REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Mohamed Boudiaf de M'sila
Faculté des Mathématiques et de l'Informatique
Département des Mathématiques

Mémoire de Master

Domaine : Mathématiques et Informatique
Filière : Mathématiques
Option : Analyse Mathématiques et numérique

Thème

Résolution d'un problème d'optimisation quadratique sans contraintes

---

Présentée par :
ABDELLI SAMIRA

Soutenu publiquement le : 06/2024

Devant le jury composé de :

BELAALA Maatougui	M.C.B,	Université de M'sila	Président
SAADI Khalil	Prof,	Université de M'sila	Encadreur
SEGHIRI Fakher eddine	M.A.A,	Université de M'sila	Examinateur

Année universitaire 2023/2024.

Remerciement

Tout d'abord, nous remercions Allah, notre Créateur, qui nous a donné la force, la Volonté et le courage d'accomplir ce travail important et utile dans Le domaine d'optimisation.

Je tiens à remercier profondément mon directeur de thèse, Professeur. SAADI Khalil, qui a proposé le sujet de cette mémoire et pour son précieux soutien, ses Conseils et son Assistance tout au long de ce projet.

Nous remercions également les membres du comité de jugement et tous les professeurs du département de mathématiques pour toutes les connaissances qu'ils nous avons apporté au cours de notre parcours académique.

Je tiens également à remercier ma famille et mes amis pour leur soutien et leurs Encouragements continus pendant que je travaillais sur ce mémoire.

Enfin, je remercie tous ceux qui ont contribué directement et indirectement à la Réalisation de ce travail.

الملخص

الهدف من هذه الأطروحة هو استكشاف متعمق لتقنيات حل مشاكل التحسين التربيعي دون قيود، من خلال تحليل المبادئ النظرية وكذلك الجوانب العملية لتنفيذها وتتمحور هذه الأطروحة حول ثلاثة فصول

الفصل الأول: العموميات والمفاهيم الأساسية

الفصل الثاني: المعالجة التحليلية لمشكلة التحسين تربيعية بدون قيود

الفصل الثالث: الحل العددي لمشكلة التحسين تربيعية بدون قيود

Abstract

The objective of this dissertation is to explore in depth the techniques for solving quadratic optimization problems without constraints, by analyzing the theoretical principles as well as the practical aspects of their implementation.

This dissertation is structured around three chapters:

- **Chapter 1: Generalities and basic concepts.**
- **Chapter 2: Analytical treatment of an optimization problem quadratic without constraints.**
- **Chapter 3: Numerical resolution of an optimization problem quadratic without constraints.**

Résumé

L’objectif de ce mémoire est d’explorer en profondeur les techniques de résolution des problèmes d’optimisation quadratique sans contraintes, en analysant les principes théoriques ainsi que les aspects pratiques de leur mise en œuvre.

Ce mémoire s’articule autour de trois chapitres :

- **Chapitre 1 : Généralités et notions de bases.**
- **Chapitre 2 : Traitement analytique d’un problème d’optimisation quadratique sans contraintes.**
- **Chapitre 3 : Résolution numérique d’un problème d’optimisation quadratique sans contraintes.**

Table des matières

Introduction

L'optimisation quadratique sans contraintes est une branche particulière de l'optimisation mathématique, se concentrant sur les problèmes où l'objectif est de minimiser ou de maximiser une fonction quadratique sans contraintes. Les problèmes d'optimisation quadratique sont couramment formulés comme suit :

$$\begin{cases} \min f(x) = \frac{1}{2}x^T A x - b^T x + c \\ x \in \mathbb{R}^n \end{cases}$$

où A est une matrice symétrique de format $n \times n$, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

Ces problèmes se rencontrent fréquemment dans divers domaines tels que l'économie, la finance, l'ingénierie, et les sciences appliquées. La structure quadratique de la fonction objectif permet l'application de méthodes analytiques et numériques spécifiques pour trouver les points optimaux.

L'absence de contraintes simplifie l'analyse théorique et la mise en œuvre algorithmique, mais elle ne réduit pas l'importance de la compréhension approfondie des propriétés de la fonction quadratique. Les méthodes classiques utilisées pour résoudre ces problèmes incluent la méthode du gradient, la méthode des gradients conjugués et la méthode de Newton. Chacune de ces méthodes offre des avantages en termes de convergence et d'efficacité, en fonction des caractéristiques spécifiques du problème.

L'objectif de ce mémoire est d'explorer en profondeur les techniques de résolution des problèmes d'optimisation quadratique sans contraintes, en analysant les principes théoriques ainsi que les aspects pratiques de leur mise en œuvre.

Ce mémoire s'articule autour de trois chapitres :

Chapitre 1 : Dans ce chapitre, nous rappelons quelques définitions et notions fondamentales des fonctions de plusieurs variables. Nous aborderons la différentiabilité des fonctions de plusieurs variables ainsi que les ensembles et fonctions convexes. Ces concepts seront essentiels pour l'étude et la résolution d'un problème d'optimisation quadratique sans contraintes.

Chapitre 2 : Nous étudierons analytiquement les problèmes d'optimisation sans contraintes en général. L'accent sera mis sur les formes quadratiques et sur les méthodes qui permettent d'obtenir des solutions pour ce type de problèmes.

Chapitre 3 : Ce chapitre traite de la résolution numérique d'un problème d'optimisation sans contrainte, en se basant principalement sur la condition du premier ordre. Tous les algorithmes d'optimisation fournissent une solution approximative correspondant aux zéros du vecteur gradient $\nabla f(x)$. Avec des conditions de convergence adéquates, ces zéros constitueront des solutions approximatives au problème d'optimisation.

Trois méthodes sont abordées dans ce chapitre : la méthode du gradient, la méthode des gradients conjugués et la méthode de Newton. Pour chaque méthode, nous examinerons le principe général ainsi que les résultats de convergence.

Chapitre 1

Généralités et notions de bases

Dans ce chapitre, nous allons rappeler quelques défnitions et notions d'une fonction de plusieurs variables. On va faire un rappel sur la différentiabilté des fonctions de plusieurs variables ainsi que les ensembles et les fonctions convexes. Ceci nous sera utile dans l'étude et résolution d'un problème d'optimisation quadratique sans contraintes .

1.1 Définition d'une fonction de plusieurs variables

Définition 1.1.1 *Une fonction de plusieurs variables est une fonction numérique de n variables réelles, elle est notée :*

$$\begin{array}{ccc} f:\mathbb{R}^n & \to & \mathbb{R} \\ (x_1,...,x_n) & \mapsto & f(x_1,...,x_n) \end{array}.$$

Si f est une fonction de deux variables, alors le graphe de f est l'ensemble des points $(x,y,z)\in\mathbb{R}^3$ avec $z=f(x,y)$. Le graphe de la fonction f se présente sous la forme d'une surface.

Exemple 1.1.1 *1. $f(x,y)=ax+by$ f est linéaire en x et en y Le graphe de f*

est un plan .

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

1.2 Dérivée partielle d'une fonction de plusieurs variables

Définition 1.2.1 *(Dérivées partielles en un point). Soit f une fonction de deux veriables x et y. On se demande comment varie f si on ne modifie qu'une seule variable x ou y. La réponse est donnée par la dérivée partielle de f par apport à cette variable .Si on fait varier x en gardant une valeur constante de y (par exemple $y = b$) alors on peut voir f comme une fonction d'une seule veriable x.On peut écrire*

$$g(x) = f(x, b).$$

Si g est dérivable en a (pour $x = a$), alors on appelle cette dérivée, la dérivée partielle de f par rapport à x en (a, b) que l'on note $f_x(a, b)$

$$f_x(a, b) = g\acute{}(a)$$

Par définition de la dérivée, on a :

$$g\acute{}(a) = \lim_{h \longrightarrow 0} \frac{g(a+h) - g(a)}{h}$$

Ou bien :

$$f_x(a,b) = \lim_{h \longrightarrow 0} \frac{f(a+h,b) - f(a,b)}{h}$$

Définition 1.2.2 *(Dérivée partielle). Si on fait varier le point (a, b), f_x et f_y de-*

viennent des fonctions de deux variables. Si f est une fonction de deux veriables, alors ses dérivées partielles sont les fonctions f_x et f_y définies de la façon suivante :

$$f_x(x,y) = \lim_{h \longrightarrow 0} \frac{f(x+h,y) - f(x,y)}{h}$$

$$f_y(x,y) = \lim_{h \longrightarrow 0} \frac{f(x,y+h) - f(x,y)}{h}$$

Les règles de calcul des dérivée partielles sont celles des fonctions d'une seule veriables.

Notation de Leibnitz

$*$) $f_x(x,y)$ est notée $\frac{\partial}{\partial x} f(x,y)$.

$*$) $f_y(x,y)$ est notée $\frac{\partial}{\partial y} f(x,y)$.

