

REPUBLIQUE ALGERIENNE DEMOCRATIQUE et POPULAIRE.  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

UNIVERSITE MOULOU D MAMMERE, TIZI-OUZOU  
Faculté des Sciences  
Département de Mathématiques



**Mémoire de Master**  
en  
**RECHERCHE OPÉRATIONNELLE**

Thème

**Optimisation mono-critère et multi-critère non linéaire**

Présenté par  
ISSAD Samia  
LASSACI Lilia

Encadré par  
OUKACHA Brahim (UMMTO)

Devant le jury d'examen composé de :

AIDENE Mohamed	Professeur	UMMTO	Président
OUKACHA Brahim	MCA	UMMTO	Rapporteur
LOUADJ Kahina	MCB	UMMTO	Examinatrice
TALEB Youcef	MAA	UMMTO	Examineur

Soutenu le 15 / 10 / 2012

## *Remerciements*

Nous remercions Dieu Clément et Miséricordieux, 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 notre travail à terme.

Nous tenons à exprimer notre gratitude à notre promoteur, que *M<sup>r</sup> Oukacha . B* reçoit l'expression de toute notre reconnaissance pour ses précieux conseils et orientations.

Nos plus sincères remerciements vont aux membres du jury de nous avoir honoré en acceptant d'évaluer notre travail.

Nous ne saurons oublier le grand mérite des enseignants qui ont contribué à notre cursus particulièrement ceux du département "mathématiques" et qu'ils trouvent ici le témoignage de notre profonde reconnaissance.

Nous remercions aussi tous nos camarades du département mathématiques en particulier nos amis de la promotion. On leur exprime notre profonde sympathie et leur souhaitent beaucoup de réussite.

Finalement, nous remercions toute personne ayant contribué de près ou de loin à l'élaboration de ce mémoire.

## *Dédicaces*

*A mes chers parents, que nulle dédicace ne puisse exprimer ce que je leurs dois, pour leur bienveillance, leur affection et leur soutien... Trésors de bonté, de générosité et de tendresse, en témoignage de mon profond amour et ma grande reconnaissance ” Que Dieu vous garde ”.*

*A mes chers frères M’Hamed et Boussad et mes très chères sœurs Dahbia, son mari Djaffar sans oublier ses adorables fils Mohamed et Adam, Saida, son mari Mokrane et Dalila.*

*A toute la famille Lassaci et Khemmar .*

*A mon très cher futur mari Nacer.*

*Je remercie très chaleureusement ma binôme et grande amie Samia pour cette splendide année de travail, ses grandes qualités tant sur le plan professionnel qu’humain, sa générosité, sa gentillesse, son soutien et sa patience.*

*A tous ceux qui m’aiment.*

*LILIA.*

## *Dédicaces*

*A mes chers parents, que nulle dédicace ne puisse exprimer ce que je leurs dois, pour leur bienveillance, leur affection et leur soutien... Trésors bonté, de générosité et de tendresse, en témoignage de mon profond amour et ma grande reconnaissance ” Que Dieu vous garde ”.*

*A mes chers frères Hocine, Hamou et le petit ange Lyes que je remercie pour leurs encouragements et leur soutien.*

*A mes très chères sœurs Siham et Mellissa pour tous leur amour et soutien.*

*A mes oncles et à mes tantes, leur mari et leurs enfants.*

*A la mémoire de mes grands pères. A mes grands mères que Dieu les gardes pour moi.*

*Je remercie très chaleureusement ma binôme et grande amie Lilia pour cette splendide année de travail, ses grandes qualités tant sur le plan professionnel qu’humain, sa générosité, sa gentillesse, son soutien et sa patience.*

*Je tiens également à remercier tous ceux qui m’ont encouragé et motivé tout au long de l’année: Ahcene, Samia, Gaya, Syria, Fatima, Mériem.*

*A tous ceux qui m’aiment.*

*SAMIA*

# Table des matières

<b>1</b>	<b>Rappels et quelques définitions:</b>	<b>5</b>
1.1	Convexité: . . . . .	5
1.1.1	Ensemble convexe: . . . . .	5
1.1.2	Combinaison convexe: . . . . .	6
1.1.3	problème convexe: . . . . .	8
1.2	Continuité: . . . . .	8
1.2.1	Fonction continue en un point: . . . . .	8
1.2.2	Fonction continue sur un intervalle: . . . . .	9
1.3	Dérivabilité: . . . . .	9
1.4	Les dérivées partielles: . . . . .	9
1.4.1	Dérivées partielles d'ordre 1 . . . . .	9
1.4.2	Dérivées partielles d'ordre supérieure à 1 . . . . .	10
1.5	Le gradient: . . . . .	10
1.6	La différentiabilité: . . . . .	10
<b>2</b>	<b>optimisation mono-critère non linéaire</b>	<b>11</b>
2.0.1	Différents types de problème d'optimisation: . . . . .	12
2.0.2	Définition des solutions: . . . . .	13
2.0.3	Conditions d'optimalités: . . . . .	13
2.1	Cas de problème sans contraintes . . . . .	19
2.1.1	La méthode de descente basée sur le gradient . . . . .	19
2.1.2	Méthode du gradient conjugué: . . . . .	21
2.1.3	Algorithme du gradient conjugué: . . . . .	22
2.1.4	Méthode de Newton: . . . . .	23
2.2	Étude comparative des méthodes d'optimisation sans contraintes . . . . .	25
2.3	Cas de problème avec contraintes . . . . .	28

2.3.1	Les méthodes de directions admissibles . . . . .	28
2.3.2	les méthodes de pénalité extérieure: . . . . .	36
2.4	Étude comparative des méthodes d'optimisation avec contraintes . . . . .	39
<b>3</b>	<b>Optimisation multicritère</b>	<b>42</b>
3.1	Position de problème: . . . . .	42
3.2	Solution d'un problème d'optimisation multi-critères: . . . . .	43
3.2.1	Optimalité de Pareto: . . . . .	43
3.2.2	Optimalité de Slater: . . . . .	44
3.2.3	Optimalité de Geoffrion: . . . . .	44
3.3	Conditions d'optimalité: . . . . .	45
3.4	Méthodes de résolution d'un problème d'optimisation multicritères . . . . .	47
3.4.1	Méthodes sans articulation du décideur: . . . . .	47
3.4.2	Méthodes avec articulation des préférences à posteriori: . . . . .	50
3.4.3	Méthodes avec articulation des préférences à priori: . . . . .	52
3.4.4	Méthode d'articulation progressive des préférences(Interactive): . . . . .	55

# Introduction

Les problèmes que rencontrent les gestionnaires des plus hauts niveaux jusqu'aux petites entreprises, se présentent sous forme de données, de contraintes, dont on doit tenir compte et d'un objectif ou plusieurs objectifs à atteindre.

On doit commencer par interpréter tous les paramètres et les transformer sous des formes qu'on peut gérer, en cherchant des approches mathématique pour les résoudre.

L'optimisation est un outil mathématique très puissant pour résoudre les problèmes d'aide à la décision. Au début, le modèle mono-critère a été utilisé de manière intense . Ce dernier revient à représenter le problème d'aide à la décision sous la forme d'un problème de minimisation d'une seule fonction objectif, vu que les problèmes réels de décision ont un caractère multi-critère et il est souvent difficile de construire une fonction unique qui puisse représenter fidèlement des préférences du décideur sur l'ensemble des décisions. Ainsi, dans les années 70, pour mieux représenter la réalité, l'approche multi-critère a été adoptée.

Dans les problèmes d'optimisation multi-critère, en général, les critères ou une partie de ces critères sont contradictoires, par conséquent dans ce type de problème, on parle de solution de compromis plutôt que de solution optimale car il est rare de trouver une décision qui satisfait tout les critères en même temps.

Les problèmes d'optimisation se présentent sous forme de données, de contraintes dont on doit tenir compte et un ou plusieurs objectifs à atteindre selon la nature de la fonction économique, on distingue deux classes de problèmes: les problèmes d'optimisation linéaire dont on dispose d'un arsenal de méthodes de résolution et programmation non linéaire.

La programmation non-linéaire regroupe un ensemble de sujets dans l'étude de problèmes

d'optimisation. Une fonction objectif, parfois nommée critère, est donnée, et le problème consiste à trouver (caractériser) et calculer un point minimisant (ou maximisant) cette fonction. Parfois, tous les points de  $\mathbb{R}^n$  sont candidats, parfois, des contraintes limitent le domaine de recherche.

Les fonctions utilisées (fonction objectif, contraintes) sont continues, même différentiables. On utilise des résultats d'analyse mathématique pour caractériser les points candidats ; un premier pas consiste donc à obtenir des conditions véritables satisfaites par les minima ou maxima recherchés. Lorsqu'un point ne satisfait pas à ces conditions d'optimalité, on en déduit une manière de calculer un point meilleur, et généralement un algorithme itératif réduisant (pour la minimisation) progressivement la fonction objectif.

Dans ce présent document, nous allons nous intéresser à l'étude de problèmes d'optimisation mono-critère et multi-critère non linéaire, pour ce faire, nous avons opté pour le plan du travail suivant:

- Le premier chapitre consiste en un rappels et quelques définitions sur la convexité.
- Le deuxième chapitre est consacré à l'optimisation mono-critère et méthodes de résolution.
- Le troisième chapitre présente l'optimisation multi-objectif et méthodes de résolutions.

En fin, nous terminons notre travail par une conclusion générale et la bibliographie utilisée.

# Chapitre 1

## Rappels et quelques définitions:

### 1.1 Convexité:

Pour étudier les problèmes d'optimisation, il est nécessaire de recourir à des outils scientifiques dont l'étude est basée sur l'analyse convexe. En effet, l'hypothèse de convexité va jouer un rôle très important pour la plupart des algorithmes que nous décrirons, la convergence vers l'optimum ne pourra être démontrée qu'avec cette hypothèse.

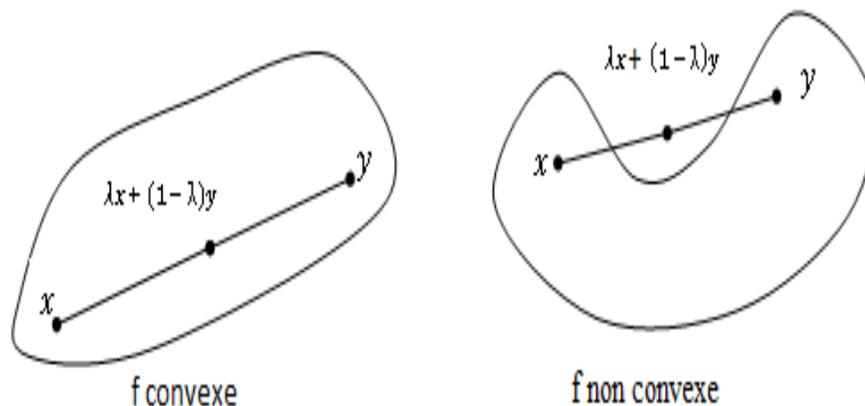
Nous allons ici rappeler quelques notions de convexité importantes auxquelles nous ferons appel par la suite, ainsi que quelques propriétés.

#### 1.1.1 Ensemble convexe:

Un ensemble  $A \subset \mathbb{R}^n$  est dit convexe si pour tous points  $x$  et  $y$  de  $A$ , le segment  $[x,y]$  est inclus dans  $A$ ,

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

Autrement dit, un ensemble convexe contient toujours le segment  $[x,y]$  joignant deux de ces points  $x$  et  $y$ , une interprétation "optique" consiste à dire dans une pièce convexe, deux personnes peuvent toujours s'apercevoir.

FIG. 1.1 – *Interprétation géométrique d'ensembles convexe et non convexe*– **Propriété des ensembles convexes:**

1. Soit  $X_1, X_2, \dots, X_m \subset \mathbb{R}^n$ ,  $m$  ensembles convexes alors on a  $Y = \bigcap_{i=1}^m X_i$  est un ensemble convexe.
2. La réunion d'ensembles convexes n'est pas forcément convexe.
3. Soit  $X_1, X_2, \dots, X_m \subset \mathbb{R}^n$ ,  $m$  ensembles convexes,  $\forall (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}_+^m$ ,  
 $Y = \sum_{i=1}^m \lambda_i X_i = \{ \sum_{i=1}^m \lambda_i x_i, x_i \in X_i, i = \overline{1, m} \}$

**1.1.2 Combinaison convexe:**

On appelle combinaison convexe de  $n$  points  $x_1, x_2, \dots, x_n$ , tout point obtenu par la formule :

$$Y = \sum_{i=1}^{i=n} \lambda_i x_i, \quad \text{avec} \quad \sum_{i=1}^{i=n} \lambda_i = 1$$

**Théorème 1.[10]:**

Un ensemble  $A$  est convexe si et seulement si toute combinaison convexe des points de  $A$  appartient à  $A$ .

**Enveloppe convexe:**

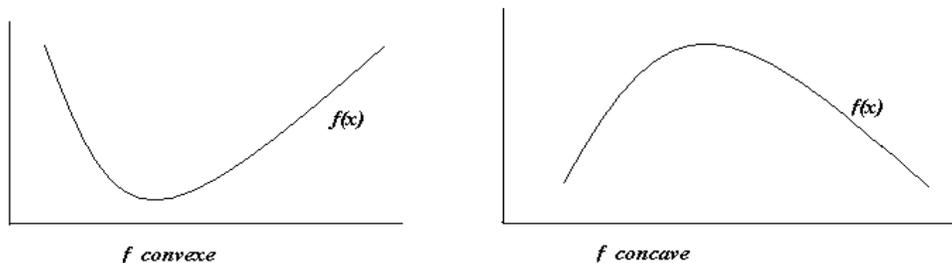
L'enveloppe convexe d'un sous ensemble  $A \subset X$  quelconque est le plus petit convexe contenant  $A$ , elle est notée **conv(A)**.

**Fonction convexe:**

On dit qu'une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est convexe, si elle vérifie  $\forall x \in \mathbb{R}^n, \forall \lambda \in [0,1]$ ,

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

\*  $f$  est concave si  $-f$  est convexe.

**– Propriété des fonctions convexes:**

1. Soit  $f : X \rightarrow \mathbb{R}^n$  avec  $X$  un ensemble convexe.

Si  $f$  est une fonction convexe, alors:

$$f\left(\sum_{i=1}^{i=n} \lambda_i x_i\right) \leq \sum_{i=1}^{i=n} \lambda_i f(x_i)$$

avec  $\lambda_i \geq 0, \sum_{i=1}^{i=n} \lambda_i = 1, x_i \in X, i = \overline{1, n}$ .

2. Soit  $f_1, f_2, \dots, f_n$   $n$  fonctions convexes définies sur un ensemble convexe  $X$ , alors:

$f(x) = \sum_{i=1}^{i=n} \lambda_i f_i(x)$  est une fonction convexe sur  $X$ , avec  $\lambda_i \geq 0, i = \overline{1, n}$ .

En d'autres termes, une combinaison linéaire à coefficients positifs de fonctions convexes est une fonction convexe.

