الجمهوريـة الجزائـريـة الديمقراطيـة الشعبيـة
République Algérienne Démocratique et Populaire
وزارة التعليـم العالـي والبحـث العلمـي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Nº Réf :……………

Centre Universitaire
Abd elhafid boussouf Mila

Institut des sciences et de la technologie

Département de Mathématiques et Informatique

Mémoire préparé En vue de l'obtention du diplôme de Master

En : Mathématiques

Spécialité : Mathématiques appliquées

Méthodes d'optimisation avec contraintes

Préparé par : Aissa boudab
Abdelkrim Benhammada

Soutenue devant le jury

Boubelouta Khadidja	MCB	C. U. Abdelhafid Boussouf, Mila	Président
Fadel Wahida	MAA	C. U. Abdelhafid Boussouf, Mila	Rapporteur
Laib Hafida	MCA	C. U. Abdelhafid Boussouf, Mila	Examinateur

Année universitaire : 2021/2022

Remerciements

Nous tenons à remercier tout d'abord DIEU Allah qui nous a donné durant toutes ces années la santés, le courage et la patience pour en arriver là.

Nous tenons à exprimer nos profondes gratitude et nos sentiments les plus sincères à notre encadreur ***Madame Fadel Wahida*** *maitre de centre universitaire Abdelhafid Boussouf-Mila pour avoir dirigé nos travaux. Merçi pour vos échanges scientifiques, vos conseils et votre rigueur et pour tous les efforts que vous avez fournis pour faciliter et aider à accomplir notre travail de fin d'étude.*

Nous voudrions aussi remercier tous les membres du jury pour nous avoir honoré par leur présence et pour avoir accepté d'évaluer ce travail de mémoire.

Aussi nous adressons nos sincères remerciements à l'ensemble des enseignants et tout le personnel du département de Mathématique et Informatique.

Enfin .nous remercions nos parents. nos frères et soeurs, sans oublier nos amis, qui de près ou loin nous ont soutenus et encouragés tout au long de ces années.

Dédicaces

Je dédie ce travail :

à ma mère, pour son amour, ses encouragements et ses sacrifices.

à mon père, pour son soutien, son affection.

à mes frère, ma sœur, et son enfant et à tous les membres de ma famille,

à tous mes amis.

Je profite cette occasion pour dédier ce mémoire à toute ma promo.

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

****Abdelkarim****

Dédicaces

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,

A mes chères sœurs pour leurs encouragements permanents, et leur soutien moral,

A mes chers frères pour leur appui et leur encouragement,

A toute ma famille pour leur soutien tout au long de mon parcours universitaire,

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fuit de votre soutien infaillible,

Merci d'être toujours là pour moi.

****Aissa****

Résumé

Notre objectif dans ce mémoire est de faire une étude générale du problème d'optimisation avec contraintes et c'est à travers l'étude de l'existence et de l'unicité de la solution en premier lieu, puis nous avons présenté les conditions nécessaires et suffisantes d'optimalité des premier et de seconde ordre pour les contraintes d'égalité et d'inégalité, et enfin nous avons défini quelques méthodes qui permettent de trouver la solution optimale :

Méthode du Gradient Projeté, Méthode de Lagrange-Newton pour les contraintes en égalité, Méthode de Newton projeté pour les contraintes de borne, Méthode de Pénalisation, Méthode de programmation Quadratique Successive (S.Q.P), Méthodes d'UZAWA .

Mots clés : *optimisation avec contraintes, Méthodes de résolution d'un problème d'optimisation avec contraintes.*

Abstract

Our objective in this memory is a general study of an optimization problem with constraints and it is through the study of the existence and the uniqueness of the solution first, then we presented the necessary and sufficient conditions optimality for first- and second-order for the equality and inequality constraints, and finally we defined some methods which allow to find the optimal solution as :

Projected Gradient Method, Lagrange-Newton Method for Equality Constraints, Projected Newton Method for Bound Constraints, Penalization Method, Method of Successive Quadratic Programming (S.Q.P),Duality Method : UZAWA Methods.

Keywords *: constraints optimization, solution Method for constrained optimization problem.*

ملخص

هدفنا في هذا العمل هو عمل دراسة عامة لمشكل المثالية بقيود وذلك من خلال دراسة وجود ووحدانية الحل في المقام الأول بعدما قمنا باستعراض الشروط اللازمة والكافية للمثالية من الدرجة الأولى و الثانية لنوعي القيود بالمساواة وعدم المساواة ، و في الأخير قدمنا بعض الطرق التي تسمح بإيجاد الحل الأدنى وهي :

طريقة التدرج المسقط ، طريقة لاجرانج نيوتن من أجل القيود بمساواة ، طريقة نيوتن المسقطة من أجل القيود بحد ، طريقة الجزاء ، طريقة البرمجة التربيعية المتتالية (S.Q.P) ، طريقة UZAWA .

الكلمات المفتاحية : مشكل المثالية بقيود ، طرق حل مشكلة المثالية بقيود.

