République Algerienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la Recherche scientifique
Université Mouloud Mammeri de Tizi Ouzou
Faculté des Sciences
Département de Mathématiques
Laboratoire LAROMAD

Mémoire de Master

Filière : Mathématiques
Spécialité : Recherche Opérationnelle

Présenté par :

Edjekouane Khadidja et Iamranene Kenza

OPTIMISATION DIFFERENTIABLE ET NON DIFFERENTIABLE : METHODES, ALGORITHMES ET APPLICATIONS

Soutenue le septembre 2022 devant le jury :

Dr.	K. KASDI	UMMTO	Président du jury
M.A.A	S. TICHERFATINE	UMMTO	Examinateur
Pr.	M.OUANES	UMMTO	Encadreur

Année Universitaire : 2021/2022

Dédicaces & Remerciements

Remerciements

Nous remercions "ALLAH" le clément, tout puissant d'avoir guidé Nos pas, Vers les portes du savoir tout en illuminant notre chemin, et nous avoir donné suffisamment de courage et de persévérance pour mener à terme ce travail.

Nous tenons à remercier notre encadreur monsieur Ouanes Mohand, qui nous a encouragé en nous faisant part d'observations constructives et pour ses précieux conseils.

Nos remerciements les membre de Jury d'avoir accepté d'examiner notre travail.

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

En fin, Nous tenons à remercier nos familles et nos amis au département mathématiques et informatique, et ailleurs pour leur disponibilité et leur soutien.

Dédicaces

Je dédie ce modeste travail qui est le fruit de mon études pendant cinq ans, premièrement aux personnes les personnes les plus chers à mon coeur :

Ma très chère mère "Malika" le symbole de la tendresse et de la gentillesse..

Celui qui nous a partagé et nous a appris comment vivre, mon très cher père "Essaid".

Mes frères "Marzouk", "Hamza", "Ghiles", "Mahmoud", ma soeur "Lilia" et ma famille.

Je profite de cette occasion pour dédier ce mémoire à toute ma promo, et tous mes amis.

Enfin, à tout ceux qui me donnent de l'énergie et la confiance pour réaliser toujours les meilleurs résultats.

*** Kenza ***

Dédicaces

J'ai le grand plaisir de dédier ce modeste travail :

A la lumière de mes jours, la source de mes efforts, ma vie et mon âme, la prunelle de mes yeux, à la femme qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie, ma chère maman "Faroudja".

A l'homme, mon précieux offre de dieu qui doit ma vie, ma réussite et tout mon respect mon chère père "Ahcene".

A mes frères "Karim", "Walid", "Aissa", "rachid" et mon âme soeur "HAYAT" qui m'avez toujours soutenu et encouragé durant ces années d'études.

A ma famille, mon mari, mes proches et à ceux qui me donnent de l'amour et de la vivacité.

A mes amis qui m'ont toujours encouragé et à qui je souhaiterais plus de succès.

*** Khadidja ***

Liste des abréviation

C^1 : L'ensemble des fonctions différentiables et les dérivées premières continue.

σ_d : La dérivée directionnelle dans la direction de d.

$\|.\|$: Désigne la norme euclidienne dans $\mathbb{R}$.

$\delta f(x)$: Sous-différentiel de f en x.

$x = (x_1...x_n)^t$: Vecteur d'éléments en colonne

x^* : la solution optimale.

P_u : Opérateur de projection sur l'ensemble U.

γ : Sous-gradient.

Table des matières

Liste des figures

Liste des tableaux

Introduction générale

Nous faisons tous de l'optimisation, dans notre vie quotidienne, nous cherchons à optimiser notre travail, nos espaces de rangement, ou encore le trajet que nous aurons à parcourir pour nous rendre quelque part, etc.

On cherche une meilleure solution aux problèmes qui jalonnent notre existence.

De manière générale, l'optimisation va donc consister à trouver cette meilleure solution. Comme nous le rappelle, l'adage populaire selon lequel : " Les mathématiques permettent de mettre le monde en équation, il peut être tracé un parallèle entre l'optimisation quotidienne et celle la plus technique que l'on retrouve en sciences. En mathématique, la meilleure solution Se recherche au sein d'un domaine initial.

L'optimisation recouvre l'étude des critères d'optimalité pour les différents problèmes, la détermination des méthodes algorithmiques pour résoudre les problèmes.

Le contenu est divisé en quatre chapitres :

Le premier chapitre est un chapitre de généralités qui permet de définir l'ensemble de notations, dont nous aurons besoin dans la suite. Nous présentons les principales définitions, éléments d'analyse convexe et d'algèbre, matrice, différentiabilité...

Le deuxième chapitre est consacré à la présentation des différentes méthodes d'optimisation pour les problèmes différentiables.

Le troisième chapitre est consacré à l'optimisation non différentiable

À la fin, dans le dernier chapitre, on terminera par quelques applications sur les différentes méthodes.

Notions principales de base 1

1.1 Introduction

Dans ce chapitre nous identifions les propriétés de la fonction objectif, on va présenter quelque notions de base essentielles d'analyse et d'algèbre qui seront utiles dans le développement de la théorie et des algorithmes et on va introduire une notion qui joue un rôle très important, celle de la convexité des fonctions et des ensembles.

1.2 Convexité

1.2.1 Ensembles convexes

Une partie $C \subset V$ est convexe si : $\forall(x, y) \in C^2, \forall\lambda \in [0, 1]$:

$$\lambda x + (1 - \lambda)y \in C$$

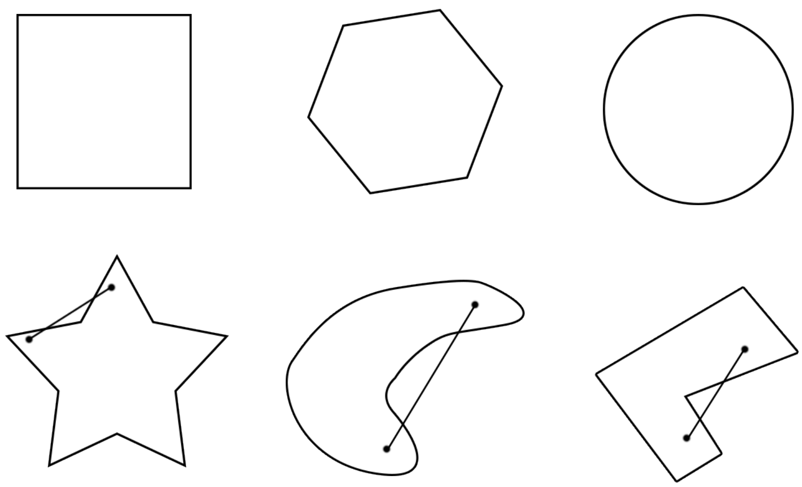


Figure 1.1 – *En haut : Quelques exemples d'ensembles convexes. En bas : Quelques Exemples non convexes*

1.2.2 Fonctions convexes

Définition 1.2.2.1

On appelle une fonction de plusieurs variables à valeurs réelles une fonction :
$f : \mathbb{R}^n \longrightarrow \mathbb{R}$ pour $x = (x_1, .., x_n)^t \in \mathbb{R}^n, f(x) = f(x_1, .., x_n) \in \mathbb{R}$

Définition 1.2.2.2

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est continue au point $x = (x_1, .., x_n)^t \in \mathbb{R}^n$ si et seulement si pour tout $\varepsilon > 0$ il existe $\alpha > 0$ tel que pour tout :

$$y = (y_1, .., y_n)^t \in \mathbb{R}^n$$

$$\|y - x\| \leq \alpha \implies |f(y) - f(x)| \leq \varepsilon$$

Définition 1.2.2.3

Une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est dite convexe sur un ensemble D convexe si elle vérifie : $\forall (x, y) \in D$

$$\forall \lambda \in [0, 1]$$

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

1.3 Différentiabilité

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue.

1.3.1 Dérivée partielle

La dérivée partielle de f en x par rapport a x_i également notée $\partial f(x)/\partial x_i$ si elle existe est

$$f_{x_i} = \frac{\partial f(x)}{\partial x_i} = \lim_{\alpha_i \to 0} \frac{f(x_1, .., x_i + \alpha_i, ...x_n) - f(x_1, x_2, ..x_i, ...x_n)}{\alpha_i}$$

1.3.2 Gradient

Si F est différentiable alors on définie le gradient de F en x par :

$$\nabla F(x) = g(x)$$

$$g(x) : \mathbb{R}^n \longrightarrow \mathbb{R}$$

Si toutes les dérivées partielles existent alors :

$$g(x) = [\frac{\partial F(x)}{\partial x_i}]_{i=1,..,n} = \begin{bmatrix} \frac{\partial F(x)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x)}{\partial x_n} \end{bmatrix}$$

Exemple :

Le gradient de la fonction $F(x_1, x_2, x_3) = 3x_1^2 + 4x_3x_2 + x_2^2$

$$\nabla F(x) = (6x_1; 4x_3 + 2x_2; 4x_2)$$

1.3.3 Dérivée directionnelle

Définition 1.3.1 :

La dérivée directionnelle de f en x dans la direction de $d \in \mathbb{R}^n$ est :

$$f_d(x) = \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha}$$

1.3.4 Fonction différentiable :

Définition 1.3.2 :

Soit $a \in U$ ouver de E. On dit que f est différentable en a s'il existe un voisinage V de O dans E, une application linéaire $L : E \longrightarrow F$ est une application $\varepsilon : V \longrightarrow F$ vérifiant $\lim_{h \to 0} \varepsilon(h) = 0$ tels que, pour tout $h \in V$:

$$f(a + h) = f(a) + L(h) + \|h\|\varepsilon(h)$$

L'application L, si elle existe est unique et s'appelle la différentielle de f en a (Ou encore application linéaire tangente), on la note df_a.

1.3.5 Matrice jacobienne :

Définition 1.3.5.1.[1]

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ telle que $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ est différentiable pour tout $i = 1..,m$. Dans ce cas f est différentiable et la fonction $Jf(x) : \mathbb{R}^{n\times m}$ est appelée matrice gradient et est définie par :

$$J_{f(x)} = (\nabla f_1(x)...\nabla f_m(x))$$

$$J_{f(x)} = \begin{pmatrix} \frac{\delta f_1(x)}{\delta x_1} & \frac{\delta f_2(x)}{\delta x_2} \cdots & \frac{\delta f_m}{\delta x_1} \\ \vdots & \vdots & \vdots \\ \frac{\delta f_1(x)}{\delta x_n} & \frac{\delta f_2(x)}{\delta x_n} \cdots & \frac{\delta f_m}{\delta x_n} \end{pmatrix}$$

1.3.6 Matrice hessienne :

La hessienne de f en x est la matrice carrée, notée $H(x)$:

$$H(x) = \nabla^2 f(x)$$

$$f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}$$

$$H(x) = \left[\frac{\delta^2 f(x)}{\delta x_i \delta x_j}\right]_{i=1..n_j=1..n}$$

$$= \begin{bmatrix} \frac{\delta^2 f(x)}{\delta x_1^2} & \cdots & \frac{\delta^2 f(x)}{\delta x_1 \delta x_n} \\ \vdots & \vdots & \vdots \\ \frac{\delta^2 f(x)}{\delta x_n \delta x_1} & \cdots & \frac{\delta^2 f(x)}{\delta x_n^2} \end{bmatrix}$$

1.3.7 Fonction de classe C^1 :

Soit f une fonction définie sur un ouvert U de $\mathbb{R}^n$. On dit que f est de classe C^1 si toutes les dérivées partielles de f existent et sont continues sur U.

1.4 Les minimums locaux et globaux :

Définition 1.4.1 (Minimum global)

Soit $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction
x^* est le minimum global de F si $F(x^*) \leq F(x) \quad \forall x \in U$ On dit que x^* est un minimiseur global de F.

Définition 1.4.2 (Minimum local)

Soit $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, x^* est le minumum local de F si
$\exists \varepsilon > 0, \forall x \in]x^* - \varepsilon, x^* + \varepsilon[, F(x^*) \leq F(x) \quad x^*$ est un minimiseur local de F.

1.5 Existance et unicité de la solution :

Théorème 1.5.1 :(Weirstrass)[2]

Si D l'ensemble des solutions est compact non vide de $\mathbb{R}^n$ et si f est continue sur D alors la fonction f admet au moins une solution optimale globale $x^* \in D$.

1.6 Conditions d'optimalité (sans contraintes)

Soit le programme non linéaire et sans contraintes [2]

$$(P_1) \begin{cases} Min f(x) \\ x \in D \subseteq \mathbb{R}^n \end{cases}$$

Oû la fonction $f(x)$ est au moins deux fois continument différentiable.

