

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

**UNIVERSITÉ MOHAMED KHIDER, BISKRA**

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

**DÉPARTEMENT DE MATHÉMATIQUES**



Mémoire présenté en vue de l'obtention du Diplôme :

**MASTER en Mathématiques**

Option : **Analyse**

Par

**Hana Belkahla**

Titre :

# Optimisation non différentiable méthode de faisceaux

Membres du Comité d'Examen :

Dr. **Fatma Kaci**      UMKB      Encadreur

Dr. Laadjal Baya      UMKB      Examineur

Dr. Adouane Saida      UMKB      Examineur

Septembre 2020

## DÉDICACE

Je dédie ce travail à :

Ma chère maman pour toute sa tendresse, amour et affection

Mes chers frères et mes sœurs

Toute ma famille et mes amies

Tous mes formateurs pour leurs amabilités et tous qui m'ont aidé de pré ou de loin pour la réalisation de ce travail.

## REMERCIEMENTS

Je tiens tout d'abord à remercier "Dieu" le tout puissant et miséricordieux, qui m'a donné la force et la patience d'accomplir ce travail.

Avec une profonde reconnaissance et considération particulière que je remercie mon encadreur "**Kaci Fatma**" pour la sollicitude avec laquelle elle a suivi et guidé ce travail.

Un grand merci également pour les membres du jury.

Je présente tous mes remerciements aux enseignants du département, a ceux qui m'ont soutenue et m'ont aidé tout au le long de cette étude, ainsi à toutes les personnes qui ont contribué directement ou indirectement à ce travail.

# Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Liste des figures	v
Introduction	1
<b>1 Motivations et notations principales</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Motivations et notations principales . . . . .	3
1.2.1 Eléments d'analyse convexe . . . . .	3
1.2.2 Problème d'optimisation . . . . .	5
1.2.3 Rappels de calcul différentiel . . . . .	6
1.2.4 Existence et unicité d'un point de minimum . . . . .	8
1.2.5 Sous-gradient et Sous-différentiel . . . . .	12
1.2.6 Condition d'optimalité . . . . .	14
<b>2 Méthode de faisceaux</b>	<b>16</b>
2.1 Intoduction . . . . .	16
2.2 Pourquoi des méthodes spéciales? . . . . .	16
2.3 Méthodes de descente . . . . .	17

2.3.1	<b>Modèle de descente</b>	18
2.4	Instabilité de la méthode de la plus profonde descente	18
2.5	Méthodes Black-box "boite noire"	20
2.5.1	Méthode de sous-gradient	20
2.5.2	Méthode de plans sécants	20
2.5.3	Quelques remarques sur l'algorithme	23
2.5.4	Stabilité de la méthode	25
2.6	Les méthodes des faisceaux	25
2.6.1	Stabilisation	25
2.6.2	Quelques exemples de problèmes stabilisés	27
2.6.3	Implémentation	28
	<b>Conclusion</b>	<b>29</b>
	<b>Bibliographie</b>	<b>31</b>
	<b>Annexe B : Abréviations et Notations</b>	<b>32</b>

# Table des figures

1.1	Ensemble convexe $S$ . . . . .	3
1.2	Ensemble non convexe $S'$ . . . . .	4
1.3	Fonction convexe . . . . .	4
1.4	Fonction strictement convexe . . . . .	4
1.5	Fonction concave . . . . .	5
1.6	Exemple de fonction coercive $f(x) = \ x\ _2$ . . . . .	8
1.7	Minimum et maximum locaux et globaux . . . . .	9
1.8	la fonction $x \rightarrow x^2$ présente un minimum global strict en 0 . . . . .	9
1.9	Exemple de sous-gradient d'une fonctionnelle . . . . .	12
1.10	$f(x)= x $ . . . . .	13
1.11	Direction du sous-gradient . . . . .	14
2.1	Directions de descente . . . . .	17
2.2	Fonction minimax . . . . .	19
2.3	Trajection du zigzag . . . . .	19
2.4	Black-box . . . . .	20
2.5	Exemple d'approximation tangentielle . . . . .	22
2.6	Itérations de plans sécants . . . . .	23
2.7	Itérations de plans de coupe près d'un minimiseur . . . . .	24

# Introduction

L'optimisation vient du latin optimum qui signifie le meilleur. L'optimisation est une branche des mathématiques, cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes qui consistent à minimiser ou maximiser une fonction sur un ensemble.

L'optimisation joue un rôle important en recherche opérationnelle (domaine à la frontière entre l'informatique, les mathématiques et l'économie), dans les mathématiques appliquées (fondamentales pour l'industrie et l'ingénierie), en analyse et en analyse numérique, en statistique pour l'estimation du maximum de vraisemblance d'une distribution, pour la recherche de stratégies dans le cadre de la théorie des jeux, ou encore en théorie du contrôle et de la commande.

On s'intéresse dans ce mémoire au problème suivant :

$$\min_{u \in U} f(u)$$

où  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$

Parmi les méthodes les plus utilisées pour résoudre des problèmes de ce type, on peut citer la méthode de faisceaux. Cette méthode est apparue pour résoudre les problèmes de minimisation d'une fonction non continuellement différentiable.

Le contenu de ce mémoire est divisé en deux chapitres :

Le premier chapitre est un chapitre de généralités qui permet de définir l'ensemble de notations et définitions que nous aurons besoin dans la suite.

Le deuxième chapitre est consacré aux méthodes d'optimisation non différentiable de type "Boîte noire" et particulièrement à la méthode de faisceaux.



# Chapitre 1

## Motivations et notations principales

### 1.1 Introduction

Dans ce chapitre, on rappelle les outils de l'optimisation convexe pour une fonction différentiable. [16],[10],[6]

### 1.2 Motivations et notations principales

#### 1.2.1 Eléments d'analyse convexe

- Un ensemble  $S$  de  $\mathbb{R}^n$  est **convexe** si et seulement si,

$$\forall (x, y) \in S^2 : \lambda x + (1 - \lambda)y \in S, \forall \lambda \in [0, 1]$$

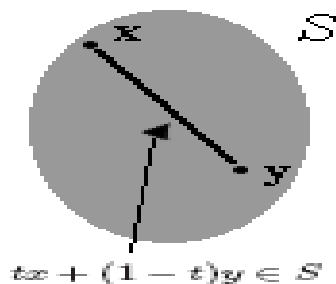


FIG. 1.1 – Ensemble convexe  $S$

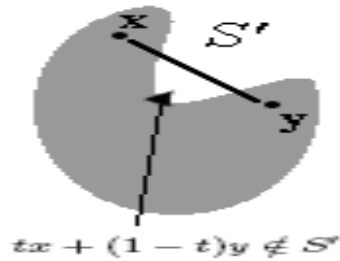


FIG. 1.2 – Ensemble non convexe  $S'$

- Une fonction  $f$  de  $\mathbb{R}^n$  dans  $\mathbb{R}$  est convexe si pour tout  $x, y \in \mathbb{R}^n$  et  $\lambda \in [0, 1]$  l'inégalité suivante est vérifiée :

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

Soit  $f$  dérivable sur un intervalle  $I$

- La fonction  $f$  est convexe sur  $I$  si, sur l'intervalle  $I$ , sa courbe représentative est entièrement située au-dessus de chacune de ses tangentes.