Table des matières

Introduction

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

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

Beaucoup de systèmes susceptibles d'être décrits par un modèle mathématique sont optimisés. La qualité des résultats et des prédictions dépend de la pertinence du modèle, du bon choix des variables que l'on cherche à optimiser, de l'efficacité de l'algorithme et des moyens pour le traitement numérique.

L'optimisation consiste en la recherche du minimum (ou du maximum) d'une certaine quantité, appelée coût ou objectif. Dans ce mémoire, on supposera que le coût dépend de n *variables réelles, rassemblées en un vecteur* $x = (x_1, \ , \ , x_n) \in \mathbb{R}^n$*, et qui fournissent une valeur* $f(x)$ *où* f *est une fonction de* $\mathbb{R}^n$ *dans* $\mathbb{R}$*. En général, les variables* $x = (x_1, \ , \ , x_n)$ *ne seront pas autorisées à prendre n'importe quelle valeur, mais devront satisfaire des contraintes que l'on représentera par un sous ensemble* $C \subset \mathbb{R}^n$*. On écrira les problèmes d'optimisation sous la forme générale suivante :*

$$(\mathcal{P}) \quad \min_{x \in C} f(x)$$

On dit que problème $(\mathcal{P})$ *admet une solution s'il existe un choix de variables* $x^* \in C$ *tel que*

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

On dit alors que x^* *est un minimiseur (ou point de minimum) de* f *sur* C*, et que* $f(x^*)$ *est un minimum de* f *sur* C*.*

Parmi les classification des problèmes d'optimisation on trouve :
Optimisation programmation linéaire.
Optimisation convexe.
Optimisation différentielle (ou lisse).
Optimisation SDP (semi-définie positive).
Optimisation non-différentielles (ou non lisse).
Optimisation multicritère (ou multiobjectif).
Optimisation en dimension infinie.
Optimisation discrète (ou combinatoire).
Sans contraintes. Avec contraintes.

On s'intéresse dans ce mémoire au problème d'optimisation avec contraintes à sa forme générale et nous avons opté le plan du travail suivant :

— *Le premier chapitre commence par présenter des notions mathématiques fondamentales, on définit et on introduit les outils fonctionnels de base nécessaire pour l'optimisation avec contraintes.*

— *Dans le deuxième chapitre nous nous intéressons à l'étude de quelque résultats théoriques (existence et unicité des solutions) aussi les conditions d'optimalité.*

— *Dans le dernier chapitre, nous consacrerons de quelques méthode de résolution classique les plus utilisées dans la pratique.*

Chapitre 1

Généralités et outils de base

Dans ce premier chapitre, nous apportons le bagage mathématique nécessaire Pour le traitement du problème considéré dans ce mémoire. Il s'agit de quelques définitions de topologie, calcul différentiel et de la convexité.

1.1 Rappel de calcul différentiel

Quelques Notations

[5]

1. *Pour tout* $n \in \mathbb{N}^*$, $\mathbb{R}^n$ *désigne l'espace euclidien* $\mathbb{R} \times \mathbb{R} \times \dots \mathbb{R}$ *("produit n fois").*
   En général un vecteur $x \in \mathbb{R}^n$ *sera noté* $x = (x_1, x_2, \dots x_n)^T$ *(vecteur colonne).*
2. *On note* $e_1, e_2, \dots e_n$ *les éléments de la base canonique de* $\mathbb{R}^n$*, où* e_i *est le vecteur de* $\mathbb{R}^n$ *donné par :*

$$(e_i)_j = \delta_{ij} = \begin{cases} 0 & si \quad j \neq i \\ 1 & si \quad j = i \end{cases}, \forall i, j = 1, 2 \dots n \tag{1.1}$$

(symboles de Kronecker).

3. *Pour tous* $x, y \in \mathbb{R}^n$ *on note par* $< x, y > \in \mathbb{R}$ *le produit scalaire de* x *et* y*, qui est donné par*
$$< x, y >= \sum_{i=1}^{n} x_i y_i.$$
4. *Pour tout* $x \in \mathbb{R}^n$ *on note par* $\|x\| \geq 0$ *la norme euclidienne de* x*, donnée*
$$\|x\| = \sqrt{< x, x >} = \sqrt{\sum_{i=1}^{n} x_i^2}$$
   Pour tous $x \in \mathbb{R}^n$

   et $r > 0$ *on notera par* $B(x, r)$ *la boule ouverte du centre* x *et rayon* r*, donnée par*
$$B(x, r) = \{y \in \mathbb{R}^n, \|y - x\| < r\}.$$
5. *Si* $a, b \in \mathbb{R}^n$ *on note* $[a, b]$ *le sous-ensemble de* $\mathbb{R}^n$ *donné par*
$$[a, b] = \{a + t(b - a) = (1 - t)a + tb, t \in [0, 1]\}.$$
   L'ensemble [a, b] est aussi appelé le segment reliant a à b.