Définition 1.6.1 :

Soit $x \in D$ on dit que $y \in D$ est une direction admissible s'il existe $\alpha > 0$ tel que $x + ty \in D, \forall t \in [0, \alpha]$
Soit x^* un mimumum local de $f(x)$, on a slors nécessairement pour tout $t > 0$ assez petit
$f(x^* + ty) - f(x^*) \geq 0, \forall y$ admissible ceci implique :

$$\lim_{t \longrightarrow 0} \frac{f(x^* + ty) - f(x^*)}{t} = \nabla f(x^*) y \geq 0$$

En particulier, la direction $y = -\nabla f(x^*)$ est admissible et on déduit

$$-\|\nabla f(x^*)\|^2 \geq 0 \implies \|\nabla f(x^*)\|^2 \leq 0 \implies \nabla f(x^*) = 0$$

Théorème 1.6.1 : (Condition nécessaire du 1er ordre)[3]

Soit x^* minimum local pour (P_1), on a alors $\nabla f(x^*) = 0$.
Un point x satisfaisant cette condition est appelé un point stationnaire.
Au deuxième ordre, on obtient :

$$f(x^*) \leq f(x^* + ty) = f(x^*) + t\nabla f(x^*)^t y + \frac{1}{2}t^2 y^t Hf(x^*)y + 0(t^2)$$

Quand t tend vers 0 .
Comme x^* est un minimum local on a :

$$\lim_{t \longrightarrow 0} \frac{f(x^* + ty) - f(x^*)}{t^2} = \frac{1}{2}y^t Hf(x^*)y + \lim_{t \longrightarrow 0} \frac{0(t^2)}{t^2} = \frac{1}{2}y^t Hf(x^*)y \geq 0$$

Théorème 1.6.2 : (Condition nécessaire du 2nd ordre)[2]

Soit x^* un minimum local pour (P_1) on a alors :

$$\begin{cases} \nabla f(x^*) = 0 \\ y^t Hf(x^*)y \geq 0 \quad \forall y \in D \end{cases}$$

La matrice hessienne au point x^* est donc semi-définie positive. Ces conditions ne sont pas suffisantes.

Théorème 1.6.3 : (Condition suffisante du 2nd ordre)[2]

Soit x^* un point tel que :

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

$$y^t Hf(x^*)y > 0 \quad \forall y \in D$$

Alors x^* est un minimum local (Strict) .

Exemple : (Condition d'optimalité sans contrainte)

$$f(x_1, x_2) = x_1^2 - x_2^3$$

Gradient : $\nabla f(x_1, x_2) = \begin{pmatrix} 2x_1 \\ -3x_2^2 \end{pmatrix}$

Hessienne : $\nabla^2 f(x_1, x_2) = \begin{pmatrix} 2 & 0 \\ 0 & -6x_2 \end{pmatrix}$

Point stationnaire $\nabla f(x_1, x_2) = 0$

$$\begin{cases} 2x_1 = 0 \\ -3x_2^2 = 0 \end{cases} \Longrightarrow \begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases} \Longrightarrow x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Valeurs propres du hessienne

$$\nabla^2 f(x^*) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Longrightarrow \begin{cases} \alpha_1 = 2 \\ \alpha_2 = 0 \end{cases}$$

$\nabla^2 f(x)$ est semi-défini positif.

- x^* vérifie les conditions de minumum local.
- x^* ne vérifie pas les conditions suffisante de minumum local.

$\Longrightarrow$ x^* n'est pas un minimum local.

1.7 Conditions d'optimalités (Avec contraintes)

Soit le programme mathématique (PM)

$$\begin{cases} Min f(x) \\ g_i(x) \leq 0 \quad i = 1..m \\ h_j(x) = 0 \quad j = 1..p \\ x \in \mathbb{R}^n \end{cases}$$

Ou les fonctions f, g_i, h_j sont au moins deux fois continûment différentiables. '

Définition 1.7.1

Le lagrangien du programme mathématique (PM) est défini par :

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

'

Théorème 1.7.1 (Karush-Kuhn-Tucker)[2]

Soit x^* un minimum local régulier de (PM). Alors il existe les multiplicateurs $\lambda^* \in \mathbb{R}^m_+$ et $\mu^* \in \mathbb{R}^p$ tels que

$$\nabla\mathbb{L}(x^*, \lambda^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0$$

Avec
$(\nabla\mathbb{L}(x^*, \lambda^*, \mu^*) = 0)$ (Condition d'optimalité)
$\lambda_i^* \nabla g_i(x^*) = 0 \quad (i = 1..m)$ (Condition de complémentarité)
$\lambda_i^* \geq 0 \quad (j = 1..p)$ (KkT)
$g_i(x^*) \leq 0 \quad (i = 1..m)$
$h_i(x^*) = 0 \quad (j = 1..p)$
Sont les conditions nécessaires du premier ordre.

Remarque :

1. Si les contraintes ne sont pas qualifiées en x^*, les conditions de KKT ne s'appliquent pas (x^* peut être optimale sans vérifier ces conditions).
2. Si (PM) est convexe, les conditions de KKT sont à la fois nécessaires et suffisantes pour que x^* soit minimum global.

Soit $I(x^*) = \{i | i = 1..m, g_i(x^*) = 0\}$ l'ensemble des indices des contraintes d'inégalités actives au point x^*.

Définition 1.7.2 :

On définit un espace tangent $T(x^*)$ au point x^* d'un problème (PM) par :

$$\mathbb{T}(x^*) = \{y \in \mathbb{R}^n : \nabla g_i(x^*)^t y \leq 0, i \in I(x^*) \quad et \quad \nabla h_j(x^*)^t y = 0 \quad j = 1..p\}$$

Théorème 1.7.2 (Conditions nécessaires du 2nd ordre)[2]

Soit x^* un minimum local régulier de (PM) . Alors, il existe des multiplicateurs $\lambda^* \in \mathbb{R}^m_+$ et $\mu^* \in \mathbb{R}^p$ tel que :
Les conditions de KKT sont satisfaites .
$y^t H\mathbb{L}(x^*, \lambda^*, \mu^*)y \geq 0, \forall y \in \mathbb{T}(x^*)$(Hessienne semi-définie positive).

Théorème 1.7.3 : (Conditions suffisante du 2nd ordre)[2]

S'il existe des multiplicateurs $\lambda^* \in \mathbb{R}^m_+$ et $\mu^* \in \mathbb{R}^p$ tel que :
Les conditions de KKT sont satisfaites .
$y^t H\mathbb{L}(x^*, \lambda^*, \mu^*)y > 0 \quad \forall y \in T(x^*)$.
$T(x^*)$ est toujours un espace tangent du problème (PM) au point alors x^* qui est un minimum local strict de (PM).

Exemple : (Condition d'optimalité avec contrainte)

Soit P le problème de minimisation avec les contraintes suivantes :

$$\min_{x \in U} \quad (x_1 - 1)^2 + x_2 - 2$$

$$U = \{x \in \mathbb{R}^2 / x_2 - x_1 - 1 = 0, x_1 + x_2 - 4 \leq 0\}$$

1 Vérifier les conditions de 1er ordre et donner la solution optimale.
On a le langrangien usuel associé à P est :

$$\mathbb{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x)$$

$$\mathbb{L}(x, \lambda, \mu) = (x_1 - 1)^2 + x_2 - 2 + \lambda_1(x_2 - x_1 - 1) + \mu_1(x_1 + x_2 - 4)$$

$$\nabla_x \quad \mathbb{L}(x, \lambda, \mu) = \begin{pmatrix} 2(x_1 - 1) \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\nabla_x \quad \mathbb{L}(x, \lambda, \mu) = \begin{pmatrix} 2(x_1 - 1) - \lambda_1 + \mu_1 \\ 1 + \lambda_1 + \mu_1 \end{pmatrix}$$

$$\nabla^2_{xx} \quad \mathbb{L}(x, \lambda, \mu) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = H\mathbb{L}$$

Les KKT sont :

Cas 1 :

Les contraintes $x_1 + x_2 - 4 < 0 \implies \textit{inactive} \quad \mu_1^* = 0$

$$x_2 - x_1 - 1 = 0 \implies \textit{active}$$

$$\begin{cases} 2(x_1^* - 1) - \lambda_1^* = 0 \quad .. \quad 1 \\ 1 + \lambda_1^* = 0 \quad .. \quad 2 \\ x_2^* - x_1^* - 1 = 0 \quad .. \quad 3 \end{cases}$$

$$\begin{cases} 2x_1^* - 2 - \lambda_1^* = 0 \quad .. \quad 1 \\ 1 + \lambda_1^* = 0 \quad .. \quad 2 \\ x_2^* - x_1^* - 1 = 0 \quad .. \quad 3 \end{cases} \Longrightarrow \begin{cases} 2x_1^* = 2 + 1 \quad .. \quad 1 \\ \lambda_1^* = -1 \quad .. \quad 2 \\ x_2^* = x_1^* + 1 \quad .. \quad 3 \end{cases}$$

$1 \Longrightarrow 2x_1^* = 1 \Longrightarrow x_1^* = \frac{1}{2}$
$3 \Longrightarrow 2x_2^* = \frac{1}{2} + 1 = \frac{3}{2}$

$$x_1^* = \frac{1}{2} \quad ; \quad x_2^* = \frac{3}{2}; \lambda_1^* = -1 \quad ; \quad \mu_1^* = 0$$

Cas 2 :

Les contraintes $x_1 + x_2 - 4 = 0$ active $\mu_1^* \geq 0$

$$x_2 - x_1 - 1 = 0$$

On obtient le système suivant :

$$\begin{cases} 2x_1^* - 2 - \lambda_1^* + \mu_1^* = 0 \quad .. \quad 1 \\ 1 + \lambda_1^* + \mu_1^* = 0 \quad .. \quad 2 \\ x_2^* - x_1^* - 1 = 0 \quad .. \quad 3 \\ x_1^* + x_2^* - 4 = 0 \quad .. \quad 4 \end{cases}$$

$1 + 2 \Longrightarrow 2x_1^* - 2 + 2\mu_1^* + 1 = 0 \Longrightarrow 2x_1^* - 1 + 2\mu_1^* = 0 \Longrightarrow 2\mu_1^* = 1 - 2x_1^* \Longrightarrow$

$$\mu_1^* = \frac{1}{2} - x_1^* \quad .. \quad 5$$

$$3 + 4 \Longrightarrow 2x_2^* - 5 = 0 \Longrightarrow x_2^* = \frac{5}{2}$$

$$4 \Longrightarrow x_1^* = -x_2^* + 4 \Longrightarrow x_1^* = \frac{-5}{2} + 4 \Longrightarrow x_1^* = \frac{3}{2}$$

$$5 \Longrightarrow \mu_1^* = \frac{1}{2} - \frac{3}{2}$$

$$\Longrightarrow \mu_1^* = -1$$

$$2 \Longrightarrow \lambda_1^* = -\mu_1^* - 1 \Longrightarrow \lambda_1^* = 1 - 1 = 0$$

Donc $x_1^* = \frac{3}{2}$ $x_2^* = \frac{5}{2}$; $\lambda_1^* = 0, \mu_1^* = -1 \Longrightarrow$ (Impossible car $\mu_1^* \geq 0$)
On a : $f(x) = (x_1 + 1)^2 + x_2 - 2$
$Hf = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ $D.P$ donc f est convexe.
$h(x) = x_2 - x_1 - 1$ est convexe (Car $h(x)$ est linéaire)
$g(x) = x_1 + x_2 - 4$ est convexe (Car $g(x)$ est linéaire)
Donc KKT sont suffisantes pour que x^* soit un minimum de f sur U donc $x^* = \begin{bmatrix} \frac{1}{2} \\ \frac{3}{2} \end{bmatrix}$ est solution optimale de P et la valeur optimale de P est :

$$f(x^*) = f \begin{bmatrix} \frac{1}{2} \\ \frac{3}{2} \end{bmatrix} = [\tfrac{1}{2} - 1]^2 + [\tfrac{3}{2}] - 2 = -\frac{1}{4}$$

QUELQUES MÉTHODES ET ALGORITHMES D'OPTIMISATION DIFFÉRENTIABLE

2

INTRODUCTION

Dans ce chapitre nous allons présenter quelques méthodes d'optimisation ainsi que leur algorithmes les plus utilisés dans la pratique. Parmi les méthodes on trouve la méthode de gradient et de Newton, qui sert à résoudre les problèmes d'optimisation différentable.