Exemple 1.2.1 *1.*

$$\begin{aligned} f(x,y) &= ax + by \\ \frac{\partial}{\partial x} f(x,y) &= a \\ \frac{\partial}{\partial y} f(x,y) &= b \end{aligned}$$

2.

$$\begin{aligned} g(x,y) &= xy \\ \frac{\partial}{\partial x} f(x,y) &= y \\ \frac{\partial}{\partial y} f(x,y) &= x \end{aligned}$$

1.3 Dérivée directionnelle

Définition 1.3.1 *On souhaite déterminer la variation de $f(x,y)$ lorsqu' on se déplace dans la directionnelle du vecteur*

$$u=\begin{bmatrix} a \\ b \end{bmatrix}.$$

Cette veriation est appelée dérivée directionnelle de la fonction f, dans la direction du vecteur u. La dérivée directionnelle est notée $\nabla_u f(x,y)$, elle est définie de la façon suivante

$$\nabla_u f(x,y)=\frac{\partial}{\partial x}f(x,y)\,a+\frac{\partial}{\partial y}f(x,y)\,b$$

1.4 Vecteur gradient

Définition 1.4.1 *Soit $f:\mathbb{R}^n\longrightarrow\mathbb{R}$ une fonction de classe C^1. Le vacteur gradient de la fonction f est défini par*

$$g(x)=\nabla f(x)=\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ .. \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

où $\frac{\partial f}{\partial x_i}$ est la dérivée partielle de f par apport à x_i, $i=1,...,n$.

1.5 Matrice hessienne

Définition 1.5.1 *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de classe C^2. La matrice hessienne de la fonction f est définie par*

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

où $\frac{\partial^2 f}{\partial^2 x_i}$ est la dérivée partielle d'ordre 2 de f.

1.6 Relation entre dérivée directionnelle et vecteur gradient

La dérivée directionnelle peut être vue comme le produit scalaire de deux vecteurs

$$\begin{aligned} \nabla_u f(x,y) &= \frac{\partial}{\partial x} f(x,y)\, a + \frac{\partial}{\partial y} f(x,y)\, b \\ &= \begin{bmatrix} \frac{\partial}{\partial x} f(x,y) \\ \frac{\partial}{\partial y} f(x,y) \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x} f(x,y) \\ \frac{\partial}{\partial y} f(x,y) \end{bmatrix} \cdot u \end{aligned}$$

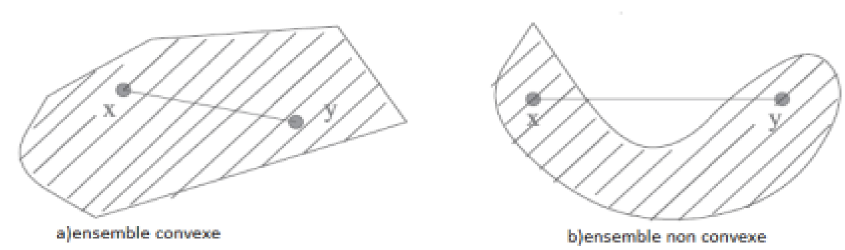
Le premier vecteur est le vecteur gradient de la fonction $f(x,y)$. La dérivée directionnelle s'écrit donc

$$\nabla_u f(x,y) = \nabla f(x,y) \cdot u$$

1.7 Ensemble convexe

Définition 1.7.1 *Soit C un sous-ensemble de $\mathbb{R}^n$. On dit que C est convexe si pour tout $x, y \in C$ le segment $[x, y]$ est inclus dans C, c'est à dire,*

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



Exemple d'ensemble convexe et non convexe

a exemple d'ensemble convexe.

b exemple d'ensemble non convexe (noter qu'il existe de segment dont les extrémités appartiennent à l'ensemble, qui ne sont pas entièrement contenus dans le ensemble).

1.8 Fonction convexe (strictement convexe)

Définition 1.8.1 *Soit C un sous-ensemble convexe de $\mathbb{R}^n$ et $f : C \longrightarrow \mathbb{R}$ une fonction.*

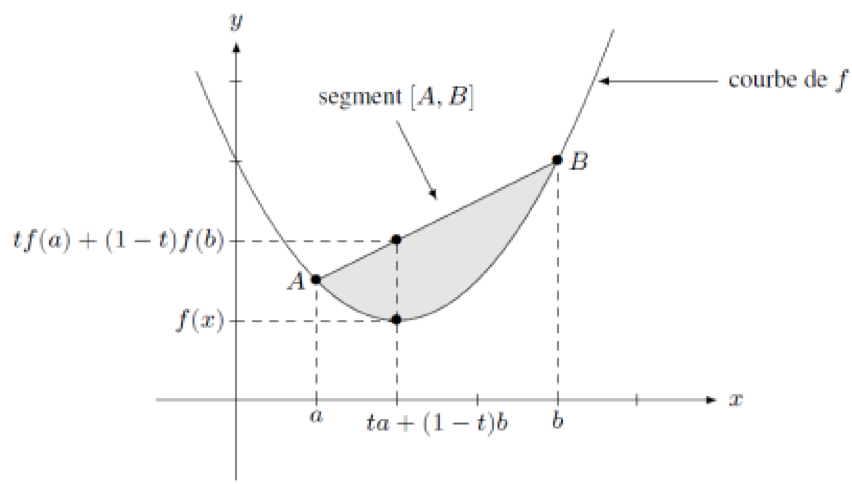


FIG. 1.1 – Fonction convexe

1. *On dit que f est convexe sur C si*

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

2. *On dit que f est strictement convexe sur C si*

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

la définition de f convexe signifie que si l'on prend deux points sur la courbe de f, alors le segment défini par ces deux points reste au dessus de la courbe de f.

1.9 Relation entre la convexité et différentiabilité

Théorème 1.9.1 *(Inégalité de gradient). Soient C sous-ensemble convexe de $\mathbb{R}^n$ et $f : C \longrightarrow \mathbb{R}$ une fonction différentiable. Alors, f est convexe sur C si et seulement si*

$$f\,(y) \geq f\,(x) + \langle \nabla f\,(x)\,, y - x\rangle \text{ pour tout } x, y \in C.$$

Preuve. On suppose que f est convexe sur C. Soient $x, y \in C$ et $\lambda \in]0,1]$. Alors,

1. Si $x = y$, alors l'inégalité de gradient est triviale.

2. Si $x \neq y$, nous avons

$$f[(1-\lambda)x + \lambda y] \leq (1-\lambda) f(x) + \lambda f(y)$$

et donc

$$f(y) \geq f(x) + \frac{f[x + \lambda(y-x)] - f(x)}{\lambda}$$

D'après la différentiabilité de f et par passage à la limite $\lambda \longrightarrow 0^+$,on obtient l'inégalité de gradient. Réciproquement, on suppose que f est vérifiée l'inégalité de gradient . Soient $x, y \in C$ et $\lambda \in]0,1[$. On pose

$$z = (1-\lambda)x + \lambda y.$$

Alors, on a

$$x - z = \frac{z - \lambda y}{1-\lambda} - z = -\frac{\lambda}{1-\lambda}(y-z).$$

En utilisant l'inégalité de gradient pour x, z et y, z , nous obtenos

$$f(x) \geq f(z) + \langle \nabla f(z), x - z \rangle.$$

$$f(y) \geq f(z) - \frac{1-\lambda}{\lambda} \langle \nabla f(z), x - z \rangle.$$

En multipliant la premier inégalité par $\frac{1-\lambda}{\lambda}$ et en l'ajoutant à la seconde , on obtient

$$\frac{f(z)}{\lambda} \leq \frac{1-\lambda}{\lambda} f(x) + f(y)$$

En multipliant l'négalité précédent par λ, on obtient

$$f(z) \leq (1-\lambda) f(x) + \lambda f(y).$$

d'où f est convexe.

■

Une légère modification de la preuve ci-dessus montrera qu'une fonction est strictement convexe si et seulement si l'inégalité de gradient est satisfaite avec une inégalité strict pour tout $x \neq y$.

Théor■me 1.9.2 (L'inégalité de gradient pour la fonction strictement convexe) *Soient C un sous-ensemble convexe de $\mathbb{R}^n$ et $f : C \longrightarrow \mathbb{R}$ une fonction différentiable. Alors, f est strictement convexe sur C si et seulement si*

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle,$$

pour tout $x, y \in C$ et $x \neq y$.

Chapitre 2

Traitement analytique d'un problème d'optimisation quadratique sans contraintes

Dans ce chapitre, on va étudier analytiquement des problemes d'optimisation sans contraintes en général. On se concentre sur les formes quadratiques et sur tout ce qui nous aide à obtenu la solution pour ce type de problèmes.

2.1 Classification d'un problème d'optimisation

Le problème d'optimisation peut être classé de plusieurs façon, voici quelques classification courantes :

Caractéristiques	Proprietétés	Classification
Nombre de variables	une seule variable	Monovariable
	plus d'une variable	Multivariable
Type de variables	Réelles	Continue
	Entières	Discrète
	Réelles et entières	Mixte
	Entières avec permutation	Combinatoire
Type de fonction objectif	Linéaire en fonction des variables	Linéaire
	Quadratique en fonction des variables	Quadratique
	Non linéaire en fonction des variables	Non linéaire
Formulation du problème	Soumis à des limitation	Avec contraine
	Pas de limitations	Sans contraine

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de plusieurs variables . Un problème d'optimisation se formule de la façon suivante :

- Problème avec contraintes

$$\begin{cases} \min f(x) \\ x \in D \end{cases}$$

L'ensemble $D \in \mathbb{R}^n$ s'appelle l'ensemble des contraintes. La fonction f s'appelle fonction objectif ou fonction de cout .

- Problème sans contraintes

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

Ici on essaie de trouver un vecteur $x^*\in\mathbb{R}^n$, tel que

$$f(x^*)\leq f(x)$$

pour tout $x\in\mathbb{R}^n$. Le vecteur x^* s'appelle un minimum global. Sinon on se content d'un minimum local.