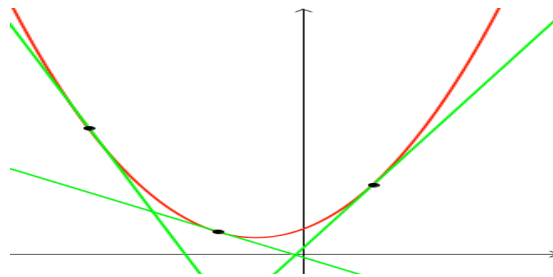


FIG. 1.3 – Fonction convexe

-  $f$  est dit strictement convexe si pour tout  $x, y \in \mathbb{R}^n$ ,  $x \neq y$ , et pour tout  $\lambda \in [0, 1]$ , on a

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

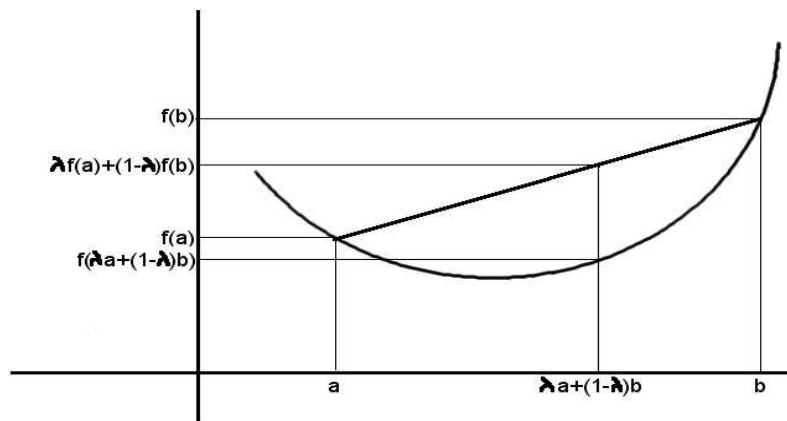


FIG. 1.4 – Fonction strictement convexe

- $f$  est dit **concave** si  $-f$  est une fonction convexe.
- $f$  est **concave** sur  $I$  ( $f$  dérivable sur  $I$ ) si, sur l'intervalle  $I$ , sa courbe représentative est entièrement située en dessous de chacune de ses tangentes.

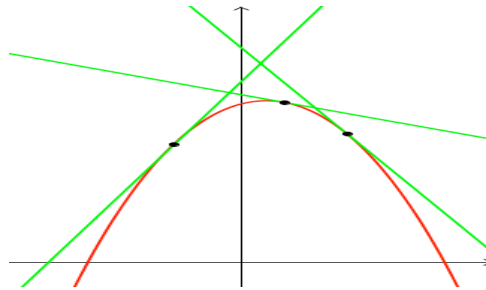


FIG. 1.5 – Fonction concave

## 1.2.2 Problème d'optimisation

Soit  $U \subset \mathbb{R}^n$ , et  $f : U \rightarrow \mathbb{R}$  une fonction de  $U$  dans  $\mathbb{R}$ .

Un **problème d'optimisation** est tout problème de la forme :

Trouver  $x^*$  telque :

$$x^* \in U \quad \text{et} \quad f(x^*) = \min_{x \in U} f(x) \quad (1,1)$$

- Si  $U$  et  $f$  sont convexes alors le problème est convexe.

L'optimisation intervient dans de nombreux domaines :

- En recherche opérationnelle (problème de transport, économie, gestion de stocks...).
- En analyse numérique (approximation, résolution de systèmes linéaires, non linéaires...).
- En automatique (modélisation de systèmes, filtrage...).
- En ingénierie (dimensionnement de structures, conception optimale de systèmes (réseaux, ordinateurs...)).

### 1.2.3 Rappels de calcul différentiel

#### Dérivée partielle

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue. la fonction notée  $\partial f(x) / \partial x_i$  est appelée  $i^{\text{ième}}$  dérivée partielle de  $f$  et est définie par

$$\frac{\partial f(x)}{\partial x_i} = \lim_{\alpha \rightarrow 0} \frac{f(x_1, \dots, x_i + \alpha, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{\alpha}$$

#### Gradient

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue . La fonction notée  $\nabla f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  est appelée le gradient de  $f$  et est définie par

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

Par exemple, si  $f(x_1, x_2, x_3) = e^{x_1} + x_1^2 x_3 - x_1 x_2 x_3$ , le gradient de  $f$  est donné par

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} e^{x_1} + 2x_1 x_3 - x_2 x_3 \\ -x_1 x_3 \\ x_1^2 - x_1 x_2 \end{pmatrix}$$

#### Dérivée directionnelle

La dérivée directionnelle de  $f$  en  $x$  dans la direction  $d$  est donnée par :

$$\lim_{\alpha \searrow 0} \frac{f(x + \alpha d) - f(x)}{\alpha}$$

La dérivée directionnelle est le produit scalaire entre le gradient de  $f$  et la direction  $d \in \mathbb{R}^n$ , c'est-à-dire  $\nabla f(x)^T d$ .

### Fonction différentiable

Soit  $x_0 \in U$ . On dit que  $J$  est différentiable en  $x_0$  s'il existe une application linéaire  $L : \mathbb{R}^n \rightarrow \mathbb{R}$  tel que au voisinage de  $x_0$  l'on ait

$$J(x_0 + h) - J(x_0) = L(h) + o(|h|)$$

on la note par  $DJ(x_0)$ .

-Si la dérivée directionnelle de  $f$  dans la direction  $d$  existe, alors la fonction  $f$  est dite différentiable.

### Matrice hessienne

La matrice hessienne de  $f$  en  $x$  est définie par

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

La matrice hessienne est toujours symétrique.

Par exemple, si  $f(x_1, x_2, x_3) = e^{x_1} + x_1^2 x_3 - x_1 x_2 x_3$ , le hessien de  $f$  est donné par

$$\nabla^2 f(x_1, x_2, x_3) = \begin{pmatrix} e^{x_1} + 2x_3 & -x_3 & 2x_1 - x_2 \\ -x_3 & 0 & -x_1 \\ 2x_1 - x_2 & -x_1 & 0 \end{pmatrix}$$

### Fonction de classe $C^1$

Soit  $U$  un ouvert de  $\mathbb{R}^n$ ,  $J : U \rightarrow \mathbb{R}$ . Si toutes les dérivées partielles de  $J$  existent et continues sur  $U$  on dit que  $J$  est de classe  $C^1$  sur  $U$ , et on écrit  $J \in C^1(U)$ .

Soit  $U$  un ouvert de  $\mathbb{R}^n$ ,  $J : U \rightarrow \mathbb{R}$  et  $x_0 \in U$ . Si  $J$  est différentiable en  $x_0$  alors :

- a)  $J$  est continue en  $x_0$

b)  $J$  admet toutes les dérivées directionnelles en  $x_0$  et sa différentielle est donnée par :

$$DJ(x) : \mathbb{R}^n \rightarrow \mathbb{R}$$

$$: h = (h_1, \dots, h_n) \mapsto \partial x_1 J(x_1) h_1 + \dots + \partial x_n J(x_n) h_n = \nabla J(x_0) \cdot h$$

## 1.2.4 Existence et unicité d'un point de minimum

### Fonctions coercives

Soit  $\Omega$  un ensemble non borné et  $f : \Omega \rightarrow \mathbb{R}$ . On dit que  $f$  est coercive sur  $\Omega$  si on a

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

$$x \in \Omega, \|x\| \rightarrow +\infty$$

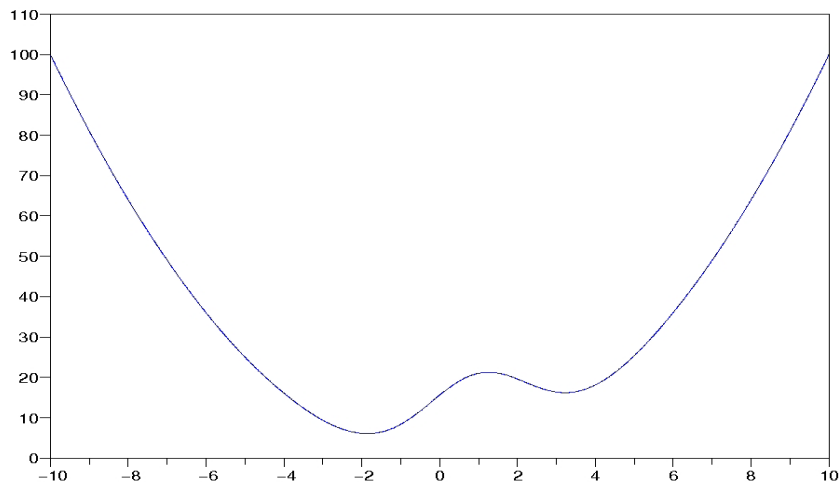


FIG. 1.6 – Exemple de fonction coercive  $f(x) = \|x\|_2$

### Les minimums locaux et globaux

Soit  $U \subset \mathbb{R}^n$ ,  $x^* \in U$  et  $f : U \rightarrow \mathbb{R}$ .