### 1.1.3 problème convexe:

Soit (P) le problème défini par:

$$(P) \begin{cases} \text{minimiser } f(x) \\ x \in X \end{cases}$$

On dit que (P) est un problème convexe si  $X$  est convexe et  $f(x)$  est une fonction convexe sur  $X$ .

#### Propriétés:

1. Si (P) est convexe alors toute solution locale de ce problème est une solution globale, c'est-à-dire les notions d'optimums global et local coïncident.
2. Dans les problèmes convexes certaines conditions nécessaires d'optimalités deviennent des conditions suffisantes.

#### Remarque 1:

– La convexité d'une fonction est une caractéristique très importante en optimisation, en effet, lorsque la fonction n'est pas convexe, il est pratiquement impossible d'identifier un optimum global pour un problème d'optimisation. L'importance de la convexité est liée aux problèmes de minimisation .

Lorsqu'on étudie les problèmes de maximisation , on utilise la notion de concavité.

– Notons que les notions de convexité et concavité ne sont pas des propriétés complémentaires. Une fonction peut n'être ni convexe ni concave.

## 1.2 Continuité:

### 1.2.1 Fonction continue en un point:

#### Définition 1:

On dit qu'une fonction  $f$  est continue en  $x = a$ , si les trois conditions suivantes sont satisfaites :

- $f(a)$  existe dans  $\mathbb{R}$  .
- les limites à gauche et à droite existent dans  $\mathbb{R}$  et sont égales.

- les limites à gauche et à droite sont égales à  $f(a)$ .

Autrement dit:

$f$  est continue en  $a$  si  $\lim_{x \rightarrow a} f(x) = f(a)$ .

## 1.2.2 Fonction continue sur un intervalle:

### Définition 2:

On dit qu'une fonction est continue sur un intervalle si elle est continue en tout point de l'intervalle.

## 1.3 Dérivabilité:

### Définition 3:

Soient  $x_0 \in \Omega$  et  $f : \Omega \subset \mathbb{R} \rightarrow \mathbb{R}$ . On dit que  $f$  est dérivable en  $x_0$  si et seulement si:

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \text{ existe}$$

Dans ce cas, la limite est appelée dérivée de  $f$  en  $x_0$ , elle est notée  $f'(x_0)$ .

## 1.4 Les dérivées partielles:

### 1.4.1 Dérivées partielles d'ordre 1

#### Définition 4:

On définit la dérivée partielle par rapport à  $x_i$  de la fonction  $f$  au point  $X = (x_1, \dots, x_n)$  par la limite :

$$\lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_{i+h}, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_n)}{h}$$

quand elle existe et elle est notée:  $\frac{\partial f}{\partial x_i}(x)$

#### Remarque 2:

Pour calculer la dérivée partielle par rapport à  $x_i$  de  $f$  on dérive  $f$  par rapport à  $x_i$  en supposant les autres variables constantes.

### 1.4.2 Dérivées partielles d'ordre supérieure à 1

**Définition 5:**

Soit  $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$

Si les dérivées partielles de  $f$  existent et sont dérivables sur  $D$ , leurs dérivées partielles sont pour  $f$  les dérivées partielles secondes et on écrit:

dérivée mixte:

$$\frac{\partial}{\partial x_j} \left( \frac{\partial f}{\partial x_i} (x) \right) = \frac{\partial^2 f}{\partial x_i \partial x_j} (x)$$

dérivée secondes :

$$\frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_i} (x) \right) = \frac{\partial^2 f}{\partial x_i^2} (x)$$

## 1.5 Le gradient:

**Définition 5:**

On définit le gradient de  $f$  par:

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

## 1.6 La différentiabilité:

**Définition 6:**

On dit que  $f$  est différentiable en  $x^0$  si:

$$\lim_{x \rightarrow x^0} \frac{f(x) - f(x^0) - \nabla f(x^0)(x - x^0)}{\|x - x^0\|} = 0$$

**Remarque 3:**

Si  $f$  est différentiable en chaque point de  $D$ , on dit qu'elle est différentiable sur  $D$ .

**Définition 7:**

Si les dérivées partielles de  $f$  sont continues, on dit que  $f$  est continûment différentiable ou bien de classe  $C^1$

## Chapitre 2

# optimisation mono-critère non linéaire

### Introduction:

Dans cette partie, notre objectif est de présenter les définitions et les étapes nécessaires à la formulation et la résolution d'un problème d'optimisation non linéaire avec ou sans contraintes. Nous nous intéressons qu'au cas de la minimisation, ceci n'est pas restrictif dans la mesure où la recherche du maximum d'une fonction  $f$  se ramène immédiatement au problème de la minimisation de la fonction  $g = -f$ .

### Définition 1:

Le problème d'optimisation non linéaire est donné sous la forme suivante:

$$\begin{cases} \text{Minimiser } f(x) \\ \text{Sous les contraintes} \\ x \in X, \text{ avec, } X \subseteq \mathbb{R}^n \end{cases}$$

où  $f(x)$  est appelé fonction objectif ou fonction économique définie par:

$$\begin{aligned} f: X &\longrightarrow \mathbb{R} \\ x &\longrightarrow f(x) \end{aligned}$$

$X$  est appelé ensemble des solutions réalisable ou admissible, il est appelé aussi ensemble des décisions.

### 2.0.1 Différents types de problème d'optimisation:

Donnons quelques types de problèmes d'optimisation les plus étudiées:

#### 1. Problème d'optimisation avec contraintes:

On définit ce type de problème comme suit:

$$\begin{cases} \text{Minimiser } f(x) \\ \text{Sous les contraintes} \\ x \in X, \quad X \subseteq \mathbb{R}^n \text{ et } X \neq \mathbb{R}^n \end{cases}$$

##### Définition 2:

Un problème de programmation mathématique se formule de la manière suivante:

$$\begin{cases} \text{Minimiser } f(x) \\ \text{Sous les contraintes} \\ x \in X \text{ tels que } X = \{x \in \mathbb{R}^n, \text{ tels que } g_i(x) \leq 0, \text{ pour } i = 1, \dots, m\} \end{cases}$$

où  $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, \dots, m$ .

#### 2. Problème d'optimisation sans contraintes:

##### Définition 3:

Ce type de programme est défini comme suit:

$$\begin{cases} \text{Minimiser } f(x) \\ \text{avec } x \in \mathbb{R}^n \end{cases}$$

$f(x)$  peut être une fonction non linéaire.

#### 3. Problème quadratique:

##### Définition 4:

La fonction objectif à minimiser est une fonction quadratique de type:

$$F(x) = x^t A x + b^t x + c$$

avec

$A$ : une matrice définie positive

$b \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$  avec des contraintes linéaires, le problème s'écrit sous la forme

$$\begin{cases} \text{Minimiser } F(x) \\ \text{Sous les contraintes} \\ D * x \leq d \end{cases}$$

avec  $D$  une matrice  $m * n$ ,  $x$  et  $d$  sont deux vecteurs de  $\mathbb{R}^n$

### 2.0.2 Définition des solutions:

Dans un problème d'optimisation non linéaire on distingue deux types de solutions: le minimum local et le minimum global.

Soient l'ensemble  $S \in \mathbb{R}^n$  et une fonction  $f : S \rightarrow \mathbb{R}$ , les minima locaux et globaux de  $f$  sur  $S$  sont définis de la manière suivante:

- **Minimum local:** Intuitivement, un vecteur  $x^* \in S$  est un minimum local de  $f$  sur  $S$  s'il a un coût plus faible que celui de ses voisins. Formellement,  $x^*$  est un minimum local de  $f$  sur  $S$  si:

$$\exists \epsilon > 0 \text{ tels que: } f(x^*) \leq f(x); \quad \forall x \in S, \text{ avec } \|x - x^*\| < \epsilon.$$

où  $\| \cdot \|$  désigne la norme.

Le minimum local est stricte si:  $f(x^*) < f(x), \quad \forall x \in S, \text{ avec } \|x - x^*\| < \epsilon.$

- **Minimum global:** Un vecteur  $x^* \in S$  est un minimum global de  $f$  sur  $S$  s'il a un coût plus faible que celui de tous les autres vecteurs dans  $S$ . Formellement,  $x^*$  est un minimum global de  $f$  sur  $S$  si:

$$f(x^*) \leq f(x); \quad \forall x \in S.$$

### 2.0.3 Conditions d'optimalités:

Avant de développer des algorithmes permettant d'identifier des solutions d'un problème d'optimisation, il faut être capable de décider si un point donné est optimal ou non. On distingue deux types de conditions d'optimalité:

1. Les conditions nécessaire d'optimalité.
2. Les condition suffisantes d'optimalité.

#### L'utilité des conditions nécessaire et suffisante:

On s'intéresse à ces conditions pour les raisons suivantes:

- Elles permettent dans certains cas de calculer explicitement les solutions optimales.

- Elles constituent une base pour les méthodes qualitatives, c'est-à-dire les méthodes qu'étudient les propriétés des solutions du problème d'optimisation.
- Elles sont utilisées pour l'élaboration des méthodes numériques d'optimisation.

### 1. Cas d'un problème d'optimisation sans contraintes:

Considérons le problème d'optimisation sans contraintes suivant:

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

On cherche à résoudre (P); donc il s'agit de déterminer un point  $x^*$  de  $\mathbb{R}^n$  tels que:

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

C'est-à-dire un minimum global de  $f$  sur  $\mathbb{R}^n$ , mais pour beaucoup de problème d'optimisation sans contraintes, les principales méthodes de résolutions connues ne permettent pas la détermination d'un minimum global, il faut alors se contenter d'un minimum local, c'est-à-dire des points qui vérifient la relation (\*) seulement dans un voisinage de  $x^*$ .

Nous allons voir maintenant comment de tels points peuvent être caractérisés.

#### (a) Conditions nécessaire d'optimalité:

On suppose que  $f(x)$  est continue et à des dérivées partielles première  $\partial f / \partial x_i$  et seconde  $\partial^2 f / \partial x_i \partial x_j$  continues pour tout  $x \in \mathbb{R}^n$ .

#### ***Théorème 1.[17]: Conditions nécessaire d'optimalité***

*Une condition nécessaire pour que  $x^*$  soit un minimum local de  $f$  est:*

- a)  $\nabla f(x^*) = 0$ .
- b) *Si  $f$  est deux fois différentiable, alors:*  
 $\nabla^2 f(x^*) = [\partial^2 f / \partial x_i \partial x_j(x^*)]$  *est une matrice semi-définie positive.*

La condition a) est appelée condition nécessaire du premier ordre.

La condition b) est appelée condition nécessaire du second ordre.

#### **Remarque 1:**

En pratique, la condition nécessaire du second ordre est facile à vérifier systématiquement, car elle exige de calculer les dérivées secondes et d'analyser les valeurs propres

de la matrice hessienne. la condition nécessaire d'optimalité du premier ordre joue un rôle central en optimisation. les vecteurs  $x$  qui vérifient cette condition sont appelés des points critiques ou points stationnaires. Parmi eux, il y a des minima locaux, des maxima locaux et des points qui ne sont ni l'un ni l'autre. Ces derniers sont appelés points selle.

(b) **Conditions suffisantes d'optimalité:**

***Théorème 2.*** [17]: ***Condition suffisante d'optimalité local***

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction continue. Soit  $x^*$  un optimum local de  $f$ . Soit  $f$  est:

- i.  $\nabla f(x^*) = 0$ .
- ii. La hessienne  $\nabla^2 f(x^*)$  est une matrice définie positive.

***Théorème 3.*** [14]: ***Condition suffisante d'optimalité global***

Soit  $f$  une fonction continue  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  et  $x^* \in \mathbb{R}^n$  un minimum local de  $f$ .

- Si  $f$  est une fonction convexe, alors  $x^*$  est un minimum global de  $f$ .
- Si de plus  $f$  est strictement convexe,  $x^*$  est l'unique minimum global de  $f$ .

2. **Cas d'un problème d'optimisation avec contraintes:**

Nous allons maintenant discuter des conditions d'optimalité pour le problème de minimisation de  $f(x)$  sous contrainte  $x \in X \subset \mathbb{R}^n$ , où  $X$  est défini par une collection de  $m$  inégalités et de  $r$  égalités :

$$\begin{cases} \text{minimiser } f(x) \\ \text{sous contraintes } g_1(x) \leq 0, \dots, g_m(x) \leq 0 \\ h_1(x) = 0, \dots, h_r(x) = 0 \end{cases}$$

Donnons tout d'abord la définition de la notion de direction admissible, puis nous énoncerons un théorème qui établira une première condition nécessaire d'optimalité pour un tel problème.

**Définition 5: (Direction admissible)**

Soient un ensemble  $X \subset \mathbb{R}^n$ , avec  $X \neq \mathbb{R}^n$  et  $x^* \in X$ .

Une direction admissible de  $X$  en  $x^*$  est un vecteur  $d$  tel que:

$d \neq 0$  et  $x^* + \alpha d \in X$ ,  $\forall \alpha \in [0, \alpha_{max}]$  pour un certain  $\alpha_{max} > 0$ .

Les conditions d'optimalité énoncées plus bas s'obtiennent suite au constat suivant :  
Si  $x^*$  est un minimum local de  $f$  sur  $X$ , alors il ne peut y avoir aucune direction de descente dans l'ensemble des directions admissibles de  $X$  en  $x^*$ .

La condition telle que nous venons de l'énoncer ne présente malheureusement que peu d'intérêt en vue d'un usage en pratique, car si l'ensemble des directions de descente de  $f$  en  $x^*$ , disons  $D = \{d/\nabla f(x^*)d < 0\}$ , peut être exprimé en fonction du gradient de la fonction objectif, ce n'est pas nécessairement le cas de l'ensemble de toutes les directions admissibles de  $X$  en  $x^*$ ,

$A = \{d/d \neq 0 \text{ et } x^* + \alpha d \in X, \forall \alpha \in [0, \alpha_{max}]\}$  pour un certain  $\alpha_{max} > 0$ .

**La condition d'optimalité de Fritz John**

La condition donnée ci-dessus peut être également exprimée par les relations établies au théorème suivant, attribué à Fritz John (1948).

**Théorème 4. [5] : (condition nécessaire de Fritz John)**

Soient un ensemble  $X \subset \mathbb{R}^n$  avec  $X \neq \emptyset$ , une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  et un vecteur  $x^* \in X$ .

Nous admettons que  $X$  est défini par une collection de  $m$  inégalités  $g_i(x) \leq 0$  et de  $r$  égalités  $h_i(x) = 0$  avec:

$$g_i : \mathbb{R}^n \rightarrow \mathbb{R}, \forall i \in \{1, \dots, m\}$$

et

$$h_i : \mathbb{R}^n \rightarrow \mathbb{R}, \forall i \in \{1, \dots, r\}$$

Supposons, de plus, que  $f$ , les  $g_i$  et les  $h_i$  sont continuellement différentiables pour tout  $i$ .