Remarque 2.1.1 *Nous avons la relation suivante*

$$\max f(x) = -\min -f(x)$$

c'est à dire, tout problème de maximisation peut être transformé en un problème de minimisation

$$\begin{cases}\max f(x)\\ x\in\mathbb{R}^n\end{cases}\Longleftrightarrow\begin{cases}\min -f(x)\\ x\in\mathbb{R}^n\end{cases}$$

2.2 Résultats d'existence (Théorème de Weierstrass)

Théor■me 2.2.1 *(Weierstrass). Soient E un sous-ensemble compact (fermé et borné) non vide de $\mathbb{R}^n$ et $f:E\longrightarrow\mathbb{R}$ une fonction continue sur E. Alors, f est bornée et atteint ses bornes. Autrement dit, il existe $x\in E$ point de minimum global de f sur E i.e :*

$$\forall y\in E: f(x)\leq f(y).$$

De la même façon, il existe un point de maximum global de f sur E.

Preuve. *Soit (x_n) une suite minimisante dans $f(E)$,i.e : d'éléments de E telle que*

$$\lim_{n \longrightarrow +\infty} f(x_n) = \inf f(E).$$

Comme E est fermé borné, il existe une sous-suite extraite $\left(x_{\varphi(n)}\right)$ qui converge vers un $x \in E$. Cette suite extraite vérifie :

$$x_{\varphi(n)} \underset{n \longrightarrow +\infty}{\longrightarrow} x \text{ et } f\left(x_{\varphi(n)}\right) \underset{n \longrightarrow +\infty}{\longrightarrow} \inf f(E).$$

Or f est continue, d'où par unicité de la limite, il s'en suit

$$f(x) = \inf_{y \in E} f(y) = \min_{y \in E} f(y).$$

et f réalise son minimum sur E. ■

2.3 Fonction coercive (Théorème d'existence d'un minimum global)

Définition 2.3.1 *(Fonction coercive). On dit que $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est coercive si*

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

C'est à dire

$$\forall A \in \mathbb{R}, \exists B > 0 / \forall x \in \mathbb{R}^n : [\|x\| \geq B \Longrightarrow f(x) \geq A].$$

Ici $\|.\|$ désigne une norme quelconque de $\mathbb{R}^n$.

Exemple 2.3.1 *$f : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par $f(x,y) = x^2 + y^2$ est coercive.*
$g : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par $g(x,y) = x^2 - y^2$ n'est pas coercive . En effet, la suite de terme général. En effet, la suite de terme général $x_n = (0,n), n \in N$ est telle que

$$\lim_{n \longrightarrow +\infty} \|x_n\| = \lim_{n \longrightarrow +\infty} n = +\infty.$$

Mais

$$\lim_{n \longrightarrow +\infty} g(x_n) = \lim_{n \longrightarrow +\infty} (-n^2) = -\infty.$$

Proposition 2.3.1 *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction et soit $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction vérifiant*

$$f(x) \geq g(\|x\|), \text{ avec } \lim_{t \longrightarrow +\infty} g(t) = +\infty.$$

Alors, f est coercive.

Preuve. Comme $g \underset{t \longrightarrow +\infty}{\longrightarrow} +\infty$, on a

$$\forall A \in \mathbb{R}, \exists B > 0 / \forall t \in \mathbb{R}, t \geq B \Longrightarrow g(t) \geq A.$$

Proposition 2.3.2 *On pose $t = \|x\|$, on obtient*

$$\forall A \in \mathbb{R}, \exists B > 0 / \forall t \in \mathbb{R}, t \geq B \Longrightarrow f(x) \geq g(\|x\|) \geq A.$$

■

Théor■me 2.3.1 *(Existance). Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue et coercive. Alors, f admet au moins un point de minimum global sur $\mathbb{R}^n$.*

Preuve. Soit $\alpha \in \mathbb{R}^n$ et considérons l'ensemble

$$E = \{x \in \mathbb{R}^n / f(x) \leq f(\alpha)\}.$$

Il est facile de montrer :

1. E est fermé, car :

$$E = f^{-1}(]-\infty, f(a)]),$$

   Donc E est l'image inverse d'un intervalle fermé par une fonction continue.

2. E est borné : supposons le contraire, alors il existe une suite (x_k) de E avec $x_k \underset{k \longrightarrow +\infty}{\longrightarrow} +\infty$. Comme f est coercive sur $\mathbb{R}^n$, ceci entraine $f(x_k) \underset{k \longrightarrow +\infty}{\longrightarrow} +\infty$. donc contradiction, car :

$$f(x_k) \leq f(\alpha), \forall k \in \mathbb{N}.$$

   On déduit alors que E est un ensemble compact dans $\mathbb{R}^n$.

■

Théor■me 2.3.2 *(Condition suffisante d'optimalité globale). Soient* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction et* x^* *un point de minimum local de* f *sur* $\mathbb{R}^n$*. Alors, nous avons :*

1. *Si* f *est convexe, alors* x^* *est un point de minimum global de* f *sur* $\mathbb{R}^n$*.*

2. *Si* f *est strictement convexe, alors* x^* *est l'unique point de minimum global de* f *sur* $\mathbb{R}^n$*.*

2.4 Conditions d'optimalité d'un extrémum local

2.4.1 Conditions nécessaires du premier ordre

Théor■me 2.4.1 *(Condition nécessaire d'optimalité de premier ordre). Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction différentiable en* $x^* \in \mathbb{R}^n$ *est un point de minimum (est un point de maximum local) de* f *sur* $\mathbb{R}^n$*,alors*

$$\nabla f\left(x^*\right) = 0$$

Preuve. Si x^* est un point de minimum local de f sur $\mathbb{R}^n$, alors

$$\forall x \in B\left(x^*, \rho\right), f\left(x^*\right) \leq f\left(x\right)$$

Où $B\left(x^*, \rho\right)$ est une boule ouverte de rayon ■ centrée en x^* On suppose que $\nabla f\left(x^*\right) \neq 0$. Soit $h \in \mathbb{R}^n, h \neq 0$, on peut trouver $t_h = \frac{\rho}{\|h\|} > 0$ tel que

$$\forall t \in \left]0, t_h\right[, x^* + th, x^* - th \in B\left(x^*, \rho\right)$$

et donc

$$\forall t \in \left]0, t_h\right[, f\left(x^*\right) \leq f\left(x^* + th\right) \quad \text{et } f\left(x^*\right) \leq f\left(x^* - th\right)$$

Ou f est différentiable en x^*, donc

$$\lim_{x \longrightarrow 0^+} \frac{f\left(x^* + th\right) - f\left(x^*\right)}{t} = \left\langle \nabla f\left(x^*\right), h\right\rangle \text{ et } \lim_{x \longrightarrow 0^+} \frac{f\left(x^* - th\right) - f\left(x^*\right)}{t} = \left\langle \nabla f x^*, -h\right\rangle$$

Donc

$$\forall h \in \mathbb{R}^n \left\langle \nabla f\left(x^*\right), h\right\rangle \geq 0 \text{ et } \left\langle \nabla f\left(x^*\right), -h\right\rangle \geq 0.$$

C'est à dire

$$\forall h \in \mathbb{R}^n \setminus \{0\}, \langle \nabla f(x^*), h \rangle = 0$$

On pose $h = \nabla f(x^*)$, alors on a

$$\|\nabla f(x^*)\|^2 = 0 \Longleftrightarrow \nabla f(x^*) = 0.$$

■

Remarque 2.4.1 *1. Ce théorème n'a pas de sens si la fonction f n'est pas différentiable.*

2. Le théorème précédent donne une condition nécessaire mais non suffisante. En effet $\nabla f(x^*) = 0.$ n'entraîne pas que f atteigne un minimum (ou un maximum) même local,en x^*.

Prendre par exemple $x^* = 0.$ et la fonction f définie par $f(x) = x^3$ pour s'en convaincre.

Définition 2.4.1 *Un point $x^* \in \mathbb{R}^n$ vérifiant $\nabla f(x^*) = 0$ est appelé point critique ou point stationaire. La relation*

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

est aussi appelée équation d'Euler.

Théor■me 2.4.2 *(CNS du premier ordre dans le cas convexe). Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction différentiable et convexe sur $\mathbb{R}^n$. Le point $x^* \in \mathbb{R}^n$ est un point de minimum global de f sur $\mathbb{R}^n$ si et seulement si $\nabla f(x^*) = 0$.*

Preuve. On a vu que la condition est toujours nécessaire. Montrons qu'elle est suffisante. Soit $x^* \in \mathbb{R}^n$, tel que $\nabla f(x^*) = 0$.Comme f est convexe, on obtient

$$\forall x \in \mathbb{R}^n, f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*).$$

On a donc immédiatement le fait que x^* est un point de minimum global de f sur $\mathbb{R}^n$. ■

2.4.2 Conditions du deuxième ordre

Théor■me 2.4.3 (Condition nécessaire du second ordre) *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de class C^2 sur $\mathbb{R}^n$ et $x^* \in \mathbb{R}^n$. Nous avons :*

1. *Si x^*est un point de minimum local de f, alors :*
   (a) $\nabla f(x^*) = 0$
   (b) $\nabla^2 f(x^*)$ *est semi-définie positive.*
2. *Si x^*est un point de maximum local de f, alors :*
   (a) $\nabla f(x^*) = 0$
   (b) $\nabla^2 f(x^*)$ *est semi-définie négative.*

Preuve. (1. a déjà été vue. Montrons 2.). Soit $x \in \mathbb{R}^n$ Appliquons la formule de Taylor-Youg à la fonction $\varphi(t) = f(x^* + tx)$. Comme $\nabla f(x^*) = 0$, on obtient

$$0 \leq f(x^* + tx) - f(x^*) = \frac{t^2}{2} \langle \nabla^2 f(x^* + tx) x, x \rangle + o\left(t^2\right).$$

Après division par t^2, on fait tendre t vers 0 et on a la résultat voulu . ■

Exemple 2.4.1 *La réciproque du théorème précédent est fausse. Il suffit de penser des exemples suivants*

1. *Soit f la fonction définie sur $\mathbb{R}$ par $f(t) = t^3$ .Nous avons*

$$f'(t) = 3t^2 \text{ et } f''(t) = 6t$$

Donc $f'(0) = f''(0) = 0$.Alors,0 est un point critique de f mais la dérivée seconde change le signeau voisinage de 0, donc 0 ni minimum ni maximum donc est un point d'inflexion.