On dit que  $x^*$  est un point de minimum global de  $f$  sur  $U$  si

$$f(x) \geq f(x^*), \quad \forall x \in U$$

On dit que  $x^*$  est un point de minimum local de  $f$  sur  $U$  si  $\exists V$  voisinage de  $x^*$  dans  $\mathbb{R}^n$ , tel que

$$f(x) \geq f(x^*), \quad \forall x \in U \cap V$$

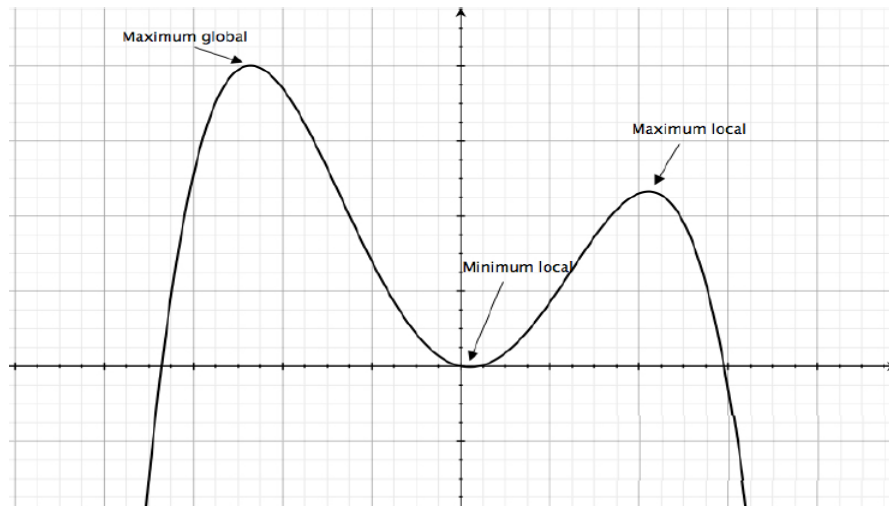


FIG. 1.7 – Minimum et maximum locaux et globaux

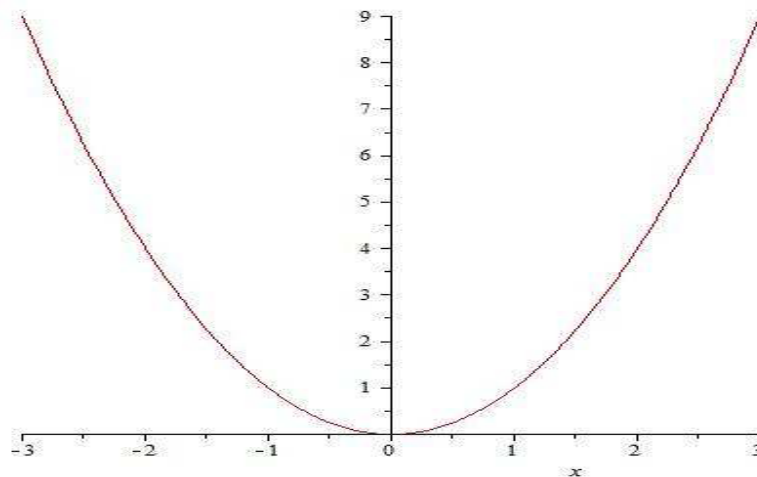


FIG. 1.8 – la fonction  $x \rightarrow x^2$  présente un minimum global strict en 0

**Théorème 1.1** ( *Théorème de Weierstrass* )

Soit  $f$  une fonction continue sur  $U \subset \mathbb{R}^n$  avec  $U$  fermé borné. Alors

$$\exists x_0 \in U \text{ tel que } f(x_0) = \min_{x \in U} f(x)$$

$\exists x_1 \in U$  tel que  $f(x_1) = \max_{x \in U} f(x)$

Cela signifie que  $f$  est bornée sur  $U$  ( $f(x_0) \leq f(x) \leq f(x_1)$ ) et atteint ses bornes. Le maximum et le minimum sont atteints dans ce cadre-là.

**Théorème 1.2** (Existence)

Soit  $U \in \mathbb{R}^n$  un ensemble non-vide et fermé et  $f : U \rightarrow \mathbb{R}$  une fonction continue. On suppose

-Soit  $U$  est borné

-Soit  $U$  est non borné et  $f$  est une fonction coercive.

Alors il existe au moins un point de minimum de  $f$  sur  $U$  (c'est à dire,  $\exists x^* \in U$  tel que  $f(x^*) \leq f(x), \forall x \in U$ )

**Preuve.** On distingue deux cas :

**Cas 1 :** L'ensemble  $U$  est borné.

Alors comme  $U$  est aussi fermé,  $U$  est compact. Comme  $f$  est continue, le Théorème de Weierstrass nous assure que  $f$  est bornée sur  $U$  et elle atteint ses bornes. Donc il existe au moins un point de minimum absolu de  $f$  sur  $U$ .

**Cas 2 :** L'ensemble  $U$  est non borné.

Soit  $a \in U$  et considérons l'ensemble

$$E = \{x \in U, f(x) \leq f(a)\} \quad \text{Remarque : } a \in E.$$

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  $x_k \in E$  avec  $\|x_k\| \rightarrow +\infty$  pour  $k \rightarrow +\infty$ . Comme  $f$  est coercive, ceci entraîne  $f(x_k) \rightarrow +\infty$  ce qui est absurde, car  $f(x_k) \leq f(a), \forall k \in \mathbb{N}$ .)

On déduit alors que  $E$  est un ensemble compact dans  $\mathbb{R}^n$ . Du Théorème de Weierstrass,  $\exists u^* \in E$  tel que

$$f(x^*) \leq f(x), \quad \forall x \in E$$



Mais d'autre part, on a

$$f(x^*) \prec f(x), \quad \forall x \in U - E$$

(car  $f(x^*) \leq f(x) \prec f(x)$ ,  $\forall x \in U - E$ )

Ceci prouve que  $x^*$  est un point de minimum absolu de  $f$  sur  $U$ , ce qui finit la preuve.

**Théorème 1.3** (*Unicité*)

Soit  $U \subset \mathbb{R}^n$  un ensemble convexe et  $f : U \rightarrow \mathbb{R}$  une fonction strictement convexe. Alors il existe au plus un point de minimum de  $f$  sur  $U$ .

**Preuve.** On va raisonner par absurd. Soient  $x_1, x_2 \in U$  avec  $x_1 \neq x_2$  deux points de minimum de  $f$  sur  $U$ . Nous avons donc :

$$f(x_1) = f(x_2) \leq f(x), \quad \forall x \in U. \quad (*)$$

Comme  $f$  est strictement convexe, on a

$$f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) \prec \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) = f(x_1)$$

(car  $f(x_1) = f(x_2)$ ) et ceci contredit (\*).

**Remarques sur l'existence et l'unicité** La question de l'existence et de l'unicité d'un minimiseur  $x^*$ , pour le problème (1, 1) est délicate. On peut retenir comme principe général que la compacité fournit des résultats d'existence, et la convexité un cadre favorable pour l'unicité.

**Théorème 1.4** Si  $f$  est continue et  $A$  est compact, alors le problème (1, 1) admet au moins une solution.

**Preuve.** On considère  $(x_n)$  une suite minimisante, i.e.  $x_n \in A$  et  $f(x_n)$  converge vers la borne inférieure  $\inf_A f$  de  $f$  sur  $A$  (éventuellement égale à  $-\infty$ ). Par compacité, on peut supposer que  $(x_n)$  est convergente, quitte à extraire; notons  $x^*$  la limite. Par continuité de  $f$ , il vient que  $\inf_A f = f(x^*) = \min_A f$ .