   Remarque 1.1.1. *[5]*
   - $[a, b] = [b, a]$
   - *Si* $a, b \in: \mathbb{R}$ *avec* $a < b$ *alors on retrouve le fait que* $[a, b]$ *désigne l'intervalle des nombres* $x \in \mathbb{R}$ *tels que* $a \leq x \leq b$.
6. *On a* $< Mx, y\rangle =< x, My\rangle \ \forall x \in \mathbb{R}^n, y \in \mathbb{R}^m, M \in \mathcal{M}_{m,n}(\mathbb{R})$. *(*$\mathcal{M}_{m,n}$ *désigne l'espace des matrices de* m *lignes et* n *colonnes).*
7. *Rappelons aussi l'inégalité de Cauchy-Schwarz :*
$$| < x, y > | \leq \|x\|.\|y\| \qquad \forall x, y \in \mathbb{R}^n.$$

Gradient et Hessienne :

[5] Soit $\Omega \subset \mathbb{R}^n$ *un ouvert et* $f : \Omega \to \mathbb{R}$.

1. *On dit que f est de classe C^m sur Ω ($f \in C^m(\Omega)$) si toutes les dérivées partielles jusqu'à l'ordre m existent et sont continues.*

2. *Pour tout $x \in \Omega$ et tout $i \in 1, 2, \ldots, n$ on note (quand $\exists$)*

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0} \frac{1}{t}[f(x + te_i) - f(x)].$$

   (c'est la dérivée partielle de f en x de direction x_i.)

3. *Pour tout $x \in \Omega$ on note (quand $\exists$)*

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \ldots, \frac{\partial f}{\partial x_n})^T \in \mathbb{R}^n, \ \forall x \in \Omega$$

   (le Gradient de f en x).

   On note aussi

$$J_f(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \ldots, \frac{\partial f}{\partial x_n}) \in \mathbb{R}^n, \ \forall x \in \Omega$$

   (la Jacobienne de f en x). On a

$$\nabla f = (J_f)^T$$

4. *Pour tous $x \in \Omega$ et $h \in \mathbb{R}^n$ on note (quand $\exists$)*

$$\frac{\partial f}{\partial h}(x) = \lim_{t \to 0} \frac{1}{t}[f(x + th) - f(x)] = g'(0).$$

   (c'est la dérivée directionnelle de f en x de direction h) où on a noté :

$$g(t) = f(x + th).$$

   ***Remarque :**[5]*
   $\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x)$
   Nous rappelons aussi la formule :

$$\frac{\partial f}{\partial h}(x) = < \nabla f(x), h >, \forall x \in \Omega, \ \forall h \in \mathbb{R}^n.$$

5. *Pour* $x \in \Omega$ *on note (quand* $\exists$*)* $H_f(x) =$ *la matrice carrée* $\in \mathcal{M}_n(\mathbb{R})$ *donnée par*
$$H_f(x) = (\nabla^2 f(x))_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x),\ \forall i, j = 1, 2, \ldots n.$$
$H_f(x)$ *ou* $\nabla^2 f(x) =$ *s'appelle aussi la matrice Hessienne de* f *en* x.
Remarque : *Si* $f \in C^2(\Omega)$ *alors* $H_f(x)$ *est une matrice symétrique* $\forall x \in \Omega$ *(c.à.d* $H_f(x) = (H_f(x)^T)$*).*

Proposition 1.1.1. *[5](Gradient de la composée) Supposons qu'on a deux ouverts* $\Omega \subset \mathbb{R}^n$ *et* $U \subset \mathbb{R}$ *et deux fonctions* $f : \Omega \to \mathbb{R}$ *et* $g : U \to \mathbb{R}$ *avec en plus* $f(\Omega) \subset U$ *(on peut alors définir* $g \circ f : \Omega \to \mathbb{R}$*). Supposons que f, g sont de classe* C^1*. Alors* $g \circ f$ *est aussi de classe* C^1 *avec en plus*

$$\nabla(g \circ f)(x) = g'(f(x))\nabla f(x),\ \forall x \in \Omega.$$

Propriétés 1.1.1. *lien entre* ∇ *et* ∇^2

1. *La i - ème ligne de* $\nabla^2 f(x)$ *est la Jacobienne du i - ème élément de* ∇f*.*
2. *On a*
$$H_f(x)h = \nabla^2 f(x)h = \nabla < \nabla f(x), h >, \forall x \in \Omega, \forall h \in \mathbb{R}^n.$$

Eneffet :

$$\frac{\partial}{\partial x_i} < \nabla f(x), h >= \frac{\partial}{\partial x_i}\Big(\sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x)h_j\Big) = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x)h_j = \nabla^2 f(x)h_i = H_f(x)h_i.$$

Quelques exemples importants :

[5]

1. *Si* $f : \mathbb{R}^n \to \mathbb{R}$ *est une fonction constante alors* $\nabla f = \nabla^2 f = 0$*.*
2. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$ *définie par*

$$f(x) =< a, x >,\quad \forall x \in \mathbb{R}^n,$$

où $a \in \mathbb{R}^n$ *est un vecteur donné (c'est à dire,* f *est une fonction linéaire). Alors on calcule facilement :* $\dfrac{\partial f}{\partial x_k} = a_k$*, donc*

$$\nabla f = a$$

(le Gradient est constant).