Soit  $I$  l'ensemble des contraintes sous forme d'inégalités actives en

$$x^* : I = \{i/g_i(x^*) = 0\}$$

Si  $x^*$  est un minimum local, alors il existe des scalaires  $\mu_0, \mu_1, \dots, \mu_m$  et  $\lambda_1, \lambda_2, \dots, \lambda_r$  non tous égaux à zéro tels que :

$$\mu_0 \nabla f(x^*) + \sum_{i=1}^m \mu_i \nabla g_i(x^*) + \sum_{i=1}^r \lambda_i \nabla h_i(x^*) = 0$$

$$\mu_0, \mu_i \geq 0, \forall i \in I$$

**Remarque 2:**

Les scalaires  $\mu_i$  et  $\lambda_i$  sont appelés multiplicateurs de Lagrange et la fonction définie ci-dessus est parfois appelée fonction de Lagrange.

**La condition d'optimalité de Kuhn-Tucker**

Un inconvénient posé par la condition de Fritz John est le suivant : si  $\mu_0$  est égal à zéro, alors la condition ne comporte aucune information relative au gradient de la fonction objectif, ce qui signifie qu'elle ne fournit pas d'information relative à notre problème. Dans un tel cas, elle assure simplement qu'il existe une combinaison linéaire non triviale (et non négative si les contraintes sont toutes sous forme d'inégalités) des gradients des contraintes actives qui est égale au vecteur nul ; dans ce cas, elle ne nous est pas utile en pratique pour déterminer si un vecteur est optimal. Le cas où  $\mu_0$  est strictement supérieur à zéro est donc plus intéressant. Dans cette optique, des conditions ont été développées indépendamment par Kuhn et Tucker (en 1951), qui sont précisément les conditions de Fritz John, à qui une hypothèse y a été ajoutée impliquant que  $\mu_0$  ne puisse pas être égal à zéro.

Différentes hypothèses peuvent être posées sur les contraintes afin de garantir que  $\mu_0 > 0$  (de telles hypothèses sont appelées qualification des contraintes). Dans le théorème énoncé ci-dessous, l'on impose que les gradients des contraintes sous formes d'égalités et des contraintes sous forme d'inégalités actives au point considéré soient linéairement indépendants, ce qui signifie qu'il ne peut exister de combinaison linéaire non triviale de ceux-ci dont la somme vaut le vecteur nul. Sous cette hypothèse supplémentaire, la condition de Fritz John ne peut être satisfaite qu'avec  $\mu_0 \neq 0$ . Nous pouvons donc arbitrairement choisir  $\mu_0 = 1$  et énoncer le théorème suivant:

**Théorème 5** .[5] : *(condition nécessaire de Kuhn-Tucker:)*

Soient un ensemble  $X \subset \mathbb{R}^n$  avec  $X \neq \emptyset$ , une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  et un vecteur  $x^* \in X$ . Nous admettons que  $X$  est défini par une collection de  $m$  inégalités  $g_i(x) \leq 0$  et de  $r$  égalités  $h_i(x) = 0$  avec:

$$g_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad \forall i \in \{1, \dots, m\}$$

et

$$h_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad \forall i \in \{1, \dots, r\}.$$

Soit  $I$  l'ensemble des contraintes sous forme d'inégalités actives en

$x^* : I = \{i/g_i(x^*) = 0\}$ . Supposons, de plus, que  $f$ , les  $g_i$  et les  $h_i$  sont continuellement différentiables pour tout  $i$  et que  $\nabla g_i(x^*)$  pour  $i \in I = \{i/g_i(x^*) = 0\}$  et  $\nabla h_i(x^*)$  pour  $i \in \{1, \dots, r\}$  sont linéairement indépendants. Si  $x^*$  est un minimum local, alors il existe des scalaires  $\mu_1, \dots, \mu_m$  et  $\lambda_1, \dots, \lambda_r$  tels que :

$$\nabla f(x^*) + \sum_{i=1}^m \mu_i \nabla g_i(x^*) + \sum_{i=1}^r \lambda_i \nabla h_i(x^*) = 0$$

$$\mu_i \geq 0, \quad \forall i \in I$$

**Remarque 3:**

la condition nécessaire de Kuhn-Tucker est également suffisante si  $f$  et les  $g_i$  sont convexes.

# Méthodes de résolutions

## Introduction:

Parfois les conditions nécessaire et suffisante d'optimalité permettent de déterminer les solutions des problèmes d'optimisation. Mais dans la plus part des cas on est obligé de faire appelle aux méthodes numériques.

Dans ce chapitre, on s'intéresse à la description plus spécifique des algorithmes itératifs (ou méthodes itératives) qui permettent la résolution des problèmes d'optimisation non linéaire.

## 2.1 Cas de problème sans contraintes

### 2.1.1 La méthode de descente basée sur le gradient

Les méthodes basées sur le gradient de la fonction objectif sont des procédures parmi les plus fondamentales pour minimiser une fonction différentiable de  $\mathbb{R}^n$  dans  $\mathbb{R}$ . Comme la plupart des autres méthodes développées pour ce problème, elles reposent sur la propriété dite de descente itérative. Rappelons qu'un algorithme itératif part d'un vecteur  $x^0 \in \mathbb{R}^n$  et génère une suite de vecteur  $x^1, x^2, \dots$  de  $\mathbb{R}^n$ , la propriété de descente itérative impliquant que le coût des vecteurs ainsi générés décroisse à chaque itération :

$$f(x^{k+1}) < f(x^k), \quad \forall k \in \mathbb{N}$$

Et les vecteurs  $x^1, x^2, \dots \in \mathbb{R}^n$  sont générés de la manière suivantes:

$$x^{k+1} = x^k + \alpha^k d^k$$

Ainsi, pour assurer la propriété de descente itérative, la direction  $d^k$  choisie dans l'équation ci-dessus doit être une direction de descente. Les algorithmes de ce type sont appelés méthodes du gradient .

Certains auteurs réservent cette appellation au cas particulier où  $d^k = -\nabla f(x^k)$

Dans le présent document, l'appellation " Méthode du gradient" sera utilisée uniquement au cas où  $d$  est choisi ainsi et  $\alpha^k$  est déterminé suivant la politique dite de minimisation qui revient à calculer de façon à minimiser le problème unidimensionnelle  $q(\alpha) = f(x^k - \nabla f(x^k))$  tels que  $q$  est la semi-droite définie par le point  $x^k$  et la direction  $-\nabla f(x^k)$ .

C'est une propriété intéressante de la direction  $-\nabla f(x^k)$  qui nous conduit à ce choix : parmi toutes les directions  $d \in \mathbb{R}^n$  normalisées, il s'agit de celle qui minimise la dérivée directionnelle  $\nabla f(x^k).d$  de  $f$  en  $x^k$ . Ainsi, la direction que nous choisissons parmi toutes les directions  $d \in \mathbb{R}^n$  telles que  $\|d\| = 1$  est celle qui minimise la variation de  $f$  dans la direction  $d$  lorsque  $\alpha$  tend vers zéro.

Le problème de recherche de cette direction consiste à trouver la direction  $d$  qui minimise  $\nabla f(x).d$  sous la contrainte  $\|d\| = 1$ . La proposition ci-dessous stipule que la direction  $d = -\nabla f(x)/\|\nabla f(x)\|$  est la solution optimale de ce problème.

**Proposition:**

Soient  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  continuellement différentiable et  $x \in \mathbb{R}^n$ . Supposons que  $\nabla f(x) \neq 0$ . Alors le problème qui consiste à minimiser  $f'(x,d)$  sous la contrainte  $\|d\| = 1$  a pour solution optimale  $d^* = -\nabla f(x)/\|\nabla f(x)\|$

La direction de descente choisie sera donc, à chaque itération  $d^* = -\nabla f(x)/\|\nabla f(x)\|$ .

Les points sont ainsi successivement générés par la méthode du gradient de la manière suivante :

$$x^{k+1} = x^k - \alpha^k d^k, \quad \alpha^k > 0$$

Remarquons que la méthode s'arrête lorsque  $\nabla f(x^k) = 0$ , car dans ce cas  $x^{k+1} = x^k$ .

**Algorithme de la méthode:**

1. Initialisation  $k = 0$ :

On se donne une fonction  $f$  différentiable, un point initial  $x^0$  et un seuil de tolérance  $\varepsilon > 0$

2. Itération  $k = 1, 2, \dots$

poser  $d^* = -\nabla f(x)/\|\nabla f(x)\|$

Calculer :  $\alpha_k = \arg \min_{\alpha > 0} f(x^k + \alpha d^k)$

$$x^{k+1} = x^k + \alpha_k d^k$$

3. Critère d'arrêt:

Si:  $\|x^{k+1} - x^k\| < \varepsilon$  Stop

Si non on pose  $k = k + 1$  et aller à 2.

### Convergence de la méthode:

***Théorème 1.*** [4]: *(convergence de la méthode du gradient)*

*Soit  $\{x^k\}$  une séquence générée par la méthode du gradient. Alors tout point limite de  $\{x^k\}$  est un point stationnaire.*

### Définition 1: Point limite:

On appelle point limite d'une séquence  $\{x^k\}$  de point de  $\mathbb{R}^n$ , tout point  $x \in \mathbb{R}^n$  tels qu'il existe une sous séquence de  $\{x^k\}$  qui converge vers  $x$ .

Il peut arriver que la méthode du gradient converge de manière finie, mais ce n'est en général pas le cas. Il est donc nécessaire d'utiliser un critère permettant d'arrêter l'exécution lorsque  $\{x^k\}$  est suffisamment proche d'un point stationnaire, par exemple  $\|\nabla f(x^k)\| < \varepsilon$ , où  $\varepsilon$  est un scalaire positif arbitrairement choisi. A priori, la valeur que nous devons fixer pour  $\varepsilon$  dépend du problème considéré.

### 2.1.2 Méthode du gradient conjugué:

L'algorithme que nous présentons dans cette section possède deux intérêts. Il permet de résoudre des systèmes linéaires à coefficients strictement positif de grande taille et il sert d'algorithme de base pour la résolution des problèmes d'optimisation non linéaire.

Soit le problème de minimisation quadratique suivant:

$$f(x) = \frac{1}{2} x^t A x + b^t x + c, \text{ avec } x \in \mathbb{R}^n$$

où  $A$  est une matrice carrée d'ordre  $n$ , symétrique, définie positive et  $b \in \mathbb{R}^n$

On note:

$$\nabla f(x) = Ax + b$$

qui est non nul et  $\nabla^2 f(x) = A$

La première idée fondamentale de l'algorithme du gradient conjugué consiste à choisir chaque direction de descente conjugué à la direction de descente précédente par rapport à  $A$ . Afin de développer le gradient conjugué, introduisons la notion de direction conjugué.

**Définition 2:**

Soient  $A \in \mathbb{R}^n \times \mathbb{R}^n$  une matrice définie positive. Les vecteurs (ou directions) non nuls de  $\mathbb{R}^n$ ,  $d^1, \dots, d^k$  sont conjugués par rapport à  $A$  ( $A$  conjuguées) si:

$$(d^i)^t A d^j = 0, \quad \forall i, j \text{ tels que } : i \neq j$$

Ceci signifie que ces deux vecteurs sont orthogonaux pour le produit scalaire associé à la matrice  $A$  défini par:

$$\langle x, y \rangle_A = x^t A y; \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$$

**Théorème 2.** [6]:

Soit  $\{d^1, \dots, d^k\}$  un ensemble de directions non nulles et conjuguées par rapport à  $A$ . Alors les vecteurs  $d^1, \dots, d^k$  sont linéairement indépendants.

### 2.1.3 Algorithme du gradient conjugué:

1. Initialisation

fixer  $\varepsilon > 0$ , choisir  $x^0 \in \mathbb{R}^n$

Poser  $\nabla f(x^0) = Ax^0 + b$  et  $d^0 = -\nabla f(x^0)$

2. Itération  $k = 0, 1, \dots$

Calculer  $\alpha_k = -(d^k)^t \nabla f(x^k) / (d^k)^t A d^k$

Calculer  $x^{k+1} = x^k + \alpha_k d^k$

Calculer  $\nabla f(x^{k+1}) = \nabla f(x^k) + \alpha_k A d^k$

Calculer  $\beta^{k+1} = \|\nabla f(x^{k+1})\|^2 / \|\nabla f(x^k)\|^2$ .

Calculer  $d^{k+1} = -\nabla f(x^{k+1}) + \beta^{k+1} d^k$

3. Critère d'arrêt:

Si:  $\|x^{k+1} - x^k\| < \varepsilon$  Stop

Si non on pose  $k = k + 1$  et aller à 2.

**Convergence des méthodes du gradient conjugué:****Théorème 3.** [12] :

Soit  $x^n$  le point obtenu à la  $n^{\text{ième}}$  itération d'une méthode de directions conjuguées, en partant de  $x^0$  quelconque. Alors  $x^n$  est un minimum global de  $f$  sur  $\mathbb{R}^n$ .

La méthode converge donc de manière finie.

**2.1.4 Méthode de Newton:**

Alors que la méthode du gradient utilise une approximation linéaire pour trouver une direction de mouvement, l'idée de la méthode itérative de Newton est de minimiser à chaque itération l'approximation quadratique de  $f$  au point courant  $x^k$  donnée par le développement de Taylor d'ordre 2:

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

Une condition nécessaire pour que le minimum de  $q^k(x)$  soit atteint est  $\nabla q^k(x) = 0$  soit

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$$

Le vecteur généré à l'itération  $k + 1$  est le vecteur minimisant  $q^k(x)$ , c'est-à-dire le vecteur satisfaisant l'équation précédente, soit

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

La méthode nécessitant l'évaluation de la matrice hessienne de  $f$ , elle ne peut être utilisée que si  $f$  est deux fois continuellement différentiable

**Remarque 1:**

le méthode de Newton nécessite même l'évaluation de l'inverse de cette matrice, ce qui est coûteux en terme de calculs).

On peut remarquer aussi que la méthode s'arrête également lorsque  $\nabla f(x^k) = 0$ , car

il s'ensuit que  $x^{k+1} = x^k$ . Si, en plus,  $\nabla^2 f(x^*)$  est définie positive, alors la condition suffisante est satisfaite, impliquant que  $x^k$  soit un minimum local.

Notons que la méthode de Newton converge en une seule itération si  $f$  est quadratique.

### Algorithme de Newton:

(a) Initialisation:

$k = 0$ : On se donne une fonction  $f$  différentiable, un point initial  $x^0$  et un seuil de tolérance  $\varepsilon > 0$ .

(b) Itération  $k = 1, 2, \dots$

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

(c) Critère d'arrêt:

Si  $\|x^{k+1} - x^k\| < \varepsilon$ . STOP

Si non, on pose  $k = k + 1$  et aller à (b).

### Convergence de la méthode de Newton:

La méthode de Newton décrite ci-dessus présente plusieurs inconvénients :

(a) L'inverse de la matrice hessienne  $(\nabla^2 f(x^k))^{-1}$  peut ne pas exister, auquel cas la méthode échoue. Cela intervient typiquement lorsque la méthode atteint une région où  $f$  est linéaire (ses secondes dérivées partielles valent zéro).