2.1 MÉTHODE D'OPTIMISATION DIFFÉRENTIABLE

2.1.1 Méthode de Gradient

Définition 2.2.1.1 (Algorithmes)

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

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

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

2.2 MÉTHODE DE GRADIENT :

la méthode (Ou algorithme) du gradient fait partie de la classe la plus grande des méthodes appelées méthodes de descente. On veut minimiser une fonction f Pour cela on se donne un point de dépat arbitraire x_0 pour construire l'itération suivante x_1 il faut penser qu'on veut se rapprocher du minimum de f, on veut donc que $f(x_1) < f(x_0)$. On cherche alors x_1 sous la forme $x_1 = x_0 + \sigma_0 d_0$ où $d_0 \in \mathbb{R}^n$ et $\sigma_0 > 0$. En pratique donc, on cherche d_0 et σ_0 pour que $f(x_0 + \sigma_0 d_0) < f(x_0)$. Quand d_0 existe on dit que c'est une direction de descente et σ_0 est le pas de descente. La direction et le pas de descente peuvent être fixés ou changer d'itération. Le schéma général d'une méthode de descente est le suivant :

$$\begin{cases} x_0 \quad \text{donné dans} \quad \mathbb{R}^n \\ x_{k+1} = x_k + \sigma_k d_k, \quad \text{avec} \quad d_k \in \mathbb{R}^n et \quad \sigma_k > 0 \end{cases}$$

Où f_k et d_k choisis de telle sorte que $f(x_k + \sigma_k d_k) \leq f(x_k)$ une idée pour trouver la direction de descente est de faite un développement de Taylor de f au premier ordre au voisinage de x_{k+1} est donné par :

$$f(x_{k+1}) = f(x_k + \sigma_k d_k) = f(x_k) + \nabla f(x_k) d_k + 0(\sigma_k d_k)$$

Or puisque l'on désire avoir : $f(x_{k+1}) \leq f(x_k)$ une solution évidente consiste à prendre : $d_k = -\nabla f(x_k)$ La méthode ainsi obtenue s'appelle méthode de gradient. Donc le schéma général de la méthode est :

$$x_0 \text{ donné dans } \mathbb{R}^n$$

$$x_{k+1} = x_k - \sigma_k \nabla f(x_k) \text{ avec } \sigma_k > 0$$

Algorithme du Gradient [10]

1. Initialisation
k = 0 : choix de x_0 et de $\sigma_0 > 0$
2. Itération k
Le pas σ_k est choisi constante ou variable
$x_{k+1} = x_k - \sigma_k \nabla f(x_k)$
3. Critère d'arrêt
Si $\|x_{k+1} - x_k\| < \varepsilon$ STOP
Si non, on pose $k = k + 1$ et on retourne à 2.

Théorème 2.1 [4]

Soit f une fonction de classe C^1 de $\mathbb{R}^n$ dans $\mathbb{R}$, x^* est un minimum de f supposons que :

1. f est α-elliptique ; c'est à dire :

$$\exists \quad \alpha > 0, \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2$$

2. L'application ∇f est Lipchitzienne, c'est a dire :

$$\exists L > 0, \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

S'il existe deux réels a et b tels que $0 < a < \sigma_k < b < \frac{2\alpha^2}{L^2}$ pour tout $k \geq 0$ alors la méthode du gradient converge pour tout choix de x_0 de façon géométrique c'est à dire :

$$\exists \quad \beta \in]0, 1[, \|x_k - x^*\| \leq \beta^k \|x_0 - x^*\|$$

Méthode du gradient à pas fixe :

La méthode du gradient à pas fixe est définie par x_0 donné dans $\mathbb{R}$

$$x_{k+1} = x_k - \sigma \nabla f(x_k)$$

c'est une méthode de descente utilisant un pas fixe $\sigma_k = \sigma$ (Indépendant de k) la convergence de la méthode est assurée sous les hypothèses du théorème précédent en particulier pour un choix de pas σ vérifiant :

$$0 < \sigma < \frac{2\alpha}{L^2}$$

C'est la méthode la plus simple à utiliser.

Exemple 2.1

Soit la fonction f définie de $\mathbb{R}$ dans $\mathbb{R}$ par :

$$f(x) = 2x^2 + 3x - 1$$

L'algorithme de descente est comme suit :

$$\begin{cases} x_0 \text{ donné dans } \mathbb{R} \\ x_{k+1} = x_k + \sigma d_k \end{cases}$$

f est différentiable donc : $d_k = -\nabla f(x) = -f'(x) = -(4x+3)$ D'où :

$$\begin{cases} x_0 \text{ donné dans } \mathbb{R} \\ x_{k+1} = x_k - \sigma f'(x_k) \end{cases}$$

Calculons donc le pas σ on a :

$$|\nabla f(x) - \nabla f(y)| = |(4x+3) - (4y+3)| = |4(x-y)| \leq 4|x-y|$$

Donc ∇f est Lipchitzienne de constante L = 4 et :

$$\langle \nabla f(x) - \nabla f(y), x-y \rangle = \langle 4(x-y).(x-y) \rangle = 4(x-y)^2 \geq 4|x-y|^2$$

Donc f est α-elliptique avec $\alpha = 4$

Comme $\sigma \in]0, \frac{2\alpha}{L^2}[$ donc $\sigma \in]0, \frac{1}{2}[$ si on pose $x_0 = -0.75, \sigma = 0,1$

Itération 1 :

$$x_1 = x_0 - \sigma f'(x_0) = -0,75 - 0,1(4 \times (-0.75) + 3) = -0,75$$

D'où $x_1 = -0,75$

Critère d'arrêt : $|x_1 - x_0| = 0 < \varepsilon$ STOP.

Méthode du gradient à pas optimal [10]

ici on choisit à chaque étape σ_k de façon que :

$$f(x_k - \sigma_k \nabla f(x_k)) = \min_{\sigma \in \mathbb{R}} f(x_k - \sigma \nabla f(x_k))$$

Théorème 2.2

Si f est α-elliptique sur $\mathbb{R}^n$ si ∇f est uniformément lipchitzien de constante de lipchitz, l'algorithme de gradient à pas optimal est bien défini et converge vers la solution optimale.

Remarque 2.1

Les directions de descente sont orthogonales, c'est à dire :

$$\langle \nabla f(x_k), \nabla f(x_{k+1}) \rangle = 0$$

Cas d'une fonction quadratique :

Ici la fonctionnelle f est quadratique sur $\mathbb{R}$

$$f(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle$$

Où la matrice A est symétrique définie positive. La solution x du problème de minimisation vérifier $Ax = b$.
On appellera résidu à l'étape k la quantité $r_k = Ax_k - b$.
On prend ici une direction de descente d_k quelconque dans $\mathbb{R}^n$, non orthogonale de r_k. A chaque étape, le pas optimal σ_k est donné par :

$$\sigma_k = \frac{\langle r_k, d_k\rangle}{\langle Ad_k, d_k\rangle}$$

Et l'ont a $\langle r_{k+1}, d_k\rangle = 0$
Notons $E(v) = \frac{1}{2}\langle A(v-u), v-u\rangle$ on a alors

$$E(x_{k+1}) = (1-\gamma_k)E(x_k)$$

Avec

$$\gamma_k = \frac{1}{2}\frac{\langle r_k, d_k\rangle^2}{\langle Ad_k, d_k\rangle\langle A^{-1}r_k, r_k\rangle}$$

Puisque la quantité γ_k est par construction telle que $0 \leq \gamma_k \leq 1$ on a l'estimation suivante :

$$\langle \frac{r_k}{\|r_k\|}, \frac{d_k}{\|d_k\|}\rangle^2 \geq \mu > 0$$

Alors $\gamma_k \geq \gamma = \frac{\mu}{K(A)}$ (Où K(A) est le conditionnement de A. C'est à dire le rapport de la plus grande et la plus petite valeur propre) et donc

$$E(x_{k+1}) = (1-\gamma)E(x_k)$$

On dit que la méthode converge linéairement.

2.2.1 Méthode de Newton

L'algorithme de Newton en optimisation est une application directe de l'algorithme de Newton pour la résolution d'équation du type $f(x) = 0$. En optimisation sans contrainte l'algorithme de Newton cherche les solutions de l'équation $\nabla f(x^*) = 0$. Ceci est une équation non linéaire (Ou plutôt un système d'équations non linéaires) dans $\mathbb{R}$ et nous allons utiliser la méthode de Newton pour la résoudre. Toute fois nous n'obtiendrons que les points critiques de f : Il faudra vérifier que se sont bien les minima. ici $f = \nabla f$ est bien une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^n$. La dérivée de f n'est autre que la matrice hessienne de f : $H(x) = \nabla^2 f(x)$.

Algorithme de Newton [11]

1. Initialisation
$k = 0$: Choix de x_0
2. Itération k
$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$
3. Critère d'arrêt
Si $|x_{k+1} - x_k| < \varepsilon$ STOP.
Sinon, on pose k = k + 1 et on retourne à 2.

Remarque 2.2 :

Si le point x_1 est assez proche de la solution optimale locale x^* telle que $H(x^*)$ soit définie positive alors l'algorithme de Newton converge de façon quadratique vers la solution x^* c'est à dire :

$$\|x_{k+1} - x^*\| \leq \gamma \|x_k - x^*\|^2 \quad \gamma \geq 0$$

Exemple 2.2

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction définie par :

$$f(x_1, x_2) = x_1^2 + x_2^2 - x_1 - 2x_2$$

Soit $x^0 = \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ on trouve les itérations 1 et 2 par la méthode de Newton d'optimisation
x^0donné

$$x^{k+1} = x^k - (\nabla^2 f(x^*))^{-1} \nabla f(x^*)$$

On a

$$\nabla f(x_1^0, x_2^0) = \begin{pmatrix} 2x_1^0 - 1 \\ 2x_2^0 - 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\nabla^2 f(x_1^0, x_2^0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

On pose $d^k = -(\nabla^2 f(x_1^k, x_2^k))^{-1} \nabla f(x_1^k, x_2^k)$ alors :

$$\begin{cases} x_0 \text{ donné dans } \mathbb{R} \\ x^{k+1} = x^k + d^k \end{cases}$$

$\nabla^2 f(x_1^k, x_2^k) d^k = -\nabla f(x_1^k, x_2^k)$ donc

$$\nabla^2 f(x_1^0, x_2^0) d^0 = -\nabla f(x_1^0, x_2^0)$$

Alors

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} d_1^0 \\ d_2^0 \end{pmatrix} = \begin{pmatrix} -3 \\ -4 \end{pmatrix} \implies 2d_1^0 = -3 \quad 2d_2^0 = -4$$

D'où $(d_1^0 = -\frac{3}{2}, d_2^0 = -2) \implies d^0 = \begin{pmatrix} -\frac{3}{2} \\ -2 \end{pmatrix}$ ce qui donne :

$$x^1 = x^0 + d^0 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} -\frac{3}{2} \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$$

$$x^2 = x^1 + d^1 = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$$

Ainsi le point $\begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$ est la solution exacte de la fonction f.

2.2.2 Méthode du Gradient conjugué

Méthode du Gradient conjugué, cas linéaire [11]

La présentation de l'algorithme précédents montre qu'en réalité on ne calcule pas les extrema d'une fonction mais les points stationnaires (Ou points critiques) où F est quadratique nous avons vu que cela revient à résoudre un système linéaire : $Ax = b$. Nous allons donc présenter ici une méthode de résolution d'un système linéaire issue de la théorie de l'optimisation et convergente dans le cas des matrices symétriques définies positives. Dans ce cas l'application qui au couple (x, y) associe le produit scalaire sur $\mathbb{R}$ qu'on note $(x, y)_A$. La méthode qui suit est une méthode de descente inspirée de la méthode du gradient. La direction de descente d_k n'est plus égale au gradient $g_k = \nabla f(x_k) = Ax_k - b$: Le gradient g_k est "Corrigé "de façon que toutes les directions obtenues soient orthogonales (ou conjuguées) pour le produit scalaire $(.,.)_A$.Plus précisément on pose :

$$d_k = g_k + \alpha_k d_{k-1} \quad \text{Tel que} \quad (d_k, d_{k-1})_A = 0$$

Algorithme du Gradient conjugué [10]

1. Initialisation
$k = 0$ choix de $x_0 \in \mathbb{R}^n$ et calcule de $g_0 = Ax_0 - b$
2. Itération k
a) Si g_k = 0 STOP
b) Sinon :

- $d_k = \begin{cases} g_0 & \text{si } k = 0 \\ g_k + \alpha_k d_{k-1} & \text{si } k \geq 0 \end{cases}$ Avec $\alpha_k = \dfrac{\left\|\nabla f((a,b)^k)\right\|^2}{\left\|\nabla f((a,b)^{k+1})\right\|^2}$
- $\sigma_k = \frac{\langle g_k, d_k \rangle}{\langle Ad_k, d_k \rangle}$
- $x_{k+1} = x_k - \sigma_k d_k$
- $g_{k+1} = Ax_{k+1} - b$

c) $k = k + 1$

Theoreme 2.3 [5]

La méthode du gradient conjugué trouve le minimum d'une fonction quadratique F, où A est symétrique definie positive, en au plus n itérations ou n est l'ordre de A

2.2.3 Méthode de quasi Newton

1.2.4.1 Méthode de Broyden, Fletcher, Goldfarb et Shanno (BFGS) [1]

La formule de mise à jour de Broyden, Fletcher, Goldfarb et Shanno est une formule de correction de rang deux, qui s'obtient à partir de la formule DFP en intervertissant les rôles de S_k et y_k. La formule obtenue permet de mettre à jour B_k de hessienne lui même et non de son inverse.Exigera que posée dans les mêmes propriétés, à savoir B_{k+1} reste définie positive si B_k l'est et bien sur l'équation d'approximation de quasi Newton doit être vérifiée, c'est à dire :

$$B_{k+1} S_k = Y_k$$

On obtient donc

$$B_{k+1} = B_k + \frac{Y_k Y_k^T}{Y_k^T S_k} - \frac{B_k S_k S_k^T B_k}{S_k^T B_k S_k}$$

Algorithme de quasi Newton

initialisation :
k =0
itérations :
.Calculer $d_k = -B_k \nabla f(x_k)$ *avec* $B_0 = I$
. Déterminer α_k en appliquant la recherche linéaire avec $\alpha_0 = 1$
. $x_{k+1} = x_k + \alpha_k d_k$
.k=k+1
Mettre à jour

$$B_{k+1} = (I - \frac{S_k Y_k^T}{S_k^T y_k}) B_k (I - \frac{S_k Y_k^T}{S_k^T y_k}) + \frac{S_k S_k^T}{S_k y_k}$$

Avec

$$S_k = x_{k+1} - x_k \quad et$$

$$Y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

critère d'arrêt
Si $|\nabla f(x_k)| \leq \epsilon$ alors $x^* = x_k$

3 Quelques méthodes d'optimisation non différentiables

3.1 Introduction

Dans ce chapitre nous nous intéressons aux méthodes d'optimisation adaptées aux cas où la fonction f à minimiser est convexe mais non différentiable .
De manière générale les méthodes pour l'optimisation non différentiables sont : La méthode de sous gradient, la méthode de sous gradient projeté, et les méthodes de faisceaux.