Ceci nous donne

$$H_f = 0.$$

3. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$ *donnée par*

$$f(x) = < Ax, x >, \quad \forall x \in \mathbb{R}^n,$$

où $A \in \mathcal{M}_n(\mathbb{R})$ *est un matrice carrée, réelle, de taille* n *(c'est à dire,* f *est la fonction quadratique associée à la matrice* A*). Alors pour un* $p \in \{1, 2, \dots n\}$ *fixé, on peut écrire*

$f(x) = \sum\limits_{i,j=1}^{n} A_{ij} x_i x_j = A_{pp} x_p^2 + \sum\limits_{j=1, j \neq p}^{n} A_{pj} x_p x_j + \sum\limits_{i=1, i \neq p}^{n} A_{ip} x_i x_p + \sum\limits_{i,j=1, i \neq p, j \neq p}^{n} A_{ij} x_i x_j$

ce qui nous donne

$\dfrac{\partial f}{\partial x_p} = 2A_{pp} x_p + \sum\limits_{j=1, j \neq p}^{n} A_{pj} x_j + \sum\limits_{i=1, i \neq p}^{n} A_{ip} x_i = \sum\limits_{j=1}^{n} A_{pj} x_j + \sum\limits_{i=1}^{n} A_{ip} x_i = (Ax)_p + (A^T x)_p.$

Nous avons donc obtenu :

$$\nabla f(x) = (A + A^T)x, \quad \forall x \in \mathbb{R}^n.$$

On peut aussi écrire

$$\frac{\partial f}{\partial x_i} = \sum_{k=1}^{n} (A + A^T)_{ik} x_k, \quad \forall i = 1, \dots, n.$$

On a alors immédiatement :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = (A + A^T)_{ij}, \quad \forall i, j = 1, \dots, n.$$

c'est à dire

$$H_f(x) = A + A^T, \quad \forall x \in \mathbb{R}^n$$

(donc la Hessienne de f est constante).

Remarque *En particulier, si A est symmétrique alors*

$$\nabla < Ax, x >= 2Ax, \ \forall x \in \mathbb{R}^n.$$

$$\nabla^2 < Ax, x >= 2A, \ \forall x \in \mathbb{R}^n.$$

formules de Taylor

Proposition 1.1.2. *[5]*

Soit $\Omega \subset \mathbb{R}^n$ ouvert, $f : \Omega \to \mathbb{R}$, $a \in \Omega$ et $h \in \mathbb{R}^n$ tels que $[a, a+h] \subset \Omega$. Alors :

1. *Si $f \in C^1(\Omega)$ alors*

   i) $f(a+h) = f(a) + \int_0^1 < \nabla f(a+th), h > dt$
   (formule de Taylor à l'ordre 1 avec reste intégral).

   ii) $f(a+h) = f(a)+ < \nabla f(a+\theta h), h >$ *avec* $0 < \theta < 1$
   (formule de Taylor - Maclaurin à l'ordre 1)

   iii) $f(a+h) = f(a)+ < \nabla f(a), h > +o(\|h\|)$
   (formule de Taylor - Young à l'ordre 1)

2. *Si $f \in C^2(\Omega)$ alors*

   i) $f(a+h) = f(a)+ < \nabla f(a), h > + \int_0^1 (1-t) < Hf(a+th)h, h > dt$ *(formule de Taylor à l'ordre 2 avec reste intégral).*

   ii) $f(a+h) = f(a)+ < \nabla f(a), h > +\frac{1}{2} < Hf(a+\theta h)h, h >$ *avec* $0 < \theta < 1$
   (formule de Taylor - Maclaurin à l'ordre 2)

   iii) $f(a+h) = f(a)+ < \nabla f(a), h > +\frac{1}{2} < Hf(a)h, h > +o(\|h\|^2)$
   (formule de Taylor - Young à l'ordre 2).

Remarque : *Dans la proposition précédente la notation $o(\|h\|^k)$ pour $k \in \mathbb{N}^*$ signifie une expression qui tend vers 0 plus vite que $\|h\|^k$ (c'est à dire, si on la divise par $\|h\|^k$, le résultat tend vers 0 quand $\|h\|$ tend vers 0).*

Dérivée au sens de Gateaux

Définition 1.1.1. *[10] Soient $\mathbb{E}$ et $\mathbb{F}$ deux espaces vectoriels normés*
Soient v un vecteur unitaire de $\mathbb{E}$ et f une fonction définie sur $\mathbb{E}$ à valeurs dans $\mathbb{F}$
On dit que f est dérivable au sens de Gateaux au point $x \in \mathbb{E}$ suivant le vecteur v si

- $\lim \dfrac{f(x+tv)-f(x)}{t}$ *existe lorsque $t \to 0$ (on notera cette limite $df_v(x)$ qui est un élément de $\mathbb{F}$ et non une application linéaire de $\mathbb{E}$ dans $\mathbb{F}$).*
- *$v \to df_v(x)$ est linéaire continue.*