2. *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par*

$$f(x,y) = 3x^4 - 4x^2y + y^2.$$

Preuve. Nous avons :

$$\nabla f(x,y) = 0 \Longleftrightarrow \begin{cases} 3x^3 - 2xy = 0 \\ -2x^2 + y = 0 \end{cases} \Longleftrightarrow (x,y) = (0,0).$$

■

Exemple 2.4.2 *Alors $(0,0)$ est un point critique de f.*

On a :

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}.$$

La matrice Hessienne $\nabla^2 f(0,0)$ est semi-définie positive puisque ses valeurs propres sont 0 et 2.

D'autre part, nous avons :

$$-6025 = f\left(\frac{1}{2}, \frac{1}{2}\right) \leq f(0,0) \leq f(0,1) = 1.$$

Donc, le point critique $(0,0)$ ni un minimum ni un maximum.

On peut aussi remarquer que f n'est pas coercive puisque $f(n,n^2) = 0$ si n tend *vers* $+\infty$.

Nous pouvons toutefois donner une réciproque sous forme de condition suffisante du second ordre plus forte (pour un résultat plus faible) de ce qui précède.

Théor■me 2.4.4 (Condition suffisante du second ordre) *Soient* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction de classe* C^2 *sur* $\mathbb{R}^n$ *et* $x^* \in \mathbb{R}^n$.

1. *Si*
$$\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ est définie positive .} \end{cases}$$
alors, x^* *est un point de minimum local strict de* f *sur* $\mathbb{R}^n$.

2. *Si*
$$\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ est définie négative.} \end{cases}$$
alors, x^* *est un point de maximum local strict de* f *sur* $\mathbb{R}^n$.

Preuve. Soit $x \in \mathbb{R}^n$. On utilise de nouveau la formule de Taylor appliquée à la fonction $\varphi : t \longrightarrow \varphi(t) = f(x^* + tx)$. Nous avons

$$f(x^* + tx) - f(x^*) = \frac{t^2}{2}\langle \nabla^2 f(x^*)\, x, x\rangle + o\left(t^2\right) \geq \frac{t^2}{2}\alpha \|x\|^2 + o\left(t^2\right).$$

Ceci montre que x^*réalise un minimum local strict de f . ■

2.5 Forme quadratique (forme matricielle)

Définition 2.5.1 *On dit fonction quadratique, toute fonction* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *définie par*

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c.$$

où A est une matrice carré d'ordre $n, b \in \mathbb{R}^n, c \in \mathbb{R}$ et $\langle .,. \rangle$ est le produit scalaire Euclidien.

Proposition 2.5.1 *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction quadratique définie par*

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c.$$

Alors,

1. $f \in C^\infty$ sur $\mathbb{R}^n$
2. Pour tout $x \in \mathbb{R}^n$,

$$\nabla f(x) = \frac{1}{2}\left(A + A^T\right)x - b \text{ } et \text{ } \nabla^2 f(x) = \frac{1}{2}\left(A + A^T\right).$$

3. Si A est symétrique, on a :

$$\nabla f(x) = Ax - b \text{ } et \text{ } \nabla^2 f(x) = A.$$

4. Si A est symétrique et semi-définie positive , alors f est une fonction convexe.
5. Si A est symétrique et semi-définie négative , alors $f$est une fonction concave.
6. Si A est symétrique et définie positive , alors f admet une unique point de minimum global vérifiant le système linéaires suivant :

$$Ax = b.$$

7. Si A est symétrique et définie négative , alors f admet une unique point de maximum global vérifiant le système linéaires suivant :

$$Ax = b.$$

2.6 Méthode générale pour résoudre un problème d'optimisation sans contraintes

La résolution d'un probléme d'optimisation sans contraintes s'interesse à trouver les points où une fonction $f(x)$ attient ses valeurs extrêmes (minimales ou maximales) sans restriction sur les variables x. Voici une méthode générale pour aborder ce type de problème :

1- Formulation du problème

On cherche à minimiser ou maximiser une fonction $f(x)$, où x est un vecteur de variables de décision.

2- Calculer le gradient

Le gradient de la fonction $f(x)$ est un vecteur des dérivées partielles de f par rapport à chacune des variables :

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}\right)^T.$$

3- Trouver les points critiques

Les points critiques de $f(x)$ sont trouvés en résolvant le systéme d'équations suivants :

$$\nabla f(x) = 0.$$

4- Analyser la nature des points critiques

Pour déterminer si chaque point critique est un minimum, un maximum ou un point selle, on utilise la matrice Hessienne de $f(x)$, qui est la matrice carrée des dérivées secondes :

Proposition 2.6.1 *(Méthode pratique pour $\mathbb{R}^2$). Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction* *de class* C^2 *et* $(x^*, y^*) \in \mathbb{R}^2$ *un point critique de* f *. On pose*

$$r = \frac{\partial^2 f}{\partial x^2}(x^*, y^*), t = \frac{\partial^2 f}{\partial y^2}(x^*, y^*), s = \frac{\partial^2 f}{\partial x \partial y}(x^*, y^*).$$

nous avons :

1. *Si* $\det[\nabla^2 f(x^*, y^*)] = rt - s^2 > 0$ *et* $tr(\nabla^2 f(x^*, y^*)) = r + t > 0$, *alors* (x^*, y^*) *est un point de minimum local strict de* f *sur* $\mathbb{R}^n$.
2. *Si* $\det[\nabla^2 f(x^*, y^*)] = rt - s^2 > 0$ *et* $tr(\nabla^2 f(x^*, y^*)) = r + t < 0$, *alors* (x^*, y^*) *est un point de maximum local strict de* f *sur* $\mathbb{R}^n$.
3. *Si* $\det[\nabla^2 f(x^*, y^*)] = rt - s^2 < 0$, *alors* (x^*, y^*) *est un point selle de* f *sur* $\mathbb{R}^n$.
4. *Si* $\det[\nabla^2 f(x^*, y^*)] = rt - s^2 = 0$,*alors ne peut rien dire (il faut trouver une méthode différente pour préciser la nature de* (x^*, y^*)*).*

Exercise 2.6.1 *Soit* f *la fonction définie sur* $\mathbb{R}^2$ *par*

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

1. *Montrer que la fonction* f *n'est pas coercive.*
2. *Calculer* $\nabla f(x, y)$, *pour tout* $(x, y) \in \mathbb{R}^2$.
3. *Déterminer les quatre points critiques de* f *sur* $\mathbb{R}^2$.
4. *Donner la nature locale de tous les points critiques.*

Solution 2.6.1 *Nous avons :*

$$f(x, x) = 2x^3 - x^2 \underset{x \longrightarrow +\infty}{\longrightarrow} +\infty \quad et \quad f(x, x) = 2x^3 - x^2 \underset{x \longrightarrow +\infty}{\longrightarrow} +\infty$$

Alors, f *n'est pas coercive.*

1. *Le vecteur gradient est :*

$$\nabla f(x,y) = \begin{pmatrix} 2xy + y^2 - y \\ 2xy + x^2 - x \end{pmatrix}$$

2. *Nous avons :*

$$\nabla f(x,y) = 0 \Longleftrightarrow \begin{cases} 2xy + y^2 - y = 0 \\ 2xy + x^2 - x = 0 \end{cases} \Longleftrightarrow \begin{cases} x = 0 \ \text{ou} \ 2x + y - 1 = 0 \\ y = 0 \ \text{ou} \ 2y + x - 1 = 0 \end{cases}$$

Alors, les points critiques sont : $(0,0), (0,1), (1,0)$ *et* $\left(\frac{1}{3}, \frac{1}{3}\right)$.