3.2 Sous gradient et sous différentiel

Définition 3.2.1 [4]

Soit une fonction convexe $\phi : U \longrightarrow \mathbb{R}$ ou $U \in \mathbb{R}^n$ et un point $x^* \in U$. Un vecteur γ *est* un sous-gradient de ϕ *en* x^* si *pour tout* $x \in U$

$$\phi(x) \geq \phi(x*) + \langle \gamma, x - x* \rangle$$

L'ensemble de tout les sous gradients en $x*$ est appelé sous différentiel de ϕ en $x*$ et noté $\partial\phi(x*)$

3.2.1 Règles de calcul sous-différentiel

Lemme 3.2.1 [4]

Soit f une fonction convexe. On a :
i) Si f est différentiable sur son domaine, alors :

$$\forall x \in int(dom(f)) \quad \partial f(x) = \nabla f(x)$$

ii) Si $dom(f) \subseteq \mathbb{R}^n, A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ un opérateur linéaire et $b \in \mathbb{R}^n$, alors la fonction $\phi(x) = f(Ax + b)$ est convexe et :

$$\forall x \in int(dom(\phi)), \partial\phi(x) = A^T \partial f(Ax + b)$$

iii) Si $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x)$ avec f_1 et f_2 convexe sur $\mathbb{R}^n$
et $(\alpha_1 \alpha_2) \in \mathbb{R}_+ \times \mathbb{R}_+$ alors :

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$$

$$= \{\gamma \in \mathbb{R}^n, \exists (x_1, x_2) \in \partial f_1(x_1) \times f_2(x_2), \quad \gamma = \alpha_1 x_1 + \alpha_2 x_2\}$$

iv) Soit $f : \mathrm{R}^n \longrightarrow \mathbb{R}^n \cup \{+\infty\}$ une fonction convexe fermée et $g : \mathrm{R}^n \longrightarrow \mathbb{R}^n \cup \{+\infty\}$ une fonction convexe croissante. Notons : $h : g \circ f$
Alors $\forall x \in int(dom(f))$.

$$\partial h(x) = \{\langle \gamma_1 \gamma_2, \gamma_1 \rangle \in \partial g(f(x)), \gamma_2 \in \partial f(x)\}$$

3.3 Méthode de sous gradient

Parmi les méthodes les plus simples pour résoudre le problème d'optimisation non différentiable, on trouve le sous gradient qui est proposé pour maximiser les fonctions concaves et minimiser les fonctions convexes non nécessairement différentiables .

Considérons le problème :

$$\min_{x \in X} f(x)$$

Où $X \subset \mathbb{R}^n$ et $f : X \longrightarrow \mathbb{R}^n$ est une fonction convexe mais pas nécessairement de classe C^1.

3.3.1 Direction de descente

Les algorithmes utilisés pour résoudre des problèmes d'optimisation non-différentiables ne diffèrent pas beaucoup au premier abord, des algorithmes standards, ils consistent schématiquement en la recherche d'une direction de descente d_l et d'une longueur de pas $t_l > 0$. Le nouvel itéré s'obtient par : $x^{l+1} = x^l + t_l d^l$

3.3.2 Condition d'optimalité

Enfin, signalons comme autre caractéristique importante des problèmes non-différentiables la difficulté d'établir des tests d'arrêts qui soient applicables. Les conditions d'optimalité pour un point x^* s'écrit à l'aide de l'une des trois propriétés équivalentes :
i) $f(x) \geq f(x^*)$ pour tout $x \in \mathbb{R}^n$
ii) $0 \in \partial f(x^*)$
iii) $f'(x^*, d) \geq 0 \quad pour \quad$ tout $d \in \mathbb{R}^n$

3.3.3 Algorithme de sous-gradient

Dans cette partie, on s'intéresse aux méthodes de sous-gradient de premier ordre appliquées à la classe de problèmes suivants :

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

Où $U \subset \mathbb{R}^n$ est un ensemble convexe, f est lipschitz de constante L sur U. On note x^* un minimiseur du problème. Notons que ce problème n'a pas de raison particulière d'admettre un minimiseur unique.

Algorithme de sous-gradient dans le cas sans contrainte : [4]

• **Input** :
N : Nombre d'itérations
$x_0 \in U$: Un point de départ.
• **Output** :
x_N : Une solution approchée.
Begin
for k allant de 0 à N do
Calculer $\gamma \in \partial f(x_k)$
Recherche linéaire : Chercher un pas $T_k > 0$ tel que

$$f(x_k - T_k \frac{\gamma_k}{\|\gamma_k\|}) < f(x_k)$$

$$x_{k+1} = x_k - T_k \frac{\gamma_k}{\|\gamma_k\|}$$

End

3.3.4 Méthode de sous-gradient projeté

Revenons maintenant au problème :

$$\min_{x \in X} f(x)$$

On note P_x la projection orthogonale du point $x \in \mathbb{R}^n$ sur le domaine admissible X. Comme l'ensemble X est convexe, la projection est univaluée. Notons que la projection est facile à calculer dans le cas de contraintes de type "boite" mais très difficile à calculer dans la plupart des cas.

Algorithme 2 : Algorithme de sous-gradient dans le cas sans contrainte $(U \neq \mathbb{R}^n)$

Input
N : Nombre d'itératio n
$x_0 \in X$: Un point de départ
Output
x_N : Une solution approchée.
Begin
for k allant de 0 à N do
Calculer $\gamma_k \in \partial f(x_k)$
Recherche linéaire : Chercher un pas $T_k > 0$
tel que

$$f(x_k - T \frac{\gamma_k}{\|\gamma_k\|}) < f(x_k)$$

$$x_{k+1} = P_x(x_k - T_k \frac{\gamma_k}{\|\gamma_k\|})$$

End

Théorème 2.3 : [6]

On note $f_k^* = \min_{k\in\{0..n\}} f(x_k)$ et R le diamètre de l'ensemble X.

$$f_k^* - f^* \leq L\frac{R^2 + \Sigma_{i=0}^k h_i^2}{2\Sigma_{i=0}^k h_i}$$

En particulier si $h_k = \frac{R}{\sqrt{N+1}}$, on a un taux "optimal" et :

$$f_k^* - f^* \leq \frac{LR}{\sqrt{N+1}}$$

Preuve [6] :

Dans certains cas, nous pouvons exprimer la mise du sous-programme projeté d'une autre manière. Lorsque U est affine, c'est à dire que $U = x|Ax = b$ où A est grande et en plein rang, l'opérateur de projection est affine, et donné par :

$$P(Z) = Z - A^T(AA^T)^{-1}(AZ - b)$$

Dans ce cas nous pouvons simplifier le sous gradient par :

$$x^{(k+1)} = x^{(k)} - k(L - A^T(AA^T)^{-1}A)^{(k)}$$

3.3.5 Choix de Polyak

Polyac [7] a proposé un choix pour la longueur de pas $\alpha_k = \frac{T_k}{\|\gamma_k\|^2}$ dans les deux cas suivants :

- Cas ou f^* est connue : Il a proposé

$$\alpha_k = \frac{f(x^k) - f^k}{\|\gamma(k)\|^2}$$

- Cas ou f^* est inconnue : Il a proposé

$$\alpha_k = \frac{f(x^k) - f^k + T_k}{\|\gamma^{(k)}\|^2}$$

3.4 Méthode des plans sécants et de faisceaux

3.4.1 Les méthodes des plans sécants

La méthode des plans sécants de Kelly [8] est une méthode largement utilisé pour l'optimisation des fonctions convexes non partout différentiables. Elle consiste à maintenir une approximation de la fonction à optimiser à l'aide de linéarisations tangentielles en plusieurs points, et a chaque itération, elle détermine le minimum de la fonction approximative et génère au point où ce minimum et trouver un nouveau. plan tangent qui vient enrichir l'approximation. Le processus est poursuivi jusqu'à l'optimum cherché soit atteint. Pour le développement de cette méthode, nous considérons le problème à résoudre $min\{g(\lambda) : \lambda \geq 0\}$ avec $\lambda_i(i = 1..k)$ et pour déterminer le minimum de cette approximation on résout le programme linéaire suivant :

$$(P^s)_k min\theta_k$$

$$f(\lambda_i) + S_i^T(\lambda - \lambda_i) \leq \theta_k \quad i = 1..k$$

3.4.2 Méthode de faisceaux

Les méthodes de faisceaux sont considérées parmi les plus efficaces pour la résolution de problème non-différentiable. Historiquement, les méthode faisceaux sont basées sur la méthode des plans sécants voir [9] pour la minimisation sans contraintes de fonctionnelle convexe. Le principe est le suivant : au lieu de faire un appel à l'oracle pour générer les candidats à la descente, on se sert des informations du premier ordre de la fonction objectif pour construire un modèle (convexe) de f, plus facile à minimiser.
Étant donné un faisceau d'informations

$$\forall y \in \mathbb{R}^n, \phi_k(y) = \max_{1 \leq i \leq k} \{f_i + \langle \gamma_i ; x - x_i \rangle \leq f(y)\}$$

A chaque itération de l'algorithme, la minimisation du modèle ϕ_k donne un nouvel itéré x_{k+1} qui va permettre d'enrichir le modèle courant pour la donnée d'un nouveau plan sécant. L'information passé est ainsi conservée au fil des itérations Idée des plans sécants : Accumuler des informations des itérations précédentes : f

$$f_i = f(x_i) et \gamma_i \in \delta f(x_i); \quad 1 \leq i \leq k$$

A l'itération k on considère $\tilde{f}$ une approximation affine par morceaux de f (Plus facile à minimiser).

$$\phi_k(y) = max_{1 \leq i \leq k} \{f_i + \langle \gamma_i ; x - x_i \rangle\}$$

Notons que :

$$\phi_k \leq \phi_{k+1} \text{ et } \phi_k(x) \leq f(x), \quad \forall k \geq 0, \forall x \in \mathbb{R}^n$$

On détermine $C \in \mathbb{R}^n$ compact content un minimiseur de f et l'on fixe $\varepsilon > 0$ critère d'arrêt. Pour $k > 0$.
$\varepsilon_k = f(x_k) - \phi(x_k)$: Arrêter si $\varepsilon_k \leq \varepsilon$
$Z_{k+1} = argmin_{x \in c} \quad \phi(x)$
$d_k = -Z_{k+1} + x_k$
$Z_{k+1} = x_k - \uparrow_k d_k$

d_k n'est pas forcément une direction de descente pour f.

Application

4

4.1 Problème des moindres carrés :

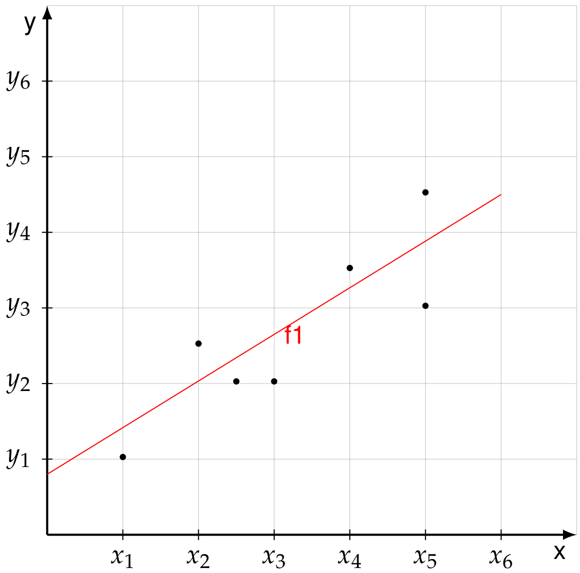
- Comment trouver l'équation d'une droite ?

Supposons qu'on a le tableau suivant donné par une expérience :

x_i	x_1	x_2	x_3	...	x_n
y_i	y_1	y_2	y_3	...	y_n

On veut utiliser les résultats d'une expérience pour établir, une relation entre x et y $\Longrightarrow$ c.à.d $y = Q(x)$

On dispose les poins (x_1, y_1) sur l'hyperplan



Trouver l'équation de la droite qui approche le mieux ces points.