(b) La méthode de Newton n'est pas une méthode de descente: il est possible que  $f(x^{k+1})$  soit supérieur à  $f(x^k)$ .

(c) Elle est attirée aussi bien par les minima que par les maxima locaux (cette propriété est liée à la précédente). En effet, la méthode, à chaque itération, recherche uniquement un point tel que le gradient de l'approximation quadratique soit égal au vecteur nul, que ce point soit un maximum, un minimum ou un point stationnaire.

La méthode ne converge donc pas en général, notamment si elle est démarrée loin d'un minimum local, pour les première et troisième raisons. Cependant, elle converge sous certaines restrictions : si elle est exécutée à partir d'un point suffisamment proche d'un minimum local et que  $\nabla^2 f(x^k)$  n'est pas singulière, alors la méthode de Newton convergera vers ce minimum (mais pas de manière finie, de sorte qu'une condition d'arrêt soit requise de façon analogue à la méthode du gradient).

## 2.2 Étude comparative des méthodes d'optimisation sans contraintes

Dans cette partie, nous présentons une étude comparative entre quelques méthodes de l'optimisation sans contraintes, en étudiant le problème quadratique suivant:

$$\text{Minimiser } f(x) = 3x_1^2 + 3x_2^2 - 6x_2 - 3x_1x_2$$

### -La méthode du gradient:

Pour résoudre ce problème on prend comme point de départ  $x^0 = (3,4)$  et la condition d'arrêt  $\|\nabla f(x^k)\| < 0,01$ .

Le résultat est donné dans le tableau suivant:

$k$	$x^k$	$f(x^k)$	$d^k = -\nabla f(x^k)$	$\alpha_k$	$\ \nabla f(x^k)\ $	$x^{k+1}$
1	(5.00 , 4.00)	15.00	(6.00 , 9.00)	0.31	10.82	(1.14 , 1.12)
2	(1.14 , 1.21)	-3.11	(3.21 , -2.14)	0.31	3,86	(0.78 , 1.46)
3	(0.78 , 1.46)	-3.96	(0.28 , 0.42)	0.11	0.51	(0.69 , 1.33)
4	(0.69 , 1.33)	-4,00	(0.15 , -0.10)	0.31	0.18	(0.67 , 1.34)
5	(0.67 , 1.34)	-4.00	(0.01 , 0.02)	0.11	0.02	(0.67 ,1.33)
6	(0.67 , 1.33)	-4.00	(0.01 , -0.00)	0.31	0.01	-

**-Exécution de la méthode du gradient pour le problème (P) depuis  $x^0 = (3,4)$**

On remarque que la méthode du gradient n'a pas eu besoin de beaucoup d'itération pour trouver la solution optimal. Cela est du au fait que la direction donné menant au minimum est proche de la direction donné par le gradient, permettant à la méthode de progresser rapidement.

Si nous résolvons le même problème en partant du point (3,25) nous obtenons le deuxième tableau.

$k$	$x^k$	$f(x^k)$	$d^k = -\nabla f(x^k)$	$\alpha_k$	$\ \nabla f(x^k)\ $	$x^{k+1}$
1	(3.00 , 25.00)	1527.00	(-57.00 , 135.00)	0.12	146.54	(9.99 , 8.44)
2	(9.99 , 8.44)	209.58	(34.66 , 14.63)	0.12	37.62	(0.99 , 4.63)
3	(0.99 , 4.63)	25.80	(-7.95 , 18.83)	0.26	20.44	(1.97 , 2.32)
4	(1.97 , 2.32)	0.16	(4.83 , 2.04)	0.12	5.25	(0.71 , 1.79)
5	(0.71 , 1.79)	-3.42	(-1.11 , 2.63)	0.26	2.85	(0.85 , 1.47)
6	(0.85 , 1.47)	-3.92	(0.67 , 0.28)	0.12	0.73	(0.67 , 1.40)
7	(0.67 , 1.40)	-3.99	(-0.15 , 0.37)	0.26	0.40	(0.69 , 1.35)
8	(0.69 , 1.35)	-4.00	(0.09 , 0.04)	0.12	0.10	(0.67 , 1.34)
9	(0.67 , 1.34)	-4.00	(-0.02 , 0.05)	0.26	0.06	(0.67 , 1.34)
10	(0.67 , 1.34)	-4.00	(0.01 , 0.01)	0.12	0.01	(0.67 , 1.34)
11	(0.67 , 1.34)	-4.00	(0.00 , 0.01)	0.26	0.01	(-)

- **Exécution de la méthode du gradient pour le problème (P) depuis  $x^0 = (3,25)$**

On remarque que la méthode a eu besoin de presque deux fois plus d'itération que lors de la première exécution. Il s'ensuit qu'il y a une forte dépendance du comportement de la méthode du gradient à son point de départ. En effet, une conséquence du raisonnement précédent est que, pour un problème donné il est possible qu'elle converge vite depuis un point initial et exerce une convergence lente depuis un autre point.

Nous exécutons maintenant la méthode du gradient conjugué et la méthode du Newton ( La condition d'arrêt est la même que précédemment  $\|\nabla f(x^k)\| < 0.01$ )

**La méthode du gradient conjugué:**

$k$	$x^k$	$\nabla f(x^k)$	$\alpha^k$	$x^{k+1}$	la direction conjuguée
1	(3.00 , 4.00)	15.00	0.31	(1.14 , -1.21)	(6.00 , 9.00)
2	(1.14 , 1.21)	-3.11	0.12	(0.67 , 1.33)	(3.21 , -2.14)
3	(0.67 , 1.33)	-4.00	-	-	(0,0)

-**Exécution de la méthode du gradient conjugué pour le problème (P) depuis  $x^0 = (3,4)$**

**La méthode du Newton:**

$k$	$x^k$	$\nabla f(x^k)$	$\nabla^2 f(x^k)$	$(\nabla^2 f(x^k))^{-1}$	$d^k$	$x^{k+1}$
1	(3.0 , 4.0)	(6.0 , 9.0)	[(6 , -3),(-3 , 6)]	[(0.22 0.11),(0.11 , 0.22)]	()	(0.67 , 1.33)
2	(0.67 , 1.33)	(0.0 , 0.0)	-	-	()	-

**- Exécution de la méthode du Newton pour le problème (P) depuis**

$$x^0 = (3,4)$$

On remarque que cette méthode converge d'une manière finie en une seule itération (et ceci indépendamment du point initial)

**Conclusion:**

En conséquence, et après avoir résolu d'autres exemples quadratiques, on conclut que chacun des algorithmes décrit précédemment peut être utilisé pour résoudre ce type de problème; cependant, la méthode de Newton et la méthode du gradient conjugué restent les mieux adaptées, de part leur convergence finie en un nombre fixe d'itérations.

## 2.3 Cas de problème avec contraintes

Nous considérons ici le problème:

$$\begin{cases} \text{minimiser } f(x) \\ \text{sous contrainte } x \in X \end{cases}$$

où  $X$  est défini par une collection d'inégalités  $g_i(x) \leq 0$  et d'égalités  $h_j(x) = 0$ , c'est-à-dire:

$$X = \{x \mid g_i(x) \leq 0; h_j(x) = 0; \text{ avec } i = 1, \dots, m \text{ et } j = 1, \dots, r\}.$$

Nous décrivons quelques méthodes itérative pour ce problème. Il s'agira des classes de méthodes suivantes :

1. Les méthodes de directions admissibles.
3. Les méthodes de pénalité intérieure.
4. Les méthodes de pénalité extérieure.

Au travers les sections suivantes, nous admettrons donc que  $X$  n'est défini que par une collection de  $m$  inégalités  $g_i(x) \leq 0$ , puisque cela n'induit aucune perte de généralité. Nous agirons de la sorte pour toutes les méthodes sauf la méthode de barrière et la méthode de Zoutendijk pour les contraintes non linéaires.

### 2.3.1 Les méthodes de directions admissibles

Cette classe de méthodes résout les problème de minimisation non linéaire en se déplaçant d'un point de  $X$  vers un autre de ses points au coût inférieur. Elles fonctionnent selon le principe suivant :

Étant donné un élément  $x^k$  de  $X$ , une direction  $d^k$  est générée telle que pour un  $\alpha^k > 0$  et suffisamment petit, les propriétés suivantes sont assurées :

- (a)  $x^k + \alpha^k d^k$  appartient toujours à  $X$ ,
- (b)  $f(x^k + \alpha^k d^k)$  est inférieur à  $f(x^k)$ ,

Une fois  $d^k$  déterminée,  $\alpha^k$  s'obtient par minimisation monodimensionnelle pour que le déplacement dans la direction  $d^k$  soit optimal, mais cette fois-ci il est nécessaire d'imposer une borne supérieure sur la valeur de  $\alpha^k$  afin de ne pas sortir de  $X$ . Cela

définit le nouveau point  $x^{k+1}$  et le processus est recommencé.

Nous décrivons deux méthodes de directions admissibles, celle de Frank et Wolfe et celle de Zoutendijk.

### La méthode de Frank et Wolfe

#### (a) Description de la méthode

Cette méthode s'applique si les contraintes sont linéaires et si  $X$  est borné. Si l'on pose  $d^k = \bar{x}^k - x^k$ , une manière simple de générer une direction  $d^k$  satisfaisant la condition de descente  $\nabla f(x^k) \cdot d^k < 0$ , est de minimiser la dérivée directionnelle de  $f$  dans la direction  $d^k$ , comme c'est le cas pour la méthode du gradient; mais il faut en plus prendre garde de ne pas sortir de l'ensemble des solutions admissibles (le point  $\bar{x}^k$  généré doit lui-même appartenir à  $X$ , de sorte que le point  $x^k + d^k$  soit lui-même admissible). Le sous-problème de recherche de  $d^k$  peut être formulé ainsi :

$$\text{minimiser } \nabla f(x^k) \cdot (x - x^k)$$

$$\text{sous contrainte } x \in X$$

et nous pouvons obtenir  $\bar{x}^k$  comme la solution optimale de ce sous-problème. Celui-ci est un programme linéaire (les contraintes du problème de départ le sont aussi) et peut être résolu par l'algorithme du simplexe, sa solution optimale  $\bar{x}^k$  se trouvera alors en l'un des points extrêmes du domaine  $X$ , de sorte que l'optimisation monodimensionnelle le long de  $d^k$  devra être faite sous la restriction  $0 < \alpha^k < 1$ . La méthode se termine si la direction générée est égale au vecteur nul.

#### (b) Convergence de la méthode:

**Théorème 1.** [8]:

*Considérons le problème de minimisation de  $f(x)$  sous contraintes  $x \in X$  où  $X$  est un polytope borné. Alors la méthode de Frank et Wolfe démarrée depuis un point initial admissible converge vers un point satisfaisant la condition de Kuhn-Tucker.*

La méthode converge vers un point de Kuhn-Tucker, mais pas nécessairement de manière finie car elle peut être sujette au zigzag, à l'instar de la méthode du gradient.

### La méthode de Zoutendijk:

La méthode de Zoutendijk fonctionne selon le même schéma : à chaque itération, elle génère une direction de descente admissible et ensuite minimise  $f$  le long de cette direction. A l'image de la méthode de Frank et Wolfe, le sous-programme de recherche d'une telle direction est linéaire. Nous commençons par décrire la méthode de Zoutendijk applicable si les contraintes sont linéaires, puis celle-ci suscitera des modifications si elles ne le sont pas afin de garantir que des directions non admissibles ne pourront être générées si la méthode se trouve en un point où une contrainte non linéaire est active.

#### – Cas des contraintes linéaires:

La méthode de Zoutendijk cherche elle aussi à trouver, à chaque itération, la direction  $d^k$  minimisant la dérivée directionnelle de  $f$  en  $x^k$ . Les variables apparaissant dans la fonction objectif du sous-programme linéaire sont les composantes de  $d^k$ , que l'on norme afin d'empêcher le problème d'être non borné. Si le point  $x^k$  courant est situé à l'intérieur de  $X$ , le problème de recherche de la direction optimale peut s'écrire :

$$\begin{cases} \text{minimiser } \nabla f(x^k).d \\ \text{sous contraintes } d_i \leq 1 \quad \forall i \in \{1, \dots, n\}; \\ \phantom{\text{sous contraintes }} d_i \geq -1 \quad \forall i \in \{1, \dots, n\}. \end{cases}$$

Si  $x^k$  se trouve maintenant sur la frontière du domaine, une ou plusieurs contraintes du problème original sont actives. Il faut donc ajouter au programme linéaire précédent des contraintes excluant les directions menant uniquement à des points non admissibles. Cela est fait simplement en faisant intervenir les gradients des contraintes actives.

Soit  $I$  l'ensemble des contraintes actives en  $x^k$  :  $I = \{i \mid g_i(x^k) = 0\}$ .

Le problème de recherche de direction final est :

$$\left\{ \begin{array}{l} \text{minimiser } \nabla f(x^k).d \\ \text{sous contraintes } \nabla g_i(x^k).d \leq 0 \quad \forall i \in I; \\ \qquad \qquad \qquad d_i \leq 1 \quad \forall i \in \{1, \dots, n\}; \\ \qquad \qquad \qquad d_i \geq -1 \quad \forall i \in \{1, \dots, n\} \end{array} \right.$$

**Théorème 2.** [8] :

*Les conditions nécessaires de Kuhn-Tucker sont satisfaites en  $x^k$  si et seulement si la valeur optimale de la fonction objectif du programme linéaire précédent est égale à zéro.*

Ainsi, la méthode de Zoutendijk se termine si la valeur optimale de la fonction objectif de ce programme linéaire est nulle, car elle se trouve alors en un point  $x^k$  qui satisfait les conditions de Kuhn-Tucker.

