

Ministère de l'enseignement Supérieur et de la Recherche Scientifique



Université Mohamed El-Bachir El-Ibrahimi de
Bordj Bouararidj.
Faculté des mathématiques et informatique.
Département de Recherche Opérationnelle.



Mémoire de Master

La méthode de direction de descente pour
l'optimisation multi-objectifs sans contraintes

Membres du jury :

Encadreur : Dr RAMDANI ZOUBIR

Présidente : Dr TOUATI HILLAL

Examineur : Dr MAZA SOFIANE

Présenté par :

Djeddou Chaima

Bennoui Rebh

septembre 2021

Au nom de Dieu, le plus miséricordieux, le plus compatissant

Remerciements et louange à Dieu Tout-Puissant pour son succès. Pour lui attribue tout le crédit pour l'achèvement de ce travail humble, et nous témoignons qu'il n'y a pas d'autre dieu que Dieu seul, qui n'a pas de partenaire. Et que les prières et la paix soient sur le Messager de Dieu, Muhammad, que les prières et la paix de Dieu soient sur lui

Mon père, j'écris ces mots et des larmes coulent de moi, ces lignes ne suffiront pas à te remercier, oh ma chose la plus précieuse, Tu as consacré ta vie à rendre ma vie meilleure et plus belle. Tu m'as donné tout ce que tu possèdes pour réaliser mes rêves et mes ambitions. Tu m'as toujours poussé en avant pour réaliser ce que je voulais. Tu m'as soutenu de moi dans tous les revers et les échecs que j'ai traversé dans ma vie.

avec fierté je le dis tu es Mon soutien... ma force... ma confiance ... ma vie ... tout l'amour pour toi mon père

À la « Reine de mon cœur »... à « La lumière de notre maison » ... à « Source de mon inspiration » ...qu'ils ont dit "Paradis est sous les pieds des mères". Tous les mots ne suffisent pas pour te décrire, ma mère. Je te remercie pour ton soutien, ton amour sans fin, ton sacrifice, et ta patience Pour moi... Maman, tu es l'ange le plus cher au cœur et aux yeux.

Je dis fièrement : Tu es ma force... mon inspiration... ma chaleur... mon amour... ma famille

"Maman et papa, merci pour tout"

Ma chère sœur, je n'oublierai pas que tu m'as porté dans mes jours les plus difficiles et m'a empêché d'effectuer les tâches ménagères pendant que je travaillais sur cette mémoire... Mon amour et ma sœur ICHRAK, tu as tout l'amour de moi, mon compagnon et âme sœur, je t'aime au-dessus de l'amour et merci pour tout.

je remercie Monsieur RAMDANI ZOUBIR me encadreur de mémoire pour sa orientation, sa confiance et sa dévouement et un apport considérable sans lequel ce travail n'aurait pas pu être mené à bon port... je tiens ensuite à remercier tout spécialement Monsieur BERKANI

pour pour son soutien et sa motivation pendant ces trois dernières années...grâce à vos encouragements constants, je crois toujours en moi et je crois que je peux...

A ma grand-mère, mon grand-père, ma grand-mère, mes oncles, les femmes de mes oncles, mes tantes, et surtout ma chère tante et ma seconde sœur, Aida je ne peux pas manquer de vous dédier ce travail et merci pour tout... Dieu sait combien je t'aime et te respecte.

mes chers amis : Kenza, Hayate, Khalissa, jojo, Chaima, Hafida, Sabrin, Meriem, Nadjima, Hanene, Samia, Fatima, Amina, Sara, Meroua, Nour el houda, Manel, Meriem, je n'oublierai pas les meilleurs jours que nous avons passés ensemble.

A tous mes amis et camarades de classe qui ont partagé cette expérience avec moi . . .

Merci à tous.

CH.Djeddou

On remercie dieu le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire. Je voudrais remercier vivement mon encadreur M. Ramdani ZOUBIR pour avoir accepté de diriger ce mémoire, pour sa gentillesse, sa bonne volonté sa disponibilité et sa patience ainsi ces orientations et ces guidances avisés et son soutien indéfectible durant la préparation de ce travail, dès le début de sa confiance à mon travail m'a donnée une énergie et une inspiration de soulever toutes les difficultés.

Je voudrais remercier vivement chaima et rebh

Mes remerciements les plus vifs vont également à tous ceux qui m'ont aidé d'une façon ou d'une autre, à tout personne ayant participé de près ou de loin à la réalisation de ce travail.

A toutes ces personnes j'exprime ma reconnaissance.

R.Bennoui

Remerciements

Nous tenons à remercier notre Créateur "ALLAH" Qui a rendu ce travail possible.

Ensuite nous tenons à exprimer notre sincère gratitude à notre superviseur Dr. " RAMDANI ZOUBIR " pour sa patience à nous guider et son dévouement continu au cours de ce voyage.

Nous remercions aussi le jury des examinateurs qui a accepté de lire notre humble travail et merci à eux pour les précieuses remarques.

Enfin, nous n'oublierons jamais de remercier sincèrement tous nos enseignants qui nous ont enseigné pendant nos études à l'université.

Table des matières

1	Optimisation non linéaire sans contraintes	8
1.1	Introduction	8
1.2	Aspects préliminaires	8
1.2.1	Normes et distances	8
1.2.2	Matrices définies	10
1.2.3	Éléments de différentiabilité	11
1.2.4	Notions de convexité	12
1.2.5	Direction de descente	14
1.3	Conditions d'optimalité	15
1.4	Méthode de direction de descente pour résoudre un programme non linéaire sans contraintes	18
2	Optimisation non linéaire multi-objectifs sans contraintes	23
2.1	Introduction	23
2.2	Notions de dominance	24
2.3	Solutions de Pareto	25
2.3.1	Aspects généraux	25
2.3.2	Front de Pareto (la surface du compromis)	28
2.3.3	Les points caractéristiques	28
2.4	Caractérisation des solutions de Pareto	30
2.4.1	Paramètres de préférence	31
2.5	Méthodes de résolution	32
2.5.1	Méthode de somme pondérée	32
2.5.2	La méthode du compromis	38

2.5.3	La méthode du Chebychev	42
2.6	Codage des méthodes avec MATLAB	44
2.6.1	Code de la méthode de somme pondérée	44
2.6.2	Code de la méthode du compromis	46
3	Méthodes de direction de descente pour l'optimisation multi-objectifs sans contraintes	49
3.1	Introduction	49
3.2	Description de la méthodes de la direction de descente	50
3.2.1	Calcul de la direction de descente	51
3.2.2	Calcul du pas optimal selon la règle d'Armijo	52
3.2.3	L'algorithme de la direction de descente	53
3.2.4	Convergence de la méthode de direction de descente	53
3.2.5	Exemple d'illustration	54
3.3	Implémentation avec MATLAB	56

Table des figures

1.1	fonction convexe-concave	13
1.2	Minimum local et global	16
2.1	Le front de Pareto	28
2.2	Le point idéal et le point nadir	30
2.3	Le principe de la méthode de pondération des fonctions objectif	33
2.4	Une difficulté insurmontable pour la méthode de pondération des fonctions objectives	34
2.5	espace de décision de l'exemple 2.5	35
2.6	espace de critère de l'exemple 2.5	36
2.7	Les points non domines trouvé pour l'exemple	37
2.8	Résolution graphique de l'exemple 2.7	40
2.9	Les solutions non dominées de l'exemple 2.7	41
2.10	Principe des méthodes de normes ou de distances pondérées	43
2.11	Le front de Pareto obtenu par MATLAB de l'exemple 2.5	46
2.12	Front de Pareto de l'exemple 2.7 avec la méthode ε -contraintes	48
3.1	Espace objectif de l'exemple 3.13	55
3.2	Front de Pareto de l'exemple 3.17 avec la méthode de pondération	57
3.3	Front de Pareto de l'exemple 3.17 avec la méthode de direction de descente	57
3.4	Front de Pareto de l'exemple 3.18 avec la méthode de direction de descente	58
3.5	Front de Pareto de l'exemple 3.18 avec la méthode de pondération	59

Liste des tableaux

2.1	Résultats pour quelques valeur de λ_2 de la méthode pondérée	37
2.2	Résultats avec la méthode ε -contrainte de l'exemple 2.7	41

Introduction générale

La recherche opérationnelle est un ensemble de techniques récentes orientées vers la recherche du meilleur choix en vue d'aboutir le meilleur résultat possible, les premiers travaux concrets de la recherche opérationnelle datent tout à fait de la Seconde Guerre mondiale lorsque la marine américaine fit appel aux mathématiciens, économistes, statisticiens et autres ingénieurs et chercheurs pour participer à la localisation et la réalisation d'une façon optimale au terme de coûts et du nombre, des stations d'anti-espionnage tout au long du territoire américain. En d'autre part la recherche opérationnelle peut être considérée comme un carrefour associant étroitement les méthodes et techniques de l'économie d'entreprise, les mathématiques et l'informatique.

L'optimisation est une branche des mathématiques appliquées et de la recherche opérationnelle. Elle possède ses origines depuis des siècles lorsque le philosophe Grec Pythagore formule son célèbre théorème, où dans un triangle rectangle, le carré de la longueur de l'hypoténuse est égale à la somme des carrés des longueurs des deux autres côtés. Puis utiliser par le célèbre mathématicien Grec Euclide dans le calcul et la minimisation des distances. En effet, L'optimisation est une branche des mathématiques cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes qui consistent à minimiser ou maximiser une fonction sur un ensemble bien définis.

L'optimisation joue un rôle important en recherche opérationnelle, domaine ou carrefour où se rencontre les trois disciplines à savoir l'informatique, les mathématiques et l'économie. L'optimisation intervient pratiquement dans tout les domaines les mathématiques fondamentales tel que l'industrie et l'ingénierie, en analyse numérique, en statistique pour l'estimation du maximum de vraisemblance d'une distribution, pour la recherche de stratégies dans le cadre de la théorie des jeux, ou encore en théorie du contrôle et de la commande.

L'optimisation multi-objectifs est un domaine d'étude important en recherche opérationnelle, à cause de la nature multi-objectifs de la plupart des problèmes réels. Les premiers travaux menés sur les problèmes multi-objectifs furent réalisés au 19^{ème} siècle sur des études en économie par Francis Y. Edgeworth (1845-1926) et elle a été utilisée de manière plus formelle par l'économiste italien Vilfredo Pareto (1848-1923). En effet, L'optimisation multi-objectifs permet de saisir de manière adéquate les caractéristiques essentielles des problèmes du monde réel et d'améliorer leur perception par les décideurs.

Dans la plupart des problèmes du monde réel, il ne s'agit pas d'optimiser seulement un seul critère mais plutôt d'optimiser simultanément plusieurs critères et qui sont généralement conflictuels, en production par exemple, on veut souvent minimiser le coût d'un produit mais aussi maximiser sa qualité. Contrairement à l'optimisation impliquant un seul critère où on cherche une solution dite optimale, la résolution d'un problème d'optimisation multi-objectifs consiste à déterminer un ensemble de solutions dites solutions efficaces où solutions non dominées ou solutions optimales au sens de Pareto qui correspondent au mieux aux préférences du décideur parmi les solutions dites solutions de bonne compromis. En effet, dans l'optimisation multi-objectifs les objectifs que nous voulons optimiser sont toujours contradictoires dans le sens où il n'y a pas de décision qui optimise toutes les fonctions d'objectifs simultanément.

Pour la résolution des problèmes multi-objectifs, il existe plusieurs méthodes numériques et méta heuristiques parmi les plus utilisées on trouve des méthodes scalaires où paramétriques où il s'agit de transformer le problème multi-objectifs en question en un problème mono-objectifs où la solution optimale de ce dernier possède une relation étroite avec la solution optimale au problème multi-objectifs. Parmi les méthodes scalaires les plus connues on peut citer la méthode de pondération des fonctions objectifs et ϵ -contraintes où plusieurs paramètres sont utilisés, ces dernières souffrent de plusieurs critiques tel que le choix des valeurs des paramètres qui pose des difficultés pour le décideurs et leur incapacité de chercher les solutions optimale dans la cas de la non-convexité des problèmes multi-objectifs.

Dans le cadre de ce mémoire nous considérons le problème d'optimisation multi-objectifs non linéaire et sans contraintes de forme $\min_{x \in \mathbb{R}^n} F(x)$ tel que $F(x)$ est un vecteur de fonctions \mathbb{R}^k qu'on note par (*MOP*). L'objectif de ce travail consiste à étudier en détail une méthode de direction de descente qui consiste à trouver une partie du front de Pareto où un sous ensemble de solutions efficaces. La méthode en question n'utilise aucun paramètre à fixer par l'utilisateur. Contrairement aux méthodes de pondération des fonctions objectifs et ϵ -contraintes où la résolution peut intégrer des sous-programmes complexes, la méthode de direction de descente consiste en résolution d'un problème quadratique semi-défini facile à résoudre. Notons aussi, dans les méthodes paramétrique la génération d'un sous-ensemble de l'ensemble Pareto peut se faire en changeant les valeurs des paramètres, par contre dans la méthode de direction de descente la détermination d'un sous-ensemble de Pareto nécessite un sous-ensemble de solutions initiales. En revanche, une modeste étude sera effectuée sur quelques exemples de (*MOP*) convexes et non convexes en utilisant des programmes des méthodes implémentés sur MATLAB et CPLEX.

Le présent travail s'organise comme suit :

- Le chapitre 1 récapitule des notions de bases et préliminaires concernant l'optimisation mono-objectif sans contraintes. Les conditions d'optimalité et quelques méthodes numériques sont aussi brièvement rappelés.
- Dans le chapitre 2, nous rappelons brièvement les notions de base concernant l'optimisation multi-objectifs non linéaire sans contraintes. Notamment, les notions d'optimalité au sens de Pareto et les deux méthodes scalaires à savoir la méthode de pondération des fonctions objectifs et la méthode ϵ -contraintes.
- Par contre, dans le 3ème chapitre nous allons étudier en détail une méthode de direction de descente développée par (Svaiter et Fliege, 2000) qui consiste à trouver un sans en-

semble du front de Pareto. Nous utilisons les logiciel MATLAB et CPLEX pour résoudre les différents exemples de test. Une modeste étude comparative sera réaliser afin de voir l'avantage des méthodes non paramétriques sur la méthode de pondération classique.

Chapitre 1

Optimisation non linéaire sans contraintes

1.1 Introduction

La programmation non linéaire sans contraintes (*PNLSC*) traite le problème d'optimisation d'une fonction objectif d'une ou de plusieurs variables sans contraintes dans l'espace \mathbb{R}^n .

Un problème d'optimisation sans contraintes se définit comme la recherche du minimum d'une fonction donnée, et écrit sous la forme suivant :

$$(\text{PNLSC}) \begin{cases} \min f(x) \\ x \in \mathbb{R}^n \end{cases} \quad (1.1)$$

Où $x = (x_1, x_2, x_3, \dots, x_n)^T$ est le vecteur de n variables de décision, $f(x)$ une fonction objectif à optimiser.

Dans l'optimisation sans contraintes la/les fonction(s) devraient être non linéaire.

L'objectif de ce chapitre est de présenter la convexités des fonctions, et les conditions d'optimalités suffisantes et nécessaires du premier et du second ordre, ainsi que dans notre travail nous présentons des Méthodes pour résolution des problèmes non linéaire sans contraintes.

1.2 Aspects préliminaires

1.2.1 Normes et distances

Les normes sont très utilisées et très importante dans l'analyse des problèmes d'optimisation. En géométrie, la norme Euclidienne est une extension de la valeur absolue des nombres aux vecteurs. En effet, la norme permet de mesurer la longueur commune à toutes les représentations d'un vecteur dans un espace affine, mais définit aussi une distance entre deux vecteurs invariants par translation.

Une manière naturelle de générer une norme sur \mathbb{R}^n est de prendre n'importe quel produit interne $\langle \cdot, \cdot \rangle$ sur \mathbb{R}^n et de définir la norme associée

$$\|\mathbf{x}\|_p \equiv \sqrt[p]{\sum_{i=1}^n |x_i|^p}$$

Pour tout $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, On définit sur \mathbb{R}^n les normes classiques bien connues suivantes :

Pour $p = 1$ on définit la norme $\|x\|_1 = \sum_{i=1}^n |x_i|$

Pour $p = 2$ on définit la norme plus utilisée est appelé la norme euclidienne

$$\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$$

Pour $p \rightarrow \infty$ on définit la norme

$$\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$$

Définition 1. (Norme Euclidienne)

Soit un vecteur $x = (x_1, x_2, \dots, x_n)$ un vecteur de \mathbb{R}^n , la norme euclidienne de x notée par $\|x\|_2$ est calculée par la formule suivante :

$$\|x\|_2 = \left[\sum_{i=1}^n (x_i)^2 \right]^{1/2} \quad (1.2)$$

Par exemple, soit $x = (2; 5; -3)$ un vecteur dans l'espace \mathbb{R}^3 , la norme Euclidienne de x est $\|x\|_2 = \sqrt{2^2 + 5^2 + (-3)^2} = \sqrt{38}$.

Propriété 1. On définit les propriétés de la norme Euclidienne :

1. $\forall x \in \mathbb{R}^n : \|x\| \geq 0$ si et seulement si $x = 0$.
2. Soit $x \in \mathbb{R}^n, \lambda \in \mathbb{R} : \|\lambda x\| = |\lambda| \times \|x\|$. En particulier, on a $\| -x \| = \|x\|$.
3. Soit $x, y \in \mathbb{R}^n : \|x + y\| \leq \|x\| + \|y\|$ et $\|x - y\| \leq \|x\| + \|y\|$.

Dans le langage courant, la distance ou l'éloignement est la longueur qui sépare deux points, ce qui est mesurée par la longueur du segment qui les relie. Dans le cas de deux points à la surface de la Terre, la distance s'entend à vol d'oiseau ou parfois par la route, en train, etc.

On appelle distance sur \mathbb{R}^n toute application $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ vérifiant les propriétés suivantes :

1. (symétrie) $d(x, y) = d(y, x)$ pour tout $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$.
2. (séparation) $d(x, y) = 0 \Leftrightarrow x = y$ pour tout $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$.
3. (inégalité triangulaire) $d(x, z) \leq d(x, y) + d(y, z)$ pour tout $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$.

Noter également que la définition précédente implique aussi $|d(x, z) - d(y, z)| \leq d(x, y)$, pour tout $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$.

Propriété 2. Si \mathbb{R}^n est un espace muni d'une norme N , alors l'application d définie sur $\mathbb{R}^n \times \mathbb{R}^n$ par $d(X, Y) = N(X - Y)$ est une distance. c'est la distance associée à la norme N .

Exemple 1. (Distances classiques)

Dans l'espace \mathbb{R} , la distance entre deux point défini par $d(x, y) = |x - y|$ est la distance usuelle. Et dans l'espace \mathbb{R}^n , à chacune des trois normes classiques de \mathbb{R}^n correspond une distance.

pour tout $x = (x_1, x_2 \dots, x_n)$ et tout $y = (y_1, y_2 \dots, y_n)$ on considère souvent les distances :

$$d_1(x, y) := \sum_{i=1}^n |x_i - y_i|, \quad d_\infty(x, y) := \max_{i=1, \dots, n} |x_i - y_i|$$

Et la distance euclidienne :

$$d_2(x, y) := \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

1.2.2 Matrices définies

Dans cette section, nous présentons l'étude de la nature des matrices carrées et ses différentes propriétés. Notons que l'étude de la nature des matrices est un aspect important pour étudier la nature des fonctions au terme de convexité.