**Théorème 1.5** Si  $f$  est continue et coercive, i.e. infinie à l'infini :

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

alors le problème (1,1) admet au moins une solution.

**Preuve.** Soit  $x_0 \in A$  et  $m = f(x_0)$ . Par coercivité de  $f$ , il existe une boule  $B$  centrée en  $x_0$  en dehors de laquelle  $f$  est strictement supérieure à  $m$ . Ainsi les bornes inférieures de  $f$  sur  $A$  et sur  $A \cap B$  sont les memes. Le résultat précédent appliqué à  $A \cap B$ , compact, permet de conclure.

**Théorème 1.6** *Si  $A$  est convexe et  $f$  strictement convexe, alors le problème (1,1) admet au plus une solution.*

**Preuve.** Supposons que  $x_1^*$  et  $x_2^*$  soient deux solutions distinctes, alors  $x^* = (x_1^* + x_2^*)/2 \in A$  et par stricte convexité de  $f$ ,

$$f(x^*) < \frac{f(x_1^*) + f(x_2^*)}{2} = \min_A f$$

qui fournit une contradiction.

### 1.2.5 Sous-gradient et Sous-différentiel

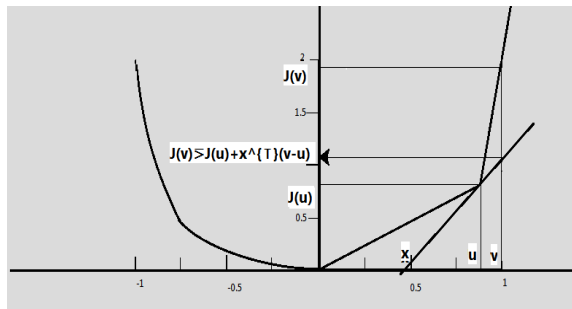


FIG. 1.9 – Exemple de sous-gradient d'une fonctionnelle

**Définition 1.1** Soit  $f : U \rightarrow \mathbb{R}$  une fonction convexe. On dit que le vecteur  $v \in \mathbb{R}^n$  est sous-gradient de  $f$  au point  $x^* \in U$  si

$$\forall x \in U, f(x) \geq f(x^*) + \langle v, x - x^* \rangle$$

L'ensemble de tous les sous-gradients en  $x^*$  est appelé sous-différentiel de  $f$  et noté par  $\partial f(x^*)$ .

**Exemple 1.1** Soit  $f : x \in \mathbb{R} \mapsto |x|$ . Calculons les sous-gradients de  $f$  en tout point  $x$  de  $\mathbb{R}$ . Dans  $\mathbb{R}^2$ , les hyperplans d'appui sont des droites et les sous-gradients associés leurs coefficients directeurs.

$$\partial f(x) = \begin{cases} \{-1\} & \text{si } x < 0 \\ \{1\} & \text{si } x > 0 \\ [-1, 1] & \text{si } x = 0 \end{cases}$$

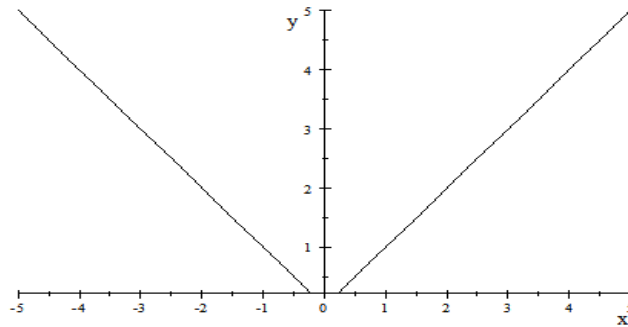


FIG. 1.10 –  $f(x)=|x|$

### Règles de calcul sous-différentiel

**Lemme 1.1** Soit  $f$  une fonction convexe. on a :

i) Si  $f$  est différentiable sur son domaine, alors :

$$\forall x \in \text{int}(\text{dom}(f)), \partial f(x) = \{\nabla f(x)\}.$$

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

$$\forall x \in \text{int}(\text{dom}(\varphi)), \quad \partial \varphi(x) = A^\top \partial f(Ax + b)$$

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

$$\begin{aligned} \partial f(x) &= \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x) \\ &= \{ \eta \in \mathbb{R}^n, \exists (x_1, x_2) \in \partial f_1(x) \times \partial f_2(x), \eta = x_1 + x_2 \}. \end{aligned}$$

iv) Soit  $g : \mathbb{R} \cup +\infty \rightarrow \mathbb{R} \cup +\infty$  une fonction convexe croissante. Notons :  $h = g \circ f$ . Alors  $\forall x \in \text{int}(\text{dom}(f))$ ,

$$\partial h(x) = \{ \eta_1 \eta_2 / \eta_1 \in \partial g(f(x)), \eta_2 \in \partial f(x) \}.$$

### Direction de descente

Les algorithmes utilisés pour résoudre des problèmes d'optimisation non-différentiable ne diffèrent pas beaucoup, au premier abord, des algorithmes standard. A l'itération  $l$ -ième, ils consistent schématiquement en la recherche d'une direction de descente  $d^l$  et d'une longueur de pas  $t_l \succ 0$ . Le nouvel itéré s'obtient ensuite par  $x^{l+1} = x^l + t_l d^l$

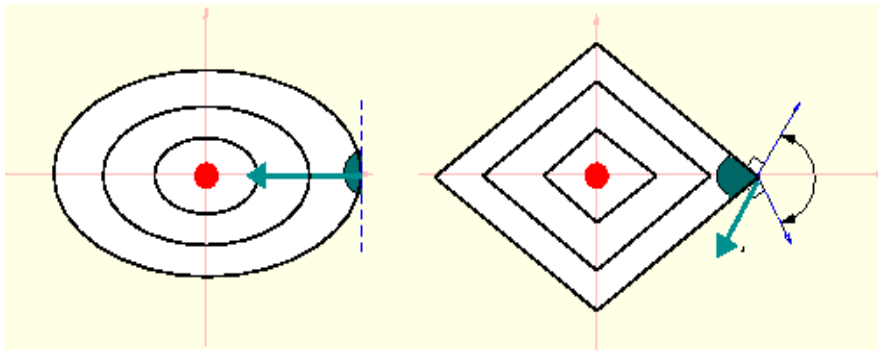


FIG. 1.11 – Direction du sous-gradient

### 1.2.6 Condition d'optimalité

La condition d'optimalité pour un point  $\bar{x}$  s'écrit à l'aide de l'une des trois propriétés équivalentes :

$$\left\{ \begin{array}{l} i) f(x) \geq f(\bar{x}) \quad \text{pour tout } x \in \mathbb{R}^N \\ ii) 0 \in \partial f(\bar{x}), \\ iii) f'(\bar{x}, d) \geq 0 \quad \text{pour tout } d \in \mathbb{R}^N \end{array} \right.$$

# Chapitre 2

## Méthode de faisceaux

### 2.1 Introduction

Dans ce chapitre, nous nous intéressons aux méthodes d'optimisation adaptées au cas où la fonction  $f$  à minimiser est convexe mais non différentiable.

Les méthodes de faisceaux sont considérées parmi les plus efficaces pour la résolution de problème non-différentiables [9],[11],[12].

### 2.2 Pourquoi des méthodes spéciales ?

L'optimisation différentiable présente des pièges dans lesquels un utilisateur non averti pourrait tomber :

**1. Piège du test d'arrêt** La condition " $\|g_k\| \leq \varepsilon, g_k \in \partial f(x)$ ", traduite directement de la condition " $\nabla f(x) \leq \varepsilon$ " dans le cas différentiable, peut ne jamais être vérifiée. Prenons par exemple :  $f(x) = |x|, x \in \mathbb{R}$ . Tant que  $x_k \neq 0$  (l'optimum global) :

$$\partial f(x_k) = \{1, -1\}.$$

**2. Piège des sous-gradients approchés** En pratique, le gradient (et même la fonction elle-même) n'est pas toujours calculé exactement : il est souvent obtenu par différences finies.

Cette approche n'est plus valable quand le sous-gradient n'est pas continu.