– **Cas des contraintes non linéaires:**

Si, au point courant  $x^k$ , une contrainte non linéaire  $g_i$  est active, le problème de recherche d'une direction présenté au paragraphe précédent peut aboutir au choix d'une direction non admissible. En effet, un vecteur  $d$  satisfaisant  $\nabla g_i(x^k).d = 0$  est tangent à la courbe  $g_i(x) = 0$ . En raison de la non linéarité de  $g_i$ , il se peut que tout déplacement dans cette direction conduise en un point non admissible. Cette approche doit donc être modifiée pour pouvoir prendre en compte ce type de contraintes. Nous énonçons ci-dessous un théorème qui permettra de surmonter cette difficulté :

**Théorème 3.** [8] : *Considérons le problème de minimisation de  $f(x)$  sous contraintes  $g_i \leq 0$  pour  $i = 1, \dots, m$ . Soient  $x^k$  une solution admissible et  $I$  l'ensemble des contraintes actives en  $x^k$  :  $I = \{i \mid g_i(x^k) = 0\}$ . Supposons, de plus, que  $f$  et les  $g_i$  sont continues et continuellement différentiables. Si  $\nabla f(x^k).d < 0$  et  $\nabla g_i(x^k).d < 0$ ,  $\forall i \in I$ , alors  $d$  est une direction de descente admissible de  $f$  en  $x^k$ .*

Ce résultat ne fait que confirmer ce à quoi nous nous attendions, c'est-à-dire que l'inégalité stricte  $\nabla g_i(x^k).d < 0$  est nécessaire pour garantir la génération

d'une direction admissible (alors que l'inégalité stricte  $\nabla f(x^k).d < 0$  doit être satisfaite pour qu'elle soit de descente). Pour trouver un vecteur  $d$  satisfaisant ces deux inégalités strictes, une possibilité est de minimiser le maximum de la valeur  $\nabla f(x^k).d$  et des valeurs  $\nabla g_i(x^k).d$  pour  $i \in I$ . En dénotant ce minimum par  $z$  et en introduisant les restrictions interdisant au problème d'être non borné. Nous obtenons le problème suivant :

$$\left\{ \begin{array}{l} \text{minimiser } z \\ \text{sous contraintes } \nabla f(x^k).d - z \leq 0; \\ \qquad \qquad \qquad \nabla g_i(x^k).d - z \leq 0 \quad \forall i \in I; \\ \qquad \qquad \qquad d_i \leq 1, \qquad \qquad \forall i \in \{1, \dots, n\}; \\ \qquad \qquad \qquad d_i \geq -1, \qquad \qquad \forall i \in \{1, \dots, n\}. \end{array} \right.$$

Si la valeur optimale de  $z$  est strictement inférieure à zéro, alors  $d$  est manifestement une direction de descente admissible. Si elle vaut zéro,  $x^k$  satisfait la condition de Fritz John.

**Théorème 4.** [8] : *Considérons le problème de minimisation de  $f(x)$  sous contraintes  $g_i \leq 0$  pour  $i = 1, \dots, m$ . Soient  $x^k$  une solution admissible et  $I = \{i \mid g_i(x^k) = 0\}$ . Alors  $x^k$  satisfait la condition de Fritz John si et seulement si la valeur optimale de la fonction objectif du programme linéaire précédent est égale à zéro.*

### Convergence de la méthode:

En général, la convergence de la méthode de Zoutendijk n'est pas garantie. Un contre exemple, attribué à Wolfe, a montré qu'elle pouvait ne pas converger vers un point satisfaisant les conditions de Kuhn-Tucker.

Observons qu'avec l'emploi de cette méthode, nous ne pouvons pas traiter de contraintes non linéaires qui soient sous forme d'égalités, car il ne serait pas possible de générer une direction de déplacement qui n'amène à sortir du domaine. Cette difficulté peut être surmontée en introduisant des mouvements correctifs destinés à revenir dans la région admissible.

**Les méthodes de pénalité intérieure :(ou méthodes de barrière)**

Nous allons maintenant décrire les méthodes de pénalité. Cette section traite des méthodes de pénalité intérieure, aussi appelées *méthodes de barrière*, Le principe de ces méthodes réside dans la transformation d'un problème contraint en une séquence de problèmes sans contraintes, en ajoutant au coût une pénalité en cas de violation de celles-ci. Un tel sous-problème est résolu à chaque itération d'une méthode de pénalité.

L'appellation "pénalité intérieure" est employée car le minimum est approché depuis l'intérieur de  $X$ .

Les méthodes de barrière s'appliquent aux problèmes dont l'ensemble admissible  $X$  est défini uniquement par une collection d'inégalités :

$$\text{minimiser } f(x)$$

$$\text{sous contraintes } g_i(x) \leq 0; \quad i = 1, \dots, m$$

où  $f$  et les  $g_i$  sont des fonction de  $\mathbb{R}^n$  dans  $\mathbb{R}$  et sont continues. En effet, ces méthodes utilisent des fonctions dites de barrière, définies uniquement à l'intérieur de  $X$ . Si des contraintes sous forme d'égalités étaient introduites, l'intérieur de cet ensemble, c'est-à-dire son sous-ensemble tel qu'en chacun de ses points aucune contrainte n'est active, serait clairement vide. C'est donc le domaine de définition de la fonction de barrière qui serait vide rendant l'utilisation de la méthode impossible. L'intérieur de l'ensemble  $X$  défini par les  $g_i$  est le suivant :

$$X_I = \{x \mid g_i(x) < 0; \forall i \in \{1, \dots, m\}\}$$

La fonction de barrière, notée  $B(x)$ , est ajoutée au coût  $f(x)$ ; elle est continue sur  $X_I$  et sa valeur tend vers l'infini lorsque la frontière de  $X$  est approchée par l'intérieur, c'est-à-dire lorsque l'un des  $g_i(x)$  approche zéro par les valeurs négatives. Une itération de la méthode consiste ensuite à minimiser la fonction  $f(x) + tB(x)$  (où  $t$  est un paramètre réel strictement positif) à l'aide d'algorithmes de minimisation directe, une fonction  $B(x)$  et un  $t$  convenablement choisis assurant que cette minimisation ne puisse nous mener à des points situés hors de  $X_I$ . La suite du processus consiste à réduire progressivement  $t$  afin de diminuer la pénalité et autoriser les algorithmes de minimisation directe à se rapprocher peu à peu de la frontière de  $X$ . Les fonctions de barrière les plus répandues sont les suivantes :

Logarithmique :

$$B(x) = - \sum_{i=1}^m \ln(-g_i(x))$$

L'inverse :

$$B(x) = - \sum_{i=1}^m \frac{1}{g_i(x)}$$

Il est important de noter que, si tous les  $g_i$  sont convexes, ces deux fonctions de barrière le sont également.

Remarquons qu'une nécessité, pour pouvoir appliquer une telle méthode, est de disposer d'un point initial situé à l'intérieur de  $X$ . La méthode de barrière est définie en introduisant la séquence de paramètres  $\{t^k\}$ ;  $k = 0, 1, \dots$  avec  $0 < t^{k+1} < t^k$  et  $t^k \rightarrow 0$  lorsque  $k \rightarrow \infty$ .

Une itération de celle-ci consiste à déterminer

$$x^k = \arg \min_{x \in \mathbb{R}^n} \{f(x) + t^k B(x)\}$$

Le fait que  $B(x)$  ne soit défini que dans l'intérieur  $X_I$  et tende vers l'infini au fur et à mesure que l'on se rapproche des bords de  $X$  assure que, même avec un algorithme de minimisation directe, le point obtenu à chaque itération appartienne lui aussi à  $X_I$ . Le fait que  $t^k$  tend vers zéro implique que le terme  $f(x) + t^k B(x)$  tend vers  $f(x)$  lorsque  $k$  tend vers l'infini.

**Théorème 5.** [16] : *Tout point limite d'une séquence  $\{x^k\}$  générée par une méthode de barrière est un minimum global du problème contraint original.*

**Algorithme de la méthode:**

(a) Initialisation:

Fixer  $\varepsilon > 0$ Choisir  $x_0$  admissible, choisir  $\eta > 1$  et ( $t_1 > 0$ ) on prend ( $t_1 = 10$ )(b) Itération  $k = 1, 2, \dots$ 

$$\Phi(x, r_k) = f(x) + (t_k \sum_{i=1}^m -\frac{1}{g_i(x)})$$

Calculer

$$\min \Phi(x, r_k)$$

(c) Critère d'arrêt:

Si  $\|x^{k+1} - x^k\| < \varepsilon$ . STOPSi non, on pose  $k = k + 1$  et on retourne à (b) avec  $t_{k+1} = \eta t_k$ .**Convergence de la méthode:**

Si  $f$  et les  $g_i$  sont convexe,  $f(x) + tB(x)$  l'est également, car elle est la somme de fonctions convexes. Tout point stationnaire d'une telle fonction étant un minimum global et nous avons la garantie que la méthode de barrière convergera vers le minimum global du problème contraint original, et ce même si l'algorithme de minimisation directe utilisé n'est voué qu'à approcher des minima locaux.

Le comportement de la méthode dépend fortement du choix du paramètre  $t^0$  initial et du facteur, disons  $\beta$ , satisfaisant  $0 < \beta < 1$ , utilisé pour décroître  $t^k$  à chaque itération par la formule  $t^{k+1} = \beta t^k$ . Il n'existe pas de règle universelle permettant d'obtenir un bon choix de  $t^0$  et de  $\beta$ . Cela dépend fortement du problème à résoudre ainsi que du point initial  $x^0$  (celui-ci se trouve-t-il ou non à proximité de la frontière du domaine?). L'utilisateur d'une méthode de barrière sera souvent condamné à exécuter la méthode plusieurs fois avec différentes valeurs de ces paramètres jusqu'à obtenir une convergence satisfaisante.

Les faits suivants peuvent néanmoins être relevés: si  $t^k$  est trop petit, et à plus forte raison si  $x^k$  est proche des bords du domaine, le terme de barrière peut se révéler trop faible ne parvenant pas à empêcher une sortie de  $X$  (il faut donc prendre garde,

durant l'exécution, et vérifier que l'on ne sorte pas de cet ensemble, et, si cela est tout de même le cas, recommencer avec un  $t^0$  ou un  $\beta$  plus grand). Dans le cas contraire, si  $t^k$  est trop grand, l'algorithme de minimisation directe ne pourra s'approcher suffisamment des bords du domaine, conduisant à une convergence globale lente. Ces paramètres doivent donc être soigneusement choisis d'après le problème et le point initial.

**Remarque 1:**

en dehors de  $x^0$ , les points obtenus à l'itération précédente sont, à chaque itération, utilisés comme points de départ de l'algorithme de minimisation directe (cela est valable pour les deux types de méthodes de pénalité).

### 2.3.2 les méthodes de pénalité extérieure:

Considérons le problème suivant:

$$(P) \begin{cases} \text{minimiser } f(x) \\ \text{sous contraintes } g_i(x) \leq 0 \quad (i = 1 \dots m) \\ x \in \mathbb{R}^n \end{cases}$$

où  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$  et  $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ .

Considérons le problème sans contraintes( problème pénalisé):

$$(P_r) \begin{cases} \text{minimiser } \Theta_r(x) \\ x \in \mathbb{R}^n \end{cases}$$

où  $\Theta_r(x)$  est obtenu en ajoutant à  $f(x)$  le terme  $rp(x)$

$$\Theta_r(x) = f(x) + rp(x)$$

avec

$$p(x) = \sum_{i=1}^m h(g_i(x)) = \sum_{i=1}^m [g_i^+(x)]^2$$

où:

$$\forall x; g_i^+(x) = \max\{0, g_i(x)\}$$

$P$  est appelée fonction de pénalisation extérieure

Le problème  $(P_1)$  est alors remplacé par le problème d'optimisation sans contraintes:

$$\begin{aligned} \text{Minimiser } \Theta(x,r) &= f(x) + r \sum_{i=1}^m \max\{0, g_i(x)\}^2 \\ x &\in \mathbb{R}^n \end{aligned}$$

où:

$r > 0$  est appelé le coefficient de pénalité, notons  $\bar{x}(r)$  un minimum de  $\Theta(x,r)$

Le choix d'une valeur appropriée du coefficient de pénalité  $r$  résulte d'un compromis: d'une part,  $r$  doit être suffisamment grand pour que le point  $\bar{x}(r)$  obtenu soit proche de l'ensemble des solutions  $X$  (autrement dit, que  $p$  soit suffisamment faible); d'autre part, si  $r$  est choisi trop grand, la fonction  $\Theta$  peut être mal conditionnée, d'où des difficultés numériques dans la recherche de l'optimum sans contraintes. Ceci explique pourquoi les méthodes de pénalités sont généralement mises en oeuvre sous forme itérative de la façon suivante:

On commence par choisir un coefficient de pénalité  $r_1$  de valeur pas trop élevée ( pour éviter les difficultés numériques) puis on résout le problème sans contraintes:

$$\min \Theta(x,r_1) = f(x) + r_1 p(x)$$

Soit  $\bar{x}(r_1)$  le point obtenu. Si la quantité  $P(\bar{x}(r_1))$  est suffisamment faible,  $\bar{x}(r_1)$  est une bonne approximation de l'optimum et les calculs sont terminés. Dans le cas contraire, c'est que la pénalité associée à la violation des contraintes n'est pas assez élevée. On choisira donc un coefficient de pénalité  $r_2 > r_1$  et on résoudra le nouveau problème sans contrainte:

$$\min \Theta(x,r_2) = f(x) + r_2 p(x)$$

On obtiendra un nouveau point  $\bar{x}(r_2)$  et ainsi de suite.

Remarquons, qu'à chaque étape  $k$  du processus précédent, il est avantageux d'utiliser le point  $\bar{x}(r_{k-1})$  obtenu à l'étape précédente comme point de départ de l'algorithme d'optimisation sans contrainte choisi.

**Algorithme de la méthode:**

(a) Initialisation:

Fixer  $\varepsilon > 0$ , choisir  $x^0$  non admissible, choisir  $\gamma > 1$  et  $(r_1 \simeq 1)$ (b) Itération  $k = 1, 2, \dots$ 

$$\Theta(x, r_k) = f(x) + r_k \sum_{i=1}^m \max\{0, g_i(x)\}^2$$

Calculer

$$\min \Theta(x, r_k)$$

(c) Critère d'arrêt:

Si  $\|x^{k+1} - x^k\| < \varepsilon$ . STOPSi non, on pose  $k = k + 1$  et on retourne à (b) avec  $t_{k+1} = \gamma t_k$ .**Remarque 2:**

- On commence toujours par un point non admissible.
- Pour le choix du paramètre de pénalité on commence par des valeurs assez petites ( $r_1 = 1$ ) ensuite on l'augmente au fur et à mesure des itérations.

**Convergence de la méthode:**

Le théorème ci-dessous étudie le comportement des solutions  $\bar{x}^r$  de  $(P_r)$  lorsque  $r$  tend vers l'infini donne des conditions pour que les solutions des problèmes pénalisés convergent vers une solution du problème original.

**Théorème 6:**[14]

Soit  $p : \mathbb{R}^n \longrightarrow \mathbb{R}$  une fonction de pénalisation extérieur vérifiant:

(a)  $p(x) \geq 0, \quad \forall x$ .(b)  $p(x) = 0 \Leftrightarrow x \in X = \{x / g_i(x) \leq 0, \quad i = 1, \dots, m\}$ .(c)  $p$  continue.