En fait, pour vérifier la nature d'une matrice carrée donnée, il existe deux techniques classiques à savoir l'utilisation des valeurs propres et la technique des mineurs principaux du critère de Sylvestre.

Définition 2.

Soit M une matrice carrée et symétrique à éléments réels d'ordre n , M est dite *définie positif* si et seulement si $x^T M x > 0, \forall x \in \mathbb{R}^n, \|x\| \neq 0$.

M est dite *définie négatif* (semi définie négatif) si son opposée $-M$ est définie positive (semi définie positif). M est dite n'est pas définie si et seulement si elle n'est pas semi définie positif et n'est pas semi définie négatif.

a) Valeurs propres

Définition 3. Soit $M \in \mathbb{R}^{n \times n}$ une matrice carrée et symétrique, un scalar λ et un vecteur v non nul, où l'équation

$$Mv = \lambda v \tag{1.3}$$

Si cette équation est vérifiée, λ s'appelle valeur propre de la matrice M et v est le vecteur propre correspondant. Pour calculer les valeurs propres de M , nous résolvons le polynôme caractéristique de la matrice M se définit comme l'équation suivante :

$$\det(M - I\lambda) = 0 \tag{1.4}$$

Les valeurs propres de M sont les racines du polynôme caractéristique.

La caractérisation de la définition d'une matrice à l'aide des valeurs propres est illustrée dans le théorème suivant.

Théorème 1.

Soit $M \in \mathbb{R}^{n \times n}$ une matrice carrée et symétrique, et λ représente le vecteur des valeurs propres de la matrice M qui sont réelles.

- M est défini positif si et seulement si toutes ses valeurs propres λ sont positives. M est semi-défini positif si et seulement si toutes ses valeurs propres λ sont non négatives.
- M est défini négatif ou semi-défini négatif si et seulement si toutes ses valeurs propres λ sont négatives ou non positives.
- M est indéfini si et seulement s'il a à la fois des valeurs propres positives et négatives.

b) Critère de Sylvestre

Le critère de Sylvestre fournit une autre méthode simple permettant d'étudier la nature d'une matrice carrée.

Définition 4. (Mineur principal) Soit $M \in \mathbb{R}^{n \times n}$ une matrice carrée et symétrique, on note $A^k, k = 1, 2, \dots, n$, la sous-matrice $k \times k$ prise dans le coin supérieur gauche de M , où

$$A^k = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix}$$

En particulier, $A^1 = a_{11}$ et $A^n = M$. Un déterminant de A^k est appelé un mineur principal d'ordre k de la matrice M et note par $\Delta_k = \det(A^k)$.

La caractérisation de la définition d'une matrice à l'aide du critère de Sylvestre est illustrée dans le théorème suivant.

Théorème 2. Soit $M \in \mathbb{R}^{n \times n}$ une matrice symétrique.

1. M est semi définie positif si et seulement si $\Delta_k \geq 0$, pour tous les mineurs principaux.
2. M est définie positif si et seulement si $\Delta_k > 0$, pour tous les mineurs principaux.
3. M est définie négatif si et seulement si Δ_k a le même signe que $(-1)^k, \forall k = 1, 2, \dots, n$.

1.2.3 Éléments de différentiabilité

Dans cette section, nous donnons quelques caractérisations des fonctions différentiables. La notion de différentiabilité généralise celle de dérivées de fonctions réelles de variable réelle.

Étant donné une fonction f à plusieurs variables $f : \mathbb{R}^n \rightarrow \mathbb{R}$, les définitions suivantes nous donnent les dérivés partielles du premier ordre et de deuxième ordre.

Définition 5. On dit que f est différentiable au point a s'il existe une application linéaire continue $L : E \rightarrow F$ telle que

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

c'est-à-dire telle que l'application $\varepsilon : E \setminus \{0\} \rightarrow F$

$$\forall h \in E \quad f(a + h) = f(a) + L(h) + \|h\|\varepsilon(h)$$

vérifie :

$$\lim_{h \rightarrow 0} \varepsilon = 0$$

Si f n'est pas définie sur E tout entier mais seulement sur voisinage de a , on adopte la même définition, après avoir prolongé f à E de façon arbitraire.

Définition 6. (Gradient de fonction) Une fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dit être continuellement différentiable en $x \in \mathbb{R}^n$, si $\frac{\partial f(x)}{\partial x_i}$ existe et contenue, $i = 1, \dots, n$. Le gradient de f en x est défini comme

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right]^T \quad (1.5)$$

Si f est continuellement différentiable en tout point de \mathbb{R}^n , alors f est dit continuellement différentiable sur \mathbb{R}^n et noté par $f \in C^{(1)}(\mathbb{R}^n)$.

Définition 7. (Hessien de fonction)

Une fonction continuellement différentiable $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est appelé deux fois continuellement différentiable en $x \in \mathbb{R}^n$, si $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ existe et contenue, $i, j = 1, 2, \dots, n$. Le Hessien de f est défini comme la matrice symétrique $n \times n$ avec des éléments

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x), 1 \leq i, j \leq n \quad (1.6)$$

Si f est deux fois continuellement différentiable en tout point de \mathbb{R}^n , alors f est dit deux fois continuellement différentiable sur \mathbb{R}^n et noté par $f \in C^{(2)}(\mathbb{R}^n)$.

1.2.4 Notions de convexité

La théorie de la convexité est d'une grande importance dans l'étude des problèmes d'optimisation. Dans cette section, nous introduisons les concepts fondamentaux et les résultats des fonctions convexes.

Définition 8. (Fonction convexe)

On dit qu'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est convexe sur l'espace \mathbb{R}^n , si pour tout $x^1, x^2 \in \mathbb{R}^n$ et tout $\alpha \in [0, 1]$ on a

$$f(\alpha x^1 + (1 - \alpha)x^2) \leq \alpha f(x^1) + (1 - \alpha)f(x^2) \quad (1.7)$$

Ensuite, une fonction f est dite strictement convexe sur \mathbb{R}^n , si l'inégalité ci-dessus est vraie en tant que inégalité est stricte pour tout $x^1 \neq x^2$ et tout $\alpha \in]0, 1[$, c'est-à-dire,

$$f(\alpha x^1 + (1 - \alpha)x^2) < \alpha f(x^1) + (1 - \alpha)f(x^2) \quad (1.8)$$

Alors f est appelée une fonction uniformément (ou fortement) convexe sur \mathbb{R}^n . Si $-f$ est une fonction convexe (strictement convexe) sur \mathbb{R}^n , alors on dit que f est une fonction concave (strictement concave).

La figure 1.1 montre des exemples de fonction convexe, concave et ni convexe ni concave.

L'interprétation géométrique de cette définition est que le graphe d'une fonction convexe est toujours en dessous du segment reliant les points $(x^1, f(x^1))$ et $(x^2, f(x^2))$.

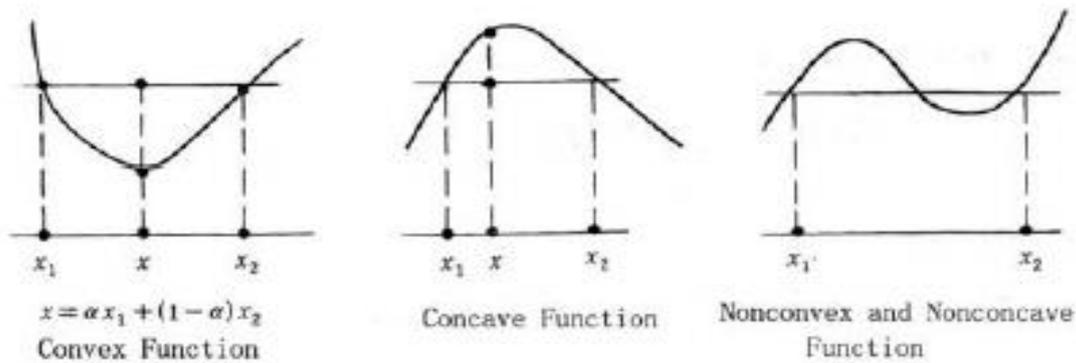


FIGURE 1.1 – fonction convexe-concave

- Théorème 3.**
1. Soient f une fonction convexe sur l'espace \mathbb{R}^n et un nombre réel $\alpha \geq 0$, alors αf est aussi une fonction convexe sur \mathbb{R}^n .
 2. Soient f_1, f_2 deux fonctions convexes sur l'espace \mathbb{R}^n . Alors $f_1 + f_2$ est aussi une fonction convexe sur \mathbb{R}^n .
 3. Soient f_1, f_2, \dots, f_m des fonctions convexes sur l'espace \mathbb{R}^n et $\alpha_1, \alpha_2, \dots, \alpha_m \geq 0$, des nombres réels. Alors $\sum_{i=1}^m \alpha_i f_i$ est aussi une fonction convexe sur \mathbb{R}^n .

Dans les théorèmes suivants nous donnons les caractérisations et les relations importantes entre la différentiabilité et les fonctions convexes.

Théorème 4. soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et différentiable sur \mathbb{R}^n . Alors f est convexe si et seulement si

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1), \forall x_1, x_2 \in \mathbb{R}^n$$

De même, f est strictement convexe sur \mathbb{R}^n si et seulement si

$$f(x_2) > f(x_1) + \nabla f(x_1)^T (x_2 - x_1), \forall x_1, x_2 \in \mathbb{R}^n, x_1 \neq x_2$$

Le théorème suivant donne une autre caractérisation nécessaire et suffisante des fonctions convexes différentiables.

Théorème 5. Supposons que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et différentiable sur l'espace \mathbb{R}^n . Alors f est convexe sur \mathbb{R}^n si et seulement si son gradient ∇f est monotone, c'est-à-dire,

$$\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle \geq 0, \forall x_1, x_2 \in \mathbb{R}^n$$

De même, f est strictement convexe sur \mathbb{R}^n si et seulement si son gradient ∇f est strictement monotone, c'est-à-dire,

$$\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2 \rangle > 0, \forall x_1, x_2 \in \mathbb{R}^n, x_1 \neq x_2$$

Les fonctions convexes et concaves ont de nombreuses propriétés spéciales et importantes. Le théorème suivant est utile pour vérifier la convexité ou la concavité d'une fonction deux fois différentiable.

Théorème 6. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et deux fois différentiable. Alors

1. f est une fonction convexe si et seulement si sa matrice Hessien est semi-définie positive en tout point de \mathbb{R}^n .
2. f est strictement convexe si et seulement si sa matrice Hessien est définie positive en tout point de \mathbb{R}^n .
3. f est une fonction concave si et seulement si sa matrice Hessien est semi définie négatif en tout point de \mathbb{R}^n , De plus, f est strictement concave si et seulement si sa matrice Hessien est définie négatif.

Il y a des fonctions à une seul variable que nous ne l'utilisons pas la matrice Hessien pour vérifier si la fonction est convexe ou pas.

Théorème 7. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ est deux fois dérivable sur \mathbb{R} , alors f est convexe si et seulement si $f''(x) \geq 0, \forall x \in \mathbb{R}$. De plus, f est une fonction concave sur \mathbb{R} si et seulement si $f''(x) \leq 0, \forall x \in \mathbb{R}$.

1.2.5 Direction de descente

Dans cette section nous donnant une classe importante d'algorithmes de résolution des problèmes d'optimisation sans contrainte. Le concept central est celui de direction de descente.

Définition 9. (direction de descente).

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continument différentiable sur \mathbb{R}^n . Un vecteur $d = (d_1; d_2; \dots; d_n)$ est appelé une direction de descente de f en $x \in \mathbb{R}^n$ si la dérivée directionnelle $f'(x; d)$ est négative, ce qui signifie que

$$f'(x; d) = \nabla f(x)^T d < 0, d \neq 0$$

La propriété la plus importante des directions de descente est que faire des pas suffisamment petits le long de ces directions conduit à une diminution de la fonction objectif.

Dans les lemmes suivants nous présente les propriétés de descente des directions de descente.

Lemme 1. Soit f une fonction continument différentiable sur \mathbb{R}^n , et soit $x \in \mathbb{R}^n$. Supposons que d est une direction de descente de f en $x \in \mathbb{R}^n$. Alors il existe $\varepsilon > 0$ tel que

$$f(x + td) < f(x)$$

Pour toute $t \in (0, \varepsilon]$.

Lemme 2. Soit f une fonction continument différentiable sur \mathbb{R}^n , et soit $x \in \mathbb{R}^n$. Supposons que $0 \neq d \in \mathbb{R}^n$ est une direction de descente de f en $x \in \mathbb{R}^n$ et soit $\alpha \in (0, 1)$. Alors il existe $\varepsilon > 0$ tel que l'inégalité

$$f(\mathbf{x}) - f(\mathbf{x} + t\mathbf{d}) \geq -\alpha t \nabla f(\mathbf{x})^T \mathbf{d}$$

est valable pour toute $t \in [0, \varepsilon]$.

Propriété 3. (Caractérisation des directions de descente).

Soit $f \in C^1(\mathbb{R}^n, \mathbb{R})$, $x \in \mathbb{R}^n$ et $d \in \mathbb{R}^n \setminus \{0\}$; alors

1. Si d direction de descente en x alors $d \cdot \nabla f(x) \leq 0$.
2. Si $\nabla f(x) \neq 0$ alors $d = -\nabla f(x)$ est une direction de descente stricte en x .

Définition 10. (Direction admissible).

Un vecteur $d = (d_1; d_2; \dots; d_n)$ est appelé direction admissible en point $x^* \in \mathbb{R}^n$ selon la fonction f si elle vérifie :

1. $x^* + td \in \mathbb{R}^n \quad \forall t \in [0, t_0], t_0 > 0$.
2. $f(x^* + td) \leq f(x^*)$.

Avec t est le pas de déplacement de x^* au point $x^* + td$ selon la direction d .

1.3 Conditions d'optimalité

Dans cette partie, on s'intéresse aux conditions d'optimalité d'un problème d'optimisation de la forme suivante : $\min f(x), x \in \mathbb{R}^n$ d'une fonction réelle non linéaire. En particulier, nous considérons brièvement les conditions d'optimalité du premier et du second ordre pour des problèmes sans contrainte, afin d'analyser ou de résoudre de manière efficace un problème d'optimisation et de vérifier la validité des solutions obtenues. En général, nous avons deux types de minimum : le minimum local et le minimum global. Dans ce qui suit, nous donnons leurs définitions exactes.

Définition 11. (Optimum local)

Un point x^* est appelé minimum local s'il existe $\delta > 0$ tel que $f(x^*) \leq f(x)$ pour tout $x \in \mathbb{R}^n$ satisfaisant $\|x - x^*\| < \delta$.

Un point x^* est appelé minimum local strict s'il existe $\delta > 0$ tel que $f(x^*) < f(x)$ pour tout $x \in \mathbb{R}^n$ avec $x \neq x^*$ et $\|x - x^*\| < \delta$.

Définition 12. (Optimum global)

Un point x^* s'appelle un minimum global si $f(x^*) \leq f(x)$ pour tous $x \in \mathbb{R}^n$. Un point x^* s'appelle un minimum global strict si $f(x^*) < f(x)$ pour tous $x \in \mathbb{R}^n$ avec $x \neq x^*$.

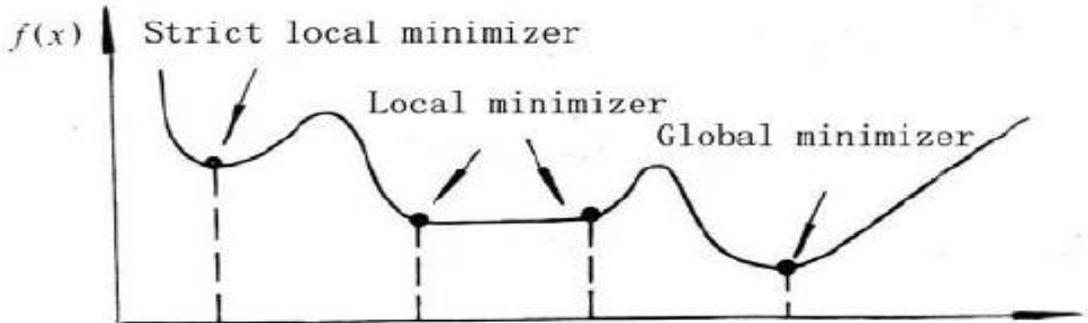


FIGURE 1.2 – Minimum local et global

Toutes les théorèmes qui sera dans cette partie proviennent de [3].

Théorème 8.

soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction convexe sur \mathbb{R}^n . Considérons le problème pour minimise $f(x)$ tel que $x \in \mathbb{R}^n$. Supposons que $x^* \in \mathbb{R}^n$ est une solution optimale locale au problème

1. Alors x^* est une solution optimale globale.
2. Si x^* est un minimum local strict ou si f est strictement convexe, alors x^* est l'unique solution optimale globale.

Nous discutons maintenant la condition nécessaire du premier ordre pour que x^* est un optimum local et la condition nécessaire du second ordre en utilisant la matrice Hessien.

Théorème 9. (Condition nécessaire du premier ordre)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuellement différentiable, si x^* est un minimum local de f alors

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

Le point x^* qui vérifient ces conditions s'appelle point stationnaire ou point critique de f .

Démonstration. (par contradiction)

Supposons que $\nabla f(x^*) \neq 0$ et $d = -\nabla f(x^*)$, le produit

$$d^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Donc, d est une direction de descente et il existe $\delta > 0$ tel que

$$f(x^* + \alpha d) < f(x^*), \forall \alpha \in (0, \delta)$$

Ce qui contradiction avec l'hypothèse que x^* est un minimum local.

Théorème 10. (Condition nécessaire du second ordre)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ deux fois continuellement différentiable et $x^* \in \mathbb{R}^n$, si x^* est un minimum local de f , alors $\nabla f(x^*) = 0$ et la matrice $\nabla^2 f(x^*)$ est semi-définie positif.

Démonstration. (par contradiction)

Supposons que $\nabla^2 f(x^*)$ n'est pas semi-défini positif, alors on peut choisir $d \in \mathbb{R}^n$ tel que $d^T \nabla^2 f(x^*) d < 0$. Pour $f \in C^2$, il existe $\delta > 0$ et on peut choisir $\varepsilon > 0$ tel que $x^* + \varepsilon d \in B(x^*, \delta)$ et $d^T \nabla^2 f(x^* + \varepsilon d) d < 0$. Par l'utilisation de $\nabla f(x^*) = 0$, il en résulte que

$$f(x^* + \varepsilon d) = f(x^*) + \frac{1}{2} \varepsilon^2 d^T \nabla^2 f(x^* + \theta \varepsilon d) d$$

Où $0 \leq \theta \leq 1$. Par conséquent $f(x^* + \varepsilon d) < f(x^*)$. Cela contredit l'hypothèse que x^* est un minimum local.

Les conditions d'optimalités indiquées dans les théorèmes précédents sont nécessaires mais n'est pas suffisantes pour l'existence d'un optimum local. En effet le théorème suivant nous assure la condition suffisante pour obtenir un minimum local.