Ex :  $f(x) = \begin{cases} x^2 & \text{si } x \geq 0 \\ -x & \text{si } x \leq 0 \end{cases}$  Par différence finie autour de 0, On obtient :

$$\forall h \succ 0, \frac{f(h) - f(0)}{h} = h \succ 0,$$

alors que  $\partial f(x) = [-1, 0]$  . Donc  $h \notin \partial f(0)$ .

**3. Malédiction du sous-différentiable** L'application  $x \mapsto \partial f(x)$  n'étant pas continue, une petite variation de  $x_k$  peut entraîner de grandes variations de  $\partial f(x_k)$ .

Basé sur l'information disponible du sous-différentiel, le calcul d'une direction de recherche  $d_k$  peut varier énormément et produire des  $x_{k+1}$  très différents.

## 2.3 Méthodes de descente

La philosophie des méthodes dites de descente est de générer une séquence  $\{x_k\}$  telle que chaque itération garantit une diminution de  $f$ . Essentiellement, pour tous les non optimaux  $x_k$  ( $0 \notin \partial f(x_k)$ ), il existe une direction de descente  $d_k$ , qui correspond à la séparation stricte des ensembles  $\{0\}$  et  $\partial f(x_k)$ .

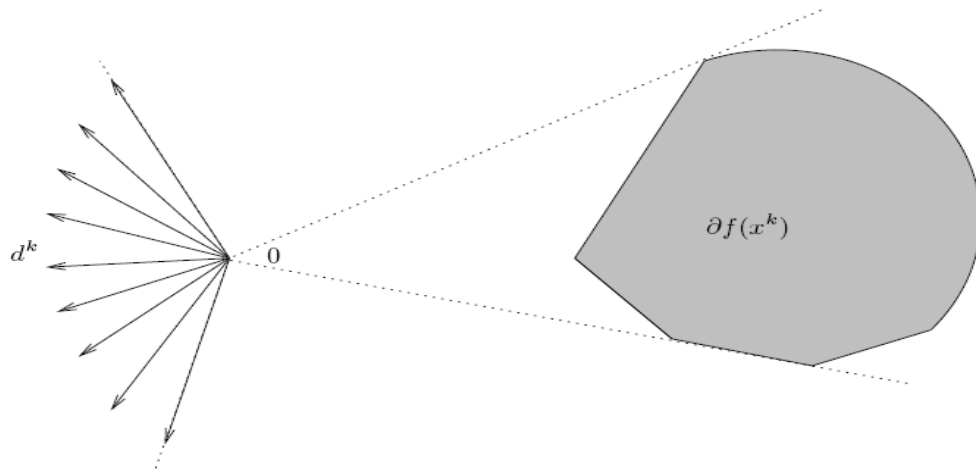


FIG. 2.1 – Directions de descente

**Shéma algorithmique** le schéma algorithmique de la plupart des méthodes de descente est le suivant :

### 2.3.1 Modèle de descente

prendre  $x^1 \in \mathbb{R}^n$  et fixe,  $k = 1$ .

**Étape 1** (test d'arrêt formel). Si  $0 \in \partial f(x^k)$ , arrêter.

**Étape 2** (descente). Trouvez une direction de descente  $d_k$  de  $f$  en  $x_k$ .

**Étape 3** (recherche en ligne). Trouvez une taille de pas  $t_k > 0$  telle que  $f(x^k + t_k d^k) < f(x^k)$ .

**Étape 4** (boucle). Définir  $x^{k+1} := x^k + t_k d^k$ . Changer  $k$  en  $k + 1$ , aller à 1.

## 2.4 Instabilité de la méthode de la plus profonde descente

Dans cette méthode, la direction de descente est just l'opposée d'un sous gradient particulier

$$d^k = -\frac{\gamma^k}{\|\gamma^k\|} \quad \text{où } \gamma^k \in \partial f(x^k)$$

Ici, l'algorithme souffre de deux désavantages ce qui les rend inefficaces :

1-Le calcul d'une direction de descente nécessite la connaissance de l'ensemble du sous-différentiable, et ceci à chaque itération.

2-La séquence  $\{x^k\}$  peut osciller et converger vers un point non optimal. Ce phénomène de zigzag est démontré dans l'exemple ci-dessous.

**Exemple 2.1 (4)** *On considère dans  $\mathbb{R}^2$  le problème min-max avec une fonction objective définie comme :*

$$f(x) = \max\{f_0(x), f_{-1}(x), f_{-2}(x), f_1(x), f_2(x)\},$$

où  $f_0(x) = -100$ ;  $f_{\pm 1}(x) = 3x_1 \pm 2x_2$ ;  $f_{\pm 2}(x) = 2x_1 \pm 5x_2$ .

La valeur optimale est  $\bar{f}(x) = -100$ , et l'ensemble (infini) de minimiseurs est

$$\{(x_1, x_2) \in \mathbb{R}^2 : x_1 \leq -50 \text{ et } |x_2| \geq 0.4x_1 + 20\},$$



C'est-à-dire, la région où  $f_0$  est active. la Figure (2.2) montre les régions où les fonctions  $f_j$  sont actives ; les lieux des plis , en gras dans la Figure, est composés des demi-lignes :

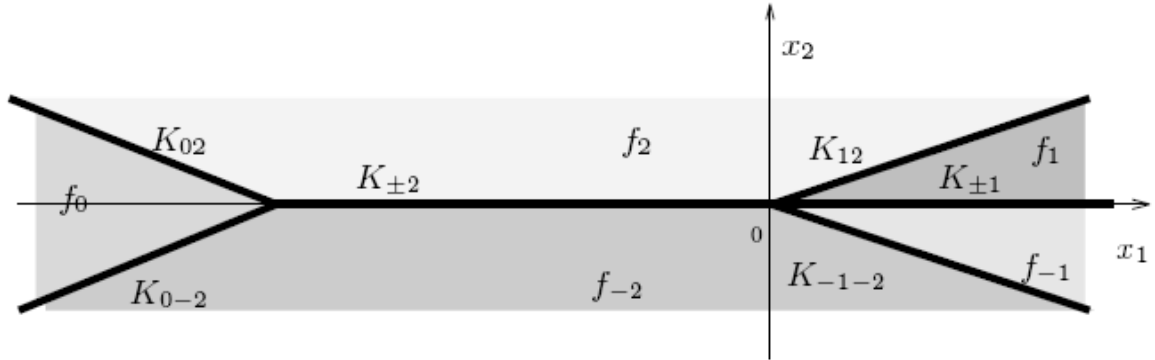


FIG. 2.2 – Fonction minimax

$$K_{02} := \{x : f(x) = f_0(x) = f_2(x)\} = \{(x_1, x_2) : x_2 = -20 - 0.4x_1\},$$

$$K_{0-2} := \{x : f(x) = f_0(x) = f_{-2}(x)\} = \{(x_1, x_2) : x_2 = 20 + 0.4x_1\},$$

$$K_{12} := \{x : f(x) = f_1(x) = f_2(x)\} = \{(x_1, x_2) : 3x_2 = x_1\},$$

$$K_{-1-2} := \{x : f(x) = f_{-1}(x) = f_{-2}(x)\} = \{(x_1, x_2) : -3x_2 = x_1\},$$

$$K_{\pm 1} := \{x : f(x) = f_1(x) = f_{-1}(x)\} = \{(x_1, x_2) : x_2 = 0, x_1 \geq 0\},$$

$$K_{\pm 2} := \{x : f(x) = f_2(x) = f_{-2}(x)\} = \{(x_1, x_2) : x_2 = 0, -50 \leq x_1 < 0\}.$$

L'application de la méthode de la plus profonde descente , nous donne une séquence  $\{x^k = (3^{3-k}, (-1)^k 3^{2-k})\}_{k \geq 1}$  qui converge très lentement vers 0 , qui est non-optimal et la Figure (2.3) nous montre le flagrant zigzag entre les deux droites  $K_{12}$  et  $K_{-1-2}$

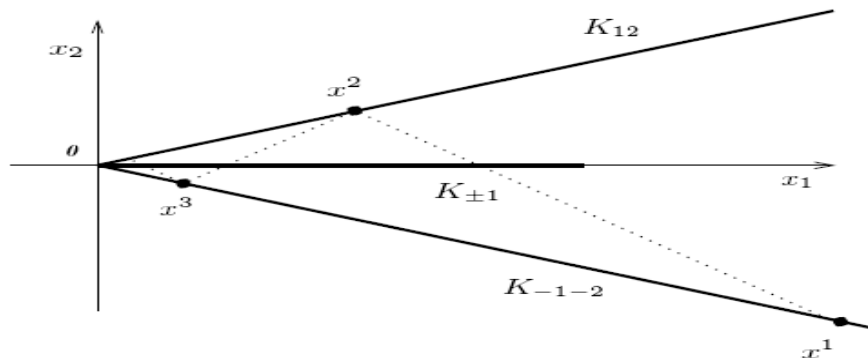


FIG. 2.3 – Trajection du zigzag

## 2.5 Méthodes Black-box "boite noire"

Les méthodes de boîte-noire sont des algorithmes définis sur la connaissance de la valeur de la fonction objective et just un élément de l'ensemble de sous-différentiable.