$l_i = y_i - (ax_i + b)$

$$\min \sum_{i=1}^{n} \quad |l_i| \Longrightarrow \text{Cette fonction n'est pas différentiable}$$

$$\min \sum_{i=1}^{n} \quad (l_i)^2 \Longrightarrow \min \sum_{i=1}^{n} \quad (y_i - (ax_i + b))^2$$

$$\text{On obtient le problème :} \begin{cases} \min \sum_{i=1}^{n} \quad (l_i)^2 \Longrightarrow \min \sum_{i=1}^{n} \quad (y_i - (ax_i + b))^2 \\ (a, b) \in \mathbb{R}^2 \end{cases}$$

C'est ce qu'on appelle le problème des moindres carrés.
Une expérience a donné le tableau suivant :

x	2	1	1
y	1	2	1

$$f(a,b) = \sum_{i=1}^{3} \quad (y_i - (ax_i + b))^2$$

$(a,b) \in \mathbb{R}^2$
$f(a,b) = (1-(2a+b))^2 + (2-(a+b))^2 + (1-(a+b))^2$
$f(a,b) = 1^2 + (2a+b)^2 + 2\times 1 \times (2a+b) + 4 + (a+b)^2 - 4(a+b) + 1 + (a+b)^2 - 2(a+b)$
$f(a,b) = 1 + (4a^2 + b^2 + 2\times 2a \times b) - (4a+2b) + 4 + (a^2+b^2+2ab) - (4a+4b) + 1 + (a^2+b^2+2ab) - (2a+2b)$
$f(a,b) = 1 + 4a^2 + b^2 + 4ab - 4a - 2b + 4 + a^2 + b^2 + 2ab - 4a - 4b + 1 + a^2 + b^2 + 2ab - 2a - 2b$
$f(a,b) = 6a^2 + 3b^2 + 8ab - 10a - 8b + 6$

4.2 Méthode de Newton

Soit $(a,b)^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \varepsilon = 10^{-2}$

$$\begin{cases} (a,b)^0 \quad \text{Donné} \\ (a,b)^{k+1} = (a,b)^k - (\nabla^2 f(a^k,b^k))^{-1} \quad \nabla f(a^k,b^k) \end{cases}$$

$\nabla f(a,b) = \begin{pmatrix} 12a+8b-10 \\ 6b+8a-8 \end{pmatrix}$

$\nabla f(1,1) = \begin{pmatrix} 12\times 1 + 8\times 1 - 10 \\ 6\times 1 + 8 \times 1 - 8 \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$

$\nabla^2 f(a,b) = \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix}$

Calculons $(\nabla^2 f(a,b))^{-1}$
det $\quad (\nabla^2 f(a,b)) = 12 \times 6 - 8 \times 8 = 8$

$$(\nabla^2 f(a,b))^{-1} = \frac{1}{8}\begin{pmatrix} 6 & -8 \\ -8 & 12 \end{pmatrix} = \begin{pmatrix} \frac{6}{8} & \frac{-8}{8} \\ \frac{-8}{8} & \frac{12}{8} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & -1 \\ -1 & \frac{3}{2} \end{pmatrix}$$

$$(a,b)^{k+1} = (a,b)^k - (\nabla^2 f(a^k,b^k))^{-1} \quad \nabla f(a^k,b^k)$$

$$(a,b)^1 = (a,b)^0 - (\nabla^2 f(a^0,b^0))^{-1} \quad \nabla f(a^0,b^0)$$

$$(a,b)^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{3}{4} & -1 \\ -1 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} 10 \\ 6 \end{pmatrix}$$

$$(a,b)^1 = \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix}$$

Critère d'arrêt :

$$\|(a,b)^{k+1} - (a,b)^k\| < \varepsilon$$

$$\left\| \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} -\frac{3}{2} \\ 1 \end{pmatrix} \right\|$$

$$= \sqrt{(-\frac{3}{2})^2 + 1^1} = \sqrt{3,25} = 1,80 \quad \geq \varepsilon$$

$$\nabla f \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \times (-\frac{1}{2}) + 8 \times 2 - 10 \\ 6 \times 2 + 8 \times (-\frac{1}{2}) - 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\left[\nabla^2 f(a,b))\right]^{-1} = \begin{pmatrix} \frac{3}{4} & -1 \\ -1 & \frac{3}{2} \end{pmatrix}$$

$$(a,b)^2 = (a,b)^1 - \left[\nabla^2 f(a^1,b^1))\right]^{-1} \nabla f(a^1,b^1)$$

$$(a,b)^2 = \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{3}{4} & -1 \\ -1 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$(a,b)^2 = \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix}$$

Donc la solution optimale $(a,b)^* = \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix}$

L'équation de la droite qui approche le mieux ces points $(x_i, y_i)_{i=1,2,3}$

$$y = -\frac{1}{2}x + 2$$

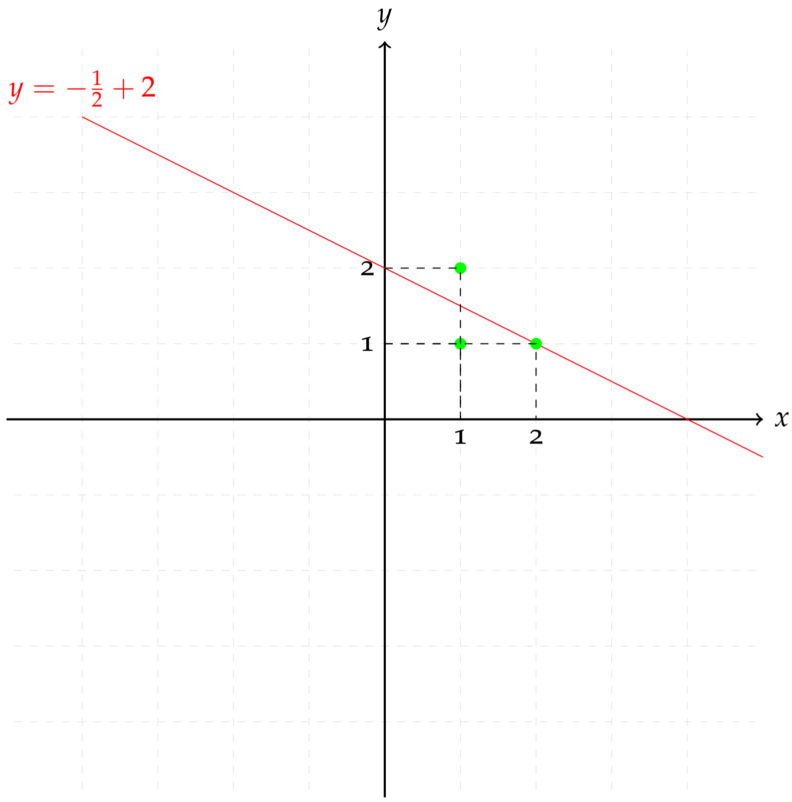


FIGURE 4.1 – *La droite qui s'approche du mieux de ces points*

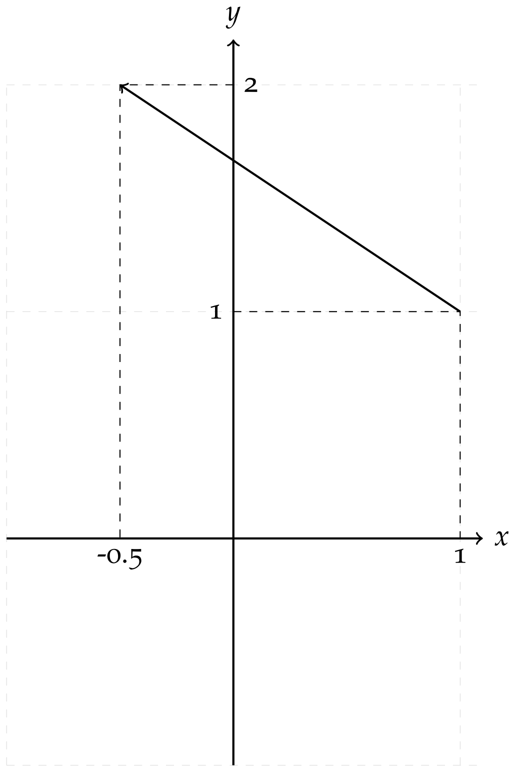


Figure 4.2 – *Itérations de la méthode de Newton*

4.3 Méthode du gradient à pas optimal :

On a $(a,b)^{k+1} = (a,b)^k - \sigma_k \nabla f((a,b)^k)$

Comme $(a,b)^0 = (1,1)$ et $\nabla f((a,b)^0) = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$

Donc $(a,b)^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \sigma_0 \begin{pmatrix} 10 \\ 6 \end{pmatrix}$

On calcule σ_0 :

$$\sigma_0 = \frac{\|\nabla f((a,b)^0)\|^2}{\nabla f((a,b)^0)^T A \nabla f((a,b)^0)} = \frac{\left\|\begin{pmatrix} 10 \\ 6 \end{pmatrix}\right\|^2}{\begin{pmatrix} 10 & 6 \end{pmatrix}\begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix}\begin{pmatrix} 10 \\ 6 \end{pmatrix}} = \frac{136}{2376} = 0,0572$$

Donc $(a,b)^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 0,0572 \begin{pmatrix} 10 \\ 6 \end{pmatrix} = \begin{pmatrix} 0,4276 \\ 0,6566 \end{pmatrix}$

Critère d'arrêt :

$$\|(a,b)^{k+1} - (a,b)^k\| = \|(a,b)^1 - (a,b)^0\| = \left\| \begin{pmatrix} 0,4276 \\ 0,6566 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|$$

$$= \sqrt{(0,4276-1)^2 + (0,6566-1)^2} = 0,6675 > \varepsilon$$

$(a,b)^2 = (a,b)^1 - \sigma_1 \nabla f((a,b)^1)$

$(a,b)^1 = \begin{pmatrix} 0,4276 \\ 0,6566 \end{pmatrix}$ et $\nabla f((a,b)^1) = \begin{pmatrix} 0,3838 \\ -0,6397 \end{pmatrix}$

Donc $(a,b)^2 = \begin{pmatrix} 0,4276 \\ 0,6566 \end{pmatrix} - \sigma_1 \begin{pmatrix} 0,3838 \\ -0,6397 \end{pmatrix}$

Calculons σ_1 :

$$\sigma_1 = \frac{\left\| \begin{pmatrix} 0,3838 \\ -0,6397 \end{pmatrix} \right\|^2}{\begin{pmatrix} 0,3838 & -0,6397 \end{pmatrix} \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 0,3838 \\ -0,6397 \end{pmatrix}}$$

$= 1,8889$

Donc $(a,b)^2 = \begin{pmatrix} 0,4276 \\ 0,6566 \end{pmatrix} - 1,8889 \begin{pmatrix} 0,3838 \\ -0,6397 \end{pmatrix} = \begin{pmatrix} -0,2974 \\ 1,8649 \end{pmatrix}$

Critère d'arrêt :

$$\left\| \begin{pmatrix} -0,2974 \\ 1,8649 \end{pmatrix} - \begin{pmatrix} 0,4276 \\ 0,6566 \end{pmatrix} \right\| = 1,4092 > \varepsilon$$

$$(a,b)^3 = \begin{pmatrix} -0,2974 \\ 1,8649 \end{pmatrix} - \sigma_2 \begin{pmatrix} 1,3505 \\ 0,8103 \end{pmatrix}$$

On a donc σ_2

$$\sigma_2 = \frac{\left\| \begin{pmatrix} 1,3505 \\ 0,8103 \end{pmatrix} \right\|^2}{\begin{pmatrix} 1,3505 \\ 0,8103 \end{pmatrix}^T \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 1,3505 \\ 0,8103 \end{pmatrix}} = 0,0572$$

$$(a,b)^3 = \begin{pmatrix} -0,2974 \\ 1,8649 \end{pmatrix} - 0,0572 \begin{pmatrix} 1,3505 \\ 0,8103 \end{pmatrix} = \begin{pmatrix} -0,3747 \\ 1,8186 \end{pmatrix}$$

Critère d'arrêt :

$$\left\| \begin{pmatrix} -0,3747 \\ 1,8186 \end{pmatrix} - \begin{pmatrix} -0,2974 \\ 1,8649 \end{pmatrix} \right\| = 0,0902 > \varepsilon$$

$$(a,b)^4 = \begin{pmatrix} -0,3747 \\ 1,8186 \end{pmatrix} - \sigma_3 \begin{pmatrix} 0,0518 \\ -0,0864 \end{pmatrix}$$

On a donc σ_3

$$\sigma_3 = \frac{\left\| \begin{pmatrix} 0,0518 \\ -0,0864 \end{pmatrix} \right\|^2}{\begin{pmatrix} 0,0518 \\ -0,0864 \end{pmatrix}^T \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} 0,0518 \\ -0,0864 \end{pmatrix}} = 1,8889$$

$$(a,b)^4 = \begin{pmatrix} -0,3747 \\ 1,8186 \end{pmatrix} - 1,8889 \begin{pmatrix} 0,0518 \\ -0,0864 \end{pmatrix} = \begin{pmatrix} -0,4726 \\ 1,9818 \end{pmatrix}$$