Théorème 11. (Condition suffisante de second ordre)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est deux fois continuellement différentiable en x^* . Si $\nabla f(x^*) = 0$ et la matrice $\nabla^2 f(x^*)$ est définie positive, alors $x^* \in \mathbb{R}^n$ est un minimum local strict.

Démonstration. Supposons que $\nabla f(x^*) = 0$ et la matrice $\nabla^2 f(x^*)$ est définie positive. Par l'expansion de Taylor, pour tout vecteur $d \in \mathbb{R}^n$ tel que $x^* + d$ se trouve dans un voisinage de x^* dans lequel $\nabla^2 f(x^* + d)$ est définie positif, on a

$$f(x^* + d) = f(x^*) + \frac{1}{2} d^T \nabla^2 f(x^* + \theta d) d$$

Où $\theta \in (0, 1)$. Ensuite, nous pouvons choisir $\delta > 0$ tel que $x^* + d \in B(x^*, \delta)$ et $d^T \nabla^2 f(x^* + \theta d) d > 0$. Par conséquent,

$$f(x^* + d) > f(x^*)$$

qui montre notre résultat.

Définition 13. (point stationnaire) On dit qu'un point $x^* \in \mathbb{R}^n$ est un point stationnaire (ou critique) pour la fonction f si $\nabla f(x^*) = 0$.

Le théorème 12 indique, pour une fonction convexe, que son minimum local est aussi un minimum global.

Théorème 12. soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ Supposons que $x^* \in \mathbb{R}^n$ est une solution optimale locale pour f

1. Si f est convexe, alors x^* est aussi un minimum global.
2. Si f est strictement convexe, alors x^* est un minimum global unique.

Le théorème 13 indique, pour une fonction convexe différentiable, que son point stationnaire est aussi un minimum global.

Théorème 13. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction convexe différentiable. Alors x^* est un minimum global si et seulement si $\nabla f(x^*) = 0$.

1.4 Méthode de direction de descente pour résoudre un programme non linéaire sans contraintes

Les méthodes de directions de descente est une famille d'algorithmes d'optimisation différentiable (l'optimisation des fonctions à variables réelles différentiable), destiné à minimiser une fonction réelle différentiable définie sur un espace Euclidien (par exemple, \mathbb{R}^n), l'espace des n -uplets de nombres réels, muni d'un produit scalaire.

Généralement, les méthodes à directions de descente qui ont été proposé pour résoudre le problème $\min_{x \in \mathbb{R}^n} f(x)$ consistent à construire une suite $\{x_k\}_{k \geq 0}$ qui converge vers le minimum local x^* vérifiant le/les conditions d'optimalités. Il convient de souligner que la plupart de ces algorithmes fonctionnent selon un schéma général consistant, à chaque itération, à se rapprocher du minimum par la résolution d'un sous-problème de minimisation.

La méthode de descente est itératif et procède donc par améliorations successives. Au point courant, un déplacement est effectué le long d'une direction de descente, de manière à faire décroître la fonction. Le déplacement le long de cette direction est déterminé par la technique numérique connue sous le nom de recherche linéaire.

Cette approche algorithmique peut être vue comme une technique de globalisation, c'est-à-dire une méthode permettant d'obtenir la convergence des itérés (sous certaines conditions) quel que soit l'itéré initial choisi. Elle s'apparente ainsi aux algorithmes à régions de confiance; ces dernières améliorent légèrement (mais parfois de manière décisive) leurs résultats de convergence mais sont plus compliquées à mettre en œuvre, ce qui limite parfois leur application.

Les algorithmes à directions de descente s'étendent aux problèmes avec contraintes simples (pourvu que la projection sur l'ensemble admissible soit aisée, peu coûteuse en temps de calcul) ou pour des problèmes avec contraintes fonctionnelles non linéaires, par l'intermédiaire de fonctions de mérite. Elles sont aussi utilisées en optimisation non lisse.

Considérons le problème d'optimisation non linéaire sans contraintes suivant :

$$\min_{x \in \mathbb{R}^n} f(x) \quad (1.9)$$

Tel que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est différentiable.

Définition 14 (Direction de descente). Un vecteur $d = (d_1; d_2; \dots; d_n)$ est appelé direction de descente au point \bar{x} si $\nabla f(\bar{x})d < 0$ tels que $f(\bar{x} + \lambda d) < f(\bar{x})$ et $\exists \delta, \lambda \in]0, \delta]$ tel que δ un réel.

Comme exemple, considérons la fonction $f(x_1, x_2) = x_1^2 + x_2^2 + x_1x_2 - x_1$ et le point $\bar{x} = (0; 0)$ la direction $d = (1; 1)$ est une direction de descente en point \bar{x} car

$$\begin{aligned} \nabla f(x_1, x_2) &= (2x_1 + x_2 - 1; 2x_2 + x_1) \\ \nabla f(\bar{x}) &= (-1; 0) \\ \nabla f(\bar{x})^t d &= (-1; 0)(1, 1) = -1 < 0 \end{aligned}$$

par contre le vecteur $d = (-1, 1)$ n'est pas une direction de descente en point \bar{x} . Le théorème suivant constitue la condition d'optimalité nécessaire d'ordre 1 qui indique qu'il n'existe aucune direction de descente de f au point optimal x^* .

Théorème 14. On suppose que le problème $\min_{x \in \mathbb{R}^n} f(x)$ admet un minimum local en x^* . Alors il n'existe aucune direction de descente de f en x^* .

Cette notion de direction de descente joue un rôle très important dans l'élaboration d'une famille de méthodes numériques appelées "méthode de directions de descente" pour résoudre $\min_{x \in \mathbb{R}^n} f(x)$. En outre, si d est une direction de descente au point \bar{x} alors $f(\bar{x} + d\lambda) \leq f(\bar{x})$.

C'est une famille de méthodes itératives qui s'appliquent sur des fonctions dérivables. Leur principe est simple, il consiste à générer une suite de points $\{x_k\}$ qui converge vers le minimum local x^* en commençant par le point x_0 , le point x_1 est donné par $x_0 + \lambda_0 d_0$ tels que d_0 est direction de descente en point x_0 et $\lambda_0 > 0$ suffisamment tel que

$$f(x_0 + \lambda_0 d_0) < f(x_0)$$

Le paramètre λ_0 est le pas à effectuer le long de la direction de descente d_0 au point x_0 .

On répète cette procédure jusqu'à avoir $\|\nabla f(x_k)\| \simeq 0$.

Un algorithme à directions de descente est donc déterminé par les paramètres d et λ

1. Pour calculer d_k il existe plusieurs techniques tels que : Méthodes de gradient, Méthodes de gradient conjugué....
2. Le pas λ peut se déterminer de différentes façons à préciser par exemple la recherche linéaire exacte, par la méthode d'Armijo ou de Goldstein et par la méthode de Wolfe.

Algorithme général de direction de descente

Initialisation : $\min_{x \in \mathbb{R}^n} f(x)$, x_0 point initial, $\epsilon > 0$ poser $k = 0$

Tant que $\|\nabla f(x_k)\| > \epsilon$ **faire**

Calculer la direction d_k

Calculer le pas λ_k tel que

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

Calculer $x_{k+1} = x_k + \lambda_k d_k$

Poser $k = k + 1$

fin tant que.

a) Méthode de recherche linéaire exacte pour calculer le pas optimal

La recherche linéaire exacte consiste à déterminer le pas optimal λ_k , c'est à dire le pas qui minimise la fonction f le long de la direction de descente d_k . Le pas λ_k est donc solution du problème unidimensionnel

$$\min_{\lambda_k > 0} f(x_k + d_k \lambda_k)$$

Il faut cependant préciser que la recherche du pas optimal par la résolution de ce problème unidimensionnel peut engendrer le cas où $f''(x_k + d_k \lambda_k) \geq 0$ dont l'existence du minimum n'est pas assurée. Ce type de recherche linéaire demande en général beaucoup de temps de calcul sans permettre d'améliorer grandement la convergence de l'algorithme. D'autres règles de calcul, moins restrictives, du pas λ_k peuvent alors être utilisées.

b) Méthodes d'Armijo

Ce type de méthodes se base sur la condition suivante

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

Si on rajoute un terme négatif β_k au second membre :

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

La façon de procéder est alors de demander à f de décroître autant qu'une proportion $\omega_1 \in]0, 1[$ de ce que ferait le modèle linéarisé de f autour du point x_k . On a alors affaire à la condition d'Armijo :

$$f(x_k + d_k \lambda_k) \leq f(x_k) + \omega_1 \nabla(f(x_k)) d_k \dots \dots \dots \text{condition d'Armijo}$$

Ici, la constante ω_1 est choisie arbitrairement et est habituellement prise très petite afin que la condition d'Armijo soit plus aisément satisfaite. Le choix de cette constante est indépendant du type de problème étudié, mais il est souvent préférable de choisir $\omega_1 < 0.5$.

Il est évident qu'un pas λ_k très petit satisfait la condition d'Armijo. Or nous avons énoncé qu'il était dangereux d'accepter un pas trop petit sous peine de converger vers un point non stationnaire. Il faut donc une méthode permettant de déterminer un pas qui ne soit pas trop petit. Deux méthodes sont couramment utilisées : la méthode d'Armijo et la méthode de Goldstein.

Algorithme d'Armijo pour le calcul du pas

(0) Initialisation : Choisir un pas λ_k^1 et un paramètre $\tau \in]0, 1[$, $i = 1$

(1) Test si λ_k^1 vérifie

$$f(x_k + d_k \lambda_k) \leq f(x_k) + \lambda_k^1 \omega_1 \nabla(f(x_k)) d_k$$

Si oui stop λ_k^1 est le pas optimal $\lambda_k^* = \lambda_k^1$, sinon aller à l'étape (2)

(2) Choisir $\lambda_k^{i+1} \in [\tau \lambda_k^i, (1 - \tau) \lambda_k^i]$

(3) On pose $i = i + 1$ et $\lambda_k = \lambda_k^i$, puis retourner à l'étape (1).

Le paramètre τ est ici laissé au choix de l'utilisateur et est habituellement choisi égal à 10^{-2} . L'étape 2 de l'algorithme précédent est souvent effectuée par interpolation. Le pas obtenu par cet algorithme est appelé pas d'Armijo.

Convergence de la méthode

Idéalement, nous souhaiterions pouvoir générer, à l'aide de la méthode du gradient, une séquence $\{x^k\}$ convergeant vers un minimum local de f .

Cependant, c'est bien sur trop demander à une telle méthode, du moins si f n'est pas convexe (en raison de la présence d'extrême locaux qui ne sont pas globaux). La méthode de la direction de descente est guidée localement selon la forme de f dans la région correspondante au point x^k , et peut ainsi être attirée par tout type de minimum, qu'il soit local ou global.

Remarquons que si, pour une quelconque raison, la méthode est démarrée depuis ou rejoint un maximum ou un point stationnaire, elle se termine en ce point. Ainsi, si f n'est pas convexe, nous pouvons au mieux, attendre de la méthode de la direction de descente qu'elle converge vers un point stationnaire. Le résultat suivant peut ainsi être démontré :

Théorème 15. (Convergence de la méthode)

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

Il peut arriver que la méthode de la direction de descente 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 ε est un scalaire positif arbitrairement choisi. A priori, la valeur que nous devons fixer pour ε dépend du problème considéré. L'inconvénient majeur de la méthode de la direction de descente survient lorsque les surfaces de cout égal de f sont "allongées", et que x^k est tel que la direction de la direction de descente y est presque orthogonale à la direction menant au minimum.

Celle-ci adopte alors le comportement bien connu du "zigzag" et progresse extrêmement lentement, comme nous aurons l'occasion de le constater lors de l'application pratique de la méthode. Il existe des moyens de surmonter ces difficultés en choisissant la direction d^k un peu différemment.

La direction de descente peut être corrigée en la multipliant par une matrice ou en y ajoutant un vecteur approprié : la performance peut être améliorée en se déplaçant dans la direction $d^k = -D\nabla f(x^k)$ ou $d^k = \nabla f(x^k) + v$, où D et v sont convenablement choisis.

Prenons l'exemple suivant en commençant par $x_0 = (2, 3)$ et $\epsilon = 0.001$

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

Utilisant la méthode de la recherche linéaire exacte pour les calculs des pas λ_k .

On calcule le gradient de $f : \nabla f(x) = (2x_1 - x_2 - 1, 2x_2 - x_1 - 1)$ le point stationnaire de f est $x^* = (1, 1)$ ce qui fait de un minimum global puisque la matrice Hessien est définie positif.

$$\nabla f(x_0) = (0, 3)$$

$$\|\nabla f(x_0)\| = 3$$

— Itération 1 : $\|\nabla f(x_0)\| = 3 > 0.01$

Soit la direction $d_0 = -\nabla f(x_0) = (0, -3)$

Cherchant λ_0 tel que λ_0 minimise la fonction suivante (fonction à une seule variable).

$$\begin{aligned} f(x_0 + \lambda_0 d_0) &= f((2, 3) + \lambda_0(0, -3)) \\ &= f(2, 3 - 3\lambda_0) \\ &= 9\lambda + (3 - 3\lambda_0)^2 - 7 \end{aligned}$$

$$f'(x_0 + \lambda_0 d_0) = 9 + 6(3 - 3\lambda_0)$$

$$f'(x_0 + \lambda_0 d_0) = 0 \Rightarrow \lambda_0 = 0.5$$

$$f''(x_0 + \lambda_0 d_0) = 18\lambda_0 = 9 > 0$$

La valeur strictement positive qui minimise $f(x_0 + \lambda_0 d_0)$ est $\lambda_0 = 0.5$ $x_1 = x_0 + \lambda_0 d_0 = (2, 3) + 0.5(0, -3) = (2, 1.5)$

Après 9 itérations, le tableau suivant récapitule les résultats

$$\begin{aligned}x_0 &= (2.0000, 3.0000) \\x_1 &= (2.0000, 1.5000) \\x_2 &= (1.2500, 1.5000) \\x_3 &= (1.2500, 1.1250) \\x_4 &= (1.0625, 1.1250) \\x_5 &= (1.0625, 1.0313) \\x_6 &= (1.0156, 1.0313) \\x_7 &= (1.0156, 1.0078) \\x_8 &= (1.0039, 1.0078) \\x_9 &= (1.0039, 1.0020)\end{aligned}$$

Et la solution optimale du problème considéré est

$$x^* = (1; 1)$$

Chapitre 2

Optimisation non linéaire multi-objectifs sans contraintes

2.1 Introduction

Dans la plupart des problèmes du monde réel, il ne s'agit pas d'optimiser seulement une seule fonction objectif (critère) mais plutôt d'optimiser simultanément plusieurs fonction (critères) et qui sont généralement conflictuelles ou contradictoires. Dans les problèmes de gestion de production, par exemple, il faut le plus souvent maximiser les bénéfices et minimiser les coûts (Achat de matières premières, salaires, économiser l'énergie...). Cette philosophie de multiplicité d'objectifs ou l'optimisation multi-objectifs dite en anglais "multiobjective optimisation" qu'on note souvent (*MOP*) semble très raisonnable et contribue efficacement à l'élaboration de nouveaux systèmes d'aide à la décision, qui visent à faciliter la tâche des décideurs en anglais "Decision Maker" qu'on note (*DM*) devant des choix complexes, qui s'avèrent souvent être conflictuels.

Un problème d'optimisation multi-objectifs (multicritères) (*MOP*) consiste à optimiser (minimiser) simultanément plusieurs fonctions objectifs f_1, f_2, \dots, f_p $p \geq 2$ sur un domaine \mathbb{R}^n

Mathématiquement, ce problème peut s'écrire comme suit :

$$(MOP) \equiv \text{"min"}_{x \in \mathbb{R}^n} F(x) = [f_1(x), f_2(x), \dots, f_p(x)]^T \quad p \geq 2 \quad (2.1)$$

Tel que :

- $x = (x_1, x_2, \dots, x_n)^T$ est le vecteur de n variables de décision.
- On pose $F(x) = (f_1(x), \dots, f_p(x))^T$ est le vecteur de $p \geq 2$ fonctions objectifs à optimiser.

$$f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, 2, \dots, p$$
$$x \rightarrow f_i(x)$$

Considérons l'application suivante :

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^p$$
$$x \mapsto F(x)$$

L'ensemble $Z = F(x)$, $x \in \mathbb{R}^n$ s'appelle espace de critères ou espace des objectifs.

Toutes les définitions qui seront dans cette partie proviennent de [1].

2.2 Notions de dominance

La difficulté principale lorsque l'on cherche à minimiser simultanément plusieurs critères est qu'il n'existe généralement pas de solution qui minimise tous les critères en même temps. En revanche, il existe (généralement) des solutions qui sont "battues" par au moins une autre solution sur tous les critères. On sait alors que de telles solutions ne peuvent pas intéresser le décideur puisqu'on peut en obtenir qui réalisent de meilleures performances sur tous les critères. La dominance de Pareto a donc été introduite avec l'idée de distinguer les solutions potentiellement intéressantes des autres en suivant le principe suivant : une solution potentiellement intéressante est une solution telle qu'on ne peut améliorer la performance sur un critère sans dégrader la performance sur au moins un autre critère. Ces solutions sont appelées solutions de compromis.

Lorsque nous résolvons notre problème d'optimisation multi-objectifs, nous obtenons une multitude de solutions. Seul un nombre restreint de ces solutions va nous intéresser. Pour qu'une solution soit intéressante, il faut qu'il existe une relation de dominance entre la solution considérée et les autres solutions, dans le sens suivant :

Définition 15. (La relation de dominance)

On dit que le vecteur $y = (y_1, y_2, \dots, y_k)$ domine le vecteur $z = (z_1, z_2, \dots, z_k)$ si

- y est au moins aussi bon que z sur tous les objectifs et,
- y est strictement meilleur que z sur au moins un objectif.

D'une manière équivalente nous avons

Définition 16. La solution y domine une autre solution z , si les conditions suivantes sont vérifiées

$$f_l(y) \leq f_l(z) \quad \forall l \in \{1, \dots, k\} \text{ et } \exists l \in \{1, \dots, k\} \text{ tq } f_l(y) < f_l(z) \quad (2.2)$$

Si la solution y domine la solution z , nous allons écrire $y \prec z$.

Notons que pour toute paire de solutions y et z , une et seulement une des affirmations suivantes est vraie

- y domine z .
- y est dominé par z .
- y et z sont équivalentes au sens de la dominance (appelées aussi solutions Pareto-équivalentes)

Lorsque l'on applique la définition de la dominance, on peut avoir quatre régions auxquelles on peut attribuer des niveaux de préférence.

Définition 17. Soient $y, z \in R^p$ on dit que le vecteur y domine le vecteur z si

- Dominance faible $\exists i$ tel que

$$i = 1, 2, \dots, p : y(i) \leq z(i)$$

— Dominance forte $\exists i$

$$i = 1, 2, \dots, p : y(i) < z(i)$$

Pour mieux comprendre ce qu'est une relation de dominance, nous considérons l'exemple suivant

Exemple 2. Par exemple, pour le cas de maximisation, considérons les trois vecteurs suivants

$$z_1 = \begin{pmatrix} 5 \\ -2 \\ 1 \end{pmatrix}; \quad z_2 = \begin{pmatrix} 3 \\ -3 \\ -2 \end{pmatrix}; \quad z_3 = \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix}$$

- Nous constatons que z_1 domine fortement z_2 et z_1 domine faiblement z_3
- Par contre, z_3 domine faiblement z_2

Propriété de la relation de dominance La relation binaire de dominance \prec telle qu'elle est définie ci-dessus :

- n'est pas réflexive, car une solution ne se domine pas elle-même
- n'est pas symétrique, car on n'a jamais $(y \prec z)$ et $(z \prec y)$;
- est transitive car $(y \prec z)$ et $(z \prec w)$ implique $(y \prec w)$

La relation de dominance est donc une relation d'ordre partiel stricte sur l'espace des objectifs.