FIG. 2.4 – Black-box

La routine de boîte-noire affichée dans la Figure 2.4 est aussi parfois appelée oracle ou simulateur. Les algorithmes seront construits sur la base de cette seule information, donnée par l'utilisateur, et ils viseront à être des schémas de descente.

Cependant, en raison du peu d'information disponible, les algorithmes peuvent ne pas toujours être en mesure de produire des directions de descente.

Le terme "boîte noire" vient du fait que les méthodes d'optimisation génèrent une séquence de minimisation indépendamment de la façon dont les calculs sont organisés pour obtenir  $f(x)$  et  $s(x)$ .

### 2.5.1 Méthode de sous-gradient

Les méthodes de sous-gradients sont l'une des méthodes de "boites noires", elles sont une généralisation naturelle des méthodes de gradient au cas non différentiable, en remplaçant le gradient par un sous-gradient arbitraire [4],[18],[17].

### 2.5.2 Méthode de plans sécants

On considère dans cette section une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , convexe propre. On rappelle qu'un élément  $r$  du sous-différentiel de la fonction  $f$  au point  $x$  vérifié par définition :

$$f(y) \succeq f(x) + \langle r; y - x \rangle \quad \forall y \in \mathbb{R}^n$$

et donc que  $y \mapsto f(x) + \langle r; y - x \rangle$  est une minorante affine exacte au point  $x$  de la fonction  $f$ . Une telle minorante affine sera appelée coupe de la fonction  $f$  au point  $x$  et est caractérisée par le couple  $(x, r)$ .

### Approximation tangentielle

Soit  $\{x^i\}_{i \in [0, p]}$  un sous-ensemble discret de points de  $\mathbb{R}^n$ , tel que la fonction  $f$  soit sous-différentiable en chacun des points  $x^i$ . Pour tout  $i \in [0, p]$ , on suppose connus la valeur  $f(x^i)$  ainsi qu'un élément  $r^i$  dans le sous-différentiel  $\partial f(x^i)$ .

**Définition 2.1** *On appelle approximation tangentielle de  $f$  basée sur les points  $\{x^i\}_{i \in [0, p]}$  la fonction  $\hat{f}_p$  définie par :*

$$\hat{f}_p(x) = \max_{i \in [0, p]} \{f(x^i) + \langle r^i, x - x^i \rangle\}$$

On note  $C_p = \{(x^0, y^0), \dots, (x^p, y^p)\}$  l'ensemble des coupes constituant cette approximation.

Une approximation tangentielle  $\tilde{f}_p$  de  $f$  est donc l'enveloppe supérieure de  $p$  coupes de la fonction  $f$ . Elle vérifie les propriétés suivantes.

**Proposition 2.1 1.** *Soit  $p \in \mathbb{N}$ . La fonction  $\tilde{f}_p$  est convexe, vérifie pour tout  $x \in \mathbb{R}^n$  :*

$$\tilde{f}_p(x) \leq f(x) \quad \text{et donc} \quad \inf_{x \in \mathbb{R}^n} \tilde{f}_p(x) \preceq \inf_{x \in \mathbb{R}^n} f(x),$$

on a :

$$\tilde{f}_p(x^i) = f(x^i) \quad \forall i \in [0, p].$$

**2.** *Soient  $p_1$  et  $p_2$  tels que  $0 \leq p_1 \leq p_2$ , avec  $C_{p_1} \in C_{p_2}$ . Alors, pour tout  $x \in \mathbb{R}^n$ , on a :*

$$\tilde{f}_{p_1}(x) \leq \tilde{f}_{p_2}(x) \quad \text{et donc} \quad \inf_{x \in \mathbb{R}^n} \tilde{f}_{p_1}(x) \leq \inf_{x \in \mathbb{R}^n} \tilde{f}_{p_2}(x).$$

**Preuve.** La preuve est une conséquence immédiate des définitions de l'approximation tangentielle et du sous-différentiel.

**Commentaires.**

1. L'approximation tangentielle permet d'approcher «par en-dessous» une fonction  $f$  convexe à l'aide d'une fonction linéaire par morceaux.
2. Plus on augmente le nombre  $p$  de coupes constituant l'approximation tangentielle, et plus la fonction  $\tilde{f}_p$  s'approche de la fonction  $f$ .

La figure suivante illustre la notion d'approximation tangentielle pour une fonction  $V$ .

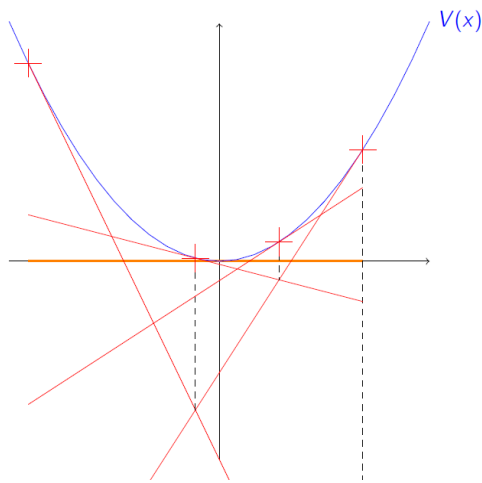


FIG. 2.5 – Exemple d'approximation tangentielle

Les méthodes des plans sécants, utilisent les valeurs

$$f_i = f(x^i) \text{ et } s^i = s(x^i), \text{ pour } i = 1, \dots, k$$

obtenues, pour construire pour  $\tilde{f}$  le modèle suivant à la pièce-affine pour  $f$  :

$$\tilde{f}_k(x) = \max_{i=1, \dots, k} \{f_i + \langle s^i, y - x^i \rangle\}.$$

La minimisation du modèle  $\tilde{f}_k$ , sur un ensemble compact convexe  $S$ , à déterminer avant de commencer la méthode, donne le nouvel itération  $x^{k+1}$ . La Figure 2.6 montre des plans sécants différents pour une fonction (lisse)  $f$ .

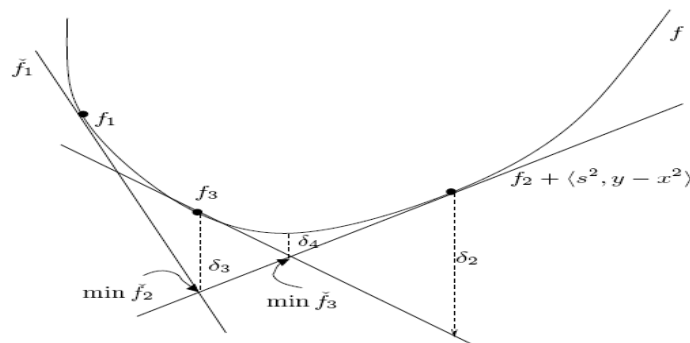


FIG. 2.6 – Itérations de plans sécants

**Algorithme 2.1** (*plans-sécants*). Soit  $\text{tol} \geq 0$  une tolérance d'arrêt donnée et soit  $S = \emptyset$  un ensemble convexe compact contenant un point minimum de  $f$ . Choisir  $x^1 \in S$  et fixer  $k = 1$ . Définir  $\tilde{f}_0 \equiv -\infty$ .