1.2 Notions de topologie

Ensemble ouvert et ensemble fermé et borné

Définition 1.2.1. *[2] Soit $C \subseteq \mathbb{R}^n$ et $v(x)$ un sous ensemble ouvert de $\mathbb{R}^n$ contient C, on dit que $v(x)$ est un voisinage de C.*

Définition 1.2.2. *[4](point intérieur)*
Soit $C \subseteq \mathbb{R}^n$ et $x \in C$: On dit que x est intérieur à C s'il existe un voisinage de x contenu dans C. D'une manière équivalente, x est intérieur à C s'il existe $\varepsilon > 0$ tel que :

$$\forall y \in C, \quad \|x-y\| \leq \varepsilon.$$

L'ensemble des points intérieur de C est appelé intérieur de C et noté $Int(C)$ ou C^0.

Définition 1.2.3. *[2](Ensemble ouvert)*
Un ensemble $C \subset \mathbb{R}^n$ est un ouvert si et seulement si $C^0 = C$

Définition 1.2.4. *[2](L'adhérence)*
On dit que $C \in \mathbb{R}^n$ est un point d'adhérence d'un sous ensemble $C \subset \mathbb{R}^n$:

$$\forall v(x), v(x) \cap C \neq \phi.$$

L'ensemble de tous les points d'adhérence d'un ensemble est appelé fermeture ou clôture noté cl(C) ou $\overline{C}$.

Définition 1.2.5. *[2](Ensemble fermé)*
$C \subset \mathbb{R}^n$ *est un ensemble fermé si et seulement si* $C = \overline{C}$.

Définition 1.2.6. *[2](Ensemble compact)*

Soit C un sous-ensemble de $\mathbb{R}^n$. C est compact si pour toute suite $(x_k)_k$ d'élément de C, il existe une sous-suite convergente vers un élément de C.

C est compact si et seulement si C est fermé et borné.

Définition 1.2.7. *[1](Ensemble borné)*
Soit $C \subset \mathbb{R}^n$ est un ensemble borné si : $\exists M > 0$ tel que : $\forall x \in C$, $\|x\| \leq M$.

1.3 convexité

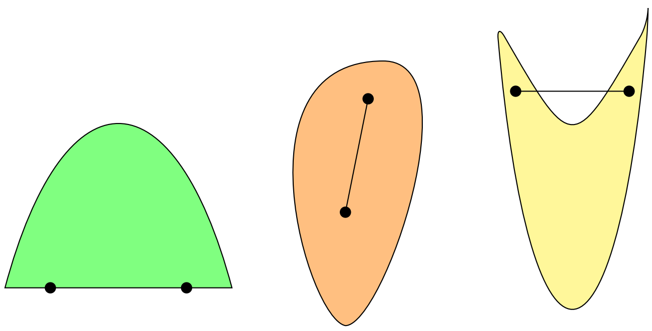
1.3.1 Définitions

Ensembles convexes

[5] Une partie C de $\mathbb{R}^n$ est convexe si pour tout couple de points x et y de C le segment $[x, y]$ est entièrement dans C, formellement :

Une partie $C \subset \mathbb{R}^n$ est (un) convexe si

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



convexe *strictement convexe* *non-convexe*

Proposition

[11]
On a les propriétés élémentaires suivantes :

1. *Si C_1, C_2 sont deux convexes de $\mathbb{R}^n$ et λ_1, λ_2 deux réels, alors $\lambda_1 C_1 + \lambda_2 C_2$ est un convexe de $\mathbb{R}^n$.*
2. *Si $(C_j)_{j \in J}$ est une famille quelconque de convexes de $\mathbb{R}^n$, alors $\cap_{j \in J} C_j$ est un convexe de $\mathbb{R}^n$.*
3. *Si C est un convexe de $\mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}^n$ est une application affine, i.e. $f(x) = Ax + b$ pour certains $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ et $b \in \mathcal{M}_m(\mathbb{R})$; alors $f(C)$ est un convexe de $\mathbb{R}^n$.*

Combinaison convexe

[11]
Soient $x_1, ..., x_k$ un nombre fini de points de $\mathbb{R}^n$, et $\lambda_1, ..., \lambda_k$ des réels tels que

$$\lambda_j \geq 0 \quad \forall j = \overline{1, k} \quad \textit{et} \; \sum_{j=1}^{k} \lambda_j = 1$$

On dit que

$$x = \sum_{j=1}^{k} \lambda_j x_j$$

est une combinaison convexe des points $x_1, ..., x_k$. Plus généralement, si $S \subseteq \mathbb{R}^n$ est un sous-ensemble quelconque, on dit que $x \in \mathbb{R}^n$ est une combinaison convexe de points de S s'il existe un nombre fini de points de S dont x soit une combinaison convexe.

Remarque

Dans le cas particulier de deux points x_1 et x_2, toute combinaison convexe des $x_1; x_2$ peut s'écrire sous la forme

$$x = \lambda x_1 + (1 - \lambda) x_2, \quad \lambda \in [0, 1]$$

qui intervient dans la définition de l'ensembles convexes.