Propriété 4. Les solutions qui dominent les autres mais ne se dominent pas entre elles sont appelées solutions non dominées ou solutions optimales au sens de Pareto ou l'ensemble de solutions efficaces.

2.3 Solutions de Pareto

A la fin du 19^{ème} siècle, l'économiste Vilfredo Pareto formule le concept d'optimalité qui porte son nom (optimalité au sens de Pareto), qui constitue les origines de la recherche sur l'optimisation multi-objectif où la notion d'optimalité dans l'optimisation mono-objectif n'a aucun sens. En effet, elle n'existe aucune solution réalisable x qui fasse diminuer un objectif sans augmenter dans le même temps au moins un autre objectif.

Cependant, dans la plupart des cas, l'optimum de Pareto n'est pas constitué d'une seule solution mais d'un ensemble de solutions de "meilleur compromis" appelées solutions efficaces ou solutions non-dominées ou bien optimale au sens de Pareto.

Lorsque nous voulons résoudre un problème d'optimisation multi objectifs, il s'agit de déterminer un ensemble de solutions efficaces où l'ensemble optimale au sens de Pareto où l'ensemble de solution non dominées.

2.3.1 Aspects généraux

Définition 18. Un vecteur de décision $x^* \in \mathbb{R}^n$ est une solution optimale de Pareto (efficace ou non dominée), s'il n'existe pas d'autre vecteur de décision $x \in \mathbb{R}^n$ tel que $f_i(x) \leq f_i(x^*)$

pour tout $i = 1 \dots p$ et $f_j(x) < f_j(x^*)$ pour au moins une fonction objectif f_j

Un vecteur critère $z^* \in \mathbb{Z}$ est optimal de Pareto, s'il n'existe pas d'autre vecteur critère $z \in Z$ tel que $z_i \leq z_i^*$ pour tout $i = 1 \dots p$ et $z_j < z_j^*$ pour au moins une composant z_j ; ou de manière équivalente, z^* est optimal de Pareto si son vecteur de décision est optimal de Pareto.

On peut parler d'un ensemble de solutions optimales de Pareto ou d'un ensemble optimal de Pareto. Cet ensemble peut être non convexe.

En plus, dans l'optimalité de Pareto plusieurs autres termes sont parfois utilisés pour le concept d'optimalité décrit ci-dessus. Ces termes sont, par exemple, non-infériorité, efficacité et non-dominance.

Une définition équivalente de l'efficacité est :

Définition 19. Une solution $x^* \in \mathbb{R}^n$ est dite solution optimale de Pareto (solution efficace) si et seulement s'il n'existe pas de solution $x \in \mathbb{R}^n$ telle que $f_i(x)$ domine $f_i(x^*)$ pour tout $i = 1 \dots p$

Définition 20. (Ensemble Pareto optimal)

L'ensemble Pareto optimal de \mathbb{R}^n (ou l'ensemble efficients), est défini par l'ensemble χ

$$\chi = \{x \in \mathbb{R}^n / \nexists x^* \in \mathbb{R}^n, f(x^*) \text{ domine } f(x)\}$$

Définition 21. (Solution Pareto-équivalentes)

Les solutions $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$ sont dites solutions Pareto-équivalentes (on note $x \sim y$) si

$$x \not\prec y \text{ et } y \not\prec x$$

Définition 22. (optimalité locale au sens de Pareto) Un vecteur de décision $x^* \in \mathbb{R}^n$ est optimal localement au sens de Pareto s'il existe un réel $\delta > 0$ tel qu'il n'y ait pas de vecteur de décision x qui domine le vecteur de décision x^* avec $x \in S \cap B(x^*, \delta)$, où $B(x^*, \delta)$ représente une boule de centre x^* et de rayon δ . Un vecteur objectif $z^* \in \mathbb{Z}$ est localement optimal de Pareto si le vecteur de décision qui lui correspond est localement optimal de Pareto.

Définition 23. (optimalité globale au sens de Pareto) Un vecteur de décision $x^* \in \mathbb{R}^n$ est optimal globalement au sens de Pareto (ou optimal au sens de Pareto) s'il n'existe pas de vecteur de décision $x \in \mathbb{R}^n$ tel que x domine le vecteur x^*

Un vecteur objectif $z^* \in Z$ est globalement optimal de Pareto si le vecteur de décision qui lui correspond est globalement optimal de Pareto

Naturellement, toute solution optimale globalement Pareto est localement optimale de Pareto. L'inverse est valable pour les problèmes d'optimisation multi-objectifs convexes.

Théorème 16. Soit le problème d'optimisation multi-objectif est convexe. Alors chaque solution optimale localement Pareto est aussi globalement optimale de Pareto.

Un vecteur est faiblement optimal de Pareto s'il n'existe aucun autre vecteur pour lequel toutes les composantes sont meilleures. Plus formellement, cela signifie ce qui suit :

Définition 24. Un vecteur de décision $x^* \in \mathbb{R}^n$ est faiblement optimal de Pareto s'il n'existe pas d'autre vecteur de décision $x \in \mathbb{R}^n$ tel que $f_i(x) < f_i(x^*)$ pour tout $i = 1 \dots p$

Un vecteur objectif $z^* \in Z$ est faiblement optimal de Pareto s'il n'existe pas d'autre vecteur objectif $z \in Z$ tel que $z_i < z_i^*$ pour tout $i = 1 \dots p$, ou de façon équivalente, si le vecteur de décision correspondant est faiblement optimal de Pareto.

Remarque. Bien que la programmation mono-objectif est étudiée dans l'espace de décision, la programmation multi-objectif est étudiée dans l'espace des critères. Une des raisons est que l'espace des critères est souvent de dimension plus petite que celle de l'espace de décision. Une autre raison est que les décideurs sont souvent intéressés par les valeurs des critères.

Le décideur prend ses décisions en se basant sur une fonction. Cette fonction est appelée la fonction utilité.

Définition 25. Une fonction $U : \mathbb{R}^n \rightarrow \mathbb{R}$ représentant les préférences du décideur parmi les vecteurs critères est appelée fonction utilité.

Soit Z_1 et $Z_2 \in Z$ deux vecteurs objectifs différents. Si $U(Z_1) > U(Z_2)$, alors le décideur préfère Z_1 à Z_2 . Si $U(Z_1) = U(Z_2)$, alors le décideur trouve les vecteurs objectifs également souhaitables, c'est-à-dire qu'ils sont indifférents.

Il faut souligner que la fonction utilité est un concept totalement dépendant du décideur. Différents décideurs peuvent avoir des fonctions utilité différentes pour le même problème.

Il est possible de définir l'optimalité dans un contexte multi-objectif de manière plus générale que par l'optimalité de Pareto ou de Pareto faible. Soit un cône convexe pointu D défini dans \mathbb{R}^n . Ce cône D est appelé cône de commande et il est utilisé pour induire un ordre partiel sur Z . Ayons deux vecteurs objectifs, Z_1 et $Z_2 \in Z$. Un vecteur objectif Z_1 domine Z_2 , noté par $Z_1 \leq_D Z_2$, si $Z_2 - Z_1 \in D$ et $Z_1 \neq Z_2$, c'est-à-dire $Z_2 - Z_1 \in D_0$.

Nous pouvons maintenant présenter une définition de l'optimalité basée sur la dominance, qui est une alternative aux définitions données précédemment. Lorsqu'un cône de commande est utilisé pour définir l'optimalité, le terme efficacité sera utilisé dans ce qui suit.

Définition 26. Soit D un cône convexe pointu. Un vecteur de décision $x^* \in \mathbb{R}^n$ est efficace (par rapport à D) s'il n'existe pas d'autre vecteur de décision $x \in \mathbb{R}^n$ tel que $f(x) \leq_D f(x^*)$. Un vecteur objectif $z^* \in Z$ est efficace s'il n'existe pas d'autre vecteur objectif $z \in Z$ tel que $z \leq_D z^*$

Cette définition signifie qu'un vecteur est efficace (non dominé) s'il n'est pas dominé par aucun autre vecteur réalisable. La définition ci-dessus peut être formulée de plusieurs manières. Si on remplace \leq_D par sa définition, on aura la condition sous la forme $0 \neq z^* - z \in D$ ou $z^* - z \in D_0$.

Les vecteurs de décision et d'objectifs faiblement efficaces peuvent être définis de manière correspondante aux vecteurs efficaces. Si l'ensemble Z des vecteurs objectifs est ordonné par un cône d'ordre D , les vecteurs faiblement efficaces peuvent être caractérisés de la manière suivante :

Définition 27. Soit D un cône convexe pointu. Un vecteur de décision $x^* \in \mathbb{R}^n$ est faiblement efficace (par rapport à D) s'il n'existe pas d'autre vecteur de décision $x \in \mathbb{R}^n$ tel que $(x^*) \in f(x) + \text{int } D$, c'est-à-dire $(f(x^*) - \text{int } D) \cap Z = 0$.

Un vecteur objectif $z^* \in Z$ est faiblement efficace s'il n'existe pas d'autres vecteurs objectif $z \in Z$ tel que $z^* \in z + \text{int } D$, c'est $(z^* - \text{int } D) \cap Z = 0$.

Une formulation alternative est qu'un vecteur objectif $z^* \in Z$ est faiblement efficace si $(z - z^*) \cap (-\text{int } D) = 0$.

2.3.2 Front de Pareto (la surface du compromis)

Par définition, le front de Pareto est l'ensemble convexe dans le cas continu (non convexe dans le cas discret) constitué par l'image des solutions efficaces dans l'espace des critères. En effet, le front de Pareto est l'ensemble $Z_{ND} = F(X_E)$. La figure 2.1 montre le front de Pareto pour les différentes situations envisageables pour un problème bi-critères. Notons que, le front de Pareto de chaque cas et la partie en gras de l'ensemble.

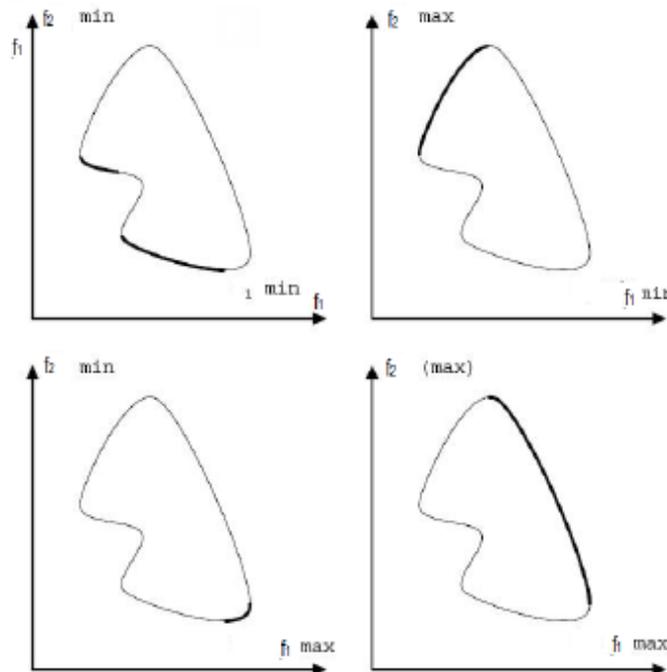


FIGURE 2.1 – Le front de Pareto

2.3.3 Les points caractéristiques

Dans le cadre de l'optimisation multi-objectifs, le plus souvent le décideur raisonne plutôt en termes d'évaluation d'une solution sur chaque critère et se place naturellement dans l'espace des critères. Dans cette section, nous définissons le point idéal et le point nadir comme des

limites inférieure et supérieure dans le front de Pareto, et nous présentons aussi la matrice des gains.

Définition 28. (point idéal)

Un vecteur objectif minimisant chacune des fonctions objectives est appelé vecteur objectif idéal.

Les composantes z_i^* du vecteur critère idéal $z^* \in \mathbb{R}^p$ sont obtenues en minimisant chacune des fonctions objectifs individuellement, non soumises pas aux contraintes, c'est-à-dire, en résolvant $\min f_i(x)$ avec $x \in \mathbb{R}^n$ pour $i = 1, 2, \dots, n$.

Il est évident que si le vecteur critère idéal est admissible c'est-à-dire $z^* \in Z$ alors il serait une solution du problème d'optimisation multi-objectifs (et l'ensemble optimal de Pareto serait réduit). Cela n'est pas possible en général car il y a une certaine contradiction entre les objectifs. Même si le vecteur objectif idéal n'est pas réalisable, il peut être considéré comme un point de référence. A partir du vecteur objectif idéal, nous obtenons les bornes inférieures de l'ensemble optimal de Pareto pour chaque fonction objectif.

Parfois, nous avons aussi besoin d'un vecteur strictement meilleur. En d'autres termes, qui domine strictement toute solution optimale de Pareto.

Remarque. Le point idéal est généralement une solution utopique, dans le sens où il n'appartient pas à l'espace objectif réalisable.

Dans certains cas, le décideur peut le définir comme un point de référence exprimant le but qu'il veut atteindre pour chaque objectif. Le point idéal définit la borne supérieure (cas de maximisation) et la borne inférieures (cas de minimisation) où les points non dominés peuvent atteindre. Il est souvent employé comme points de référence dans la programmation du compromis ou dans des méthodes interactives dont le but est de trouver une solution la plus préférée pour le décideur.

Les bornes supérieure de l'ensemble optimal de Pareto correspondent aux composantes du point appelé nadir, et noté Z^{nad} , et se définit comme suit :

Définition 29. (Point Nadir). Les coordonnées du point nadir correspondent aux pires valeurs obtenues par chaque fonction objectif sur l'ensemble de solutions non dominées. Le point Nadir définit la borne supérieure (cas de minimisation) où les points non dominés peuvent atteindre. Le point nadir $Z^{\text{nad}} = (Z_1^{\text{nad}}, Z_2^{\text{nad}}, \dots, Z_p^{\text{nad}})$ donné par $Z_K^{\text{nad}} := \max_{x \in X_E} f_K(x)$, $K = 1, \dots, p$

Le point nadir est souvent utilisé dans les méthodes d'optimisation interactives. La figure 2.2 représente le point idéal et le point nadir pour un problème bi-critères.

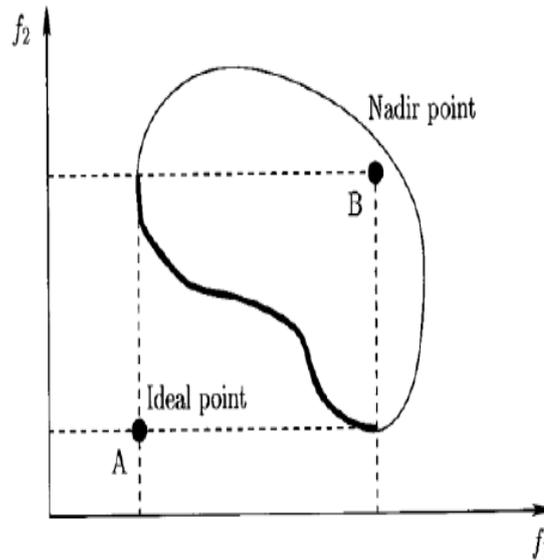


FIGURE 2.2 – Le point idéal et le point nadir

Définition 30. (Matrice des gains)

soit \hat{x}^j une solution optimale du critère f_j . La matrice des gains $p \times p$ est formée des éléments $z_{ij} = f_i(\hat{x}^j)$ tels que

$$G = \begin{bmatrix} z_1^* & \dots & z_{1j} & \dots & z_{1j} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{i1} & \dots & z_i^* & \dots & z_{ik} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{p1} & \dots & z_{kj} & \dots & z_p^* \end{bmatrix}$$

Les coordonnées du point idéal apparaissent sur la diagonale principale de cette matrice et les coordonnées du point Nadir apparaissent sur la diagonale opposée. Lorsqu'un critère 1 possède plusieurs solutions optimales, la colonne j de la matrice des gains dépendra de la solution \hat{x}^j choisie (la matrice des gains est univoquement déterminée si, pour tous les critères j , la solution \hat{x}^j est unique).

2.4 Caractérisation des solutions de Pareto

Dans certaines situations, l'ensemble de la solution efficace, qui est une partie de l'ensemble des solutions admissibles, soit très vaste et parfois infini, il est souvent impossible d'énumérer toutes les solutions efficaces. Et même lorsque c'est possible, il est nécessaire d'aider le décideur à faire son choix parmi les solutions efficaces. La sélection d'une solution efficace spécifique du meilleur compromis nécessite donc une certaine connaissance de la structure de préférence du décideur. Les méthodes proposées pour aider un décideur à faire son choix parmi les solutions efficaces peuvent être classées en plusieurs catégories selon la façon dont le décideur articule ou incorpore ses préférences. Chaque modèle de préférence est généralement explicite en fixant un

certain nombre de paramètres. Citons brièvement quelques paramètres qui sont fréquemment utilisés.

2.4.1 Paramètres de préférence

- (a) Un vecteur $\lambda \in \mathbb{R}^{p+}$ Appelé vecteur de coefficients d'importance des p fonctions critères ou paramètres de préférences.
- (b) des vecteurs de performances ayant des significations particuliers, comme le point de préférence qui est défini par des réservations (valeurs souhaitables) sur chaque critère et le point de réservation qui est défini par des niveaux des réservations (valeurs non souhaitables) sur chaque critère.

Fonction scalarisante

Etant donné un ensemble de paramètres de préférence $\Lambda = \{(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p\}$ ou $\lambda_i, i = 1, 2, \dots, p$ sont les coefficients d'importance des p critères. La fonction scalarisant des certaines, croissante et agrège les valeurs des critères pour chaque solution d'agrégation les plus employées sont décrites ci-dessous. **La somme pondérée** : largement utilisée en optimisation linéaire multiobjectifs ou problème de programmation paramétrique. Parmi les sommes les plus utilisées, nous citons :

- $s_1(z, \lambda) = \sum_{i=1}^p \lambda_i z_i, \lambda \in \Lambda, \sum_{i=1}^p \lambda_i = 1$
- $s_2(z, \lambda) = \sum_{i=1}^p \lambda_i |z_i^* - z_i|$ Où z_i^* est la i -ème composante du vecteur idéal et $|\cdot|$ représente la norme de deux vecteurs.