**Étape 1** (*Appel à la boîte-noire - test d'arrêt*). Appeler la boîte noire de la Figure 2.4 avec  $x = x^k$  Calculer

$$\delta_k = f(x^k) - \tilde{f}_{k-1}(x^k). \quad (1.1)$$

Si  $\delta_k \leq \text{tol}$ , stop

**Étape 2** (*Direction de descente du candidat*). Trouver

$$d^k \in \text{Arg min}_{x^k + d \in S} \tilde{f}_k(x^k + d)$$

**Étape 3** (*taille de pas*). Prendre  $t_k = 1$ .

**Étape 4** (*boucle*). Définir  $x^{k+1} = x^k + t_k d^k$ . Changer  $k$  en  $k + 1$ , aller en 1.

### 2.5.3 Quelques remarques sur l'algorithme

- Pour simplifier, L'étape 2 peut être remplacée par

$$x^{k+1} \in \text{Arg min}_{y \in S} \tilde{f}_k(y) \quad (1.2)$$

et l'étape 3 peut être supprimée

- Comme dans les méthodes de sous-gradient, la séquence générée par l'algorithme n'a pas nécessairement des valeurs d'objectif décroissantes  $f(x^k)$ . Un exemple simple de cette situation apparaît à la Figure 2.7, qui illustre un zoom de la Figure 2.6. Fondamentalement, plus un optimum est  $x^k$  ( $x^4$  dans la figure), le pire sera la prochaine itération. En conséquence, le minimiseur  $x^{k+1}$  du modèle est éloigné du minimum de  $f$  ( $x^5$  dans la Figure).

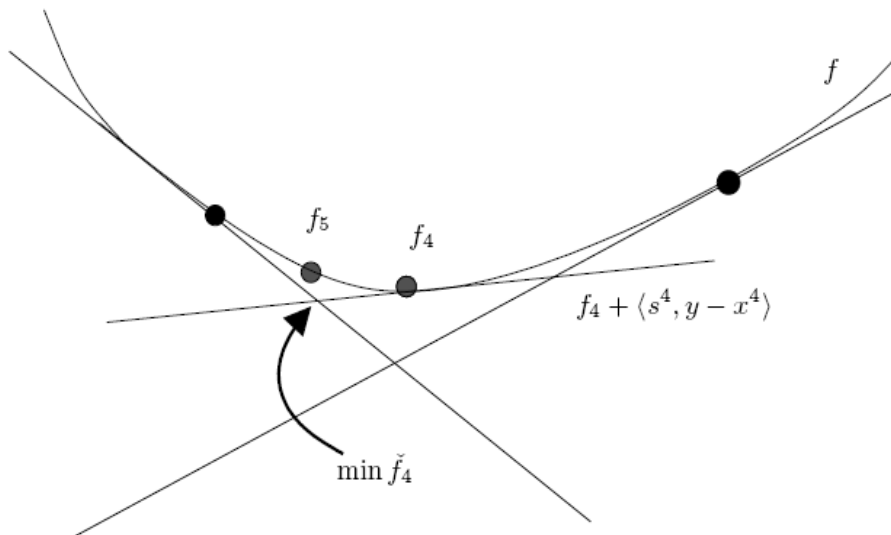


FIG. 2.7 – Itérations de plans de coupe près d'un minimiseur

Les méthodes des faisceaux, proposent des variantes pour forcer la diminution de  $f$ , tout en conservant l'idée du modèle, importante pour arrêter les tests.

**Théorème 2.1 (4)** *Considérons la séquence  $\{x^k\}$  générée par l'algorithme 2.1.*

(i) *Si l'algorithme boucle pour toujours ( $k \rightarrow \infty$ ), alors*

$$\lim_{k \rightarrow \infty} \tilde{f}_{k-1}(x^k) = \bar{f} = \lim_{k \rightarrow \infty} \bar{f}_k = \liminf_{k \rightarrow \infty} f(x^k).$$

*Par conséquent, l'algorithme 2.1 est garanti de s'arrêter à une certaine  $k_{last}$  d'itération chaque fois que  $tol > 0$ .*

(ii) *Si la séquence est finie, son dernier élément,  $x^{k_{last}}$  est optimal :*

$$f(x^{k_{last}}) \leq \bar{f} + tol.$$

## 2.5.4 Stabilité de la méthode

- Les méthodes des plans-sécants ont de bonnes propriétés de convergence pour les fonctions dont le graphe est en forme de V, ou des fonctions pointues. En particulier, pour les fonctions polyédriques.
- Pour les fonctions générales, cependant, l'algorithme des plans-sécants peut présenter des instabilités et un mauvais comportement numérique. Comme le montre graphiquement la Figure 2.5.

## 2.6 Les méthodes des faisceaux

Aucune des méthodes de la boîte noire considérées jusqu'à présent ne sont des schémas de descente. Seule la méthode la plus profonde descente garantit une diminution de la fonction objective à chaque itération. Cependant, il nécessite une connaissance complète de  $\partial f$ , et peut être piégé à des plis non optimaux.

Nous concentrons maintenant notre attention sur une famille de méthodes de boîte noire qui combinent les propriétés de descente et de stabilité, appelées méthodes des faisceaux, parce qu'elles gardent la mémoire des itérations passées dans un faisceau d'informations :

$$B = \{f_i, y^i, s^i, i = 1, \dots, k\} \text{ et } x^k, \text{ le point avec la meilleure valeur objective.}$$

### 2.6.1 Stabilisation

Afin de générer des points candidats et de sélectionner des centres de stabilité, nous définissons des sous-problèmes en modifiant (1.2) dans l'algorithme 2.1, en suivant certains principes de stabilisation déterminés par :

- (i) Le choix d'un modèle  $\varphi_k$ , qui se rapproche de  $f$  (par exemple,  $\varphi_k = \tilde{f}_k$ ).
- (ii) Le choix d'un centre de stabilité  $x^k$ , pour lequel  $f(x^k)$  est la "meilleure" valeur obtenue jusqu'ici et à partir de laquelle la diminution de  $f$  sera mesurée.
- (iii) Le choix de la normalisation  $|\cdot|_k$  pour éviter de grandes oscillations.

Nous utilisons ces éléments pour définir un sous-problème stabilisé, dont la solution  $y^{k+1}$  n'est considérée que comme un candidat à la diminution.

**Algorithme 2.2** (*des faisceaux*)

Soient  $tol \geq 0$  et  $m \in ]0, 1[$  deux paramètres. Choisir  $x^1$ , appelle la boîte noire de la Figure 2.4 avec  $x = x^1$ , construis le modèle  $\varphi_1$ , et définir les paramètres de l'algorithme, tels que  $|\cdot|_1$ . Définir  $k = 1$  et  $\delta_1 = \infty$ .

**Étape 1** (*test d'arrêt réalisable*). Si  $\delta_k \leq tol$ , arrêter.

**Étape 2** (*candidat*). Résoudre

$$y^{k+1} \in \text{Argmin} \quad \text{problème stable} \quad (\varphi_k, x^k, |\cdot|_k), \quad \text{et} \quad (1.3)$$

Définir  $\delta_{k+1} = \delta(\varphi_k, x^k, |\cdot|_k, y^{k+1}) \geq 0$ .

**Étape 3** (*appeler la boîte noire - évaluer le candidat*). Appeler la boîte noire de la figure 2.4 avec  $x = y^{k+1}$ .

**Test de descente :**

$$f(x^k) - f(y^{k+1}) \geq m\delta_{k+1} ? \begin{cases} \text{Oui} : x^{k+1} := y^{k+1} (\text{étape de descente}) \\ \text{Non} : x^{k+1} := x^k (\text{pas nul}) . \end{cases}$$

**Étape 4** (*amélioration de la boucle du modèle*). Annexer  $y^{k+1}$  au modèle, c.-à-d. Définir les paramètres de l'algorithme pour l'itération suivante, tels que  $|\cdot|_{k+1}$ . Passer  $k$  à  $k + 1$ , passer à l'étape 1.