3. *La matrice hessienne de* f *sur* $\mathbb{R}^2$ *est donnée par :*

$$\nabla^2(x,y) = \begin{bmatrix} 2y & 2x + 2y - 1 \\ 2y + 2x - 1 & 2x \end{bmatrix}$$

   (a) *pour* $(x,y) = (0,0)$ *,on a :*

$$\det\left(\nabla^2 f(0,0)\right) = \left| \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \right| = -1$$

   donc, $(0,0)$ *est un point selle.*

   (b) *pour* $(x,y) = (0,1)$ *,on a :*

$$\det\left(\nabla^2 f(0,1)\right) = \left| \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \right| = -1$$

   donc, $(0,1)$ *est un point selle.*

(c) *pour* $(x,y) = (1,0)$ *,on a :*

$$\det\left(\nabla^2 f\,(1,0)\right) = \left|\begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}\right| = -1$$

donc, $(1,0)$ *est un point selle.*

(d) *pour* $(x,y) = \left(\frac{1}{3},\frac{1}{3}\right)$ *,on a :*

$$\det\left(\nabla^2 f\left(\frac{1}{3},\frac{1}{3}\right)\right) = \left|\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}\right| = \frac{1}{3} > 0 \ \textit{et}\ r = t = \frac{2}{3} > 0.$$

donc, $\left(\frac{1}{3},\frac{1}{3}\right)$ *est un point de minimum local .*

Chapitre 3

Résolution numérique d'un problème d'optimisation quadratique sans contraintes

La résolution numérique d'un problème d'optimisation sans contrainte s'appuie principalement sur la condition du premier ordre. Tous les algorithmes d'optimisation fournissent une solution approximative qui correspond au zéro du vecteur gradient $\nabla f(x)$. Avec des conditions de convergence adèquates, ces zèros constitueront des solutions approchèes à ce problème d'optimisation.

Nous intéressons dans ce chapitre à la conception de méthodes numériques pour la recherche des points $x^* \in \mathbb{R}^n$ qui réalisent le minimum d'une fonction

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

où f est supposée au moins différentiable sur $\mathbb{R}^n$. On parle d'optimisation sans contrainte.

3.1 Principe général des méthodes d'optimisation

Les métaheuristiques sont une classe d'algorithmes d'optimisation qui tentent d'obtenir une valeur approchée de l'optimum global dans le cas de problèmes d'optimisation di cile. Elles ne donnent cependant aucune garantie sur la abilité du résultat. On supposera que x^* existe (éventuellement qu'il est unique) et on se propose de trouver une approximation numérique de x, en construisant une suite

$$\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$$

telle que $x^k \longrightarrow x^*$ pour $k \longrightarrow +\infty$. Le principe est de construire un algorithme itératif de la forme

$$x_{k+1} = x_k + \rho_k d_k. \tag{3.1}$$

d_k est la direction de descente, ρ_k est le pas. Il est, soit xé, éventuellement le même pour toutes les étapes (on parle alors de méthode à pas fixe), soit calculé à chaque étape de faon à minimiser f dans la direction d_k(on parle alors de méthode à pas optimal). Pour s'approcher de la solution optimale du problème (3.1) (dans le cas général, c'est un point en lequel ont lieu peut tre 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.

Définition 3.1.1 *(Algorithme) Un algorithme est défini par une application A de $\mathbb{R}^n$ dans $\mathbb{R}^n$ permettant la génération d'une suite d'éléments de $\mathbb{R}^n$ par la formule :*

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné, } k = 0, \\ x_{k+1} = A(x_k). \end{cases}$$

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

$\{x_k\}_{k\in\mathbb{N}}$.

3.2 Méthode à direction de descentes

La méthode (ou algorithme) du Gradient fait partie d'une classe plus grande de méthodes numériques appelées méthodes de descente.

Définition 3.2.1 *Soient $\nleq$ un ouvert non vide de $\mathbb{R}^n$ et $f:\mathbb{R}^n \longrightarrow \mathbb{R}$ un fonction de plusiurs variables.*

1. *On dit que $d\in\mathbb{R}^n$ avec $d\neq 0$ est une direction de descente en $x\in\mathbb{R}^n$ s'il existe $\rho_0>0$ tel que*

$$f(x+\rho d)\leq f(x), \forall\rho\in[0,\rho_0].$$

2. *On dit que $d\in\mathbb{R}^n$ avec $d\neq 0$ 0 est une direction de descente stricte en $x\in\mathbb{R}^n$ s'il existe $\rho_0>0$ tel que*

$$f(x+\rho d)<f(x), \forall\rho\in[0,\rho_0].$$

3. *Le principe d'une "méthode de descente" consiste à faire les itérations suivantes*

$$x_{k+1}=x_k+\rho_k d_k, \rho_k>0.$$

tout en assurant la propriété :

$$f(x_{k+1})<f(x_k).$$

On peut caractériser les directions de descente en x_k à l'aide du gradient :

Proposition 3.2.1 *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de classe C^1 et $d \in \mathbb{R}^n$ avec $d \neq 0$. alors , on a :*

1. *Si d direction de descente en $x \in \mathbb{R}^n$, alors $\langle d, \nabla f(x) \rangle < 0$.*
2. *Si $\nabla f(x) \neq 0$ alors $d = -\nabla f(x)$ est une direction de descente stricte en $x \in \mathbb{R}^n$.*

Définition 3.2.2 *(Algorithme général des méthodes à direction de descentes)*

1. *On se fixe un point initial $x_0 \in \mathbb{R}^n$.*
2. *À chaque étape (notons k le numéro de l'étape) :*
   - *On choisit une direction de discente d_k*
   - *On choisit un pas ρ_k tel que $f(x_k + \rho_k d_k) < f(x_k)$*
   - *On pose $x_{k+1} = x_k + a_k d_k$.*

Les différentes méthodes de descente correspondent à des choix différents pour les directions de descente et les pas.

3.3 Méthode de gradient

Définition 3.3.1 *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de classe C^1. Une méthode de gradient, est une méthode de descente définie par la suite récurrente suivantes :*

$$\begin{cases} x \in \mathbb{R}^n \\ x_{k+1} = x_k - \rho_k \nabla f(x) \end{cases} \tag{3.2}$$

De plus, le choix du pas ρ_k peut être effectué de la manière suivante :

1. *Soit $\rho_k = \rho > 0$ est fixé , c'est ce que l'on appelle la méthode du gradient à pas fixe ou constant,*

2. *Soit ρ_k est choisi comme le minimum de la fonction $\varphi(\rho) = f(x_k - \rho \nabla f(x))$. . C'est ce que l'on appelle la méthode du gradient à pas optimal.*

(Algorithme du Gradient).

1. Poser $k = 0$.
2. Choisir $x_0, \varepsilon \succ 0$ et $k_{\max}$.
3. Tant que $\|x_{k+1} - x_k\| \geq \varepsilon$ et $k \leq k_{\max}$ faire.
4. Calculer $d_k = -\nabla f(x_k)$.
5. Calculer ρ_k.
6. Poser $x_{k+1} = x_k + \rho_k d_k$.

Proposition 3.3.1 *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de classe C^1 et (x_k) la suite de la méthode du gradient à pas optimal définie par (3.2) Alors, pour tout $k \in \mathbb{N}$ le vecteur $x_{k+1} - x_k$ est orthogonal sur le vecteur $x_{k+2} - x_{k+1}$.*

Preuve. En utilisant la suite de la méthode du gradient , on obtient :

$$\langle x_{k+1} - x_k, x_{k+2} - x_{k+1} \rangle = \rho_k \rho_{k+1} \langle \nabla f(x_k), \nabla f(x_{k+1}) \rangle$$

On introduit la fonction suivante :

$$\varphi(\rho) = f(x_k - \rho \nabla f(x_k)),$$

on a :

$$\acute{\varphi}(\rho) = -\langle \nabla f(x_k - \rho \nabla f(x_k)), \nabla f(x_k) \rangle.$$

et puisque $\blacksquare_K$ est le point de minimum de φ, donc d'après la condition nécessaire de premier ordre on a :

$$0 = \varphi'(\rho_k) = -\langle \nabla f\left(x_k - \rho \nabla f\left(x_k\right)\right), \nabla f\left(x_k\right)\rangle = -\langle \nabla f\left(x_{k+1}\right), \nabla f\left(x_k\right)\rangle.$$

D'où

$$\langle x_{k+1} - x_k, x_{k+2} - x_{k+1}\rangle = 0.$$

$\blacksquare$

Remarque 3.3.1 *1. Grâce Proposition (3.2), le vecteur gradient $\nabla f\left(x_k\right)$ est parallèle au vecteur tangente de plan du niveau $\{x, f\left(x\right) = f\left(x_{k+1}\right)\}$ au point x_{k+1}.*

2. La suite $\left(f\left(x_k\right)\right)$ est strictement décroissante.

Proposition 3.3.2 *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de classe C^1 et (x_k) la suite de la méthode du gradient à pas optimal définie par (3.2)*
Si $\nabla f\left(x_k\right) \neq 0$, alors $f\left(x_{k+1}\right) < f\left(x_k\right)$.

Remarque 3.3.2 *Grâce Proposition (3.3), nous avons les critères d'arrêts suivants :*

1. $\|\nabla f\left(x_k\right)\| \leq \varepsilon$.
2. $|f\left(x_{k+1}\right) - f\left(x_k\right)| \leq \varepsilon$.
3. $\|x_{k+1} - x_k\| \leq \varepsilon$.

3.3.1 Méthode de Gradient à pas fixe

On peut utiliser un pas fixé a priori $\rho > 0$ $\forall k$ on obtient alors la méthode de gradient simple :

$$\begin{cases} d_k = -\nabla f(x_k) \\ x_{k+1} = x_k + \rho_k d_k \end{cases}$$

pour $f \in C^1$ cette méthode converge si ρ est choisi assez petit.

Le choix du pas

- Un pas bien choisi donne des résultats à ceux obtenus par la plus profonde descente.
- Un pas plus petit atténue les zigzags des itérés mais augmente signi cativement le nombre d'itérations.
- Un pas trop grand fait diverger la méthode.