La somme pondérée $s_2(z, \lambda)$ mesure la déviation qui sépare l'évaluation des solutions proposées, qui sont généralement des points efficaces ou faiblement efficaces, du point d'aspiration. Cette déviation peut être mesurée par d'autres normes parmi les quelles citons

- (a) La norme pondérée

$$s_3^q(z, \lambda) = \left[\sum_{i=1}^p \lambda_i |z_i^* - z_i| \right]^{\frac{1}{q}}, q \in \mathbb{Z}_+^*$$

- (b) La norme pondérée de Chebychev

$$s_4(z, \lambda) = \max_{i=1,2,\dots,p} \{\lambda_i |z_i^* - z_i|\}$$

- (c) La norme composée (augmentée) de Chebychev

$$s_5(z, \lambda) = \max_{i=1,2,\dots,p} \{\lambda_i |z_i^* - z_i|\} + \rho \sum_{i=1}^p \lambda_i |z_i^* - z_i|, \rho > 0$$

Caractérisation à l'aide de point cibles

- (a) Niveaux d'aspiration

$$s_6^q(z, \lambda) = \left[\sum_{i=1}^p \lambda_i |\hat{z}_i - z_i| \right]^{\frac{1}{q}}, q \in \mathbb{Z}_+^*$$

Le point $\hat{z} = \{\hat{z}_1, \hat{z}_2, \dots, \hat{z}_p\} \in \mathbb{Z}$ représente le point cible qu'on en souhaite s'approcher autant que possible.

(b) Niveaux de réservation

$$s_7^q(z, \lambda) = \left[\sum_{i=1}^p \lambda_i |z_i - z'_i| \right]^{\frac{1}{q}}, q \in \mathbb{Z}_+^*$$

Tel que $z'_i = z'_1, z'_2, \dots, z'_p \in \mathbb{Z}$ représente le point qu'on en souhaite s'éloigner le plus que possible.

Théorème 17. (Isermann 1974) Soit \hat{x} un point arbitraire de \mathbb{R}^n , \hat{x} est une solution efficace du problème multi-objectifs (MOP) si et seulement si la valeur optimale de la fonction objectif W est nulle dans le programme linéaire suivant :

$$(P_{\hat{x}}) \begin{cases} \min_i W = \sum_{i=1}^p \varphi_i \\ \text{s.C } f_i(x) - \varphi_i = f_i(\hat{x}), \quad i = 1, \dots, p \\ x \in \mathbb{R}^n \quad \text{ET } \varphi_i \in \mathbb{R}_+^*, i = 1, \dots, p \end{cases}$$

Le problème $(P_{\hat{x}})$ est souvent utilise pour tester l'efficacité d'une solution donnée.

Théorème 18. (Ecker et Kouada, 1978)

Si $(P_{\hat{x}})$ n'admet pas solution optimale finie, alors l'ensemble des solutions efficaces du problème (MOP) est vide.

2.5 Méthodes de résolution

Nous présenterons quelques méthodes appartenant à la classe de méthodes paramétriques qui permettent essentiellement de trouver une ou plusieurs solutions efficaces d'un problème multi-objectifs. Cette famille de méthodes en question consiste à transformer le problème multi-objectifs vers un problème mono-objectif.

Ces méthodes fonctionnent essentiellement en présence de certains paramètres ou scalaires qu'on devrait fixer a priori et elles sont appelées aussi méthodes scalaires. En effet, nous allons présenter les méthodes les plus connues et les plus utilisées dans la littérature à savoir

1. la méthode de pondération des fonctions objectives (de somme pondérée)
2. la méthode ε -contraintes (compromis).
3. la méthode de distance pondérée (méthode de Chebychev).

2.5.1 Méthode de somme pondérée

L'idée principale de la méthode de pondération des fonctions objectif revient à (Gass and Saaty, 1955) et (Zadeh, 1963). Cette approche de la résolution d'un problème d'optimisation multi-objectifs est la plus évidente, elle fait partie de la famille des méthodes scalaires " *Weighted Suon Scalacization Méthod*" en anglais. Cette méthode s'appelle aussi *La méthode de scalarisation des fonctions objectif*.

Le principe de cette méthode consiste à transformer le problème d'optimisation multi-objectifs en question vers un problème d'optimisation mono-objectif en associant des coefficients de pondération $\beta_i, i = 1, 2, \dots, k$ aux fonctions objectif $f_i, i = 1, 2, \dots, k$, avec $\beta_i \geq 0$, et $\sum_{i=1}^k \beta_i = 1$ et à faire la somme des fonctions objectif pondérées.

On obtient ainsi un nouveau problème d'optimisation à une seule fonction objectif de la forme suivante.

$$(P_\beta) \begin{cases} \min F(x) = \sum_{i=1}^k \beta_i f_i(x) \\ x \in \mathbb{R}^n \\ \text{avec } \beta_i \geq 0, i = 1, 2, \dots, k \text{ et } \sum_{i=1}^k \beta_i = 1 \end{cases} \quad (2.3)$$

Généralement, on appelle le vecteur $\beta \in \mathbb{R}^k = (\beta_1; \beta_2; \dots; \beta_k)$ *vecteur de poids ou vecteur de paramètres* et on appelle le problème mono-objectif (P_β) *le problème paramétrique ou problème pondéré*.

On peut représenter graphiquement le fonctionnement de la méthode sur un problème à deux objectifs. Le problème est le suivant :

$$(MOP) \begin{cases} \text{"min"} f_1(x) \\ \text{"min"} f_2(x) \\ x \in \mathbb{R}^n \end{cases}$$

Le problème paramétrique est le suivant :

$$(P_\beta) \begin{cases} \min F(x) = \beta_1 f_1(x) + \beta_2 f_2(x) \\ x \in \mathbb{R}^n \\ \text{avec } \beta_1, \beta_2 \geq 0, \text{ et } \beta_1 + \beta_2 = 1 \end{cases} \quad (2.4)$$

Notre nouvelle fonction objective aura pour expression :

$$F(x) = \beta_1 f_1(x) + \beta_2 f_2(x)$$

Ceci est l'expression d'une droite dans le plan f_1, f_2 .

En effet, si l'on cherche à minimiser $F(x)$ on cherche en fait une constante C de l'équation de droite suivante la plus petite possible :

$$f_2(x) = -\frac{\beta_1}{\beta_2} f_1(x) + C$$

Sur la figure 2.3 l'ensemble Y correspond à l'ensemble réalisable dans l'espace des objectifs.

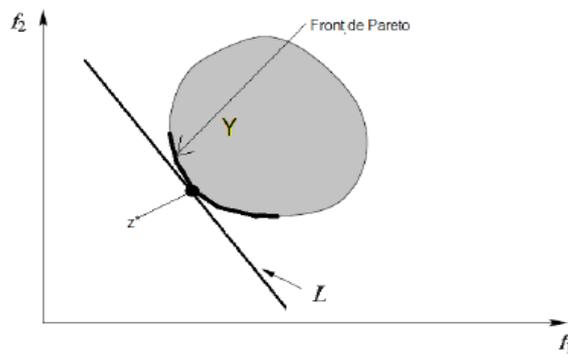


FIGURE 2.3 – Le principe de la méthode de pondération des fonctions objectif

La droite L est le tracé de la droite $f_2 = -\frac{\beta_1}{\beta_2}f_1 + C$ pour certaines valeurs de β_1 et β_2 . Le point non dominé z^* est donné par l'intersection de la droite L et le front de Pareto. En effet, z^* est la solution optimale du problème paramétrique (P_β) . Si l'on répète ce processus pour plusieurs valeurs des coefficients de pondération β , les différentes solutions trouvées forment la surface de compromis.

Cette méthode n'est applicable qu'à des ensembles Y convexes. Dans le cas contraire, elle ne permet pas de trouver la totalité de la surface de compromis. Par exemple, la portion en trait gras de la surface de compromis de la figure 2.4 ne peut être obtenue par cette méthode.

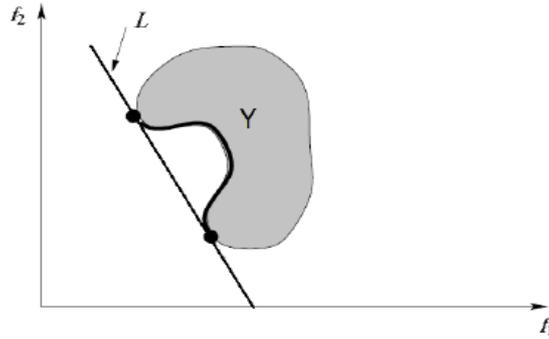


FIGURE 2.4 – Une difficulté insurmontable pour la méthode de pondération des fonctions objectives

Théorème 19. Si la solution optimale de problème paramétrique (P_β) est unique alors elle est efficace pour le problème $(MOP)2.1$.

Démonstration. Ici on montre

$$x^* \in \mathbb{R}^k \text{ Optimale pour } (P_\beta) \text{ est unique} \implies x^* \text{ est efficace pour } (MOP) 2.1.$$

Par contradiction, supposons que $x^* \in \mathbb{R}^k$ l'unique solution optimale de P_β pour un certain vecteur de paramètres β et x^* n'est pas efficace pour le problème (MOP) .

En effet, il existe une autre solution réalisable $x \in \mathbb{R}^k$ $f_i(x) \leq f_i(x^*)$ et il existe un certain j tel que $f_j(x) < f_j(x^*)$. Sachant que le vecteur β est non négatif alors on aura

$$\sum_{i=1}^k \beta_i f_i(x) < \sum_{i=1}^k \beta_i f_i(x^*)$$

. Puisque, x^* est optimale et unique alors $\sum_{i=1}^k \beta_i f_i(x^*) < \sum_{i=1}^k \beta_i f_i(x)$ pour tout $x \in \mathbb{R}^k$. En effet, les deux inégalités sont contradictoires.

Théorème 20. La solution optimale de problème paramétrique P_β est efficace pour le problème (MOP) si $\forall i, \beta_i > 0, i = 1, 2, \dots, k$

Exemple pour la méthode de pondération

Pour illustrer la méthode de pondération des fonctions objectif, considérons l'exemple qui est un problème multi-objectifs non linéaire (MONLP)

$$(\text{MONLP}) \begin{cases} \text{" min " } f_1(x) = x_1^2 + x_2^2 \\ \text{" min " } f_2(x) = (x_1 - 5)^2 + (x_2 - 5)^2 \\ -5 \leq x_1 \leq 10 \\ -5 \leq x_2 \leq 10 \end{cases} \quad (2.5)$$

On peut voir facilement que l'exemple 2.5 est un problème quadratique convexe car les deux fonctions objectif f_1 et f_2 sont des équations de cercles qui peuvent s'écrire sous forme quadratique strictement convexe et l'ensemble de décision qui est un quadrant qui est donné par la figure 2.5

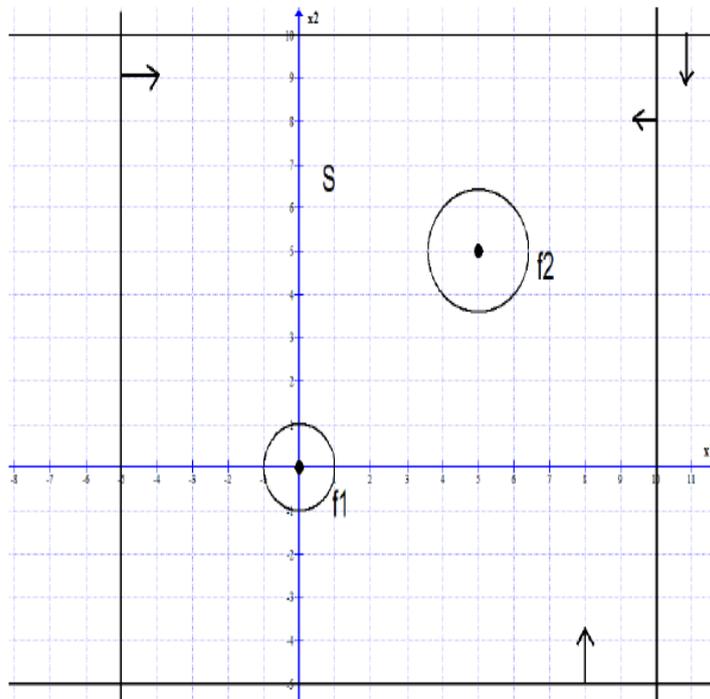


FIGURE 2.5 – espace de décision de l'exemple 2.5

Utilisons le script suivant écrit sous MATLAB pour représenter graphiquement le domaine réalisable dans l'espace des critères.

```
f1 = []
f2 = []
for i = -5 : 0.2 : 10
for j = -5 : 0.2 : 10
f1 = [f1; i.2 + j.2]
f2 = [f2; (i - 5) . 2 + (j - 5) . 2]
end
```

end

plot (f1, f2, '*')

La fonction objectif du problème paramétrique (P_λ) peut s'écrire comme suit

$$\begin{aligned} \min \varphi &= \lambda_1 f_1(x) + \lambda_2 f_2(x) \\ &= \lambda_1 (x_1^2 + x_2^2) + \lambda_2 ((x_1 - 5)^2 + (x_2 - 5)^2) \end{aligned}$$

Avec $\lambda_1, \lambda_2 \geq 0$ et $\lambda_1 + \lambda_2 = 1$ Si on pose $\lambda_1 = 1 - \lambda_2$ où $0 \leq \lambda_1 \leq 1$ alors la fonction objectif φ devient

$$\varphi = x_1^2 + x_2^2 - 10\lambda_2(x_1 + x_2) + 50\lambda_2$$

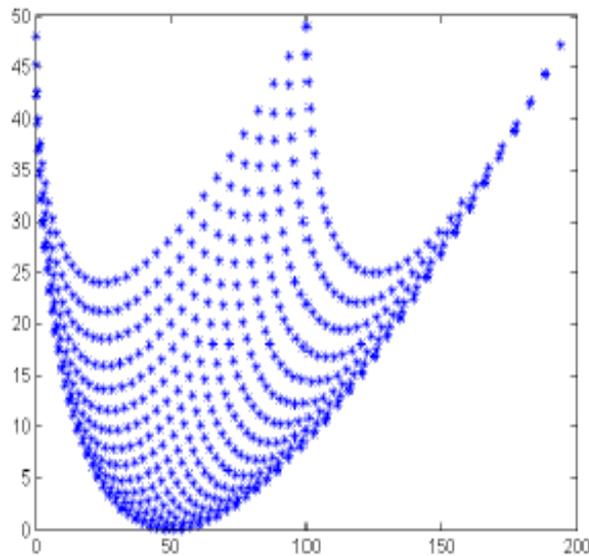


FIGURE 2.6 – espace de critère de l'exemple 2.5

On cherche maintenant l'optimum de cette nouvelle fonction objectif φ : Le gradient de la fonction φ est le vecteur

$$\begin{aligned} \nabla \varphi &= \begin{pmatrix} \frac{\delta \varphi}{\delta x_1} & \frac{\delta \varphi}{\delta x_2} \end{pmatrix} \\ &= (2x_1 - 10\lambda_2; 2x_2 - 10\lambda_2) \end{aligned}$$

Le Hessian de la fonction φ est la matrice

$$\nabla^2 \varphi = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Comme on peut le vérifier facilement, le Hessian est une matrice définie positif alors le minimum de la fonction φ est donné par la résolution du système :

$$\begin{cases} 2x_1 - 10\lambda_2 = 0; \\ 2x_2 - 10\lambda_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1^* = 5\lambda_2 \\ x_2^* = 5\lambda_2 \end{cases}$$

Cependant, toutes les solutions efficaces de problème multi objectif peuvent s'écrire

$$x = 5\lambda_2(1; 1) \text{ avec } 0 \leq \lambda_2 \leq 1$$

Si on remplace $x_1^* = 5\lambda_2, x_2^* = 5\lambda_2$, on obtient alors les fonctions objectif suivantes :

$$\varphi = 50\lambda(1 - \lambda_2)$$

$$f_1 = 50\lambda_2^2$$

$$f_2 = 50(1 - \lambda_2)^2$$

On peut alors calculer quelques solutions du front de Pareto pour un certain nombre de valeurs du coefficient λ_2 . Ces résultats sont réunis dans le tableau 2.1 l'allure de front de Pareto est alors celle de la figure 2.7.

λ_2	φ	f_1	f_2
0	0	0	50
0.1	4.5	0.5	40.5
0.2	8	2	32
0.3	10.5	4.5	24.5
0.4	12	8	18
0.5	12.5	12.5	12.5
0.6	12	18	8
0.7	10.5	24.5	4.5
0.8	8	32	2
0.9	4.5	40.5	0.5
1	0	50	0

TABLE 2.1 – Résultats pour quelques valeur de λ_2 de la méthode pondérée

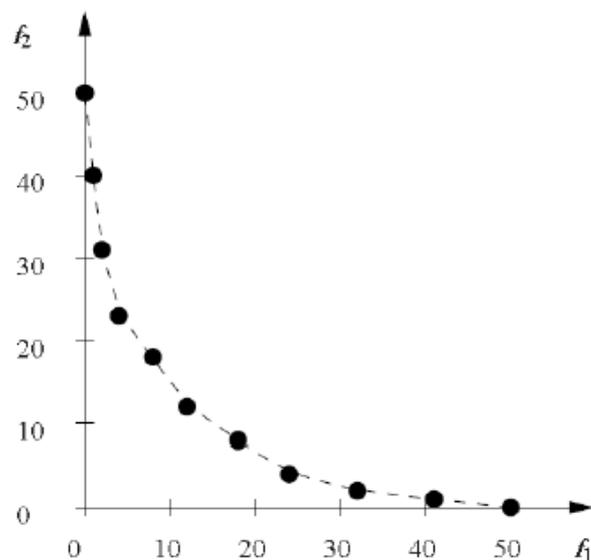


FIGURE 2.7 – Les points non dominés trouvés pour l'exemple

Cette méthode a été l'une des premières utilisées. D'un point de vue algorithmique, elle est très efficace. On peut, en faisant varier les coefficients de pondération, retrouver la surface de compromis, si le domaine réalisable est convexe. Cependant, cette méthode ne permet pas de trouver les solutions enfermées dans une concavité.

2.5.2 La méthode du compromis

Après avoir étudié les méthodes qui permettent de fusionner les fonctions objectives en un seul, nous allons étudier des méthodes qui permettent de transformer un problème d'optimisation multi-objectifs en un problème d'optimisation mono-objectif qu'on note (P) comportant quelques contraintes supplémentaires. La démarche est la suivante

1. on choisit un objectif à optimiser prioritairement soit f_k ;
2. on choisit un vecteur de contraintes initial $\varepsilon_i, i = 1, 2, \dots, p, i \neq k, \varepsilon_i \geq 0$;
3. on transforme le problème en conservant l'objectif prioritaire et en transformant les autres objectifs en contraintes d'inégalité de type

$$f_i \leq \varepsilon_i, i = 1, 2, \dots, p, i \neq k$$

On appelle aussi cette méthode la méthode de la contrainte ε

Mathématiquement, le problème d'optimisation mono-objectif (P) peut s'écrire comme suite :

$$(P) \begin{cases} \min f_k(x) \\ f_i(x) \leq \varepsilon_i, i = 1, 2, \dots, p, i \neq k \\ x \in \mathbb{R}^n \end{cases} \quad (2.6)$$

La résolution du problème d'optimisation mono-objectif ε -contraintes noté par (P) peut donner des solutions efficace ou des solutions faiblement efficaces pour le problème multi-objectifs (MOP) 2.1.

Citons les théorèmes suivants :

Théorème 21. La solution optimale du problème ε -contraintes (P) est faiblement efficace (MOP) 2.1 .