Proposition

[11]

Un sous-ensemble C de $\mathbb{R}^n$ est convexe si et seulement si il contient toutes les combinaisons convexes de points de C.

Remarque

Lorsqu'un ensemble de $\mathbb{R}^n$ n'est pas convexe, on peut considérer le plus petit ensemble convexe le contenant, c'est la notion importante suivante d'enveloppe convexe.

Enveloppe convexe

[11]

L'enveloppe convexe d'une partie $E \subset \mathbb{R}^n$ est le plus petit convexe contenant E. C'est aussi l'ensemble des combinaisons convexes des points de E. les points de l'enveloppe convexe sont combinaison convexe d'au plus n + 1 points de E.

Polyèdre convexe

[5]

On appelle polyèdre l'intersection d'un nombre fini de demi-espaces de $\mathbb{R}^n$, i.e. l'ensemble des points vérifiant un système d'inéquations linéaires toutes dans le même sens.

Un polyèdre borné est aussi l'enveloppe convexe d'un nombre fini de points, appelés sommets (nous admettrons cette équivalence, non triviale). Les solutions d'un système de p inéquations linéaires à n inconnues définissent donc un polyèdre dans $\mathbb{R}^n$. d'où l'importance de la théorie des polyèdres pour l'optimisation sous contraintes *d'inégalité.*

1.3.2 Points extrémaux

Définition 1.3.1. *[5] Un point $x \in C$ d'un convexe C est dit extrémal si il n'est pas le milieu de deux points inclus dans C.*

Exemples :[5]

- *Tous les points du cercle bordant un disque.*
- *Les sommets d'un polyèdre (i.e. on montre que les points extrémaux d'un polyèdre dans $\mathbb{R}^n$ sont intersection d'au moins n hyperplans frontières).*
- *Les points extrémaux de l'ensemble des matrices bistochastiques sont les matrices de permutation, i.e. les matrices formées de 0 et 1 qui ont un seul terme non nul sur chaque ligne et chaque colonne.*

  Le résultat suivant est essentiel pour l'optimisation sous contraintes d'inégalité d'une fonction linéaire :

Proposition 1.3.1. *Le maximum d'une fonction convexe sur un convexe compact est atteint en au moins un point extrémal.*
(résultat admis). En particulier le maximum d'une fonction linéaire sur un polyèdre borné est atteint en un point sommet.

1.3.3 Fonctions convexes

Définitions

* *Fonctions convexes[11]*

Soit $C \subseteq \mathbb{R}^n$, un ensemble convexe.

La fonction $f : C \to \mathbb{R}^n$ est dite convexe si, et seulement si

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

* *Fonctions strictement convexes*[11]

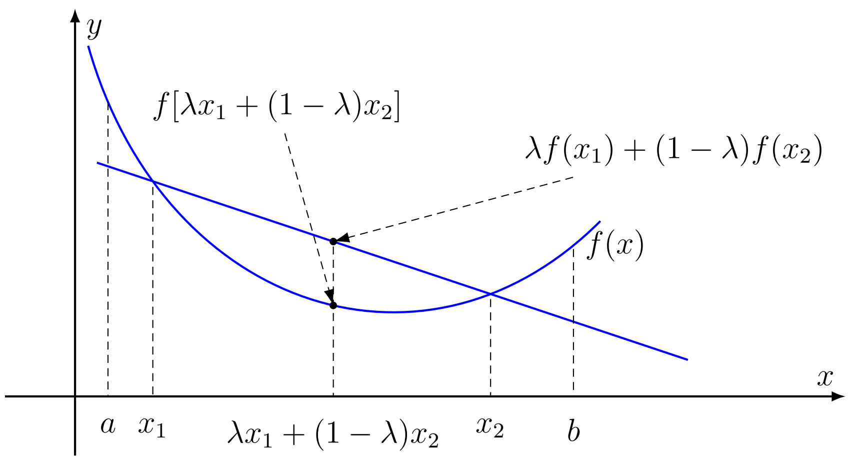
$$\forall x, y, x \neq y, \in C, \forall \lambda \in]0,1[: \quad f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

* *Une fonction est (strictement) convexe si et seulement si sa restriction à tout segment inclus dans C est (strictement) convexe : pour établir la convexité on peut donc toujours se ramener à une fonction d'une variable.*

* *Si $f(x)$ est convexe l'ensemble $\{x/f(x) \leq c\}$ est convexe (la réciproque est fausse).*

* *$f(x)$ est convexe si son épigraphe $\{(x,r), x \in \mathbb{R}^n, r \in \mathbb{R}/r \geq f(x)\}$ est un convexe de $\mathbb{R}^n \times \mathbb{R}$ (ce qui revient à dire intuitivement que son graphe est le bord d'un convexe)..*

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



la convexité de fonction f.

Exemple 1.3.1. *La fonction f définie sur ${\mathbb{R}_+^*}^2$ par :*

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

est une fonction strictement convexe.