Critère d'arrêt :

$$\|(a,b)^4 - (a,b)^3\| = \left\| \begin{pmatrix} -0,4726 \\ 1,9818 \end{pmatrix} - \begin{pmatrix} -0,3747 \\ 1,8186 \end{pmatrix} \right\|$$

$= 0,1903 > \varepsilon$

$$(a,b)^5 = (a,b)^4 - \sigma_4 \nabla f((a,b)^4)$$

$$(a,b)^5 = \begin{pmatrix} -0,4726 \\ 1,9818 \end{pmatrix} - \sigma_4 \begin{pmatrix} 0,1824 \\ 0,1094 \end{pmatrix}$$

On a donc σ_4

$$\sigma_4 = \frac{\left\|\begin{pmatrix}0,1824\\0,1094\end{pmatrix}\right\|^2}{\begin{pmatrix}0,1824\\0,1094\end{pmatrix}^T\begin{pmatrix}12 & 8\\8 & 6\end{pmatrix}\begin{pmatrix}0,1824\\0,1094\end{pmatrix}} = 0,0572$$

$$(a,b)^5 = \begin{pmatrix}-0,4726\\1,9818\end{pmatrix} - 0,0572\begin{pmatrix}0,1824\\0,1094\end{pmatrix} = \begin{pmatrix}-0,4831\\1,9755\end{pmatrix}$$

Critère d'arrêt :

$$\left\|(a,b)^5 - (a,b)^4\right\| = \left\|\begin{pmatrix}-0,4726\\1,9818\end{pmatrix} - \begin{pmatrix}-0,4831\\1,9755\end{pmatrix}\right\|$$

$$= 0,0122 > \varepsilon$$

$$(a,b)^6 = (a,b)^5 - \sigma_5 \nabla f((a,b)^5)$$

$$(a,b)^6 = \begin{pmatrix}-0,4831\\1,9755\end{pmatrix} - \sigma_5\begin{pmatrix}0,0070\\-0,0117\end{pmatrix}$$

On a donc σ_5

$$\sigma_5 = \frac{\left\|\begin{pmatrix}0,0070\\-0,0117\end{pmatrix}\right\|^2}{\begin{pmatrix}0,0070\\-0,0117\end{pmatrix}^T\begin{pmatrix}12 & 8\\8 & 6\end{pmatrix}\begin{pmatrix}0,0070\\-0,0117\end{pmatrix}} = 1,8889$$

$$(a,b)^6 = \begin{pmatrix}-0,4831\\1,9755\end{pmatrix} - 1,8889\begin{pmatrix}0,0070\\-0,0117\end{pmatrix} = \begin{pmatrix}-0,4963\\1,9975\end{pmatrix}$$

Critère d'arrêt :

$$\left\|(a,b)^6 - (a,b)^5\right\| = \left\|\begin{pmatrix}-0,4963\\1,9975\end{pmatrix}\begin{pmatrix}-0,4831\\1,9755\end{pmatrix}\right\|$$

$$= 0,0257 > \varepsilon$$

$$(a,b)^7 = (a,b)^6 - \sigma_6 \nabla f((a,b)^6)$$

$$(a,b)^7 = \begin{pmatrix}-0,4963\\1,9975\end{pmatrix} - \sigma_6\begin{pmatrix}0,0246\\0,0148\end{pmatrix}$$

On a donc σ_6

$$\sigma_6 = \frac{\left\|\begin{pmatrix}0,0246\\0,0148\end{pmatrix}\right\|^2}{\begin{pmatrix}0,0246\\0,01482\end{pmatrix}^T\begin{pmatrix}12 & 8\\8 & 6\end{pmatrix}\begin{pmatrix}0,0246\\0,0148\end{pmatrix}} = 0,0572$$

$$(a,b)^7 = \begin{pmatrix}-0,4963\\1,9975\end{pmatrix} - 0,0572\begin{pmatrix}0,0246\\0,0148\end{pmatrix} = \begin{pmatrix}-0,4977\\1,9967\end{pmatrix}$$

Critère d'arrêt :

$$\left\|(a,b)^7 - (a,b)^6\right\| = \left\|\begin{pmatrix}-0,4977\\1,9967\end{pmatrix}\begin{pmatrix}-0,4963\\1,9975\end{pmatrix}\right\|$$

$= 0,0016 < \varepsilon$ STOP

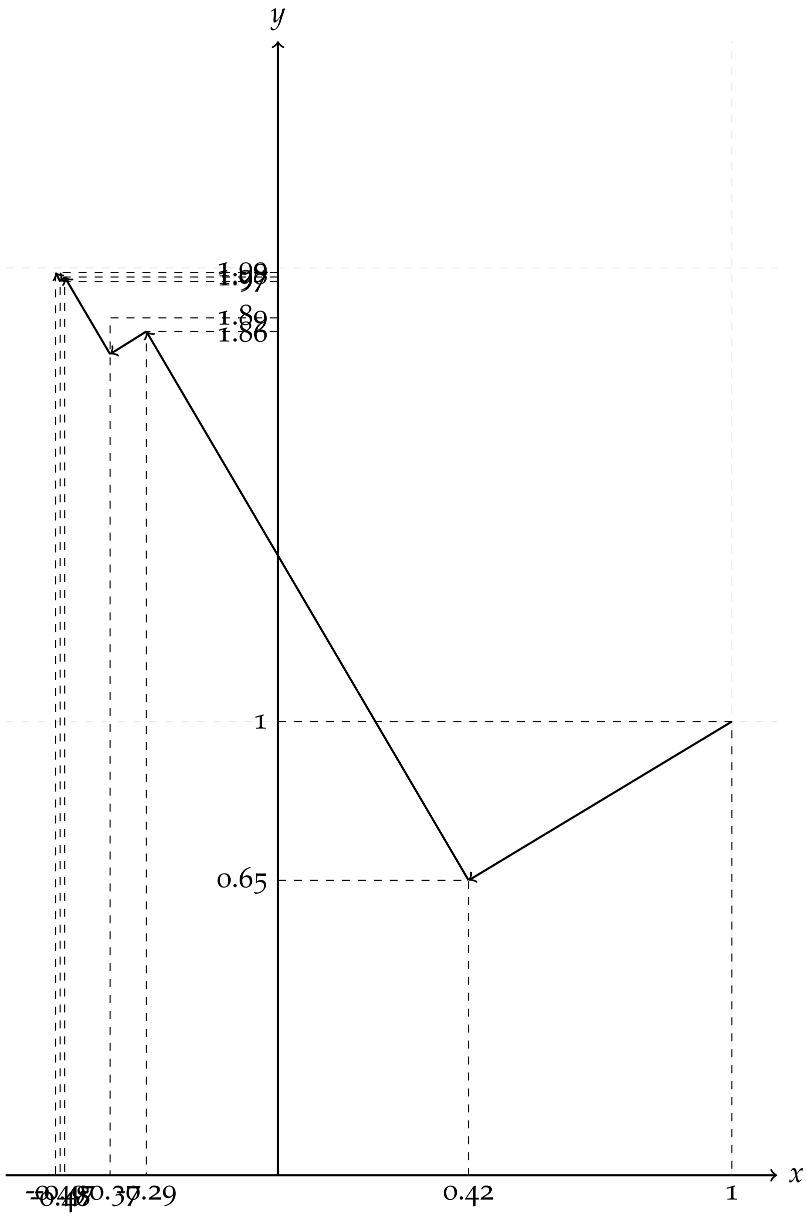


Figure 4.3 – *Itération de la méthode du gradient à pas optimal*

4.4 Méthode du gradient conjugué

On a $(a,b)^0 = \begin{pmatrix}1\\1\end{pmatrix}$, $A = \begin{pmatrix}12 & 8\\8 & 6\end{pmatrix}$ $\quad C = \begin{pmatrix}10\\8\end{pmatrix}$

$$\begin{aligned} d^0 &= \nabla f((a,b)^0) = A(a,b)^0 - C \\ &= \begin{pmatrix}12 & 8\\8 & 6\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}10\\8\end{pmatrix} \\ &= \begin{pmatrix}10\\6\end{pmatrix} \end{aligned}$$

Donc

$$\begin{aligned} \sigma_0 = \frac{\langle \nabla f((a,b)^0), d^0 \rangle}{\langle Ad^0, d^0 \rangle} &= \frac{\left\langle \begin{pmatrix}10\\6\end{pmatrix}, \begin{pmatrix}10\\6\end{pmatrix} \right\rangle}{\left\langle \begin{pmatrix}12 & 8\\8 & 6\end{pmatrix}\begin{pmatrix}10\\6\end{pmatrix}, \begin{pmatrix}10\\6\end{pmatrix} \right\rangle} \\ &= \frac{\left\langle \begin{pmatrix}10\\6\end{pmatrix}, \begin{pmatrix}10\\6\end{pmatrix} \right\rangle}{\left\langle \begin{pmatrix}168\\116\end{pmatrix}, \begin{pmatrix}10\\6\end{pmatrix} \right\rangle} = \frac{10 \times 10 + 6 \times 6}{168 \times 10 + 116 \times 6} \\ &= \frac{136}{2376} = \frac{17}{297} = 0,057239057 \end{aligned}$$

Donc $(a,b)^1 = (a,b)^0 - \sigma_0 d^0$

$$(a,b)^1 = \begin{pmatrix}1\\1\end{pmatrix} - \frac{17}{297}\begin{pmatrix}10\\6\end{pmatrix}$$

$$(a,b)^1 = \begin{pmatrix}1\\1\end{pmatrix} - 0,057239057\begin{pmatrix}10\\6\end{pmatrix} = \begin{pmatrix}0,427609473\\0,656565657\end{pmatrix}$$

On a $\nabla f((a,b)^1) = A(a,b)^1 - C$

$$\begin{pmatrix}12 & 8\\8 & 6\end{pmatrix}\begin{pmatrix}0,427609473\\0,656565657\end{pmatrix} - \begin{pmatrix}10\\8\end{pmatrix} = \begin{pmatrix}0,38383893\\-0,639730274\end{pmatrix}$$

On a donc

$$\begin{aligned} \alpha_0 = \frac{\|\nabla f(a^1,b^1)\|^2}{\|\nabla f(a^0,b^0)\|^2} &= \frac{\left\| \begin{pmatrix}0,38383893\\-0,639730274\end{pmatrix} \right\|^2}{\left\| \begin{pmatrix}10\\6\end{pmatrix} \right\|^2} \\ &= 0,004092552551 \end{aligned}$$

Donc $d^1 = \nabla f((a,b)^1) + \alpha_0 d^0$

$$d^1 = \begin{pmatrix}0,38383893\\-0,639730274\end{pmatrix} + 0,004092552551 \times \begin{pmatrix}10\\6\end{pmatrix}$$

$$d^1 = \begin{pmatrix}0,424764455\\-0,615174959\end{pmatrix}$$

On pose alors $k = k + 1 = 0 + 1 = 1$
Comme $\nabla f((a,b)^1) \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
Donc calculons σ_1

$$\sigma_1 = \frac{\langle \nabla f((a,b)^1), d^1 \rangle}{\langle Ad^1, d^1 \rangle} = \frac{0,556587179}{0,254868188} = 2,183823659$$

Donc $(a,b)^2 = (a,b)^1 - \sigma_1 d^1$

$$(a,b)^2 = \begin{pmatrix} 0,427609473 \\ 0,656565657 \end{pmatrix} - 2,183823659 \begin{pmatrix} 0,424764455 \\ -0,615174959 \end{pmatrix} = \begin{pmatrix} -0,500001193 \\ 1,999999278 \end{pmatrix}$$

On a $\nabla f((a,b)^2) = A(a,b)^1 - C$

$$\nabla f((a,b)^2) = \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} -0,500001193 \\ 1,999999278 \end{pmatrix} - \begin{pmatrix} 10 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