3.3.2 Pas optimal : Cas quadratique

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction quadratique définie par :

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c.$$

où A est une matrice carré d'ordre n, symétrique et définie positive. $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$.D'autre part, on a :

$$\nabla f(x) = Ax - b \text{ et } \nabla^2 f(x) = A.$$

Soit $\varphi(\rho) = f(x_k - \rho g_k)$ avec $g_k = \nabla f(x_k)$.Le pas optimal ρ_k est caractérisé par :

$$\varphi'(\rho_k) = 0$$

donc, on a :

$$\langle \nabla f(x_k - \rho_k g_k), g_k\rangle = 0.$$

donc, on obtient :

$$\rho_k = \frac{\langle g_k, g_k \rangle}{\langle A g_k, g_k \rangle} = \frac{\|g_k\|^2}{g_k^T A g_k} > 0.$$

La méthode du gradient à pas optimal peut donc s'écrire (dans le cas quadratique) :

$$\begin{cases} g_k = A x_k - b. \\ \rho_k = \frac{\|g_k\|^2}{g_k^T A g_k}. \\ x_{k+1} = x_k - \rho_k g_k. \end{cases} \tag{3.3}$$

Exemple 3.3.1 *Résoudre le problème suivant :*

$$\min_{(x,y)\in\mathbb{R}^2} f(x,y) = x^2 + y^2$$

avec point de départ $(x, y) = (1, 1)$*. Nous avons :*

$$\nabla f(x,y) = (2x, 2y)^T, \nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Donc, $g_0 = (2,2)^T, \rho_0 = \|g_0\|^2 / g_0^T \nabla^2 f(1,1) g_0 = 1/2$ *et*

$$(x_1, y_1)^T = (x_0, y_0)^T - \rho_0 g_0 = (1,1)^T - \frac{1}{2}(2,2)^T = (0,0)^T.$$

La méthode converge vers le point de minimum global $(0,0)$ *pour une seule itération.*

3.3.3 Convergence

Pour étudier la convergence de la suite (3,3) nous avons le lemme suivante :

Lemme 3.3.1 *Soit* $f : \mathbb{R} \longrightarrow \mathbb{R}^n$ *une fonction quadratique définie par :*

$$f(x) = \frac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle + c.$$

où A est une matrice carré d'ordre n, symétrique et définie positive. Donc, la fonction f admet un unique point de minimum global x^* tel que $Ax^* = b$. Soit (x_k) la suite de la méthode du gradient à pas optimal définie par (3.2). On pose

$$E(x_k) = f(x_k) - f(x^*) = \frac{1}{2}\langle f(x_k - x^*), x_k - x^*\rangle.$$

Nous avons :

$$E(x_{k+1}) = \left(1 - \frac{\left(g_k^T g_k\right)^2}{(g_k^T A g_k)(g_k^T A^{-1} g_k)}\right) E(x_k) \tag{3.4}$$

Preuve. On pose $y_k = x_k - x^*$. Nous avons :

$$\frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{2\rho_k g_k^T A y_k - \rho_k^2 g_k^T A g_k}{y_k^T A y_k}.$$

Donc, $Ay_k = Ax_k - Ax^* = Ax_k - b = g_k$, nous avons :

$$\frac{E(x_k) - E(x_{k+1})}{E(x_k)} = \frac{\frac{2\left(g_k^T g_k\right)^2}{g_k^T A g_k} - \frac{\left(g_k^T g_k\right)^2}{g_k^T A g_k}}{g_k^T A^{-1} g_k} = \frac{\left(g_k^T g_k\right)^2}{(g_k^T A g_k(g_k^T A^{-1} g_k))}.$$

Pour déterminer le taux de convergence, nous avons le lemme suivant : ■

Lemme 3.3.2 *(Inégalité de Kantorovich). Soit A une matrice carré d'ordre n, symétrique et définie positive.*

Preuve. Pour tout vecteur $x \neq 0$, nous avons :

$$\frac{\left(x^T x\right)^2}{(x^T A x)(x^T A^{-1} x)} \geq \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}.$$

Où $■_1$ et $■_n$ sont respectivement les plus petites et les plus grandes valeurs propres de A.

Soient $0 < \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ les valeurs propres de la matrice A. Avec changement derepère, la matrice A devienne une matrice diagonale telles que les éléments diagonaux sont les valeurs propres du matrice A c'est à dire $A = diag\left(\lambda_1, \lambda_2, ..., \lambda_n\right)$. Donc, nous avons :

$$\frac{\left(x^T x\right)^2}{\left(x^T A x\right)\left(x^T A^{-1} x\right)} = \frac{\left(\sum\limits_{i=1}^{n} x_i^2\right)^2}{\left(\sum\limits_{i=1}^{n} \lambda_i x_i^2\right) - \left(\sum\limits_{i=1}^{n} x_i^2/\lambda_i\right)} = \frac{1/\sum\limits_{i=1}^{n} \lambda_i \alpha_i}{\sum\limits_{i=1}^{n} \alpha_i/\lambda_i}$$

où $\alpha_i = x_i^2/\sum\limits_{i=1}^{n} x_i^2$. D'autre part, on a :

$$\sum_{i=1}^{n} \lambda_i \alpha_i = \beta_1 \lambda_1 + \beta_n \lambda_n et \sum_{i=1}^{n} \alpha_i/\lambda_i = \frac{\beta_1}{\lambda_1} + \frac{\beta_n}{\lambda_n} = \frac{\lambda_1 + \lambda_n - \beta_1 \lambda_1 - \beta_n \lambda_n}{\lambda_1 \lambda_n} avec \beta_1 + \beta_n = 1.$$

On pose : $\beta_1 \lambda_1 + \beta_n \lambda_n = \lambda$, on obtient :

$$\frac{\left(x^T x\right)^2}{\left(x^T A x\right)\left(x^T A^{-1} x\right)} \geq \min_{\lambda_1 \leq \lambda \leq \lambda_n} \frac{\lambda_1 \lambda_n}{\lambda\left(\lambda_1 + \lambda_n - \lambda\right)} = \frac{4\lambda_1 \lambda_n}{\left(\lambda_1 + \lambda_n\right)^2}.$$

Avec lemme (3.1) et lemme (3.2) , nous avons le théorème de convergence suivant :

■

Théor■me 3.3.1 *(Convergence).*

1. *Nous avons :*

$$E\left(x_{k+1}\right) \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 E\left(x_k\right).$$

2. *Pour tout $x_0 \in \mathbb{R}^n$, la méthode du gradient à pas optimal (3.2) pour les fonctions quadratiques converge vers au point de minimum global x^* de f.*

Preuve. En utilisant Lemme (3.1) et Lemme (3.2), on obtient :

$$E\left(x_{k+1}\right) \leq \left(1 - \frac{4\lambda_1 \lambda_n}{\left(\lambda_1 + \lambda_n\right)^2}\right) E\left(x_k\right) = \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 E\left(x_k\right).$$

1. On peut démontrer par récurrence que :

$$E(x_k) \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^{2k} E(x_0)$$

A est définie positive, alors on a :

$$\lambda_1 \|x_k - x^*\|^2 \leq E(x_k) \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^{2k} E(x_0)$$

Par passage à la limite, on obtient

$$x_k \underset{n \to +\infty}{\longrightarrow} x^*.$$

■

3.4 Méthode du gradient conjugué

Définition 3.4.1 *(Vecteurs conjugués). Soit $A \in M_n(\mathbb{R})$ une matrice symétrique définie positive.*

1. *Deux vecteurs u et v de $\mathbb{R}^n / \{0\}$ sont dits A-conjugués si :*

$$\langle Au, v \rangle = \langle u, Av \rangle = 0.$$

2. *Une famille $\{d_1, ..., d_m\}$ de $\mathbb{R}^n / \{0\}$ est dite A-conjugués si :*

$$\langle d_i, Ad_j \rangle = 0.$$

pour tout $i, j = 1, ..., m$ tel que $i \neq j$.

Lemme 3.4.1 *Soit $A \in M_n(\mathbb{R})$ une matrice symétrique définie positive. Si les vec-*

teurs $d_0, d_2, ..., d_m, m \leq n-1$, *sont des vecteurs non nul A-conjugués, alors ils sont linéairement indépendants.*

Exemple 3.4.1 *Soit la matrice suivante :*

$$\begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

On remarque que : A est symétrique et définie positive puisque toutes les mineurs principaux sontstrictement positive c'est à dire :

$$\Delta_1 = 3 > 0, \Delta_2 = \left| \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \right| = 12 > 0, \Delta_3 = \left| \begin{bmatrix} 3 & 0 & 1 \\ 0 & 5 & 2 \\ 1 & 2 & 3 \end{bmatrix} \right| = 20 > 0.$$

Notre but est de construire un ensemble de vecteurs $A-$*conjugués* d_1, d_2 *et* d_3.
Soit $d_0 = (1,0,0)^T, d_1 = (x,y,z)^T$ *et* $d_2 = (\alpha, \beta, \gamma)^T$. *Si* $d_0^T A d_1 = 0$, *alors on a :*

$$d_0^T A d_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 3x + z.$$

Soient $x = 1, y = 0,$*et* $z = -3$. *Alors* $d_1 = (1,0,3)^T$.
Pour trouver le troisième vecteur $d_2 = (\alpha, \beta, \gamma)^T$ *qui serait* $A-$*conjugué avec* d_0 *et* d_1, *nous avons :*

$$\begin{cases} d_0^T A d_2 = 3\alpha + \gamma = 0, \\ d_1^T A d_2 = -6\beta - 8\gamma = 0. \end{cases} \Longleftrightarrow d_2 = (1, 4, -3)^T.$$

Remarque 3.4.1 *La méthode ci-dessus pour trouver des vecteurs A-conjugués n'est* *pas efficace.*

Une procédure systématique pour trouver des vecteurs $A-$conjugués peut être conçue en utilisant l'idée sous-adjacente au processus de Gram-Schmidt de transformer une base donnée de $\mathbb{R}^n$ en une base orthonormée de $\mathbb{R}^n$.