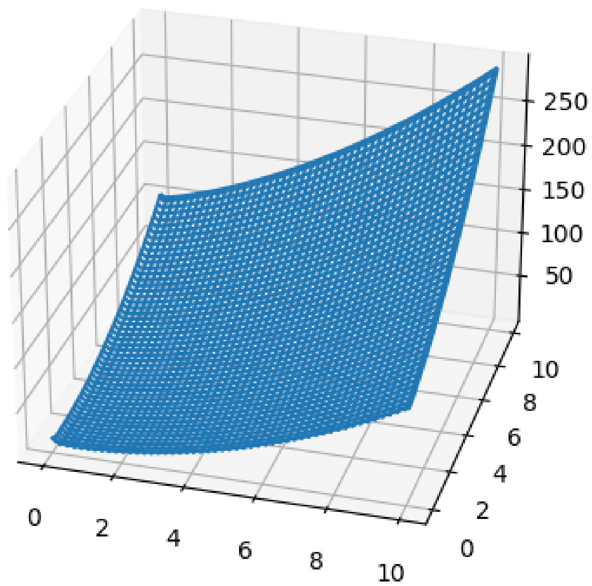


FIGURE 1.1 – la fonction f.

1.3.4 Optimisation et convexité

[5] Rappelons d'abord quelques propriétés d'existence des minimums d'une fonction continue.

Définition 1.3.2. *[5]Une fonction est coercive si* $\lim_{\|x\|\to+\infty} f(x) = +\infty$
Une fonction continue admet un minimum sur tout ensemble compact (donc en particulier sur les parties fermées bornées de $\mathbb{R}^n$*). On en déduit que sur un espace de dimension finie une fonction continue coercive admet au moins un minimum, mais elle peut admettre plusieurs minimums et d'autres types d'extrémums. Sur un espace de dimension finie une fonction convexe est continue, si elle est coercive elle admet donc au moins un minimum. Plus généralement le fait que la fonction est convexe permet de préciser l'existence et la nature de ses extrémums.*

Proposition 1.3.2. *[5] Si* $f(x)$ *est une fonction convexe différentiable sur* $\mathbb{R}^n$

$$\nabla f(x) = 0 \Leftrightarrow \forall y \in \mathbb{R}^n \quad f(x) \leq f(y)$$

Le fait que la différentielle est nulle en un point, condition nécessaire pour le minimum d'une fonction quelconque, est donc ici une condition suffisante d'existence d'un minimum global.

Proposition 1.3.3. *[5]Un minimum local d'une fonction convexe est global. L'ensemble des minimums d'une fonction convexe est convexe.*

Proposition 1.3.4. *[5] Une fonction strictement convexe sur un convexe* C *a au plus un minimum sur* C*.*
En résumé, la propriété que nous utiliserons le plus

Théorème 1.3.1. *[5] Une fonction strictement convexe sur un convexe compact* C *admet un minimum unique.*

Théorème 1.3.2. *[5] Une fonction strictement convexe et coercive sur un espace de dimension finie admet un minimum unique.*

1.4 points critiques

Définition 1.4.1. *[13] (Points critiques)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une application différentiable. Tout point $x \in \mathbb{R}^n$ vérifiant :

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

est appelé point critique (ou point stationnaire) de f.

Théorème 1.4.1. *[13] Soit f une fonction de classe $\mathcal{C}^2$. Soit $X^* = (x_0,\ y_0,\ z_0, \dots) \in \mathbb{R}^n$ un point critique de $f(X) \Rightarrow \nabla f(X^*) = 0$. La matrice Hessienne $H_f(x^*)$ au point $X^* = (x_0,\ y_0,\ z_0, \dots)$ est symétrique dont les valeurs propres $\{\lambda_i\}_{i=1,n} \in \mathbb{R}$ et $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p \leq \cdots \leq \lambda_n$.*

- *si $\{\lambda_i\}_{i=1,n} > 0 \Rightarrow f$ admet un minimum local au point critique $X^* = (x_0,\ y_0,\ z_0,\ \dots)$. la matrice $H_f(x^*)$ est dite définie positive.*
- *si $\{\lambda_i\}_{i=1,n} < 0 \Rightarrow f$ admet un maximum local au point critique $X^* = (x_0,\ y_0,\ z_0,\ \dots)$. la matrice $H_f(x^*)$ est dite définie négative.*
- *si $\{\lambda_i\}_{i=1,n} = 0 \Rightarrow$ il faudra analyser la série de Taylor a des ordres ≤ 3 pour pouvoir conclure sur le comportement de f au voisinage du point critique.*
- *si $\forall j \neq i$ tel que $\lambda_j = 0$, pour conclure sur la nature du point critique, il faudra déterminer le signe de la trace soit : T_{H_f}.*
  Comme il a été mentionné précédemment, la matrice Hessienne $H_f(x)$, est symétrique donc elle est diagonalisable. Le signe est conservé par changement de base. Cela signifie concrètement que les signes de $\{\lambda_i\}_{i=1,n}$ de la matrice diagonale sont identiques à ceux des éléments diagonaux de la Hessienne. Par conséquent, il n'est pas nécessaire de calculer les valeurs propres pour conclure sur la nature d'un point critique. Du fait de la conservation du signe, après changement de base, nous nous baserons uniquement sur la matrice $H_f(x^*)$. Nous en déduisons les cas de figure ci-dessous :

i) si $|H_f(x^*)| > 0 \Rightarrow$ les $\{\lambda_i\}_{i=1,n}$ ont le même signe.