On suppose, d'autre part, que  $f$  est une fonction continue, que  $X$  est fermé, et que l'une des deux conditions suivantes est vérifiée:

(a)  $p(x) \geq 0, \quad \forall x$ .(b)  $X$  est borné et  $p(x) \longrightarrow +\infty$  quand  $\|x\| \longrightarrow +\infty$

Alors, lorsque le coefficient de pénalité  $r$  tend vers  $+\infty$ :

- La suite  $\{\bar{x}(r)\}$  admet au moins un point d'accumulation, et tout point d'accumulation de cette suite est une solution optimale (globale) du problème (P).
- $X$  est borné et  $P(x) \rightarrow +\infty$  quand  $\|x\| \rightarrow +\infty$ ,
- $p(\bar{x}(r)) \rightarrow 0$

## 2.4 Étude comparative des méthodes d'optimisation avec contraintes

Dans cette partie, nous présentons une étude comparative entre les méthodes de pénalité intérieur et pénalité extérieur. Bien que les méthodes de pénalité sont très bien supportées par la théorie, elle souffre de faiblesse dans le cadre de l'utilisation pratique particulièrement lorsque la frontière de la région admissible est approchée. Ces méthodes ont été exécutées en utilisant le problème suivant:

$$(P') \left\{ \begin{array}{l} \min f(x) = (x_1 - 3)^2 + (x_1 - 4)^2 \\ -x_1 - 2x_2 \leq -5 \\ x_1 - x_2 \leq 3 \\ x_1 + 2x_2 \leq 9 \\ 2x_1 - x_2 \leq 4 \\ x_1, x_2 \geq 0 \end{array} \right.$$

### 1. La méthode de pénalité extérieur:

En démarrant du point non admissible (6,9), avec un paramètre initial  $r = 0.5$ , et un facteur d'augmentation associé  $\gamma = 2$  et la condition d'arrêt  $\|x^{k+1} - x^k\| < 0.01$ . Le résultat est donné dans le tableau suivant:

$k$	$x^k$	$f(x^k)$	$P(x^k, r^k)$	$r^k$	$g(x^k)$	$x^{k+1}$
1	(3.00 , 4.00 )	0	0.57	0.50	2.00	(2.67 , 3.33)
2	(2.67 , 3.33)	0.56	0.67	1.00	0.57	(2.64 , 3.27)
3	(2.64 , 3.27)	0.66	0.73	2.00	0.33	(2.62 , 3.24)
4	(2.62 , 3.24)	0.73	0.76	4.00	0.18	(2.61 , 3.22)
5	(2.61 , 3.22)	0.76	0.78	8.00	0.10	(2.60 , 3.21)
6	(2.60 , 3.21)	0.78	0.79	16.00	0.05	(2.60 , 3.20)
7	(2.60 , 3.20)	0.79	0.80	32.00	0.02	-

**-Exécution de la méthode de pénalité extérieur pour le problème ( $P'$ ) depuis le point non admissible (3,4).**

Un premier constat est que la méthode de pénalité extérieur n'as pas besoin de beaucoup d'itérations pour trouver la solution optimale de ce problème.

Les résultats du tableau nous apprend que les suites  $\{t^k\}$ ,  $\{f(x^k)\}$ , et  $\{P(x^k, t^k)\}$  croit, par contre la suite  $\{g(x^k) = \sum_{i=1}^m (\max\{0, g_i(x^k)\})^2\}$  décroît. d'où la convergence de la méthode vers un optimale globale.

## 2. La méthode de pénalité intérieur:

En démarrant du point admissible (2,2), avec un paramètre initial  $t = 10$ , et un facteur de réduction associé  $\eta = 0.7$  et la condition d'arrêt  $\|x^{k+1} - x^k\| < 0.01$ . Le résultat est donné dans le tableau suivant:

$k$	$x^k$	$f(x^k)$	$P(x^k, t^k)$	$t^k$	$x^{k+1}$
1	(2.00 , 2.00 )	5.00	26.67	10.00	(1.57 , 2.92)
2	(1.57 , 2.92)	3.23	20.17	7.00	(1.72 , 2.90)
3	(1.72 , 2.90)	2.85	15.62	4.90	(1.86 , 2.89)
4	(1.86 , 2.89)	2.52	12.43	3.43	(1.98 , 2.90)
5	(1.98 , 2.90)	2.24	10.20	2.40	(2.09 , 2.92)
6	(2.09 , 2.92)	2.00	8.64	1.68	(2.18 , 2.94)
7	(2.18 , 2.94)	1.80	7.55	7.55	(2.26 , 2.96)
8	(2.26 , 2.96)	1.63	6.78	0.82	(2.33 , 2.98)
9	(2.33 , 2.98)	1.49	6.25	0.58	(2.38 , 3.01)
10	(2.38 , 3.01)	1.37	5.87	0.40	(2.42 , 3.03)
11	(2.42 , 3.03)	1.28	5.63	0.28	(2.46 , 3.05)
12	(2.46 , 3.05)	1.20	5.43	0.20	(2.49 , 3.07)
13	(2.49 , 3.07)	1.13	5.30	0.14	(2.51 , 3.09)
14	(2.51 , 3.09)	1.08	5.21	0.10	(2.53 , 3.10)
15	(2.53 , 3.10)	1.03	5.15	0.07	(2.47 , 3.26)
16	(2.47 , 3.26)	0.82	5.10	0.05	(2.47 , 3.26)
17	(2.47 , 3.26)	0.82	5.07	0.03	-

**-Exécution de la méthode de pénalité intérieur pour le problème ( $P'$ ) depuis le point admissible (2,2).**

Il n'y a pas de technique générale pour le choix du paramètre  $t$  et du facteur de réduction, ce qui implique qu'il faille souvent redémarrer la méthode plusieurs fois. La méthode a eu besoin plus d'itération que la méthode de pénalité extérieur, pour trouver une approximation du minimum. Les suites  $\{t^k\}$ ,  $\{f(x^k)\}$ , et  $\{P(x^k, t^k)\}$  décroît, d'où la convergence de la méthode.

# Chapitre 3

## Optimisation multicritère

### Introduction

La plupart des problèmes d'optimisation rencontrés en pratique sont décrits à l'aide de plusieurs objectifs ou critères souvent contradictoires et doivent être optimisés simultanément.

Optimiser un tel problème relève de l'optimisation multi-objectifs qui cherche à optimiser plusieurs composantes d'un vecteur fonction contrairement à l'optimisation mono-objectif.

Dans ce chapitre, nous présenterons le problème de l'optimisation multi-objectifs non linéaire, en donnant les définitions de solutions et les méthodes de résolution.

### 3.1 Position de problème:

#### Définition 1:

Un problème multi-critères est de la forme:

$$\left\{ \begin{array}{l} \max (f_1(x), f_2(x), \dots, f_n(x)) \quad (n \geq 2) \\ SC \\ x \in X = \{ x \in \mathbb{R}^n \mid g_i(x) \leq 0, \quad i = 1 \dots p \} \end{array} \right. \quad (*)$$

Où

$X$  est l'ensemble des décisions

$f_i$  les fonctions critères avec  $i = 1, \dots, n$  et  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ .

On note  $f = (f_1, f_2, \dots, f_n)$  le vecteur des fonctions objectifs,

$x = (x_1, x_2, \dots, x_n) \in X$  est la variable de décision avec  $X \in \mathbb{R}^n$ .

Le mot maximiser signifie ici que l'on veut maximiser toutes les fonctions objectifs simultanément. Comme une décision vérifiant cette propriété n'existe que rarement, en raison de fait que les fonctions objectifs sont généralement conflictuelles on est amené à utiliser d'autres notions d'optimalité.

## 3.2 Solution d'un problème d'optimisation multi-critères:

Il existe différents types de solutions d'un problème d'optimisation multi-critères, on cite quelques unes:

- L'optimalité de Pareto (ou efficacité).
- L'optimalité de Slater (ou Pareto optimalité faible).
- L'optimalité selon Geoffrion.

### 3.2.1 Optimalité de Pareto:

Une décision  $\bar{x}$  est efficace si on peut améliorer l'une des composantes du vecteur  $f(\bar{x})$  sans décroître au moins l'une des autres composantes.

#### Définition 2:

Une solution  $x^p \in X$  est appelée solution optimale selon Pareto (efficace) pour le système (\*) . S'il n'existe pas de décision  $x \in X$  qui vérifie le système d'inégalité

$$f_i(x) \geq f_i(x^p); \quad i = 1, \dots, n \text{ dont au moins une est stricte.}$$

Cela peut s'écrire  $f(x) \not\geq f(x^p)$  . Le vecteur  $f(x^p)$  sera appelé maximum de Pareto.

#### Définition 3:

L'ensemble  $X^p$  de décisions maximales de Pareto du problème (\*) est:

1. Non vide, lorsque  $X$  est compact et les fonctions  $f_i$  sont continues sur  $X$
2. Extérieurement stable, c-à-d:  $\forall x \in X, \exists x^p \in X^p \text{ telque : } f_i(x) \leq f_i(x^p); \quad i = 1, \dots, n$
3. Intérieurement stable, c-à-d:  $\forall x^1, x^2 \in X^p \text{ telque : } f_i(x^1) \not\geq f_i(x^2); \quad i = 1, \dots, n$

#### Définition 4:

Dans le cas d'un problème de minimisation une solution  $x^p \in X$  est appelé décision minimale de Pareto dans le problème (\*) , s'il n'existe pas de décision  $x \in X$  qui vérifie le

système d'inégalité  $f_i(x) \leq f_i(x^p)$ ;  $i = 1, \dots, n$  dont au moins une est stricte.

Cela peut s'écrire brièvement  $\forall x \in X, f(x) \not\leq f(x^p)$ . Le vecteur  $f(x^p)$  sera appelé minimum de Pareto et  $X^p$  ensemble des décisions minimale de Pareto.

**Remarque 1:**

On a une définition équivalente à la précédente; une décision  $x^p \in X$  est maximale de Pareto, si pour toute décision  $x \in X$ , soit  $f_i(x) = f_i(x^p)$ ,  $i = 1, \dots, n$ , soit il existe un indice  $i_0(x) = 1, \dots, n$  tel que  $f_{i_0}(x) \geq f_{i_0}(x^p)$ ,  $i = 1, \dots, n$ .

On a également une autre définition équivalente: une décision  $x^p \in X$  est maximale de Pareto, si une décision  $x \in X$  vérifie le système d'inégalité  $f_i(x) \geq f_i(x^p)$ ;  $i = 1, \dots, n$  alors  $f_i(x) = f_i(x^p)$ ;  $i = 1, \dots, n$ .

### 3.2.2 Optimalité de Slater:

**Définition 5:**

On dit que  $x^s \in X$  est une décision maximale de Slater ( ou Pareto optimale faible) s'il n'existe pas de décision  $x \in X$  qui vérifie le système d'inégalité  $f_i(x) > f_i(x^s)$ ;  $i = 1, \dots, n$ . On notera cela  $f(x) \not\geq f(x^s)$ . et  $X^s$  sera appelé maximum de Slater.

L'ensemble des décisions maximale de Slater est:

1. Un sous ensemble compact non vide de  $X$  lorsque  $X$  compact et les  $f_i$  sont continues sur  $X$ .
2. Extérieurement stable, c-à-d:  $\forall x \in X, \exists x^s \in X^s$  tel que :  $f_i(x) \leq f_i(x^s)$ ;  $i = 1, \dots, n$

**Définition 6:**

On dit que  $x^s \in X$  est une décision minimale de Slater s'il n'existe pas de décision  $x \in X$  qui vérifie le système d'inégalité  $f_i(x) < f_i(x^s)$ ;  $i = 1 \dots n$ . On notera cela  $f(x) \not\leq f(x^s)$ .  $f(x^s)$  sera appelé minimum de Slater.

### 3.2.3 Optimalité de Geoffrion:

Un vecteur décision  $x^* \in X$  est paréto-optimal Geoffrion s'il est paréto-optimal et s'il existe un nombre réel  $M > 0$  tel que pour chaque  $i$  et chaque  $x \in X$  satisfaisant  $f_j(x) < f_j(x^*)$ , il existe au moins un critère  $f_j$  tels que

$$f_j(x) < f_j(x^*) \text{ et } f_i(x) - f_i(x^*) / f_j(x) - f_j(x^*) \leq M$$

**Définition 7: "Point idéal":**

Le vecteur défini par  $\bar{f}(x) = ( \max f_1(x), \dots, \max f_n(x) ) \in \mathbb{R}^n$ , constitue le point idéal.

**Définition 8: "Matrice de gains":**

La matrice carrée de dimension n suivante:

$$\begin{pmatrix} \bar{f}_1 & \cdot & \cdot & \cdot & f_{1n} \\ f_{21} & \bar{f}_2 & \cdot & \cdot & f_{2r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ f_{n1} & & \cdot & \cdot & \bar{f}_n \end{pmatrix}$$

Où:

$$\bar{f}_i = \max_{x \in X} f_i(X) = f_i(\bar{x}_j), \forall i, j = \overline{1, n}$$

**Définition 9: "point nadir":**

Le point de  $\mathbb{R}^n$  de coordonnées  $x_i = \min_{j=1, \dots, r} f_{ij}$ ,  $i = \overline{1, n}$  est appelé point nadir.

**Définition 10: Taux de substitution:**

Ils traduisent l'idée de compensation entre une perte sur un objectif et un gain sur un autre.

**Définition 11: Point de référence:**

C'est un vecteur dont les composantes sont les valeurs souhaitables qu'il faut atteindre.

### 3.3 Conditions d'optimalité:

Les conditions d'optimalité sont très importantes dans le domaine de l'optimisation. En effet, pour certaines classes de problèmes, elles permettent de trouver directement des solutions optimales; pour d'autres, elles sont utilisées dans les méthodes numériques de résolution.

Concédons le problème d'optimisation suivant:

$$(P) \begin{cases} \text{minimiser } ( f_1(x), f_2(x), \dots, f_k(x) ) \\ ( g_1(x), g_2(x), \dots, g_m )^t \leq 0 \\ x \in \mathbb{R}^n \end{cases}$$

**Théorème 1.**[13] (**Condition nécessaire d'optimalité**):

Supposons que les fonctions objectifs et les contraintes sont continûment différentiables. La condition nécessaire pour qu'un vecteur décision  $x^* \in X$  soit Pareto optimale est qu'il existe des vecteurs  $\lambda \geq 0 \in \mathbb{R}^k$  et  $\mu \geq 0 \in \mathbb{R}^m$  tels que  $(\lambda, \mu) \neq (0, 0)$  et:

1.  $\sum_i \lambda_i \nabla f_i(x^*) + \sum_i \mu_j \nabla g_i(x^*) = 0$
2.  $\mu_j g_i(x^*) = 0$  pour  $j = 1, \dots, m$

**Théorème 2.**[13] (**Condition suffisante d'optimalité**):

Supposons que le problème (P) soit convexe. Une condition suffisante pour qu'un vecteur décision  $x^* \in X$  soit Pareto optimale est qu'il existe des multiplicateurs  $\lambda \geq 0 \in \mathbb{R}^k$  et  $\mu \geq 0 \in \mathbb{R}^m$  tels que  $(\lambda, \mu) \neq (0, 0)$  et:

1.  $\sum_i \lambda_i \nabla f_i(x^*) + \sum_i \mu_j \nabla g_i(x^*) = 0$
2.  $\mu_j g_i(x^*) = 0$  pour  $j = 1, \dots, m$

**Théorème 3.**[13]

La condition énoncée précédemment est suffisante pour qu'un vecteur décision  $x^* \in X$  soit une solution Pareto-optimale faible pour  $0 \leq \lambda \in \mathbb{R}^k$  avec  $\lambda \neq 0$ .