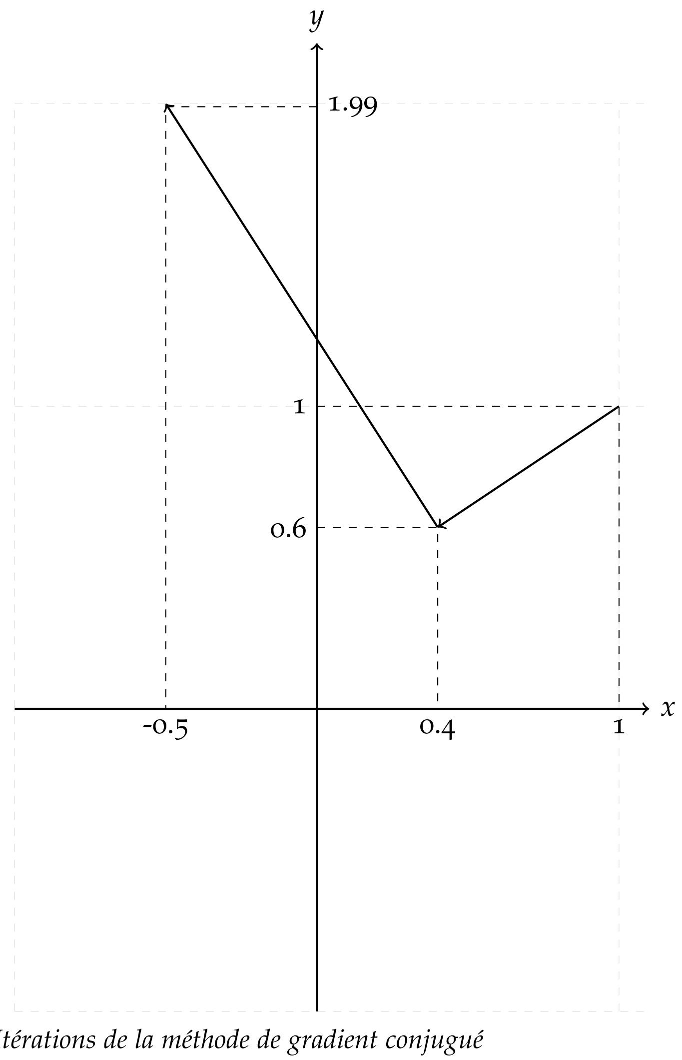


Figure 4.4 – *Itérations de la méthode de gradient conjugué*

4.5 Méthode de Quasi Newton

On a $(a,b)^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $,A = \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix}$ $C = \begin{pmatrix} 10 \\ 8 \end{pmatrix}$ $B_0 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
k=o

$(a,b)^1 = (a,b)^0 + \alpha_0 d^0$
$\nabla f((a,b)^0) = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$
On a donc $\alpha_0 = 1$

Donc $d^k = -B_k \nabla f((a,b)^k)$
$d^0 = -B_0 \nabla f((a,b)^0)$=$-\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 10 \\ 6 \end{pmatrix} = \begin{pmatrix} -10 \\ -6 \end{pmatrix}$
Donc
$(a,b)^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} -10 \\ -6 \end{pmatrix} = \begin{pmatrix} -9 \\ -5 \end{pmatrix}$
Critère d'arrêt
$\left\|\nabla f((a,b)^1)\right\| = \left\| \begin{pmatrix} -158 \\ -110 \end{pmatrix} \right\| = 192,5201 > \varepsilon$

On pose alors $k = k+1 = 0+1 = 1$
On a donc

$$\alpha_1 = \frac{\|\nabla f((a,b)^1)\|^2}{\nabla f((a,b)^1)^T A \nabla f((a,b)^1)}$$

$$= \frac{\left\| \begin{pmatrix} -158 \\ -110 \end{pmatrix} \right\|^2}{\begin{pmatrix} -158 & -110 \end{pmatrix} \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} -158 \\ -110 \end{pmatrix}}$$

$$\alpha_1 = 0,0570$$

On calculons $Y_0 et S_0$:

$S_k = (a,b)^{k+1} - (a,b)^k$
$S_0 = (a,b)^1 - (a,b)^0 = \begin{pmatrix} -9 \\ -5 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -10 \\ -6 \end{pmatrix}$

$Y_k = \nabla f(a^{k+1}, b^{k+1}) - \nabla f(a^k, b^k)$
$Y_0 = \nabla f((a,b)^1) - \nabla f((a,b)^0)$
$Y_0 = \begin{pmatrix} -158 \\ -110 \end{pmatrix} - \begin{pmatrix} 10 \\ 6 \end{pmatrix} = \begin{pmatrix} -168 \\ -116 \end{pmatrix}$
Calculons B_1

$$B_1 = (I - \frac{S_0 Y_0^t}{S_0^t Y_0}) B_0 (I - \frac{S_0 Y_0^t}{S_0^t Y_0}) + \frac{S_0 S_0^t}{S_0 Y_0}$$

$$B_1 = \begin{pmatrix} 0,3350 & -0,4630 \\ -0,3990 & 0,7222 \end{pmatrix}$$

Donc $d^1 = -B_1 \nabla f((a,b)^1)$

$$d^1 = -\begin{pmatrix} 0,3350 & -0,4630 \\ -0,3990 & 0,7222 \end{pmatrix} \begin{pmatrix} -158 \\ -110 \end{pmatrix} = \begin{pmatrix} 2,0067 \\ 16,4040 \end{pmatrix}$$

Donc
$(a,b)^2 = (a,b)^1 + \alpha_1 d^1$

$$(a,b)^2 = \begin{pmatrix} -9 \\ -5 \end{pmatrix} + 0,0570 \begin{pmatrix} 2,0067 \\ 16,4040 \end{pmatrix} = \begin{pmatrix} -8,8856 \\ -4,0650 \end{pmatrix}$$

Critère d'arrêt
$$\|\nabla f((a,b)^2)\| = \left\| \begin{pmatrix} -149,1101 \\ -103,4473 \end{pmatrix} \right\| = 181,4804 > \varepsilon$$

On pose alors $k = k + 1 = 1 + 1 = 2$

On a donc

$$\alpha_2 = \frac{\|\nabla f((a,b)^2)\|^2}{\nabla f((a,b)^2)^T A \nabla f((a,b)^2)}$$

$$= \frac{\left\| \begin{pmatrix} -149,1101 \\ -103,4473 \end{pmatrix} \right\|^2}{\begin{pmatrix} -149,1101 & -103,4473 \end{pmatrix} \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} -149,1101 \\ -103,4473 \end{pmatrix}}$$

$$\alpha_2 = 0,0570$$

On calculons $Y_1 et S_1$:

$$S_1 = (a,b)^2 - (a,b)^1 = \begin{pmatrix} -8,8856 \\ -4,0650 \end{pmatrix} - \begin{pmatrix} -9 \\ -5 \end{pmatrix} = \begin{pmatrix} 0,1148 \\ 0,9389 \end{pmatrix}$$

$Y_1 = \nabla f((a,b)^2) - \nabla f((a,b)^1)$
$$Y_1 = \begin{pmatrix} -149,1101 \\ -103,4473 \end{pmatrix} - \begin{pmatrix} -158 \\ -110 \end{pmatrix} = \begin{pmatrix} 8,8899 \\ 6,5526 \end{pmatrix}$$

Calculons B_2

$$B_2 = (I - \frac{S_1 Y_1^t}{S_1^t Y_1}) B_1 (I - \frac{S_1 Y_1^t}{S_1^t Y_1}) + \frac{S_1 S_1^t}{S_1 Y_1}$$

$$B_2 = \begin{pmatrix} 0,8433 & -0,0868 \\ -1,1145 & 0,2611 \end{pmatrix}$$

Donc $d^2 = -B_2 \nabla f((a,b)^2)$

$$d^2 = -\begin{pmatrix} 0,8433 & -0,0868 \\ -1,1145 & 0,2611 \end{pmatrix} \begin{pmatrix} -149,1101 \\ -103,4473 \end{pmatrix} = \begin{pmatrix} 115,4598 \\ -139,2032 \end{pmatrix}$$

Donc
$(a,b)^3 = (a,b)^2 + \alpha_2 d^2$

$$(a,b)^3 = \begin{pmatrix} -8,8856 \\ -4,0650 \end{pmatrix} + 0,0570 \begin{pmatrix} 115,4598 \\ -139,2032 \end{pmatrix} = \begin{pmatrix} -2,3045 \\ -11,9995 \end{pmatrix}$$

Critère d'arrêt
$$\|\nabla f((a,b)^3)\| = \left\| \begin{pmatrix} -133,6495 \\ -98,4326 \end{pmatrix} \right\| = 165,9854 > \varepsilon$$

On pose alors $k = k+1 = 2+1 = 3$

On a donc

$$\alpha_3 = \frac{\|\nabla f((a,b)^3)\|^2}{\nabla f((a,b)^3)^T A \nabla f((a,b)^3)}$$

$$= \frac{\left\| \begin{pmatrix} -133,6495 \\ -98,4326 \end{pmatrix} \right\|^2}{\begin{pmatrix} -133,6495 & -98,4326 \end{pmatrix} \begin{pmatrix} 12 & 8 \\ 8 & 6 \end{pmatrix} \begin{pmatrix} -133,6495 \\ -98,4326 \end{pmatrix}}$$

$$\alpha_3 = 0,0570$$

On calculons $Y_2 et S_2$:

$$S_2 = (a,b)^3 - (a,b)^2 = \begin{pmatrix} -2,3045 \\ -11,9995 \end{pmatrix} - \begin{pmatrix} -8,8856 \\ -4,0650 \end{pmatrix} = \begin{pmatrix} 6,5795 \\ -7,9326 \end{pmatrix}$$

$Y_2 = \nabla f((a,b)^3) - \nabla f((a,b)^2)$
$$Y_2 = \begin{pmatrix} -133,6495 \\ -98,4326 \end{pmatrix} - \begin{pmatrix} -149,1101 \\ -103,4473 \end{pmatrix} = \begin{pmatrix} 15,4938 \\ 5,0408 \end{pmatrix}$$

Calculons B_3

$$B_3 = (I - \frac{S_2 Y_2^t}{S_2^t Y_2}) B_2 (I - \frac{S_2 Y_2^t}{S_2^t Y_2}) + \frac{S_2 S_2^t}{S_2 Y_2}$$

$$B_3 = \begin{pmatrix} 0,4951 & -1,0113 \\ -0,2166 & 1,5347 \end{pmatrix}$$

Donc $d^3 = -B_3 \triangledown f((a,b)^3)$

$$d^3 = -\begin{pmatrix} 0,4951 & -1,0113 \\ -0,2166 & 1,5347 \end{pmatrix} \begin{pmatrix} -133,6495 \\ -98,4326 \end{pmatrix} = \begin{pmatrix} -33,3690 \\ 122,1103 \end{pmatrix}$$

Donc
$(a,b)^4 = (a,b)^3 + \lambda_3 d^3$

$$(a,b)^4 = \begin{pmatrix} -2,3045 \\ -11,9995 \end{pmatrix} + 0,0570 \begin{pmatrix} -33,3690 \\ 122,1103 \end{pmatrix} = \begin{pmatrix} 13,8359 \\ -26,7057 \end{pmatrix}$$

Critère d'arrêt
$\|\triangledown f((a,b)^4)\| = \left\| \begin{pmatrix} -100,7653 \\ -71,8660 \end{pmatrix} \right\| = 123,7674 > \varepsilon$

d'après 49 itérations on trouve la solution optimale :
$(a,b)^* = \begin{pmatrix} -0,4982 \\ 1,9964 \end{pmatrix}$

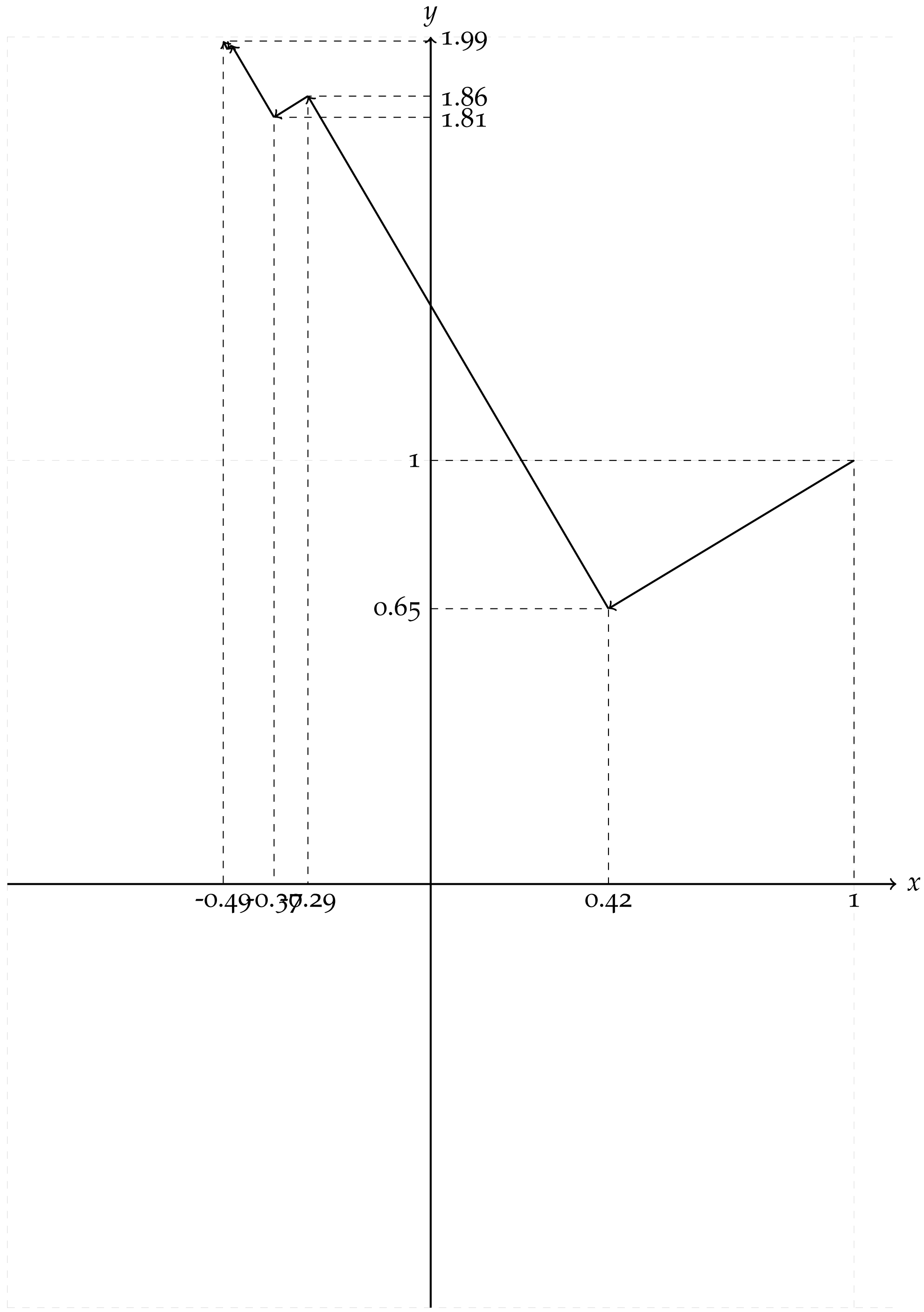


FIGURE 4.5 – *Itérations de la méthode de quasi-Newton*

Conclusion :