Démonstration. Soit $x^* \in \mathbb{R}^n$ une solution optimale pour le problème ε -contraintes (P) et supposons que x^* n'est pas faiblement efficace pour le problème (MOP) . Alors il existe une autre solution $x \in \mathbb{R}^n$ tel que $f_i(x) < f_i(x^*)$ pour tout $i = 1, 2, \dots, p$ Puisque x^* une solution réalisable de problème ε -contraintes alors elle vérifie les contraintes $f_j(x^*) \leq \varepsilon_j$ pour tout $j = 1, 2, \dots, p; j \neq k$ d'où $f_j(x) < f_j(x^*) \leq \varepsilon_j$ avec $f_k(x) < f_k(x^*)$ ce qui contredit le fait que x^* est optimale pour le problème ε -contraintes (P) .

Théorème 22. Une solution réalisable $x^* \in \mathbb{R}^n$ est efficace pour le problème (MOP) 2.1 si et seulement si elle est optimale pour le problème mono-objectif (P) avec $f_i(x^*) = \varepsilon_i, i = 1, 2, \dots, p, i \neq k$

Notons qu'avec la condition nécessaire du théorème 22 il est possible de trouver n'importe quelle solution efficace du problème multi-objectif (*MOP*) 2.1 sans se soucier de la convexité du problème.

Théorème 23. Une solution réalisable $x^* \in \mathbb{R}^n$ est efficace pour le problème (*MOP*) 2.1 si elle est optimale et unique pour le problème mono-objectif (*P*) avec $f_i(x^*) = \varepsilon_i, i = 1, 2, \dots, p, i \neq k$

Démonstration. Ici, il s'agit de montrer l'implication suivante :

x^* Solution optimale et unique pour (*P*) avec $f_i(x^*) = \varepsilon_i, i = 1, 2, \dots, p, i \neq k \Rightarrow x^*$ est efficace pour le problème (*MOP*).

Soit $x^* \in \mathbb{R}^n$ l'unique solution optimale pour le problème (*P*) pour un certain k . Supposent que x^* n'est pas efficace pour le problème (*MOP*). En d'autres termes, il existe une solution x^0 dont $f(x^0)$ domine $f(x^*)$ c'est-à-dire $f_i(x^0) \leq f_i(x^*)$ pour tout $i = 1, 2, \dots, p$ avec au moins une inégalité stricte $f_j(x^0) < f_j(x^*)$. L'unicité de la solution x^* entraîne $\forall x \in \mathbb{R}^n, f_i(x) \leq f_i(x^*), i \neq k$ et $f_k(x) < f_k(x^*)$ alors c'est contradictoire avec les précédentes inégalités et l'hypothèse x^* est efficace pour (*MOP*).

Théorème 24. L'unique solution optimale du problème mono-objectif (*P*) est efficace pour problème (*MOP*) 2.1 pour tout vecteur donné $\varepsilon = (\varepsilon_1; \varepsilon_2; \dots; \varepsilon_{i-1}; \varepsilon_{i+1}; \dots; \varepsilon_p)$.

Démonstration. Par définition, soit x^* solution optimale et unique pour le problème (*P*) ceci implique $f_k(x^*) < f_k(x)$ pour toute solution $x \in \mathbb{R}^n$ où $f_i(x^*) \leq \varepsilon_i$ pour tout $i = 1, 2, \dots, p, i \neq k$. Supposons maintenant, x^* n'est pas efficace pour le problème (*MOP*) alors il existe une autre solution x^0 dont $f(x^0)$ domine $f(x^*)$ c'est-à-dire $f_i(x^0) \leq f_i(x^*)$ pour tout $i = 1, 2, \dots, p$, avec au moins une inégalité stricte j tel que $f_j(x^0) < f_j(x^*)$ deux cas sont possibles :

Cas 1 : Si $j = k$ alors $f_k(x^0) < f_k(x^*)$ et $f_i(x^0) \leq f_i(x^*) \leq \varepsilon_i$ pour tout $i \neq k$ qui est une contradiction avec x^* une solution optimale pour (*P*)

Cas 2 : Si $j \neq k$ alors $f_j(x^0) < f_j(x^*) \leq \varepsilon_j$ et $f_i(x^0) \leq f_i(x^*) \leq \varepsilon_i$ et $f_k(x^0) < f_k(x^*)$ pour tout $i \neq j, i \neq k$ et qui est une contradiction avec x^* une solution optimale unique pour (*P*).

Exemple pour la méthode compromis

Reprenons l'exemple 2.5

$$(\text{MONLP}) \begin{cases} \text{" min " } f_1(x) = x_1^2 + x_2^2 \\ \text{" min " } f_2(x) = (x_1 - 5)^2 + (x_2 - 5)^2 \\ -5 \leq x_1 \leq 10 \\ -5 \leq x_2 \leq 10 \end{cases} \quad (2.7)$$

L'ensemble réalisable dans l'espace de décision est donné par la figure 2.5 Dans ce cas le nombre de critères est $k = 2$, choisissons la fonction $f_2(x)$ comme fonction objectif et $f_1(x)$ est

placée en contraintes du problème mono-objectif ε -contrainte qui peut s'écrire comme suit :

$$(p_\varepsilon) \begin{cases} \min f_2(x) = (x_1 - 5)^2 + (x_2 - 5)^2 \\ x_1^2 + x_2^2 \leq \varepsilon \\ -5 \leq x_1 \leq 10 \\ -5 \leq x_2 \leq 10 \end{cases} \quad (2.8)$$

Notons qu'on peut mettre $f_1(x)$ dans la fonction objectif et $f_2(x)$ dans les contraintes. Maintenant il s'agit de résoudre le problème ε -contrainte (p_ε) en choisissons quelques valeurs de ε .

Pour une certaine valeur fixe de ε ce problème appartient à une classe de problème d'optimisation qu'on appelle problèmes quadratiques avec contraintes quadratiques "quadratically constrained linear/quadratic programming problems", qui représente parfois des énormes difficultés pour sa résolution.

Nous n'allons pas aborder les méthodes spécifiques de résolution de ce type de problèmes. En effet dans notre cas, (on peut utiliser l'approche graphique pour trouver sa solution.

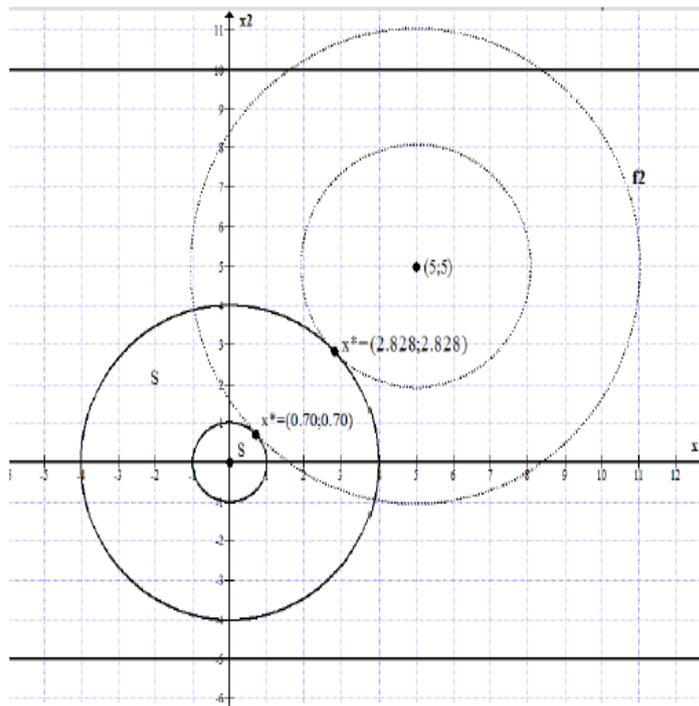


FIGURE 2.8 – Résolution graphique de l'exemple 2.7

Comme on peut le constater, l'expression des fonctions objectifs f_1 et f_2 sont respectivement des équations de cercle de centre $(0; 0)$ et $(5; 5)$ et de rayon $\sqrt{f_1}$ et $\sqrt{f_2}$. Dans l'espace de décision, la figure 2.8 montre les solutions efficaces du problème pour deux cas $\varepsilon = 1$ et $\varepsilon = 16$.

ε	x^*	$f_1(x^*)$	$f_2(x^*)$
0	(0;0)	0	50
1	(0.70;0.70)	1	36.857
4	(1.414;1.414)	4	25.715
5	(1.581;1.581)	5	23.377
9	(2.1213;2.1213)	9	16.573
16	(2.828;2.828)	16	9.431
25	(3.535;3.535)	25	4.289
36	(4.242;4.242)	36	1.147
50	(5;5)	50	0
60	(5;5)	50	0

TABLE 2.2 – Résultats avec la méthode ε -contrainte de l'exemple 2.7

Le tableau 2.2 nous donne les résultats trouvés pour quelques valeurs de ε .

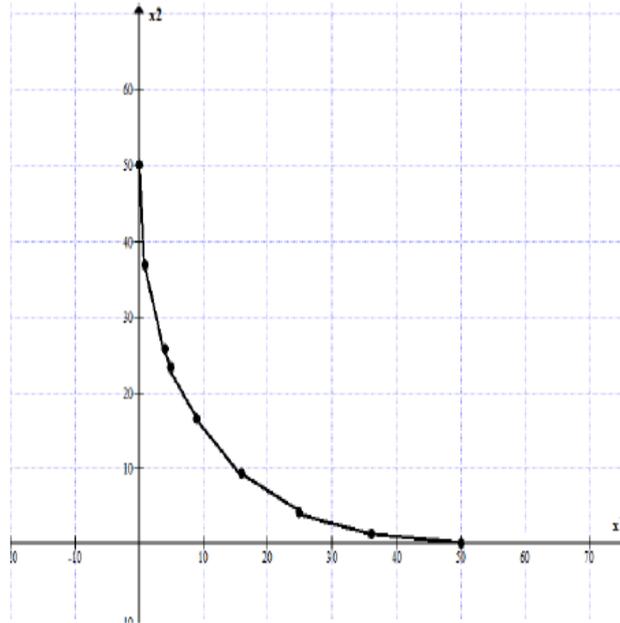


FIGURE 2.9 – Les solutions non dominées de l'exemple 2.7

D'après les résultats trouvés, nous remarquons que l'allure du front de Pareto donné par la figure 2.9 est presque la même trouvée par la méthode des sommes pondérées.

En effet, l'énoncé des théorèmes en particulier 22 et 24 sont évidemment valable dans toutes les valeurs de ε -considérées c'est-à-dire dans le tableau on voit que $f_1(x^*) = \varepsilon$ et la solution optimale du problème ε -contrainte est unique alors elle est efficace pour le problème de l'exemple 2.5

Nous constatons que la valeur $\varepsilon = 50$ représente une borne supérieure pour l'objectif $f_1 = x_1^2 + x_2^2$. Si on augmente d'avantage ε la solution efficace reste la même $x^* = (5; 5)$ et son image

le point non domine est $y^* = (0; 50)$.

On montre que cette méthode peut se ramener à la méthode de pondération des fonctions objectives moyennant une légère transformation

Les principaux inconvénients de cette méthode sont les suivants : elle est gourmande en temps de calcul et la programmation de l'algorithme peut être extrêmement difficile s'il y a trop de fonctions contraintes. Cependant, la relative simplicité de l'énoncé de la méthode l'a rendue populaire.

2.5.3 La méthode du Chebychev

Dans la suite de cette section, nous détaillons l'une des méthodes de cette famille qui est une méthode basée sur la norme pondérée de Chebychev. La méthode en question consiste à minimiser la distance pondérée entre le point de préférence et le front de Pareto. Le vecteur de paramètres de poids où de préférence w_i des critères est fixé a priori par le décideur où l'utilisateur.

Le point de référence qu'on note par z^* de coordonnées $z^* = (z_1^*, z_2^*, \dots, z_p^*)^T$ est défini par l'utilisateur où le décideur correspondent à souvent au point idéal où bien un point qui représente mieux les préférences de l'utilisateur.

La figure 2.10 illustre le fonctionnement de cette méthode, le point de référence est représenté comme étant le point point idéal de coordonnées $(z_1^*, z_2^*)^T$. La surface de compromis est donnée par l'intersection du front de Pareto et les coordonnées du point de référence z^* . Cette méthode permet donc de trouver un point dans la surface de compromis tout en minimisant une certaine distance notée par d entre le point de référence et le front de Pareto.

Cependant, le problème d'optimisation mono-objectif qui consiste à minimiser la distance en question peut se formuler comme suit :

$$\min_{x \in \mathbb{R}^n} d(x) \quad (2.9)$$

En effet, pour mesurer la distance entre le point de référence et le front de Pareto, la distance où la norme la plus utilisée est la suivante :

$$d_r(x) = \left[\sum_{i=1}^p |f_i(x) - z_i^*|^r \right]^{\frac{1}{r}} \quad \text{avec } 1 \leq r \leq \infty \quad (2.10)$$

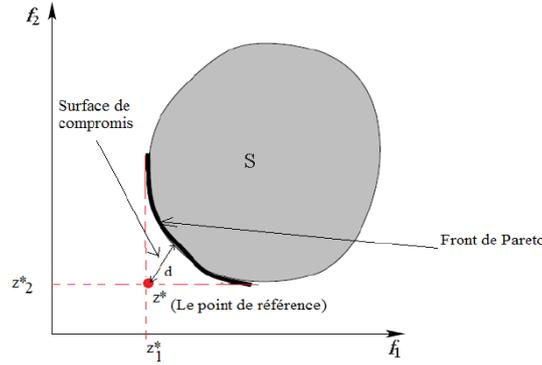


FIGURE 2.10 – Principe des méthodes de normes ou de distances pondérées

- Pour $r = 1$, le problème d'optimisation mono-objectif 2.9 peut s'écrire comme suit :

$$\min_{x \in S} d_1(x) = \sum_{i=1}^p |f_i(x) - z_i^*| \quad (2.11)$$

- Pour $r = 2$, ici on tombe sur la norme euclidienne, le problème d'optimisation mono-objectif 2.9 qui consiste à minimiser la distance en question peut se formuler comme suit :

$$\min_{x \in S} d_1(x) = \left[\sum_{i=1}^p |f_i(x) - z_i^*|^2 \right]^{\frac{1}{2}} \quad (2.12)$$

Ce problème est équivalent à

$$\min_{x \in S} d_2(x) = \sqrt{\left[\sum_{i=1}^p |f_i(x) - z_i^*|^2 \right]} \quad (2.13)$$

On peut souligner que, lorsque l'on utilise cette norme pour la recherche de solutions optimales au sens de Pareto, il n'est pas nécessaire d'ajouter le terme de la racine carrée $\sqrt{\cdot}$. Dans l'expression de la fonction objectif car elle est toujours positif (monotone). D'où le problème 2.13 sera réduit au problème suivant :

$$\min_{x \in S} d_2(x) = \left[\sum_{i=1}^p |f_i(x) - z_i^*|^2 \right] \quad (2.14)$$

- Pour $r \rightarrow \infty$, le problème d'optimisation mono-objectif 2.9 devient ainsi :

$$\min_{x \in S} d_\infty(x) = \max_{i=1,2,\dots,p} |f_i(x) - z_i^*| \quad (2.15)$$

Ce problème s'appelle la méthode "min max" ou méthode de "Chebychev". La méthode de Chebychev utilisant la distance "min max" est développée par (Stueur and Choo, 1983) est la

plus utilisée par rapport aux autre distance.

En effet, les auteurs (Stueur and Choo, 1983) ont utilisé des coefficients de pondération dans l'expression de la distance. Ces coefficients de pondération $w_i, i = 1, 2, \dots, p$ avec $w_i \geq 0, i = 1, 2, \dots, p$ et $\sum_{i=1}^p w_i = 1$ permettent de chercher un point dans une certaine direction, comme avec la méthode de pondération des fonctions objectif. On obtient alors l'expression suivante :

$$\min_{x \in \mathbb{R}^n} d_\infty(x) = \max_{i=1,2,\dots,p} (\omega_i |f_i(x) - z_i^*|) \quad (2.16)$$

Ce problème à été introduit pour la première fois par (Bowman, 1976), ce dernier est non-différentiable qui rend sa résolution impossible sous cette forme. En effet, (Bowman, 1976) a reformuler et proposer de résoudre un autre problème équivalent au problème 2.16 comme suit :

$$(P_\alpha) \begin{cases} \min \alpha \\ \alpha \geq \omega_i (f_i(x) - z_i^*), i = 1, 2, \dots, p \\ x \in S \\ \alpha \in \mathbb{R} \end{cases} \quad (2.17)$$

Le problème mono-objectif (P_α) s'appelle le problème de la norme pondérée de Chebychev. Il contient $n + 1$ variables de décision i-e $(x_1, x_2, \dots, x_n, \alpha)$.

Le théorème suivant nous donne quelques résultats théoriques concernant la relation entre la solution optimale du problème de la norme pondérée de Chebychev défini dans l'équation 2.17 et les solutions optimales de Pareto du problème multi-objectifs 2.1.

Supposons que les coefficients de pondération $w_i, i = 1, 2, \dots, p$ vérifient $w_i \geq 0, i = 1, 2, \dots, p$ et $\sum_{i=1}^p w_i = 1$

Théorème 25. 1. Le problème de la norme pondérée de Chebychev 2.17 admet au moins une solution optimale.

2. Si la solution optimale du problème de la norme pondérée de Chebychev 2.17 est unique alors elle est optimale au sens de Pareto pour le problème multi-objectifs 2.1.
3. La solution optimale du problème de la norme pondérée de Chebychev 2.17 est faiblement optimale au sens de Pareto si $w_i \geq 0, i = 1, 2, \dots, p$ pour le problème multi-objectifs 2.1

Comme remarque, comme la méthode de pondération des fonctions objectives, la méthode de la norme de Chebychev ne permet de trouver des points cachés dans des non convexités. En autre, cette méthode souffre d'autres inconvénients tel que l'ajout du p de contraintes et une variable de décision par rapport à la méthode de pondération des fonctions objectif.