**Théorème 4.**[13]

Une condition nécessaire pour qu'un vecteur décision  $x^* \in S$  soit une solution pareto-optimale Geoffrion est qu'il existe des vecteurs  $0 \leq \lambda \in \mathbb{R}^k$  et  $0 \leq \mu \in \mathbb{R}^m$  tels que:

1.  $\sum_i \lambda_i \nabla f_i(x^*) + \sum_i \mu_j \nabla g_i(x^*) = 0$
2.  $\mu_j g_i(x^*) = 0$  pour  $j = 1 \dots m$

### 3.4 Méthodes de résolution d'un problème d'optimisation multicritères

Dans cette partie, nous allons présenter quelques méthodes d'optimisation multi-objectifs. Ces méthodes consistent à trouver une partie de l'ensemble des solutions Pareto optimales ou l'ensemble tout entier. Elles peuvent être classées de plusieurs manières.

Dans notre travail, nous avons adopté la classification présentée par Hwang et Masud en 1979, vu qu'elle est basée sur le degré de participation du décideur dans le processus de recherche de la solution optimale.

Hwang et Masud ont établi quatre classes différentes:

- Les méthodes sans articulation des préférences du décideur.
- Les méthodes avec articulation des préférences du décideur à postériori.
- Les méthodes avec articulation des préférences du décideur à priori.
- Les méthodes interactives ( avec articulation progressive des préférences du décideur).

Soit le problème d'optimisation multi-objectifs suivant:

$$\left\{ \begin{array}{l} \text{minimiser } ( f_1(x), f_2(x), \dots, f_k(x) ) \\ g_i(x) \leq 0 \quad i = 1 \dots m \\ x \in \mathbb{R}^n \end{array} \right.$$

#### 3.4.1 Méthodes sans articulation du décideur:

Dans les méthodes sans articulation de préférences, le décideur ne contribue pas à la construction de la solution, le problème est résolu avec quelque méthodes relativement simple et la solution obtenue est présentée au décideur. Ces méthodes sont convenables dans les situations où le décideur n'a pas d'exigence de la solution et il peut se contenter de quelques solutions.

nous présenterons deux exemple de cette classe:la méthode du critère global qui est aussi appelée "compromise programing" et la méthode  $L_p$ - métrique poids.

### 1. Méthode du critère global:

Dans la méthode du critère global, la distance entre le point de référence et la région du critère réalisable est minimisée.

L'analyste sélectionne le point de référence et un système métrique pour mesurer les distances. Dans ce cas, le problème à résoudre est:

$$\begin{cases} \text{minimiser} & (\sum_{i=1}^k (f_i(x) - z_i^{**})^p)^{\frac{1}{p}} \\ x \in S \end{cases} \quad (1)$$

Où  $p$  est un entier,  $S$  la région réalisable.

De la définition du vecteur critère idéal  $z^{**}$ , nous savons que  $f_i(x) \geq z_i^{**}$  pour tout  $i = 1, \dots, k$  et tout  $x \in S$ . Les problèmes avec ou sans l'exposant  $\frac{1}{p}$  sont équivalents pour  $1 < p < \infty$ . De là, le problème (1) est une fonction croissante du problème correspondant sans l'exposant.

Si  $p = \infty$ , la métrique est aussi appelée métrique de Tchébycheff et le problème est de la forme:

$$\begin{aligned} \min \max_{1 \leq i \leq k} [f_i(x) - z_i^{**}] \\ x \in S \end{aligned} \quad (2)$$

Notons que la solution qu'on obtient dépend beaucoup de la valeur choisie pour  $p$ . On utilise largement le choix:  $p = 1, 2, \infty$

### Résultats théoriques:

**Théorème 5.**[13]:

*Chaque solution du problème (1) (où  $1 < p < \infty$ ) est Pareto optimale.*

**Yu** a montrer que si  $z$  est un ensemble convexe, alors pour  $1 < p < \infty$  la solution du problème (1) est unique. [9]

**Théorème 6.**[13]:

*Chaque solution du problème (2) est Pareto optimale faible.*

**Corollaire.** [13]:

*Si le problème (2) a une solution unique, celle-ci est Pareto optimale.*

**Remarque 2:**

La méthode du critère global est une méthode simple à utiliser, si le but est d'obtenir une solution pareto optimale.

Les solutions obtenues avec  $L_p$ -métrique dont ( $1 < p < \infty$ ) sont garanties à être des solutions Pareto optimales. Si la  $L_\infty$ -métrique est utilisé, la solution obtenue peut être Pareto optimale faible.

**2. Méthode  $L_p$ -métrique avec poids:**

Pour tenir compte de l'importance des critères dans la méthode du critère global, des coefficients poids  $w_i$  sont inclus dans la métrique. Dans ce paragraphe, nous présenterons la méthode brièvement. Dans ce cas, elle peut être considérée comme faisant partie de la classe des méthode à posteriori.

On suppose que  $w_i \geq 1$ , pour tout  $i = 1, \dots, k$  et  $\sum_{i=1}^k w_i = 1$ . Maintenant les problèmes pour minimiser les distances sont de la forme:

$$\begin{aligned} \text{minimiser } & \left( \sum_{i=1}^k w_i (f(x) - z_i^{**})^p \right)^{\frac{1}{p}} & (3) \\ & x \in S \end{aligned}$$

et

$$\begin{aligned} \text{minimiser } & \max_{1 \leq i \leq k} [w_i (f(x) - z_i^{**}) / z_i^{**}] & (4) \\ & x \in S \end{aligned}$$

Où  $p$  est un entier et  $S$  est la région réalisable.

**Théorème 7.**[13]:

La solution  $x \in S$  du problème (3) (lorsque  $1 < p < \infty$ ) est Pareto optimale si:

- La solution est unique
- les coefficients poids sont positifs ( $w_i > 0$  pour tout  $i = 1, \dots, k$ )

**Théorème 8.**[13]:

Chaque solution du problème (4) est Pareto optimale faible si  $w_i > 0$  pour tout  $i = 1, \dots, k$ .

**Théorème 9.**[13]:

Soit  $x^* \in S$  une solution Pareto optimale. Alors il existe un vecteur  $w$  où  $w_i > 0$  pour tout  $i = 1, \dots, k$  tel que  $x^*$  est une solution optimale du problème (3), où le poids de référence  $z^* = z^{**} + \epsilon$  (où  $\epsilon \in \mathbb{R}^k$  est un vecteur avec  $\epsilon_i > 0$  pour tout  $i = 1, \dots, k$ ).

### 3.4.2 Méthodes avec articulation des préférences à posteriori:

Les méthodes à posteriori sont utilisées pour générer des solutions Pareto optimales. Après avoir généré l'ensemble Pareto optimale (ou une partie), il est présenté au décideur qui sélectionne la solution la plus préférée.

Dans cette classe de méthodes, nous présenterons deux méthodes importantes: la méthode des poids et la méthode  $\epsilon$ -contrainte

#### 1. Méthodes des poids:

Dans la méthode des poids, présentée par Gauss et Saaty en 1955 et Zadeh en 1963[9], l'idée de base est d'associer à chaque fonction objectif, un facteur poids et de minimiser la somme des poids multipliés par les fonctions objectifs. Dans cette méthode l'ensemble des fonctions objectifs est transformé en une fonction objectif à optimiser. On suppose que les coefficients poids  $w_i$  sont des nombres réels tels que  $w_i \geq 0$ , pour tout  $i = 1, \dots, k$ , il est aussi supposé que les poids sont normalisés:  $\sum_i w_i = 1$ . Alors, le problème d'optimisation multi-objectifs est de la forme:

$$\text{minimiser } \sum_{i=1}^k w_i f_i(x) \quad (5)$$

$$x \in S$$

Où  $w_i \geq 0$  pour tout  $i = 1, \dots, k$ ,  $\sum_i w_i = 1$  et  $S$  la région réalisable.

**Résultats théoriques:**

**Théorème 10.**[13]:

Si  $x^* \in S$  une solution du problème aux poids (5), alors  $x^*$  est Pareto optimale faible.

**Théorème 11.**[13]:

Soit  $x^* \in S$  une solution du problème aux poids (5), où les coefficients poids sont positifs ( $w_i > 0$ , pour tout  $i = 1, \dots, k$ ). Alors  $x^*$  est Pareto optimale

**Théorème 12.**[13]:

Si  $x^* \in S$  une solution unique du problème aux poids (5), alors  $x^*$  est Pareto optimale

**Théorème 13.**[13]:

Supposons que le problème d'optimisation multi-objectifs et convexe. Si  $x^* \in S$  est Pareto optimale, alors il existe un vecteur poids  $w$

( $w_i \geq 0$ , pour tout  $i = 1, \dots, k$ ,  $\sum_i w_i = 1$ ) tels que  $x^*$  est une solution du problème (5)

**Théorème 14.**[13]:

Soit  $x^* \in S$  une solution du problème aux poids (5), avec tous les coefficients poids sont positifs ( $w_i > 0$ , pour tout  $i = 1, \dots, k$ ). Alors  $x^*$  est Pareto optimale au sens de Geoffrion (condition suffisante)

**Remarque 3:**

Dans la pratique, on utilise la condition ( $w_i > \varepsilon$  où  $\varepsilon > 0$  au lieu de la condition  $w_i > 0$ , pour tout  $i = 1, \dots, k$ ). Toutes les solutions Pareto optimales d'un problème convexe peuvent être trouvées si  $\varepsilon$  est très petit.

## 2. Méthode $\varepsilon$ -contrainte:

Dans la méthode  $\varepsilon$ -contrainte, introduite par Haimès, Ladson et Wismer en 1971, une des fonctions objectifs est sélectionnée pour être optimisée et toutes les autres sont converties en contraintes qui sont inférieures à leur bornes supérieures et le problème à résoudre est alors:

$$\left\{ \begin{array}{l} \text{Minimiser } f_\ell(x) \\ SC \\ f_j(x) \leq \varepsilon_j, \text{ pour } j = 1, \dots, k \text{ et } j \neq \ell \\ x \in S \end{array} \right. \quad (P)$$

avec  $S$  l'ensemble des solutions réalisables.

**Résultats théoriques:****Théorème 15.**[13]:

Soit  $x^* \in S$  une solution du problème  $(P)$ , alors elle est Pareto optimale faible.

**Théorème 16.**[13]:

Un vecteur décision  $x^* \in S$  est Pareto optimale si et seulement s'il est la solution du problème  $(P)$  pour  $\ell = 1..k$  où  $\varepsilon_j = f_j(x^*)$ ,  $j = 1, \dots, k, j \neq \ell$ .

**Théorème 17.**[13]

Si  $x^*$  est la solution unique du problème  $\varepsilon$ -contrainte  $(P)$  pour un certain  $\ell$  avec  $\varepsilon_j = f_j(x^*)$ ,  $j = 1, \dots, k, j \neq \ell$ , alors  $x^*$  est Pareto optimale.

**Remarque 4:**

La méthode  $\varepsilon$ -contrainte peut aussi être utilisée comme méthode à priori, quand le décideur spécifie  $f_\ell$  et les bornes supérieures.

La méthode  $\varepsilon$ -contrainte est plus difficile que la méthode des poids car le nombre de critères est plus grand.

**3.4.3 Méthodes avec articulation des préférences à priori:**

Dans ces méthodes, le décideur spécifie nécessairement sa ou ses préférences, ses attentes et espérances avant le processus de résolution. La difficulté est que le décideur ne connaît pas nécessairement à l'avance, ce qui est possible d'atteindre et comment réaliser ses espérances.

Dans la suite, nous présentons deux méthodes à priori: optimisation de la fonction d'utilité et la méthode de l'ordre lexicographique.

**1. Méthode de la fonction d'utilité:**

Dans cette méthode, le décideur doit préciser la forme mathématique de la fonction d'utilité  $U : \mathbb{R}^k \longrightarrow \mathbb{R}$ , qui représente sa ou ses préférences globalement. Le problème à résoudre est alors:

$$\begin{cases} \text{Maximiser } U(f_1(x), f_2(x), \dots, f_p(x)) \\ SC \\ x \in S \end{cases}$$

Ce problème est résolu par les méthode d'optimisation monocritères.

Les modèle les plus utilisé sont:

- (a) **Le modèle additif:**  $\max_{x \in S} \sum_{i=1}^p U_i f_i(x)$
- (b) **le modèle multiplicatif:**  $\max_{x \in S} \prod_{i=1}^p U_i f_i(x)$

### Remarque 5:

La méthode de la fonction d'utilité semble être une méthode très simple, mais la difficulté est liée à la spécification de l'expression mathématique de la fonction d'utilité soit par manque d'information ou par manque d'expérience. En plus, il faut que les objectifs soient commensurables.

## 2. Méthode lexicographique:

Soit le problème

$$\begin{cases} \text{Maximiser } (f_1(x), f_2(x), \dots, f_p(x)) \\ SC \\ x \in S \end{cases}$$

Dans cette méthode, le décideur doit ranger les fonctions objectifs suivant l'ordre de leur importance

$$f_1(x) > f_2(x) > \dots > f_p(x)$$

- (a) Résoudre:

$$\begin{cases} \max f_1(x) \\ x \in S \end{cases} \quad (1)$$

Si le problème à une solution unique, elle est la solution de tout le problème d'optimisation multicritère.

Dans le cas contraire on passe à l'étape 2.

Soit  $x_1^*$  la solution optimale du problème (1), avec  $f_1^* = f_1(x)$ .

(b) Résoudre:

$$\left\{ \begin{array}{l} \max f_2(x) \\ S.C \quad x \in S \\ f_1^* = f_1(x) \end{array} \right.$$

à l'étape  $k$ , il résoud:

$$\left\{ \begin{array}{l} \max f_2(x) \\ S.C \quad x \in S \\ f_1^* = f_1(x) \\ f_2^* = f_2(x) \\ \cdot \\ \cdot \\ \cdot \\ f_{k-1}^* = f_{k-1}(x) \end{array} \right.$$

**Théorème 18.**[13]:

*La solution obtenue par l'ordre lexicographique est Pareto optimale*

**Remarque 6:**

La justification de l'utilisation de la méthode de l'ordre lexicographique est sa simplicité. Toutefois, cette méthode a plusieurs inconvénients. Le décideur peut avoir des difficultés à mettre les fonctions objectives dans l'ordre absolu d'importance. D'autre part, la méthode est habituellement robuste, les fonctions objectives moins importantes ne sont pas prises en considérations: si la fonction objective la plus importante a une solution unique, les autres fonctions n'ont pas d'influence sur la solution.

### 3.4.4 Méthode d'articulation progressive des préférences(Interactive):

Une méthode interactive consiste en une suite d'étapes de calculs et de dialogue avec le décideur. La première étape de calcul fournit une première solution. Celle-ci est présentée au décideur qui réagit en apportant des informations supplémentaires sur ses préférences (étape de dialogue). Cette information injectée dans le modèle utilisé, permet de construire une nouvelle solution. Ainsi dans ce type de méthodes le décideur contribue directement à la construction de la solution.

Dans ce qui suit, nous présenterons deux méthodes interactive: méthode GDF et la méthode Nakayama.