3.4.1 Algorithme de la direction conjugué

Nous présentons maintenant l'algorithme de direction conjuguée pour minimiser la fonction quadratique de n variables

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c.$$

Où $A \in M_n(\mathbb{R}^n)$ est une matrice symétrique définie positive , $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

Définition 3.4.2 *(Algorithme de la direction conjuguée). L'algorithme de la direction conjuguée est définie par un point de départ $x_0 \in \mathbb{R}^n$ et $d_0, d_1, ..., d_{n-1}, n$ vecteur $A-$conjugués et pour tout $k \in \mathbb{N}$, nous avons :*

$$\begin{aligned} g_k &= \nabla f(x_k) = Ax_k - b, \\ \alpha_k &= -\frac{g_k^T d_k}{d_k^T A d_k}, \\ x_{k+1} &= x_k + \alpha_k d_k. \end{aligned}$$

Théorème 3.4.1 *Pour tout point de départ $x_0 \in \mathbb{R}^n$, l'algorithme de la direction conjuguée converge vers 'unique x^*(qui résout $Ax = b$) en n étapes ; c'est-à-dire $x_n = x^*$.*

Preuve. Considérons $x^* - x_0 \in \mathbb{R}^n$. Puisque les vecteurs $d_0, d_1, ..., d_{n-1}1$ sont linéairement indépendants, donc il existe des constantes $\beta_0, ..., \beta_{n-1} \in \mathbb{R}$ tels que

$$x^* - x_0 = \beta_0 d_0 + ... + \beta_{n-1} d_{n-1}. \tag{3.5}$$

En multipliant les deux côtés de l'équation (3.5) par $d_k^T A$ avec $k = 0, 1, ..., n-1$, on obtient :

$$d_k^T A (x^* - x_0) = \beta_k d_k^T A d_k,$$

où les termes $d_k^T A d_i = 0$, pour tout $k \neq i$, par la propriété $A-$conjuguée. Par conséquent,

$$\beta_k = \frac{d_k^T A(x^* - x_0)}{d_k^T A d_k}.$$

Nous pouvons écrire,

$$x_k = x_0 + \alpha_0 d_0 + \alpha_1 d_1 + ...\alpha_{k-1} d_{k-1}.$$

Donc,

$$x_k - x_0 = x_k = x_0 + \alpha_0 d_0 + \alpha_1 d_1 + ...\alpha_{k-1} d_{k-1} \tag{3.6}$$

Ainsi écrire

$$x^* - x_0 = (x^* - x_k) + (x_k - x_0),$$

et en multipliant ci-dessus par $d_k^T A$, nous obtenons :

$$d_k^T A \left(x^k - x_0\right) = d_k^T A (x^* - x_k) + d_k^T A (x_k - x_0) = -d_k^T g_k$$

car $g_k = Ax_k - b$ et $Ax^* = b$. Ainsi,

$$\beta_k = \frac{-d_k^T g_k}{d_k A d_k} = \alpha_k \tag{3.7}$$

Grâce (3.5), (3.6) et (3.7) on obtient $x_n = x^k$. ■

Exemple 3.4.2 *Trouver le point de minimum global de la fonction quadratique suivante :*

$$f(x) = \frac{1}{2}x^T \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} x - x^T \begin{bmatrix} -1 \\ 1 \end{bmatrix}, x \in \mathbb{R}^n.$$

En utilisant la méthode de la direction conjuguée avec le point initial $x_0 = (0,0)^T$ *et les directions conjuguées sont* $d_0 = (1,0)^T$ *et* $d_1 = \left(-\frac{3}{8}, \frac{3}{4}\right)^T$.

Preuve. Nous avons :

$$g_0 = Ax_0 - b = (1, -1)^T,$$

et donc

$$\alpha_0 = -\frac{g_0^T d_0}{d_0^T A d_0} = -\frac{\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}} = -\frac{1}{4}.$$

Ainsi,

$$x_1 = x_0 + \alpha_0 d_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix}.$$

Pour trouver x_2 nous calculons

$$g_1 = Ax_1 - b = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{3}{2} \end{bmatrix},$$

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T A d_1} = -\frac{\begin{bmatrix} 0 & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix}}{\begin{bmatrix} -\frac{3}{8} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix}} = 2.$$

Donc,

$$x_2 = x_1 + \alpha_1 d_1 = \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix} + 2 \begin{bmatrix} -\frac{3}{8} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} -1 \\ \frac{3}{2} \end{bmatrix}.$$

Parce que f est une fonction quadratique de deux variables, $x_2 = x^*$.

Pour une fonction quadratique de n variables, nous avons le lemme suivant : ■

Lemme 3.4.2 *Dans l'algorithme de direction conjuguée, nous avons :*

$$g_{k+1}^T d_i = 0$$

pour tout $k = 0, 1, ..., n-1,$ *et* $i = 1, ..., k.$

Preuve. On note que

$$A\left(x_{k+1} - x_k\right) = Ax_{k+1} - b - \left(Ax_k - b\right) = g_{k+1} - g_k,$$

puisque $g_k = Ax_k - b$.Ainsi,

$$g_{k+1} = g_k + \alpha_k A d_k.$$

Nous prouvons le lemme par récurrence. Pour $k = 0$, nous avons :

$$x_1 = x_0 - \left(\frac{g_0^T d_0}{d_0^T A d_0}\right) d_0.$$

Donc

$$\begin{aligned}
g_1^T d_0 &= (Ax_1 - b)^T d_0, \\
&= x_0^T A d_0 - \left(\frac{g_0^T d_0}{d_0^T A d_0} \right) d_0^T A d_0 - b^T d_0. \\
&= g_0^T d_0 - g_0^T d_0, \\
&= 0.
\end{aligned}$$

On suppose que $g_k^T d_i = 0$ pour tout $i = 0, ..., k-1$.Soit $k \geq 1$ fixé et $i = 0, ..., k-1$,nous avons :

$$g_{k+1}^T d_i = g_k^T d_i + \alpha_k d_k^T A d_j = 0$$

D'autre part, nous avons :

$$\begin{aligned}
g_{k+1}^T d_k &= (Ax_{k+1} - b)^T d_k, \\
&= \left[x_k - \frac{g_k^T d_k}{d_k^T A d_k} d_k \right]^T A d_k - b^T d_k, \\
&= (Ax_k - b)^T d_k - g_k^T d_k, \\
&= 0.
\end{aligned}$$

Par conséquent, par récurrence, pour tous $0 \leq k \leq n-1$ et $0 \leq i \leq k$, $g_{k+1}^T d_k = 0$. ■

3.4.2 Algorithme du gradient conjugué

Soit $A \in M_n(\mathbb{R}^n)$ est une matrice symétrique définie positive , $b \in \mathbb{R}^n$ $c \in \mathbb{R}$.et $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction définie par

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c.$$

L'algorithme de gradient conjugué est résumé ci-dessous.

1. Poser $k = 0$, sélectionner le point initial x_0.

2. Calculer $g_0 = Ax_0 - b$. Si $g_0 = 0$, stop, sinon poser $d_0 = -g_0$.

3. Calculer $\alpha_k = -\frac{g_k^T d_k}{d_k^T A d_k}$.

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

5. Poser $g_{k+1} = Ax_{k+1} - b$. Si $g_{k+1} = 0$, stop.

6. Calculer $\beta_k = \frac{g_{k+1} A d_k}{d_k^T A d_k}$.

7. Calculer $d_{k+1} = -g_{k+1} + \beta_k d_k$.

8. Poser $k := k + 1$; aller à l'étape 3.

Dans l'algorithme de gradient conjugué, les directions $d_0, d_1, ..., d_{n-1}$ sont conjuguées.

Exemple 3.4.3 *On considère la fonction quadratique suivante :*

$$f(x_1, x_2, x_3) = \frac{3}{2}x_1^2 + 2x_2^2 + \frac{3}{2}x_3^3 + x_1 x_3 + 2x_2 x_3 - 3x_1 - x_3$$

Nous trouvons le point de minimum par l'algorithme de gradient conjugué, en utilisant le point de départ $x_0 = (0, 0, 0)^T$.