2.6 Codage des méthodes avec MATLAB

2.6.1 Code de la méthode de somme pondérée

Programmation de l'exemple 2.5 avec MATLAB

Nous remarquons que l'exemple est un problème quadratique avec des contraintes en variables

bornées (Contraintes linéaires), nous utilisons la fonction `cplexap` de la plate-forme CPLEX qui permette de résoudre ce genre de problèmes.

```

function sommepondere ;
clc ;
format short
syms xy
f1 = (x - 5)^2 + (y - 5).^2
f2 = x^2 + y.^2
t = symvar(f1);
options = cplexoptimset ;
options.Diagnostics = 'off';
ff1 = [] ; ff2 = []
fori = -5 : 0.1 : 5 ;
forj = -5 : 0.1 : 5 ;
ff1 = [ff1; (i - 5).^2 + (j - 5).^2];
ff2 = [ff2; i.^2 + j.^2]
end
plot(ff1, ff2, 'r*')
hold on
grid
H1 = [20; 20];
H2 = [20; 20];
h1 = [-10, -10];
h2 = [0, 0];
cc=50;
A=[eye(2);-eye(2)];
b=[5*ones(2,1);5*ones(2,1)];
Xsol=[];
Fx=[];
for i=1 :100
lambda=rand;
Hd=lambda*H1+(1-lambda)*H2;
hd=lambda*h1+(1-lambda)*h2;
Hc=lambda*cc;
[xsol, fsol] = cplexqp(Hd, hd', A, b);
[xsol, ffd] = cplexqp(H2, h2', A, b, [], [], h1', H1, r);
Xsol = [Xsol, xsol];
Fx1 = subs(f1, t, xsol);
Fx2 = subs(f2, t, xsol);
Fx = [Fx; Fx1, Fx2];
plot(Fx1, Fx2, 'r*' );
end
end

```

la figure 2.11 montre le domaine réalisable dans l'espace des critères et le front de Pareto (En rouge) obtenu par MATLAB de l'exemple 2.5 pour un échantillon de 100 solutions de pareto

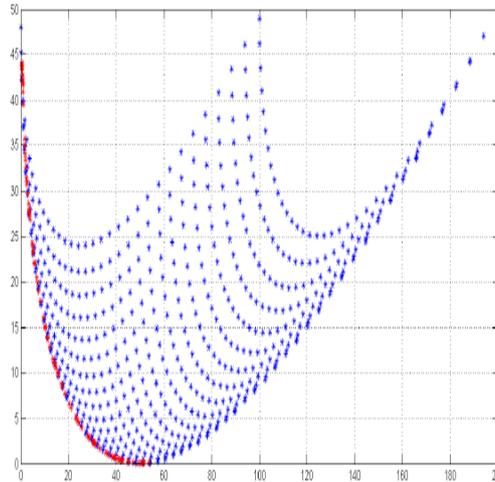


FIGURE 2.11 – Le front de Pareto obtenu par MATLAB de l'exemple 2.5

2.6.2 Code de la méthode du compromis

Programmation de l'exemple 2.7 avec MATLAB

L'exemple présenté pour illustrer le fonctionnement de la méthode ε -contraintes est un problème d'optimisation quadratique avec contraintes quadratiques. La fonction prédéfinie *cplexqcp* de la bibliothèque de *CPLEX* permet de résoudre facilement ce genre de problèmes d'optimisation.

Le programme suivant nous donne le code écrit sous *MATLAB* pour déterminer un sous-ensemble de point Pareto de 100 solutions. Le paramètre ε est positif dans notre cas puisque la contrainte $x_1^2 + x_2^2 \leq \varepsilon$ est réalisable uniquement pour des valeurs de ε positives. Nous ferons varier le paramètre ε dans l'intervalle $[0.1, 101]$ avec un pas égal à 1. Nous pouvons aussi la fonction *randint* pour générer aléatoirement les valeurs de ε dans l'intervalle $[0.1, 101]$.

```

function epsilonconstrained
clc;
format short
syms x y
f1=(x - 5).^2 + (y - 5).^2;
f2=x.^2 + y.^2;
t=symvar(f1);
options=cplexoptimset;
options.Diagnostics='off';
ff1=[];
ff2=[];
for i = -5 : 0.1 : 5;
for j = -5 : 0.1 : 5;
ff1=[ff1; (i - 5).^2 + (j - 5).^2];
ff2=[ff2; i.^2 + j.^2];
end
end
plot(ff1,ff2,'*');
hold on
grid
A=[eye(2);-eye(2)];
b=[5*ones(2,1);5*ones(2,1)];
H1=[20;02];
H2=[20;02];
h2=[-10,-10];
h1=[0,0];
Xsol=[];Fx=[];
for r = 0.1 : 1 : 101
[xsol,ffd] = cplexqcp(H2,h2',A,b,[],[],h1',H1,r);
Xsol=[Xsol xsol];
Fx1=subs(f1,t,xsol);
Fx=[Fx; Fx1, Fx2];
plot(Fx1,Fx2,'r*');
end
end

```

la figure 2.12 montre le domaine réalisable dans l'espace des critères et le front de Pareto

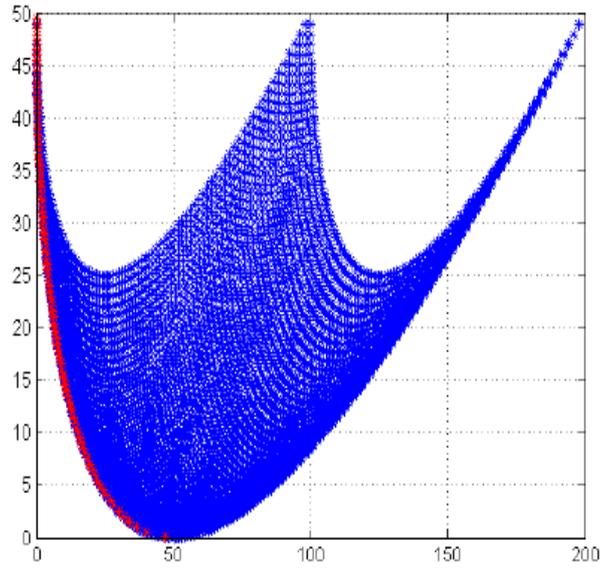


FIGURE 2.12 – Front de Pareto de l'exemple 2.7 avec la méthode ε -contraintes

(En rouge) obtenu par la méthode ε -contraintes de l'exemple 2.7 pour un échantillon de 100 solutions de Pareto.

Chapitre 3

Méthodes de direction de descente pour l'optimisation multi-objectifs sans contraintes

3.1 Introduction

Dans la littérature, la principale stratégie pour résoudre les problèmes multi-objectifs est les méthodes paramétriques (ou méthodes scalaires) parmi ces méthodes on peut citer la méthode de la somme pondérée et ε -contrainte.

Le principe général de ces méthodes consiste à utiliser un/ou plusieurs paramètres pour transformer le problème multi-objectifs (*MOP*) en question à un problème mono-objectif puis résoudre ce dernier et sous certaines conditions, il existe une relation étroite entre la solution optimale du problème mono-objectif et celle du problème multi-objectifs (*MOP*).

Cependant, il y a des nombreux inconvénients inévitables dans ces méthodes : les paramètres ne sont pas connus à l'avance, laissant au modélisateur et le décideur (*DM*) avec la charge des choix et si le (*MOP*) n'est pas convexe, il est difficile et impossible d'obtenir tous les optimums de Pareto en modifiant les paramètres. En fait, dans la méthode de pondération appliquée pour des problèmes multi-objectifs linéaire, il qu'un ne trouver pas certaines solutions Pareto pour n'importe quelle valeur du paramètre de pondération.

Afin de pallier ces inconvénients, des méthodes sans paramètres ou des méthodes non paramétriques ont été récemment développées. Ces méthodes sont des généralisations d'algorithmes classiques pour des problèmes à objectif unique. La méthode de direction de descente est l'une des méthodes ayant suscité un intérêt remarquable. Les premiers travaux sont dus à (J.Fliege et B.F. Svaiter, 2000), ayant proposé une méthode de direction de descente pour résoudre le problème multi-objectifs sans contraintes et avec contraintes.

Dans ce chapitre, nous nous intéressons entièrement à la description de méthode de direction

de descente développée par (J.Fliege et B.F. Svaiter, 2000) pour énumérer un sous ensemble du front de Pareto d'un problème multi-objectifs non linéaire et sans contraintes. La méthode donne généralement des points stationnaires qui vérifient les conditions d'optimalité nécessaires du premier ordre, sous certaines conditions de convexité du problème considéré les points stationnaires trouvés sont des points optimaux au sens de Pareto. Cette méthode est itérative, l'idée de base est simple, à chaque itération, cette méthode présente deux étapes :

1. Si la solution précédente n'est pas Pareto, une direction de descente d^k est calculée par la résolution d'un problème quadratique avec des contraintes linéaire. Un nouveau point x^{k+1} est déterminé tout au long de cette direction par l'itéré suivant

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

2. On utilise la règle d'Armijo pour déterminer le pas λ^k afin de trouver un point qui domine le point actuel tout le long de cette direction.

3.2 Description de la méthodes de la direction de descente

Le problème d'optimisation multi-objectifs non linéaire sans contraintes peut s'écrire comme suit :

$$\min_{x \in \mathbb{R}^n} F(x) \quad (3.1)$$

Où $F(x) = (f_1(x), f_2(x), \dots, f_k(x))^t$, $k \geq 2$.

Notons qu'un point $z \in \mathbb{R}^n$ est *localement optimal au sens de Pareto* ou *solution Pareto local* s'il y a un sous-ensemble de voisinage $U \subseteq \mathbb{R}^n$ de z tel qu'il n'existe pas de point $y \in U$ avec $F(y) < F(z)$ et $F(y) \neq F(z)$.

Nous Rappelons que si toutes les fonctions $f_i, i = 1, 2, \dots, k$ sont convexes alors caques point Pareto local est globalement optimale au sens de Pareto ou tout simplement *point Pareto global*.

Dans ce qui suit, nous supposons que toutes les fonctions $f_i, i = 1, 2, \dots, k$ sont continuellement différentiables en chaque point $x \in \mathbb{R}^n$. Pour chaque $x \in \mathbb{R}^n$, nous utilisons $(\nabla f_i(x))$ pour indiquer le Jacobien ou le gradient de $f_i, i = 1, 2, \dots, k$ en x et $\nabla F(x)$ est une matrice $n \times k$ avec des entrées

$$(\nabla F(x))_{i,j} = \frac{\partial F_i(x)}{\partial x_j} = (\nabla F(x)_i)_j \quad (3.2)$$

Le principe de cette méthode peut se résumer ainsi, à partir d'un point donné $x_0 \in \mathbb{R}^n$, la méthode génère une suite $\{x_k\} \in \mathbb{R}^n$ avec

$$F(x_{k+1}) < F(x_k), \forall k \geq 0 \quad (3.3)$$

Une condition nécessaire du premier ordre pour l'optimalité de Pareto locale est équivalente dire s'il n'existe pas de direction $d \in \mathbb{R}^n$ satisfaisant

$$\nabla f_i(x) d \leq 0, \forall i = 1, 2, \dots, p \quad (3.4)$$

Les points $x \in \mathbb{R}^n$ satisfaisant la condition 3.4 ci-dessus sont appelés points critiques de problème multi-objectifs 3.1.

Si x n'est pas critique du problème multi-objectifs 3.1, alors le vecteur $d \in \mathbb{R}^n$ est appelé direction de descente de F en x si $(\nabla f_i(x))d \leq 0, \forall i = 1, 2, \dots, k$.

L'idée principale simple de l'algorithme est la suivante : choisir un $x \in \mathbb{R}^n$ et vérifier la condition 3.4. Sinon, calculez une direction d par la résolution d'un sous-problème quadratique.

3.2.1 Calcul de la direction de descente

Toutes les informations qui sera dans cette partie proviennent de [8, 9, 10, 13].

Rappelons quelques définitions sur le point critique d'un problème multi-objectifs.

Définition 31. Un point $x^* \in \mathbb{R}^n$ est appelé point critique du problème multi-objectifs 3.1 si pour tout $d \in \mathbb{R}^n$ on a $\nabla f_i(x^*)^T d \geq 0, i = 1, 2, \dots, p$

Cette définition nous serve d'avantage dans le calcul des directions de descente.

Lemme 3. Soit un point $x^* \in \mathbb{R}^n$ qui n'est pas critique pour le problème multi-objectifs 3.1 alors il existe une direction de descente $d \in \mathbb{R}^n$ on a $\nabla f_i(x^*)^T d > 0, i = 1, 2, \dots, p$

Le théorème suivant récapitule la relation entre le point critique et la solution Pareto.

Théorème 26. Soit un point $x^* \in \mathbb{R}^n$ et le problème multi-objectifs 3.1, les propositions suivantes sont vérifiées

- x^* est une solution faiblement efficace alors x^* est un point critique.
- Si $f_i, i = 1, 2, \dots, p$ sont convexes et $x^* \in \mathbb{R}^n$ est un point critique du problème 3.1 alors $x^* \in \mathbb{R}^n$ est une solution faiblement Pareto du problème 3.1.
- Si $f_i, i = 1, 2, \dots, p$ sont strictement convexes et $x^* \in \mathbb{R}^n$ est un point critique du problème 3.1 alors $x^* \in \mathbb{R}^n$ est une solution Pareto du problème 3.1.

(V. Morovati, L. Pourkarimi, 2018) propose de résoudre un sous problème de forme quadratique pour déterminer une direction de descente. En effet, le théorème suivant nous donne la relation entre le point critique et une direction de descente.

Théorème 27. Considérons le problème multi-objectifs 3.1 et un point $x^* \in \mathbb{R}^n$ quelconque, les propositions suivantes sont équivalentes :

1. x^* est un point critique du problème multi-objectifs 3.1.
2. $\hat{d} = 0$ est une solution faiblement Pareto du problème multi-objectifs suivant :

$$\min_{d \in \mathbb{R}^p} \nabla f_1(x)d + \frac{1}{2} \|d\|^2 \quad \min_{d \in \mathbb{R}^p} \nabla f_2(x)d + \frac{1}{2} \|d\|^2 \quad \dots \quad \min_{d \in \mathbb{R}^p} \nabla f_p(x)d + \frac{1}{2} \|d\|^2 \quad (3.5)$$

Selon le théorème 27, si x^* n'est pas critique pour le problème 3.1 alors toute solution faiblement Pareto de sous-problème 3.5 est une direction de descente du F au point x^* . Dans le but de déterminer une direction de descente au point x^* , le sous-problème 3.5 est utilisé sous la forme de min-max scalarisation de type suivant

$$\min_{d \in \mathbb{R}^p} \left(\frac{1}{2} \|d\|^2 + \max_{i \in \{1, 2, \dots, p\}} \nabla f_i^T d \right) \quad (3.6)$$

D'après (V. Morovati, L. Pourkarimi, 2018), le problème min-max 3.6 est équivalent au sous-problème quadratique de contraintes linéaire suivant :

$$(SP) \begin{cases} \min \alpha(x) = \beta + \frac{1}{2} \|d\|^2 \\ \nabla f_i(x) d \leq \beta, \forall i = 1, 2, \dots, k \\ d \in \mathbb{R}^n, \|d\| \neq 0 \\ \alpha \in \mathbb{R} \end{cases} \quad (3.7)$$

Où (SP) est un problème quadratique convexe avec des contraintes d'inégalité linéaire qui est facile à résoudre par un algorithme simple.

Lemme 4. Le sous-problème (SP) 3.7 atteint son minimum dans \mathbb{R}^n .

Le lemme précédent nous garanti d'obtenir une solution optimale et que le sous-problème (SP) possède une seule solution optimale. Les solutions optimales du sous-problème (SP) sont notées $\beta(x)$ et $d(x)$. La valeur optimale correspondant à d pour un x fixée est notée par $\alpha(x)$.

Le sous-problème (SP) possède les propriétés suivantes.

Lemme 5. 1. Le sous-problème (SP) possède une solution unique.

2. Si x^* est critique du problème multi-objectifs 3.1, alors $d(x^*) = 0$ et $\alpha(x^*) = 0$.

3. Si x^* n'est pas critique, alors $\alpha(x^*) < 0$.

4. Les fonctions $d(x)$ et $\alpha(x)$ sont continues.

De ce Lemme, x^* est critique pour le problème multi-objectifs 3.1 si la valeur de la fonction objectif du sous-problème (SP) 3.7 est nulle en x^* c'est à dire $\alpha(x^*) = 0$.

3.2.2 Calcul du pas optimal selon la règle d'Armijo

Étant donné un point \bar{x} , une direction de descente est calculée par la résolution du sous-problème (SP) avec $\nabla F(\bar{x})^T d < 0$, un pas t devrait être déterminé par la règle d'Armijo de la manière suivante :

$$F(\bar{x} + dt) \leq F(\bar{x}) + \beta t \nabla F(\bar{x})^T d \quad (3.8)$$

Avec β est une constante choisi dans $]0, 1[$.

Au départ, on pose $t = 1$ si la condition 3.8 n'est pas satisfaite on posera $t = \frac{t}{2}$, répétons cette procédure jusqu'à la satisfaction de la condition d'Armijo.

Lemme 6. Si $\nabla F(x)^T d < 0$ alors il existe un paramètre $\epsilon > 0$ tel que

$$F(\bar{x} + dt) \leq F(\bar{x}) + \beta t F(\bar{x}) d \quad (3.9)$$

pour tout $t \in]0, \epsilon]$

Ce lemme nous indique que si une direction de descente existe au point \bar{x} alors un pas t existe tout au long de cette direction.

3.2.3 L'algorithme de la direction de descente

1. Choisir $\beta \in]0, 1]$, $\epsilon = 10^{-3}$ et un point initial $x^k \in \mathbb{R}$, posons $k = 1$.
2. calculer d^k en résolvant le sous-problème 3.7 .
3. Si $\alpha(x^k) \leq \epsilon$, x^k est critique pour le problème multi-objectifs 3.1. Arrêter.
4. Calculez un pas t^k selon la règle d'Armijo tel que indiquer dans l'équation 3.8.
5. Posons $x^{k+1} = x^k + t^k d^k$ et aller à l'étape (2).

Dans la suite nous allons montrer que l'algorithme de la direction de descente ci-dessus se termine après un nombre fini d'itérations, il se termine en un point critique du problème multi-objectifs 3.1

3.2.4 Convergence de la méthode de direction de descente

Les arguments de la convergence standard pour l'algorithme ci-dessus sont rapporté dans le théorème suivant.

Théorème 28. Chaque point de la séquence $(x^k)_k$ obtenu par l'algorithme de la méthode de la direction de descente est un point critique du problème multi-objectifs sans contrainte 3.1.

Si la fonction F des fonctions objectif a des ensembles de niveaux bornés dans le sens $\{x \in \mathbb{R}^n : F(x) \leq F(x^1)\}$ est bornée, alors la séquence $(x^k)_k$ reste bornée et possède au moins un point d'accumulation.

Démonstration Soit y un point d'accumulation de la suite $(x^k)_k$ et soit $d(y)$ et $\alpha(y)$ la solution et la valeur optimale du problème de minimisation sans contraintes en y :

$$d(y) = \operatorname{argmin}(f_y(d) + \frac{1}{2}\|d\|^2), \quad \alpha(y) = \min_d(f_y(d) + \frac{1}{2}\|d\|^2),$$

où

$$f_y(d) = \max_i (JF(y)d)_i$$

D'après le lemme 5, il suffit de prouver que $\alpha(y) = 0$. Clairement, la suite $(F(x^k))_k$ est strictement décroissante par composantes et on a

$$\lim_{k \rightarrow \infty} F(x^k) = F(y).$$

Donc,

$$\lim_{k \rightarrow \infty} \|F(x^k) - F(x^{(k+1)})\| = 0. \quad (3.10)$$

Mais

$$F(x^{(k)}) - F(x^{(k+1)}) \geq -t_k \beta JF(x^{(k)})d^{(k)} \geq 0 \quad (3.11)$$

et donc

$$\lim_{k \rightarrow \infty} t_k JF(x^{(k)})d^{(k)} = 0. \quad (3.12)$$

Remarquons $t_k \in]0; 1[$ pour tout k .

Maintenant prenons une sous-séquence $(x^{(k_u)})_u$ qui converge vers y . Nous allons considérer les deux possibilités

$$\limsup_{u \rightarrow \infty} t_{k_u} > 0$$

et

$$\limsup_{u \rightarrow \infty} t_{k_u} = 0.$$

Dans ce cas il existe une sous-suite $(x^{(k_l)})_l$ qui converge vers y et satisfaisant

$$\lim_{l \rightarrow \infty} t_{k_l} = \bar{t} > 0.$$

Utilisant l'équation 3.10, nous concluons que

$$\lim_{l \rightarrow \infty} JF(x^{(k_l)})d^{(k_l)} = 0,$$

Ce que implique aussi

$$\lim_{l \rightarrow \infty} \alpha(x^{(k_l)}) = 0.$$

Depuis $x \mapsto \alpha(x)$ est continue, nous concluons que $\alpha(y) = 0$, donc y est critique de problème multi-objectifs 3.1.