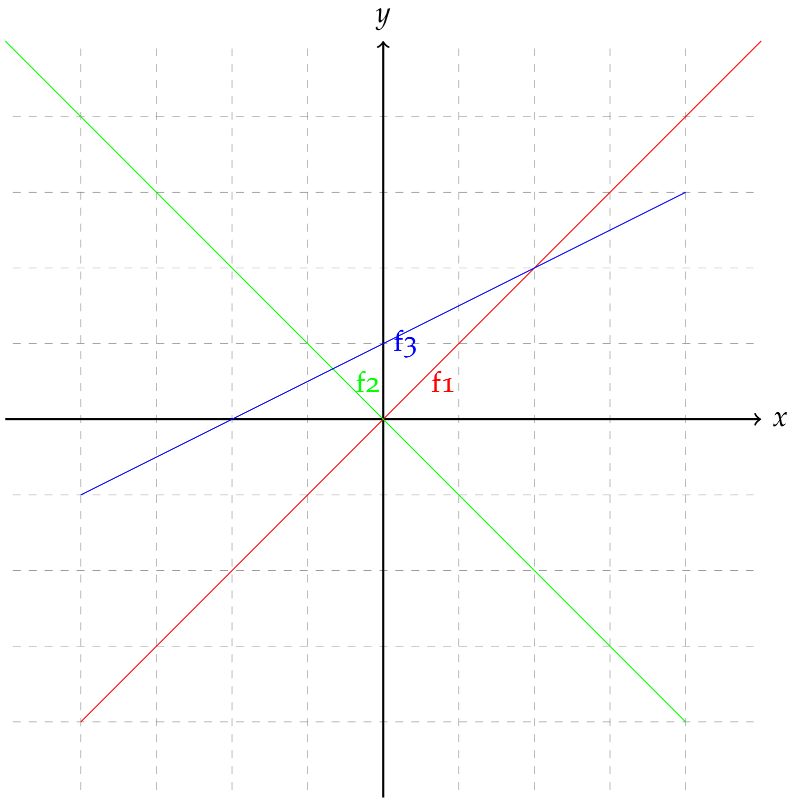
La convergence des méthodes diffère d'une méthode à une autre, certaines d'autres elles convergent plus rapidement que d'autres mais dans l'exemple les plus rapides sont Newton et gradients conjugués et la plus lente est quasi Newton.

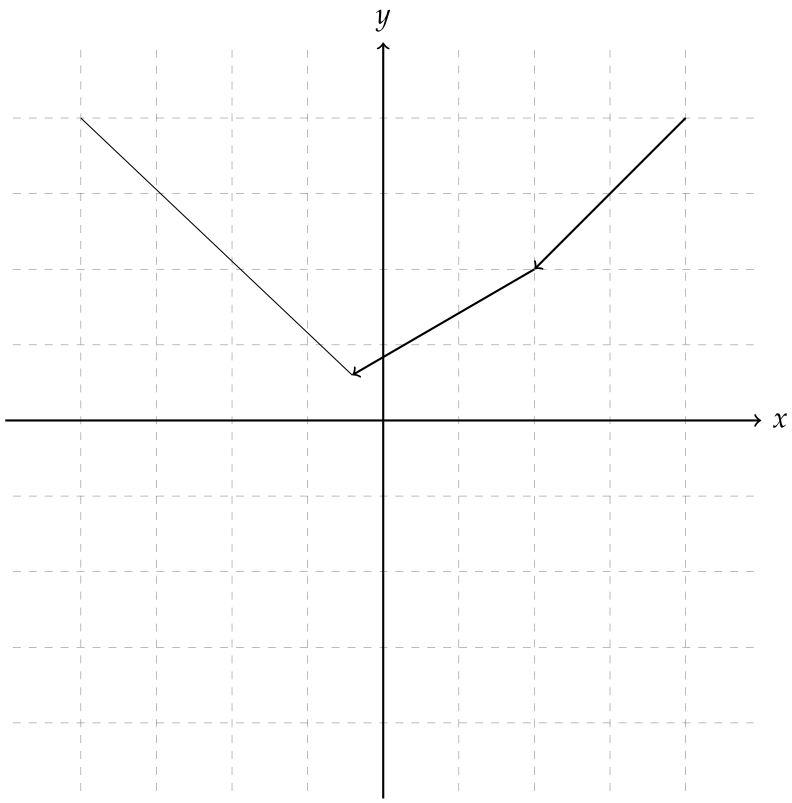
4.6 Méthode de sous gradient

Exemple :

Appliquer la méthode de sous gradient pour $f(x) = \max(x, -x, \frac{1}{2}x + 1)$

$$f(X^0) = \max_{i=1..3} \quad (a_i X^0 + b_i)$$





Ona : $f(X^0) = \max_{i=1..n} a_i x + b_i$
$\xi_0 = \alpha_0$
Test d'arrêt : $|x_i - x_{i-1}| \leq \varepsilon$ avec (ε fixé)
On prend $x_0 = \frac{1}{3}$ $\varepsilon = 0,2$ $\alpha_k = \frac{1}{k+1}$
$f(x) = \{\max f_1(x), f_2(x), f_3(x)\}$
On calcule le sous-gradient
$f(x_0) = \max\{f_1(x_0), f_2(x_0), f_3(x_0)\}$
$f(\frac{1}{3}) = \max\{\frac{1}{3}, -\frac{1}{3}, \frac{7}{6}\}$
$= \frac{7}{6} = f_3(x_0)$
Donc le sous gradient $\xi_0 = f_3'(x) = \frac{1}{2}$
$\xi_0 = \frac{1}{2}$
$x_1 = x_0 - \alpha_0 \xi_0$
$x_1 = \frac{1}{3} - \frac{1}{0+1} \times \frac{1}{2}$
$x_1 = \frac{1}{3} - \frac{1}{2} = \frac{2}{6} - \frac{3}{6} = \frac{2-3}{6} = \frac{-1}{6}$
Le critère d'arrêt :
$|x_1 - x_0| = |\frac{-1}{6} - \frac{1}{3}| = |\frac{-1}{6} - \frac{2}{6}| = |\frac{-3}{6}| = |\frac{-1}{2}|$

$$= \frac{1}{2} > \varepsilon$$

Donc on continue les itérations :
$f(x_1) = \max\{f_1(x_1), f_2(x_1), f_3(x_1)\}$
$= \max\{\frac{-1}{6}, \frac{1}{6}, \frac{11}{12}\}$
$= \frac{11}{12} = f_3(x_1)$
Donc $\xi_1 = f_3'(x) = \frac{1}{2}$
$x_2 = x_1 - \alpha_1 \xi_1$
$x_2 = -\frac{1}{6} - \frac{1}{2} \times (\frac{1}{2})$
$x_2 = -\frac{1}{6} - \frac{1}{4}$

$= \frac{-2}{12} - \frac{3}{12} = \frac{-5}{12}$
Le critère d'arrêt :
$|x_2 - x_1| = |\frac{-5}{12} + \frac{1}{6}| = |\frac{-5}{12} + \frac{2}{12}| = |\frac{-1}{4}|$
$= \frac{1}{4} > \varepsilon$
Donc on continue l'itération : $f(x_2) = \max\{f_1(x_2), f_2(x_2), f_3(x_2)\}$
$= \max\{\frac{-5}{12}, \frac{5}{12}, \frac{19}{24}\}$
$= \frac{19}{24} = f_3(x_2)$
Donc $\xi_2 = f_3^{'}(x) = \frac{1}{2}$
$x_3 = x_2 - \alpha_2 \xi_2$
$x_3 = -\frac{5}{12} - \frac{1}{3} \times (\frac{1}{2})$
$x_3 = -\frac{5}{12} - \frac{1}{6}$
$= \frac{-5}{12} - \frac{2}{12}$
$x_3 = \frac{-7}{12}$
Le critère d'arrêt :
$|x_3 - x_2| = |\frac{-7}{12} + \frac{5}{12}| = |\frac{-2}{12}| =$
$= \frac{1}{6} < \varepsilon$ STOP
Donc

$$f(\hat{x}) = \min_{i=1..3} f(x_i)$$

$$= \min\{\frac{11}{12}, \frac{19}{24}, \frac{7}{12}\}$$

$$= \frac{7}{12} = f(x_3)$$

La solution optimale est $\hat{x} = x_3 = \frac{-7}{12}$

Conclusion

Dans ce mémoire nous avons exposé beaucoup de méthodes utilisées dans la résolution des programmes d'optimisation différentiable et non différentiable. Le cadre dit convexe de tels problèmes est celui où la fonction coût (fonction objectif) est convexe, et l'ensemble admissible est convexe.

Dans le premier chapitre, on a choisi d'aborder l'étude des ensemble convexe après celle des fonctions convexes .On a utilisé les problèmes d'optimisation convexe sans et avec contraintes et on a analysé les conditions d'existences et d'unicité, et les conditions d'optimalité, ensuite on a présenté les différentes méthodes dans le cas différentiable (méthode de gradient, gradient conjuguais, Newton, Quasi Newton) et le cas non différentiable (Méthode de sous gradient, méthode de faisceaux) de résolution pour les problèmes différentiable et non différentiables .

Dans l'étude numérique on a abordé la méthode gradient, puis Newton puis la suite la méthode quasi Newton, d'autre part les méthodes on a étudiées pour l'optimisation non différentiable pourront bénéficier aux fonctions différentiable dont le gradient est mal conditionné .

Enfin, on a fini notre mémoire par des exemples concernant les méthodes citées en haut.Comme perspectives, on peut utiliser ces méthodes locales pour la recherche de l'optimum globale .

Bibliographie

[1]=[2] Michel BIERLAIR intropduction à l'optimisation différentiable ,presse polytechnique et universitaire romande 2006

[3] GS, GERAD en collaboration avec le département des mathematique et genie industrielle école polytechnique de montréal, introduction aux methodes des points interieurs, fevrier 2001.

[4] H SOULMANI ,méthode se descente en optimisation numérique .Mémoire de fin d'étude de licences univiresité sidi mohamed ben Abdellah 2016

[5] M. BERGOUNIOUX, optimisation sans contraintes, département de mathématique orleans 2003/2004

[6] Assia GUAGUI méthode d'optimisation non differentiable memoire de fin d'étude, master université Mohamed Khider Biskra 2019

[7] B Bolyak introduction to optimization software in 1987

[8] JT CHERKASKY 1979 opt pil limit analysis some observation on its use proceeding of the 16th APCOM symposium. 625 634.

[9] J.E Kelly Jn the cutting plane method for solving convexe programs, journal of the society for industrial(PP 703-712) 1960

[10] Amira SAYED technique d'optimisation, déterministe pour la résolution de problème Multidimensionnelles mémoire fin d'étude université LARBI BEN M'HIDI OUM EL BOUAGHI 2018

[11] Djamel ACHOUR , optimisation local et introduction à l'optimisation global université AKLI MOHAND OULHADJ de Bouira 2017/2018

Résumé

Nous avons présenté dans ce mémoire quelques rappels sur les notions principales de base des matrices , des ensembles convexes et des fonctions convexes .
Ensuite nous avons essayé d'exposer et de présenter dans ce mémoire plusieurs de méthodes et algorithmes utile dans la résolutions des programmes d'optimisation différentiable et non différentiable .
A la fin de mémoire, nous avons appliqué ces méthodes sur un exemple pratique

Abstract
We have presented in this thesis some reminders on the main basic notions of matrice , convex sets and convex functions .
Then we have tried to expose and present in this thesis many useful mothods and algorithms in the resolution of differentiable and non-differentiable optimization programs .
At the end of memory , we applied these methodes on a practical exemple