Nous pouvons représenter f comme :

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

où

$$A = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} \quad et \quad b = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}.$$

Nous avons :

$$g(x) = \nabla f(x) = Ax - b = (3x_1 + x_3 - 3, 4x_2 + 2x_3, x_1 + 2x_2 + 3x_3 - 1)^T.$$

Par conséquent,

$$
\begin{aligned}
g_0 &= (-3, 0, -1)^T\,, \\
d_0 &= -g_0, \\
\alpha_0 &= -\frac{g_0^T d_0}{d_0^T A d_0} = \frac{10}{36} = 0.2778.
\end{aligned}
$$

et

$$x_1 = x_0 + \alpha_0 d_0 = (0.8333, 0, 0.2778)^T\,.$$

L'étape suivante donne :

$$
\begin{aligned}
g_1 &= \nabla f(x_1) = (-0.2222, 0.5556, 0.6667)^T\,, \\
\beta_0 &= -\frac{g_1^T d_0}{d_0^T A d_0} = 0.08025.
\end{aligned}
$$

Nous pouvons maintenant calculer

$$d_1 = -g_1 + \beta_0 d_0 = (0.4630, -0.5556, -0.5864)^T\,.$$

Par conséquent,

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T A d_1} = 0.2187,$$

et

$$x_2 = x_1 + \alpha_1 d_1 = (0.9364, -0.1215, 0.1495)^T\,.$$

Pour effectuer la troisième itération, nous calculons

$$
\begin{aligned}
g_2 &= \nabla f(x_2) = 1 - 0.04673, -0.1869, 0.1402^T, \\
\beta_1 &= \frac{g_2^T d_1}{d_1^T A d_1} = 0?07075, \\
d_2 &= -g_2 + \beta_1 d_1 = (0.07948, 0.1476, -0.1817)^T\,.
\end{aligned}
$$

Par conséquent,

$$\alpha_2 = -\frac{g_2^T d_2}{d_2^T A d_2} = 0.8231,$$

et

$$x_3 = x_2 + \alpha_2 d_2 = (1, 0, 0)^T.$$

On remarque que $g_3 = \nabla f(x_3) = 0$*. Comme prévu, parce que* f *est une fonction quadratique de trois variables. Par conséquent,* $x^* = x_3$.

3.5 Méthode de Newton

Définition 3.5.1 *Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *est une fonction de classe* C^2 *. La méthode de Newton est une méthode itérative ayant la forme suivante :*

$$\begin{cases} x_0 \in \mathbb{R}^n \\ x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k), \textit{pour tout } k \in \mathbb{N}. \end{cases}$$

Où $\nabla^2 f\left(x^k\right)$ *est définie positive pour tout* k.

3.5.1 Algorithme de Newton

1. (a) x_0 est donné et $\varepsilon \succ 0$.

   (b) Calculer la direction de Newton d_k qui est la solution du système linéaire suivant :

   $$\nabla^2 f(x_k)\, d_k = -\nabla f(x_k).$$

   (c) Calculer :

   $$x_{k+1} = x_k + d_k.$$

   (d) Si $\|\nabla f(x_{k+1})\| \leq \varepsilon$, stop, sinon aller à l'étape 1.

Pour la convergence de la méthode de Newton, nous avons le lemme suivant :

Lemme 3.5.1 *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est une fonction de classe C^2 et $x^* \in \mathbb{R}^n$ est un point critique de f Nous avons :*

$$\nabla f(x) = \int_0^1 \nabla^2 f\left[x^* + t(x - x^*)\right](x - x^*)\, dt.$$

Preuve. On pose $g(t) = \nabla f\left[x^* + t(x - x^*)\right]$. Nous avons :

$$g'(t) = \nabla^2 f\left[x^* + t(x - x^*)\right](x - x^*).$$

Alors,

$$g(1) - g(0) = \int_0^1 g'(t)\, dt.$$

Donc,

$$\nabla f(x) = \int_0^1 \nabla^2 f\left[x^* + t(x - x^*)\right](x - x^*)\, dt.$$

■

Théor■me 3.5.1 *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction de classe C^2 . On suppose que :*

1. *Il existe $\alpha > 0$ tel que $h^T \nabla^2 f(x) h \geq \alpha \|h\|^2$, pour tout $x \in \mathbb{R}^n$ et $h \in \mathbb{R}^n$.*
2. *Il existe $L > 0$ tel que$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|$.*

Soient (x_k) la suite de la méthode de Newton et x^*l'unique point de minimum global de f sur $\mathbb{R}^n$. Alors , nous avons :

$$\|x_{k+1} - x^*\| \leq \frac{L}{2\alpha}\|x_k - x^*\|^2.$$

De plus, si $\|x_0 - x^*\| \leq \frac{\alpha}{L}$ alors :

$$\|x_k - x^*\| \leq \frac{2\alpha}{L}\left(\frac{1}{2}\right)^{2k}.$$

Exemple 3.5.1 *(Méthode de Newton)*

On considère le problème suivant :

$$\min_{(x_1,x_2)\in\mathbb{R}^2} f(x_1,x_2) = (x_1-2)^4(x_2-5)^4.$$

1. *On pose* $x^k = \left(x_1^k, x_2^k\right)$*. Avec la méthode Newton, montrer la formule de récurrence suivante :*
$$\forall k \in \mathbb{N}, x^{k+1} = \frac{1}{3}\left(2x^k + \begin{pmatrix}2\\3\end{pmatrix}\right).$$
2. *Soit* $x_0 = (0,0)^T$,
   (a) *Montrer que* $\forall k \in \mathbb{N}$, $x^k = \left(1-\left(\frac{2}{3}\right)^k\right)\binom{2}{3}$.
   (b) *Déduire le point de minimum de* f.

Solution 3.5.1 *Nous avons :*

$$\nabla f(x_1,x_2) = \begin{bmatrix}4(x_1-2)^3\\4(x_2-5)^3\end{bmatrix} \text{ et } \nabla^2 f(x_1,x_2) = \begin{bmatrix}12(x_1-2)^2 & 0\\ 0 & 12(x_2-5)^2\end{bmatrix}$$

1. *Pour tout* $k\in\mathbb{N}$, *nous avons :*

$$\begin{aligned}
x^{k+1} &= x^k - \left[\nabla^2 f\left(x_1^k,x_2^k\right)\right]^{-1}\nabla f\left(x_1^k,x_2^k\right)\\
&= x^k - \begin{bmatrix}\frac{1}{12(x_1-2)^2} & 0\\ 0 & \frac{1}{12(x_2-5)^2}\end{bmatrix}\begin{bmatrix}4\left(x_1^k-2\right)^3\\4\left(x_2^k-5\right)^3\end{bmatrix}\\
&= x^k - \frac{1}{3}x^k + \frac{1}{3}\begin{pmatrix}2\\3\end{pmatrix}\\
&= \frac{1}{3}\left(2x^k + \begin{pmatrix}2\\3\end{pmatrix}\right).
\end{aligned}$$

2. *On pose* $x^0 = (0,0)^T$.

(a) *montrons par récurrence. Pour $k = 0$ est trivial On suppose que :*

$$x^k = \left(1 - \begin{pmatrix} 2 \\ 3 \end{pmatrix}^k\right) \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

et on montre pour $k + 1$. Nous avons :

$$\begin{aligned} x^{k+1} &= \frac{1}{3}\left(2x^k + \begin{pmatrix} 2 \\ 5 \end{pmatrix}\right) \\ &= \frac{1}{3}\left[2\left(1 - \begin{pmatrix} 2 \\ 3 \end{pmatrix}^k\right)\begin{pmatrix} 2 \\ 5 \end{pmatrix} + \begin{pmatrix} 2 \\ 5 \end{pmatrix}\right] \\ &= \left(1 - \begin{pmatrix} 2 \\ 3 \end{pmatrix}^{k+1}\right)\begin{pmatrix} 2 \\ 5 \end{pmatrix} \end{aligned}$$

(b) *le point de minimum de f est*

$$x^* = \lim_{k \longrightarrow +\infty} x^k = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Conclusion

Ce mémoire fournit une vue d'ensemble des problèmes d'optimisation quadratique sans contraintes. En conclusion, faisons le point sur les avantages et inconvénients des méthodes abordées pour l'optimisation sans contrainte.

Méthodes de gradient : Ces méthodes nécessitent le calcul de $\nabla f(x)$. Leur convergence est linéaire, ce qui peut être lent.

Méthode des gradients conjugués : Lorsque f est quadratique (c'est-à-dire $f(x) = \frac{1}{2}x^T Ax - b^T x + c$ avec A est symétrique définie positive), cette méthode est excellente, surtout lorsqu'elle est utilisée avec un préconditionnement pour les grands n. Cependant, dans le cas général, son efficacité diminue si n est trop grand.

Méthode de Newton : La méthode de Newton présente une convergence locale quadratique, ce qui en fait une méthode très rapide. Cependant, elle nécessite le calcul de la matrice hessienne $\nabla^2 f(x)$ et son inverse. Si ce calcul est possible, cette méthode est idéale.

En résumé, chaque méthode présente des avantages et des inconvénients spécifiques. Le choix de la méthode dépend des caractéristiques du problème à résoudre, notamment la taille de n et la facilité de calcul des dérivées. Une compréhension approfondie de ces méthodes permet de choisir l'approche la plus appropriée pour chaque situation d'optimisation quadratique sans contraintes.

Bibliographie

[1] E.K.P. Chong and S.H. zak. An introduction to optimization. WILEY, 2013.

[2] J.B. Hiriart-Urruty. Optimisation et analyse convexe. EDP Sciences, 2009.

[3] G. Allaire, Analyse numérique et optimisation, Edition 2002

[4] M. Bellou , Cours d'optimisation sans contraintes, 2015

[5] G.R.Walsh, Methods of optimization, A wiley- Interscience Publication, 1975