L'ensemble des indices  $k$  pour lesquels une nouvelle descente-étape est effectuée est indiqué par  $K_s$ . En particulier, lorsqu'il y a dans la nuit beaucoup de marches de descente, la séquence in nite  $\{\delta_k\}_{k \in K_s}$  est convergente, chaque fois que (8.1) a des minimiseurs.

**Lemme 2.1** *Considérer l'algorithme 2.2 et on suppose que  $(k \rightarrow \infty)$ . on note  $f_* := \lim_{k \in K_s} f(x^k)$*



et on assume  $f_* \succ -\infty$ . alors

$$(0 \leq) \sum_{k \in K_s} \delta_k \leq \frac{f_1 - f_*}{m}.$$

Ce résultat simple est utile pour prouver la convergence chaque fois que l'algorithme génère un nombre infinte d'étape.

### 2.6.2 Quelques exemples de problèmes stabilisés

Les exemples suivants sont des formes particulières de sous-problèmes. Chaque variante est caractérisée par un paramètre, à mettre à jour, à chaque itération de l'étape 4. Nous n'allons pas expliquer ici comment choisir le modèle  $\varphi_k$  et la normalisation  $|\cdot|_k$ . Plus de détails voir le Chapitre 10 section 3 de [4].

**Exemple 1 :** (Région de confiance)

Ayant le paramètre  $k_k > 0$ . Le sous-problème stabilisé à résoudre à l'étape 2 est :

$$y^{k+1} \text{ solution à } \begin{cases} \min \varphi_k(y) \\ |y - x^k|_k^2 \leq k_k . \end{cases}$$

et  $\delta_{k+1} := f(x^k) - \varphi_k(y^{k+1})$ . La région de confiance est gérée de telle sorte que le paramètre  $k_k \rightarrow 0$  lorsque  $k \rightarrow \infty$ .

**Exemple 2 :** (Niveaux)

Avoir le paramètre  $l^k$ , le sous-problème (1.3) est le suivant :

$$y^{k+1} \text{ solution à } \begin{cases} \min \frac{1}{2} |y - x^k|_k^2 \\ \varphi_k(y) \leq l_k . \end{cases}$$

La diminution nominale  $\delta_{k+1}$  est la même que dans l'exemple précédent.

### 2.6.3 Implémentation

Le lien suivant est un lien de téléchargement de codes d'optimisation non différentiable par différentes méthodes

<http://napsu.karmita.fi/nssoftware/>

Par la méthode de faisceaux, on peut citer les codes suivants

- **PBNCGC** est une implémentation de la méthode de faisceaux la plus fréquemment utilisée [15].
- **PNEW** est un solveur de la méthode de faisceaux-Newton pour un problème sans contraintes et avec contraintes linéaire [13].
- **LMBM** est une implémentation d'une méthode de faisceaux spécifiquement développée pour les problèmes non différentiable à grande échelle [8] .

# Conclusion

Les problèmes d'optimisation non différentiable sont en général très difficiles à résoudre. Les méthodes de sous-gradients et les méthodes de faisceaux sont deux grandes familles de méthodes de résolution. En 2001, les méthodes de faisceaux sont considérées comme "l'outil le plus efficace et le plus prometteur pour l'optimisation non différentiable" [14], parce qu'elles combinent les propriétés de descente et de stabilité. Ces méthodes sont basées sur la méthode des plans sécants et cette dernière peut être très instable et avoir de mauvaises performances numériques (Figure 2.7).

# Bibliographie

- [1] Aude Rondepierre Introduction à l'optimisation convexe non différentiable  
[http://www.math.univ-toulouse.fr/~rondep/CoursTD/poly4GMM\\_nondiff.pdf](http://www.math.univ-toulouse.fr/~rondep/CoursTD/poly4GMM_nondiff.pdf)
- [2] Aude Rondepierre Méthodes numériques pour l'optimisation non linéaire déterministe.  
<http://www.math.univ-toulouse.fr/~rondep/CoursTD/polyGMM4.pdf>
- [3] Bierlaire, M. Introduction à l'optimisation différentiable. presses poly-Techniques, (2006).
- [4] BONNANS, Joseph-Frédéric, GILBERT, Jean Charles, LEMARÉCHAL, Claude, et al.  
Numerical optimization : theoretical and practical aspects. Springer Science & Business Media, (2006).
- [5] Ciuperca, S. Cours Optimisation, Université de Lyon1. <http://math.univ-lyon1.fr/~ciuperca/cours-optim-M1SAF.pdf>
- [6] Culioli, J.C . Introduction à l'optimisation. Ellipses Marketing, Science & Techniques (2012).
- [7] Emiel, G. Méthodes d'optimisation non différentiable pour la résolution de grands problèmes. Applications à la gestion à moyen-terme de la production. Doctoral dissertation , (2008).
- [8] Haarala, N., Miettinen, K., Mäkelä, M. M. : Globally convergent limited memory bundle method for large-scale nonsmooth optimization. Math. Program. 109, 181-205 (2007).
- [9] Hiriart-Urruty J.-B. and Lemaréchal C. Convex analysis and minimization algorithms. II, volume 306 of Grundlehren der Mathematischen Wissenschaften [Fundamental Prin-

- ciples of Mathematical Sciences]. Springer-Verlag. Advanced theory and bundle methods (1993).
- [10] Lagnoux, A. Chapitre 1 : Introduction à l'optimisation, Université de Toulouse (2013)  
<https://perso.math.univ-toulouse.fr/lagnoux/.les/2013/12/CoursOpt.pdf>
- [11] Lemaréchal. C. Chapter VII, Nondifferentiable optimization. In Optimization, volume 1 of Handbooks in Operations Research and Management Science, pages 529-572. Elsevier (1989).
- [12] Lemaréchal, C and Oustry, F. Nonsmooth Algorithms to Solve Semidefinite Programs. In L. El Ghaoui and S.-I. Niculescu, editors, Advances in Linear Matrix Inequality Methods in Control, chapter 3. SIAM, (2000).
- [13] Luksan, L., Viecek, J. : A bundle-Newton method for nonsmooth unconstrained minimization. Math. Program. 83, 373–391 (1998).
- [14] Mäkelä, M.M . Survey of bundle methods for nonsmooth optimization. Optim.Methods Softw.17(1) :1-29, (2002).
- [15] Mäkelä, M. M., Neittaanmäki, P : Nonsmooth Optimization : Analysis and Algorithms with Applications to Optimal Control. World Scientific Publishing Co., Singapore, (1992).
- [16] Mironescu, P. OPTIMISATION, Institut Camille Jordan, Lyon [http://math.univ-lyon1.fr/~mironescu/resources/optimisation\\_13\\_2013-2014\\_modelisation\\_fixe.pdf](http://math.univ-lyon1.fr/~mironescu/resources/optimisation_13_2013-2014_modelisation_fixe.pdf)
- [17] Polyak, B. Subgradient methods : A survey of soviet research. In Nonsmooth optimization : Proceedings of the IIASA workshop March, pages 5-30, (1977).
- [18] Shor, N. Z.. Minimization methods for nondifferentiable functions, volume 3 of Springer Series in Computational Mathematics. Springer-Verlag, Berlin, (1985). Translated from the Russian by K. C. Kiwiel and A. Ruszczyński.
- [19] Vial, G. Mini-cours d'optimisation. ENS Rennes (2008).  
<https://w3.ens-rennes.fr/math/people/gregory.vial/files/cplts/optimisation.pdf>

# Annexe B : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

$\| \cdot \|$  : Désigne la norme euclidienne dans  $\mathbb{R}^n$

$\partial f(x)$  : Sous-différentiel de la fonction  $f$  en  $x$ .

$\nabla f(x)$  : Le gradient de la fonction  $f$  en  $x$ .

$C^1$  : L'ensemble des fonction différentiable et les dirivées premiers continue.