pour la direction de Newton, cette direction est une direction de descente si :

- ★ $f'(x) \neq 0$
- ★ M est définie positive.

En effet :

$$f'(x) \cdot d = \langle \nabla f(x), d \rangle = -\langle \nabla f(x), M^{-1} \nabla f(x) \rangle \leq -\lambda_{\max}^{-1} \|\nabla f(x)\|^2 < 0$$

où λ_{\max} désigne la plus grande valeur propre de M .

3.7.5 Direction de Gauss-Newton

La direction de Gauss-Newton est utilisée pour résoudre les problèmes de moindres-carrés dont le critère est de la forme :

$$f(x) = \frac{1}{2} \|F(x)\|^2$$

où $F : E \rightarrow F$ (F est un espace hilbertien dont le produit scalaire est aussi noté $\langle \cdot, \cdot \rangle$ et $\|\cdot\|$ est la norme associée). On calcule aisément :

$$f'(x) \cdot h = \langle F(x), F'(x) \cdot h \rangle$$

$$\begin{aligned}
&= \langle F'(x)^* F(x), h \rangle \\
f''(x) \cdot (h, k) &= \langle F'(x) \cdot h, F'(x) \cdot k \rangle + \langle F(x), F''(x) \cdot (h, k) \rangle \\
&= \langle (F'(x)^* F'(x))h \cdot k \rangle + \langle F(x), F''(x) \cdot (h, k) \rangle .
\end{aligned}$$

La direction de Gauss-Newton s'obtient en ne gardant du hessien de f que son premier terme dans l'expression ci-dessus, de manière à éviter le calcul des dérivées secondes de F . C'est en réalité une solution arbitraire de l'équation normale.

$$(F'(x)^* F'(x))d = -F'(x)^* F(x).$$

On reconnaît dans le membre de droite l'opposé du gradient de f . Cette équation linéaire a en fait une solution unique si et seulement si $F'(x)$ est injective. Les directions de Gauss-Newton sont aussi les solutions du problème de moindres-carrés linéaire suivant :

$$\min_{x \in E} \frac{1}{2} \|F(x) + F'(x) \cdot d\|^2$$

Une direction de Gauss-Newton d est une direction de descente de f en x si $f'(x) \neq 0$. En effet :

$$f'(x) \cdot d = \langle \nabla f(x), d \rangle = -\langle (F'(x)^* F'(x))d \cdot d \rangle = -\|F'(x)d\|^2 < 0.$$

L'inégalité stricte vient du fait que si $F'(x)d = 0$, alors : $\nabla f(x) = F'(x)^* F(x)$ est nulle par l'équation normale, ce que nous avons supposé ne pas avoir lieu.

Bibliographie

- [1] J.F. Bonnans, J.C.Gilbert, C.Lemaréchal, C.sagastizabal , Méthode numériques d'optimisation, France (1995).
- [2] J.B Hiriart-Urruty, C. Lemaréchal, Convex analysis and minimization algorithms I : fundamentals, Springer-Verlag, Berlin Heidelberg, New York, (1993).
- [3] A. Keraghel, Analyse convexe : Théorie fondamentale et exercices, Editions Dar El'houda Ain Mlila , Algérie, (2001).
- [4] M. Minoux, Programmation mathématique : Théorie et algorithmes, T1 et T2 Dunod, (1983).
- [5] R.T. Rockafellar, Convex analysis, published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540.
- [6] M. Sakharovitch, Graphe et programmation linéaire, Hermann, Paris, (1984).
- [7] M. Simonard, programmation linéaire, dunod Paris T1 et T2, (1972).
- [8] Jean-Baptiste Hiriart-Urruty "OPTIMISATION ET ANALYSE CONVEXE" mars France 2009
- [9] A. Keraghel, Programmation mathématique différentiable, Université Ferhat Abbas, Sètif Algérie.