3.2.5 Exemple d'illustration

Pour illustrer la méthode de descente, nous considérons un MOPs dans l'exemple suivant avec deux objectifs

$$\begin{cases} \min f_1 = (x - 5)^2 + (y - 5)^2 + x - y \\ \min f_2 = x^2 + y^2 + y \\ x, y \in \mathbb{R}^2 \end{cases} \quad (3.13)$$

Les deux fonctions objectifs sont quadratiques, ils sont donc fortement convexes. De plus, chaque solution optimale localement de Pareto est globalement optimale de Pareto et chaque critique de Pareto est optimal de Pareto. Pour $x \in [-5, 5]^2$, une séquence de l'espace objectif est représentée dans la figure 3.1 Les gradients de fonctions sont $\nabla f_1(x) = (2x - 9, 2y + 11)^t$, $\nabla f_2(x) = (2x, 2y + 1)^t$

On fixe $x^0 = (-2, 2)^t$, $\varepsilon = 10^{-6}$ et le paramètre pour la longueur de pas Armijo est fixé à $\beta = 0.5$. Le sous-problème 3.5 résolu à chaque itération est un problème $\frac{1}{2}x^t H x + h^t x$ où

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, h' = (0, 0, 1)$$

Ces éléments de 3.5 deviennent inchangés à chaque itération de l'algorithme.

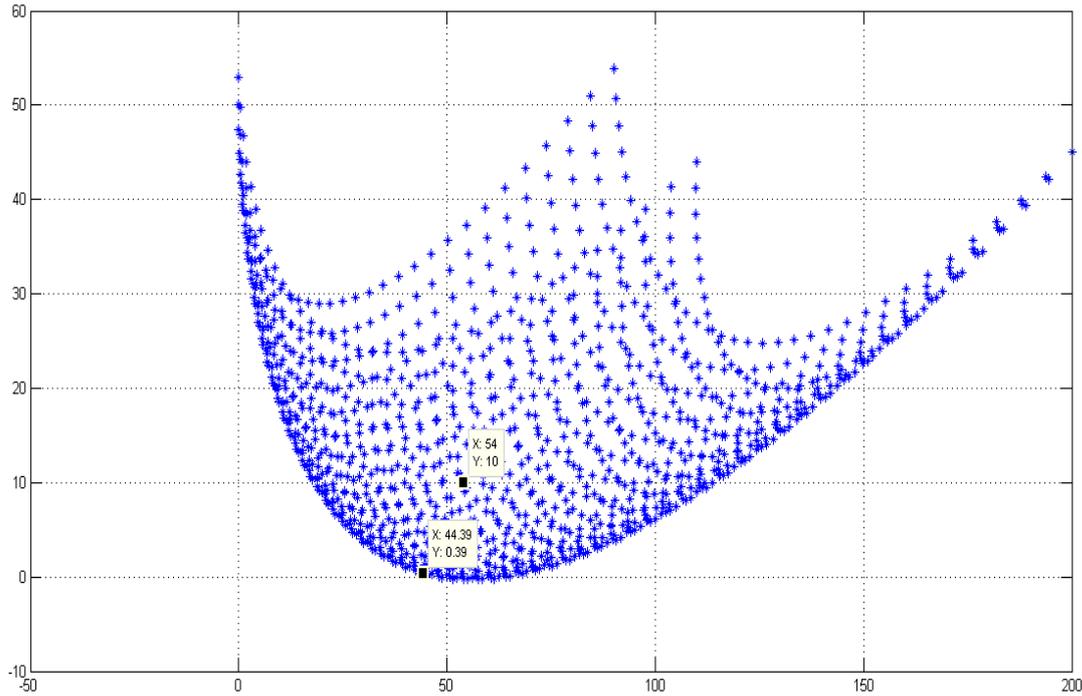


FIGURE 3.1 – Espace objectif de l'exemple 3.13

Itération 1. $\nabla f_1(x^0) = (-13, -7)^t$, $\nabla f_2(x^0) = (-4, 5)^t$.

Maintenant, nous résolvons le problème quadratique suivant

$$(SP1) \begin{cases} \min \alpha = \beta + \frac{1}{2}x^t H x + h^t x. \\ -13d_1 - 7d_2 \leq \beta. \\ -4d_1 + 5d_2 \leq \beta. \\ d_1, d_2, \beta \in \mathbb{R} \end{cases} \quad (3.14)$$

On obtient $t^1 = 0.5$, $d^{(1)} = (4.96, -3.72)^t$, $\alpha = -19.22$, $\beta = -38.44$ et la solution actuelle est

$$x^1 = (0.48, 0.14)^t \text{ and } F(x^1) = (44.39, 0.39)^r$$

Itération 2. $\nabla f_1(x^1) = (-8.04, -10.72)^t$, $\nabla f_2(x^1) = (0.96, 1.28)^t$.

$$(SP1) \begin{cases} \min \alpha = \beta + \frac{1}{2}x^t Hx + h^t x. \\ -8.04d_1 - 10.72d_2 \leq \beta. \\ 0.96d_1 + 1.28d_2 \leq \beta. \\ d_1, d_2, \beta \in \mathbb{R} \end{cases} \quad (3.15)$$

On obtient $d^2 = (0.133, 0.177)^t * 10^{-8}$, $\alpha^2 = 8.093 * 10^{-9} < \varepsilon$

L'algorithme s'arrête avec $x^* = (0.48, 0.14)^t$ comme solution optimal de Pareto pour l'exemple considéré.

3.3 Implémentation avec MATLAB

L'algorithme de la direction de descente qu'on a décrit ci-dessus est programmé sous MATLAB 7.12 (2011a), nous utilisons l'environnement CPLEX 12.2 version Étudiant (Student Version) pour résoudre le sous-problème quadratique 3.7. L'étude expérimentale qu'on va réaliser ici concerne le tracé de l'ensemble de solutions et le front de Pareto dans l'espace des critères pour cela nous sommes restreint à étudier uniquement des problèmes bi-objectifs. En effet, nous allons comparer la méthode de direction de descente avec la méthode de pondération des fonctions objectifs selon la qualité du front de Pareto.

$$f_\lambda = \lambda f_1(x) + (1 - \lambda)f_2(x), \lambda \in [0, 1] \quad (3.16)$$

Avec λ est le paramètre de pondération choisit aléatoirement dans l'intervalle $[0, 1]$. L'étude comparative des deux méthodes sera réduite à comparer la qualité du front de Pareto.

Test 1

Nous considérons le problème bi-objectifs strictement convexe suivant :

$$\begin{cases} \min f_1 = (x - 5)^2 + (y - 5)^2 + x - y \\ \min f_2 = x^2 + y^2 + y \\ x \in \mathbb{R}^2 \end{cases} \quad (3.17)$$

Dans le but de diversifier la population initiale notée par pop , nous générons aléatoirement une population de 500 points dans l'intervalle $[-5, 5]$ comme suit :

$$pop = -5 + 10 * rand(500, 2)$$

et on prendra le critère d'arrêt $|\gamma(x_k)| \leq 10^{-4}$ et la constante d'Armijo $\beta = 0.5$. Concernant, la méthode pondération un échantillon de 500 points est généré aléatoirement avec $\lambda \in [0, 1]$.

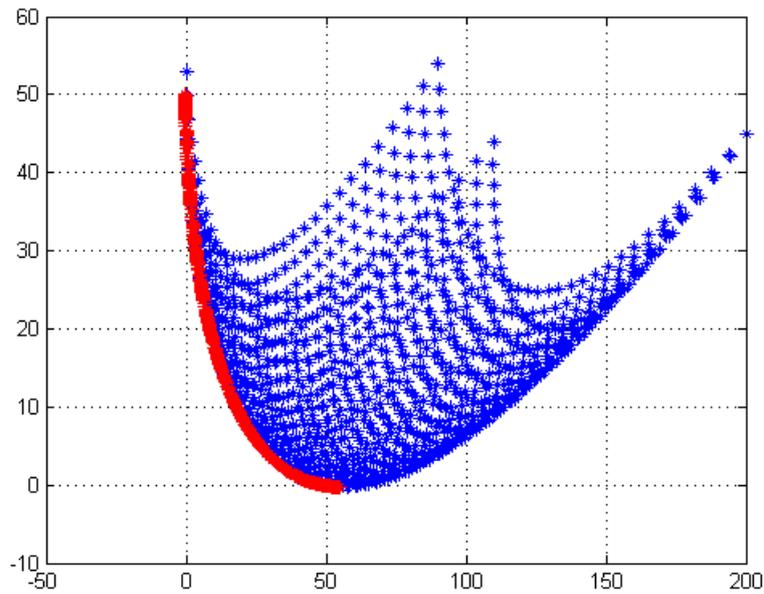


FIGURE 3.2 – Front de Pareto de l'exemple 3.17 avec la méthode de pondération

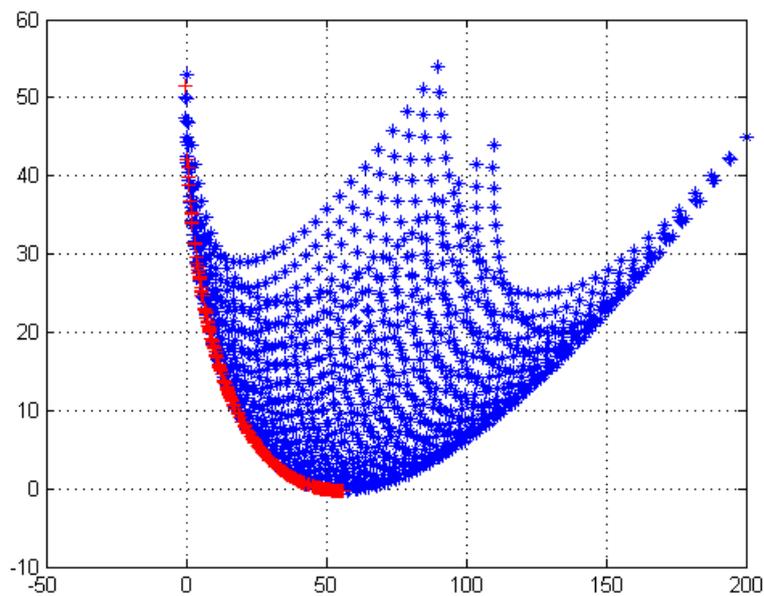


FIGURE 3.3 – Front de Pareto de l'exemple 3.17 avec la méthode de direction de descente

Les figures 3.2 et 3.3 montrent respectivement les front de Pareto des méthodes de pondération et la méthode de direction de descente. Nous constatons que les résultats trouvés par les deux méthodes sont similaires à moins que la méthode de direction de descente n'a pas déterminé les solutions Pareto dans l'intervalle $[40, 50]$ ceci est due à la technique de génération aléatoire de la population initiale.

Test 2

Nous considérons le problème bi-objectifs semi-convexe suivant :

$$\begin{cases} \min & f_1 = x.^2 - y \\ \min & f_2 = x.^2 + y - x \\ & x \in \mathbb{R}^2 \end{cases} \quad (3.18)$$

Dans ce test nous générons aléatoirement une population de 100 points dans l'intervalle $[-10, 10]$ comme suit :

$$pop = -10 + 20 * rand(100, 2)$$

et on prendra le critère d'arrêt $|\gamma(x_k)| \leq 10^{-5}$ et la constante d'Armijo $\beta = 0.75$. La figure 3.4 nous donne le front de Pareto, nous constatons que les solutions Pareto trouvées sont réparties d'une manière homogène

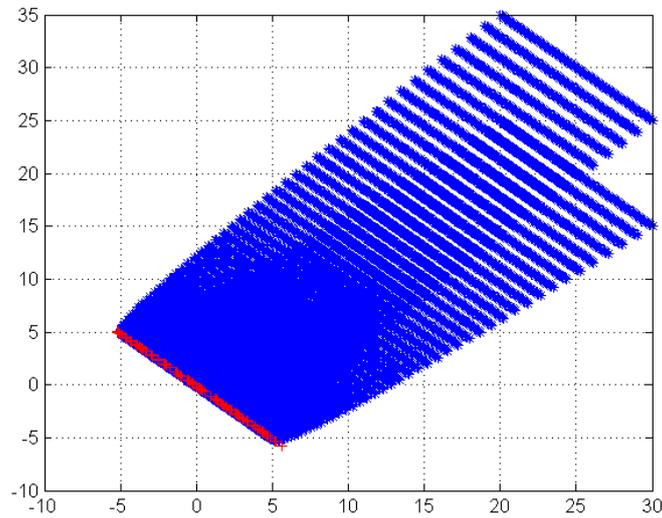


FIGURE 3.4 – Front de Pareto de l'exemple 3.18 avec la méthode de direction de descente

Pour la méthode pondération un échantillon de 100 points est généré aléatoirement avec $\lambda \in [0, 1]$. Dans cet exemple, la fonction de pondération est

$$\begin{aligned} \min & f_\lambda(x, y) = \lambda(x^2 - y) + (1 - \lambda)(x.^2 + y - x) \\ & = x^2 + (\lambda - 1)x + (1 - 2\lambda)y \end{aligned}$$

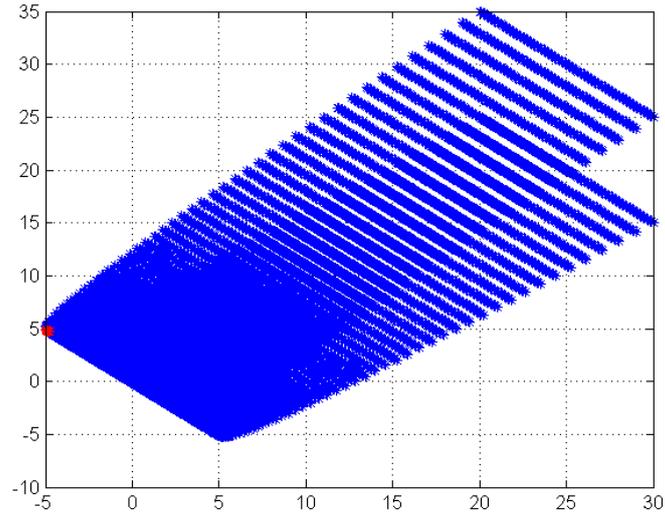


FIGURE 3.5 – Front de Pareto de l'exemple 3.18 avec la méthode de pondération

La figure 3.5 nous donne les solutions de Pareto trouvées sont concentrées autour du point $(0, 5)$.

Dans cet exemple, la méthode présente un inconvénient majeur dans la détermination des solutions Pareto diversifiées comme elles sont déterminées par la méthode de direction de descente. En effet, cet inconvénient peut être expliqué comme suit :

Calculons le gradient de la fonction de pondération

$$\nabla f_{\lambda}(x, y) = (2x - \lambda + 1, 1 - 2\lambda)$$

La condition nécessaire du premier ordre est

$$\nabla f_{\lambda}(x, y) = (0, 0) \Rightarrow x^* = -0.25, \lambda = 0.5, y^* \text{ quelconque}$$

Alors le point critique trouvé ici est le point $x^* = -0.25, y^*$ quelconque. De ce fait, nous remarquons que la méthode de pondération des fonctions ne peut pas déterminer les solutions Pareto du problème multi-objectifs 3.18 qui est semi-convexe.

Conclusion

Dans ce travail, nous avons étudié les problèmes de l'optimisation multi-objectifs non linéaire et sans contraintes. En effet, nous sommes particulièrement intéressés par la méthode de direction de descente développée par (J.Fliege et B.F. Svaiter, 2000) qui est une extension de la méthode de descente pour les problèmes mono-objectif. La méthode en question est itérative et procède essentiellement en deux étapes, la première étape consiste à déterminer une direction de descente au point courant qui peut se faire par la résolution d'un sous-programme quadratique, la règle d'Armijo est utilisée pour ajuster le pas de déplacement du point courant vers le prochain itéré tout au long de la direction.

L'algorithme de la direction de descente et la méthode de pondération des fonctions objectifs ont été programmés sous MATLAB et l'environnement CPLEX 12.2 pour résoudre les problèmes d'optimisation et les sous-problèmes quadratiques. L'étude expérimentale qu'on a réalisé sur deux tests concerne uniquement la recherche d'un sous ensemble de solutions Pareto. Les résultats montrent l'incapacité de la méthode de pondération dans la détermination d'un sous ensemble de Pareto même pour un problème multi-objectifs convexe et cela est justifié par les inconvénients classiques des méthodes scalaires dans le choix des valeurs des paramètres qui surgissent dans les méthodes. Par contre, la méthode de direction de descente nous a donné le front de Pareto approximatif et diversifié.

Enfin, nous suggérons d'enrichir ce travail par l'extension de la méthode de la direction de descente pour les problèmes multi-objectifs non linéaire avec contraintes et de réaliser une étude comparative avec d'autres méthodes scalaires telle que ϵ -contraintes.

Bibliographie

- [1] Y.Collette, P. Siarry,Multiobjective optimization : Principles and case studies,Springer Science and Business Media, 2013.
- [2] M.S. Bazaraa, H.D. Sherali and C. M. Shetty, The Nonlinear programming,Theory and Algorithms, John Willey-Sons, 2006.
- [3] W. Sun and Y.X Yuan, Optimization Theory and Methods, Nonlinear Programming, Springer Optimization and Its Applications, 2006.
- [4] K.P.E. Chong and S.H. Zak, An Introduction to Optimization, John Willey-Sons, 2001.
- [5] M.Ehrgott, Multicriteria Optimization, Springer, 2005.
- [6] M. Kaisa, Nonlinear Multiobjective Optimization, Springer Science+Business Media, LLC, 1999.
- [7] V. Morovati and L.Pourkarimi, Extension of Zoutendijk Method for Solving Constrained Multiobjective Optimization Problems, European Journal of Operational Research, 2018.
- [8] T.D.Chuong and J.C.Yao, Steepest descent methods for critical points in vector optimization problems, Applicable Analysis : An International Journal. (2012). 3 :1811-1829.
- [9] J. Fliege and B.F. Svaiter, Steepest descent methods for multicriteria optimization, Mathematical methods of operations research.(2000), 51 : 479-494.
- [10] B. Gebken, S. Peitz and M. Dellnitz, A Descent Method for Equality and Inequality Constrained Multiobjective Optimization Problems, Math O.C. (2017), 1712.03005.
- [11] M.E.Maghri and Y.Elboulqe, Reduced Jacobian Method, Journal of Optimization Theory and Applications. (2018),<https://doi.org/10.1007/s10957-018-1362-x>.
- [12] M.A.T Ansary and G. Panda, A sequential quadratically constrained quadratic programming technique for a multi-objective optimization problem, Engineering Optimization. (2018), DOI : 10.1080/0305215X.2018.1437154.
- [13] G.C. Bento,J.X. Cruz Neto, P.R. Oliveira and A. Soubeyran, The self regulation problem as an inexact steepest descent method for multicriteria optimization,European Journal of Operational Research. (2014), 235, 494-502.