#### 1. Méthode Geoffrion, Deyer et Feinberg:

La méthode GDF à été proposée par Geoffrion, Deyer et Feinberg en 1972, est l'une des première méthodes interactive. Elle est basée sur la maximisation d'une fonction d'utilité connue de manière implicite par le décideur.

A chaque itération une approximation linéaire de la fonction d'utilité est générée et maximiser. La maximisation se fait par la méthode de Frank et Wolf (FW)

#### Description de la méthode:

Le problème à résoudre est le suivant:

On suppose que:

- La fonction d'utilité  $U : \mathbb{R}^k \longrightarrow \mathbb{R}$  existe et connue implicitement par le décideur.  
De plus  $u : \mathbb{R}^k \longrightarrow \mathbb{R}$  est continûment différentiable et concave.
- Toute les fonctions objectives sont continûment différentiables.
- La région réalisable est compacte et concave.

#### Algorithme de la méthode:

**Début**

**Étape 1.**

Choisir une fonction de référence  $f_\ell$ , choisir une première solution de compromis  $x^\ell$  et poser  $h = 1$

**Étape 2.**

Demander au décideur de spécifier les taux de substitution entre la fonction  $f_\ell$  et les autres fonctions objectifs au point courant  $x^h$

**Étape 3.**

Résoudre la problème (\*). On note  $y^h \in S$  la solution de ce problème. Poser la direction  $d^h = y^h - x^h$ , si  $y^h = x^h$  allez à l'étape 6.

**Étape 4.**

Déterminer avec l'aide du décideur, le pas  $t^h$  à choisir dans la direction  $d^h$ , noter la solution correspondante par  $x^{h+1} = x^h + t^h d^h$

**Étape 5.**

$h = h + 1$ . Si le décideur veut continuer, aller à l'étape 2.

**étape 6.**

La solution finale est  $x^h$

**Fin**

**Remarque 7:**

En dépit de sa justification théorique, la méthode GDF n'est pas tellement convaincante ni dominante dans la pratique l'inconvénient de la méthode est que la solution qu'elle donne n'est pas nécessairement Paréto Optimale.

Théoriquement, l'optimalité de la solution finale est garantie si la fonction utilité est décroissante. Les taux de substitution sont cruciaux dans l'approximation de la fonction utilité et leur spécification pose de sérieux problèmes pour les décideurs.

**2. Méthode Nakayama**

La méthode de Nakayama est une méthode interactive, l'idée de cette méthode est que le niveau d'aspiration est utilisé comme point de référence. En plus du point de référence cette méthode utilise le point idéal.

Considérons le problème:

$$\begin{cases} \text{Maximiser } (f_1(x), f_2(x), \dots, f_p(x)) \\ SC \\ x \in S \end{cases}$$

avec  $S = \{g_i(x) \leq 0, j = 1, \dots, m, x \in \mathbb{R}^n\}$  nous supposons que les fonctions  $f_i, i = \overline{1, r}$  sont continûment différentiable.

Notons que le niveau d'aspiration est donné par le décideur à chaque itération, le point idéal est également donné par le décideur, il peut être différent du point idéal au sens connu.

Le problème résolu dans la méthode Nakayama est le suivant:

$$\min \max_{1 \leq i \leq r} W_i^k |f_i^* - f_i(x)| \quad (I) \\ x \in S$$

tels que:

$W_i = 1/(f_i^* - f_i^{**})$  est le poids affecté au critère  $i$ ,  $f_i^{**}$  est le point idéal,  $S$  la région réalisable.

On suppose que  $f^* < f^{**}$ ,  $f^*$  étant le niveau d'aspiration.

**Définition 12:**

Si  $f^*$  est un niveau d'aspiration du décideur alors une décision  $x^*$  est dite solution paréto optimale satisfaisante si elle est Paréto optimale et

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

Elle est dite satisfaisante si elle vérifie seulement l'inégalité précédente.

**Description de la méthode:**

Pour résoudre le problème (I), on le transforme comme suit:

$$\begin{aligned} \min F(x, z) &= z \\ h_i(x) &\leq 0, i = \overline{1, r} \\ g_j(x) &\leq 0, j = \overline{1, m} \end{aligned}$$

$g_j(x)$  sont les fonctions contraintes définissant  $S$  et  $h_j(x)$  les fonctions contraintes construites à partir du vecteur critère:  $h_i(x) = w_i(f_i^* - f_i(x))$  est le poids affecté au critère  $f_i$ .

$f^{**}$  est le point idéal,  $f^*$  est le niveau d'aspiration donné par le décideur  $f^* < f^{**}$

Après avoir obtenu la solution du problème, on la présente au décideur

On demande au décideur de classer les fonctions objectifs en trois classes:

- La classe des fonctions objectifs à améliorer.
- La classe des fonctions objectifs relaxer.
- La classe des fonctions objectifs à laisser sans changement.

Le nouveau niveau d'aspiration pour les fonctions objectifs de la classe 1 et 2 sera données par le décideur, pour la classe 3 des fonctions objectifs acceptables on garde le niveau d'aspiration précédent.

### Algorithme de la Méthode:

#### Début

#### Pas 1.

Donner ou calculer le point idéal par la méthode de résolution d'un problème d'optimisation avec ou sans contraintes.

#### Données initiales:

- $x^0 = (x_1, x_2, \dots, x_n, Z)$  point initial arbitraire du domaine de définition du problème (II)
- $k = 0$
- donner le niveau d'aspiration  $f_i^{*k}$

#### Pas 2.

Calculer les poids  $w_i = 1/(f_i^{**} - f_i^{*k}, i = \overline{1, r})$

#### Pas 3.

résoudre

$$\text{Min}F(x, z) = z, \quad w_i(f_i^{*k} - f_i(x^k) - z) \leq 0, i = \overline{1, r}, \quad g_j(x) \leq 0, j = \overline{1, m}, x^k$$

étant la solution obtenue à l'étape k.

On obtient la solution  $x^{k+1}$

**Pas 4.**

Présenter la solution  $x^{k+1}$  au décideur.

Si la solution est accepté alors stop, aller à fin

sinon demander au décideur de classer les fonctions objectives en trois classes.

- Donner celles de la classe 1 et 2, donner les niveaux d'aspiration pour cette classe.
- Donner celle de la classe 3, laisser sans changement les niveaux d'aspiration.

Retour au pas 2

**FIN**

**Résultats Théoriques:*****Théorème 19.***[15]

*Supposons que pour tout  $x \in S$ ,  $f_i^{**} \geq f_i(x)$ ,  $i = \overline{1, r}$*

*Si on pose pour un niveau d'aspiration donné  $f^*$*

$$w_i = 1/(f_i^{**} - f_i), i = \overline{1, r}$$

*Alors toute solution  $x^*$  du problème*

$$\min_{x \in S} \max_{1 \leq i \leq r} w_i |f_i^* - f_i(x)|$$

*est satisfaisante, c'est-à-dire  $f_i(x^*) \geq f_i^*$ .*

*Si les fonctions  $f_i$  sont bornées, elle est Pareto- Optimale faible.*

**Remarque 8:**

La solution obtenue par cette méthode peut -être Pareto-optimale cela dépend des hypothèses faites sur les fonctions critères.

Notons que, du point de vue pratique la classification des fonctions objectifs en trois classes et la spécification de quelques incréments et décréments des valeurs de certains de ces fonctions est la même chose que de spécifier un nouveau point de référence.

La méthode est basée sur la satisfaction du décideur, celui-ci est libre de chercher une autre solution qui le satisfait et de changer sa ou ses préférences si c'est nécessaire.

**Conclusion:**

Dans ce chapitre nous avons passé en revue quelques méthodes d'optimisation multi-critère suivant le classement défini par HUWANG et MASUD [12], ce classement est défini par rapport aux informations disponibles sur les préférence du décideur, on a vu les classes suivantes:

- Les méthodes sans articulation des préférences du décideur.
- Les méthodes avec articulation des préférences du décideur à posteriori.
- Les méthodes avec articulation des préférences du décideur à priori.
- Les méthodes avec articulation progressive des préférences du décideur (interactive).

Pour chaque classe nous avons présenté quelques résultats théoriques ainsi que quelques références. Beaucoup de méthodes pas était citer, notre objectif est de donner une idée sur chaque classe de méthode.

# Implémentation

## Implémentation de la méthode des poids:

Concédons le problème multi-critère avec contraintes suivant:

$$(P) \left\{ \begin{array}{l} \min f_1(x) = 2x_1^2 + 3x_2^2 - 4x_1 + 5 \\ \min f_2(x) = -2x_1^2 + 2x_2 + 4 \\ \min f_3(x) = x_1^2 + x_2^2 + 7 \\ S.C \\ -3x_1 - 2x_2 + 6 \leq 0 \\ -x_1 + x_2 - 3 \leq 0 \\ x_1 + x_2 - 7 \leq 0 \\ \frac{2}{3}x_1 - x_2 - \frac{4}{3} \leq 0 \\ x_1, x_2 \geq 0 \end{array} \right.$$

Afin d'utiliser la méthode des poids, nous admettons les poids  $\lambda_i \geq 0$ ,  $i = \overline{1,3}$  tels que:  $\lambda_1 = \frac{1}{4}$ ;  $\lambda_2 = \frac{1}{2}$ ;  $\lambda_3 = \frac{1}{4}$ .

Le problème mono-critère équivalent sera donc:

$$(P') \left\{ \begin{array}{l} \min f(x) = -\frac{1}{2}x_1^2 + x_2^2 - x_1 + x_2 + 5 \\ S.C \\ -3x_1 - 2x_2 + 6 \leq 0 \\ -x_1 + x_2 - 3 \leq 0 \\ x_1 + x_2 - 7 \leq 0 \\ \frac{2}{3}x_1 - x_2 - \frac{4}{3} \leq 0 \\ x_1, x_2 \geq 0 \end{array} \right.$$

Pour résoudre le problème mono-critère en utilise une méthode d'optimisation mono-critère qui est la méthode de pénalité intérieure

En démarrant du point admissible (2,2), avec un paramètre initial  $t = 8$ , et un facteur d'augmentation associé  $\gamma = 0.8$  et la condition d'arrêt  $\|x^{k+1} - x^k\| < 0.1$ .

$k$	$x^k$	$f(x^k)$	$P(x^k, t^k)$	$t^k$	$x^{k+1}$
1	(2.00 , 2.00 )	7.00	18.33	8.00	(2.92 , 1.69)
2	(2.92 , 1.69)	2.37	16.07	6.40	(3.10 , 1.70)
3	(3.10 , 1.70)	1.69	14.25	5.12	(3.29 , 1.72)
4	(3.29 , 1.72)	0.98	12.80	4.10	(3.48 , 1.75)
5	(3.48 , 1.75)	0.28	11.64	3.28	(3.66 , 1.78)
6	(3.66 , 1.78)	-0.38	10.71	2.62	(3.82 , 1.81)
7	(3.82 , 1.81)	-1.00	9.97	2.10	(3.96 , 1.84)
8	(3.96 , 1.84)	-1.57	9.38	1.68	(4.08 , 1.86)
9	(4.08 , 1.86)	-2.08	8.90	1.34	(4.19 , 1.88)
10	(4.19 , 1.88)	-2.54	8.52	1.07	(4.28 , 1.90)
11	(4.28 , 1.90)	-2.96	8.22	0.86	-

**-Exécution de la méthode des poids pour le problème (P) depuis le point admissible (2,2).**

**Remarque:**

Si on fixe d'autres valeurs aux poids  $\lambda_i$ , on aura d'autres résultat.

## Conclusion générale

Au cours de cette étude notre savoir a été enrichi en découvrant le domaine de l'optimisation non linéaire et de programmation.

Ce projet très intéressant nous a permis de concilier un travail théorique avec une tâche pratique d'implémentation. IL nous a permis d'apprécier l'efficacité des divers types de méthodes, de juger de leur convergence.

IL nous reste tant à découvrir et le sujet n'a bien sur pas été épuisé à travers ce document et de nombreux défis restent ouverts, en particulier l'implémentation de nouvelles méthodes.

Finalement, nous espérons que ce travail soit bénéfique, et avoir apporté une contribution pour les prochaines recherche.

# Bibliographie

- [1] Acherir Mohand ouahmed, "*Optimisation d'un problème de programmation mathématique non linéaire multi-critère*", Mémoire de fin d'étude d'ingénieur, UMMTO, 2006-2007.
- [2] Bachir Cherif Kahina, "*Les méthodes d'optimisation et implémentation*", Mémoire de master, UMMTO 2011.
- [3] Mokhtar S. Bazarra, Hanif D, N. Sherali, C.M. Shetty, "*Nonlinear programming, Theory and Algorithms*", John Wiley et Son, Inc, 1993.
- [4] Bertsekas Dimitri.P, "*Nonlinear programming*", Second printing, Athena scientific.
- [5] Bierlaire Michel, "*Introduction à l'Optimisation Différentiable*", Press Polytechniques et Universitaires Romandes.
- [6] "*Méthodes Numériques, Algorithmes, Analyse et Applications*", Edition Springer-Verlag Italia, Milano 2007.
- [7] Diallo Thierno, "*Etude et illustration des méthodes itératives d'optimisation non linéaire*", Mémoire de master, Ecole Polytechnique fédérale de Lausanne, 2005-2006.
- [8] Gilbert Jean charles, "*Elements d'optimisation différentiable, Théorie et Algorithmes*".
- [9] Guettaf Rabah, "*Réalisation d'un logiciel des problèmes d'optimisation multi-objectif linéaire et non linéaire*", Mémoire de magistère, UMMTO.

- [10] Guy Cohen, "*Convexité et Optimisation*", Ecole Nationale des Ponts et Chaussées et Inria. Edition (2000-2006).
- [11] Hwang C.L et Masud A.S.M, "*Multi Objective Decision Making-Methods and applications* ", A set of the art Suvey, Lecture notes in Economics and Mathematical Systems 164, Springer Verlag Berlin Edelberg,1979.
- [12] Lootsma F.A, "*Numerical Méthods for Nonlinear optimisation*", Academic Press 1971.
- [13] Miettinen K "*Nonlinear Multiobjective Optimisation* ", Kluwer Academic Publishers, Boston / London / Dordrecht, 1999.
- [14] Minoux Michel , "*Programmation Mathématique Théorie et Algorithmes* "; Collection Théchnique et scientifique des Télécommunication, a volume 1.
- [15] Nakayamma H, Sawargui Y, "*Satisficing, Trade-off Method for Multiobjective programming, In (Interactive Decision Analysis)* ", (Eds) M. Graurer et A.P Wierzbicki, spinger, pp.113-122,1984.
- [16] Osyezka .A, "*Multicriteria optimization for engineering design*". Dans J. S. Gero, éditeur, Design optimization ,Academic Press,, pages 193-227.1985 [Osy85].
- [17] Xavier Antoine, Pierre Dreyfuss et Yannick, "*Introduction à l'optimisation: Aspects Théoriques et Algorithmes*", ENSMN-ENSEM 2A (2006-2007).