- *si la trace* $T_{H_f} > 0$ *alors les* $\{\lambda_i\}_{i=1,n} > 0$. *De la formule de Taylor , il en découle* $f > 0 \Rightarrow f(x_0 + \delta a,\ y_0 + \delta b) > f(x_0,\ y_0)$ *alors la fonction* f *prend des valeurs supérieures au voisinage* $[x_0 + \delta a,\ y_0 + \delta b]$. *La fonction croît, nous concluons que le point critique* $(x_0,\ y_0)$ *est nécessairement un minimum local.*
- *si la trace* $T_{H(f)} < 0$ *alors les* $\{\lambda_i\}_{i=1,n} < 0$. *De la formule de Taylor , il en découle* $f < 0 \Rightarrow f(x_0 + \delta a,\ y_0 + \delta b) < f(x_0,\ y_0)$ *alors la fonction* f *prend des valeurs inférieures au voisinage* $[x_0 + \delta a,\ y_0 + \delta b]$. *La fonction diminue, nous concluons que le point critique* $(x_0,\ y_0)$ *est nécessairement un minimum local.*

ii) *si* $|H_f(x^*)| < 0 \Rightarrow$ *les* $\{\lambda_i\}_{i=1,n}$ *sont de signes contraires alors la fonction f admet un point selle.*

iii) *si* $|H_f(x^*)| = 0 \Rightarrow \forall j \neq i,\ \lambda_j = 0$. *Pour conclure sur la nature du point critique, il faut déterminer le signe de la trace de la Hessienne.*

- *si la trace* $T_{H_f} > 0$ *alors* f *admet un minimum local.*
- *si la trace* $T_{H_f} < 0$ *alors* f *admet un maximum local*

Pour le cas d'une fonction de trois variables $X \in \mathbb{R}^3$, *d'après la théorie développée précédemment, un déterminant* $|H_f(x^*)| > 0$ *signifie un produit positif des valeurs propres* $\prod_{i=1}^{3} \lambda_i > 0$. *Par conséquent nous pouvons dégager deux séquences de signes* $(+, +, +)$ *ou* $(-, -, +)$ *sachant que d'après le théorème ci-dessus nous obtenons toujours* $\lambda_1 \leq \lambda_2 \leq \lambda_3$. *Afin de conclure sur la nature des points critiques nous devons calculer la trace de la Hessienne*

i) *si la trace* $T_{H_f} < 0$ *alors le point critique est un point-selle.*

ii) *si la trace* $T_{H_f} > 0$ *c'est un cas ambigu. Il faudra calculer les valeurs propres.*

- *si* $\prod_{i=1}^{3} \lambda_i > 0$ *alors f admet un minimum local.*
- *si* $\prod_{i=1}^{3} \lambda_i > 0$ *alors si toutes les valeurs propres sont positives, f admet un minimum local. Si au moins une des valeurs propres est négative alors il s'agit d'un point-selle.*

Une autre manière de vérifier si $H_f(x)$ *est définie positive (existence d'un minimum* *local), ou définie négative (existence d'un maximum local) est de considérer :*

$$H_f(x) \text{ est définie positive } \Rightarrow \forall X \neq 0 \in \mathbb{R}^n : X^T H_{ij}(f) X > 0 \tag{1.2}$$

$$H_f(x) \text{ est définie négative } \Rightarrow \forall X \neq 0 \in \mathbb{R}^n : X^T H_{ij}(f) X < 0 \tag{1.3}$$

Une matrice $H_f(x) \in \mathcal{M}_{n\times n}(\mathbb{R})$ ne vérifiant aucunes de ces propriétés est dite indéfinie. Un moyen pratique de vérifier si une matrice est définie positive est de calculer les n-déterminants des sous-matrices de $H_f(x)$,
soit :

$$\triangle_1 = |a_{11}| \quad \triangle_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \triangle_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \quad \ldots\ldots \quad \triangle_n = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{2n} & a_{3n} & \ldots & a_{nn} \end{vmatrix}$$

1. *Si $\{\triangle_i\}_{i=1,n} > 0 \Rightarrow H_{ij}(f)$ est définie positive.*
2. *Si $\{\triangle_i\}_{i=1,n} < 0 \Rightarrow H_{ij}(f)$ est définie négative.*

Concluons cette analyse sur la nature d'un point critique par le théorème ci-dessous.

Remarque 1.4.1. *[13]*

— *Faudra-il encore vérifier que $H_f(x)$ est semi-définie positive $\Rightarrow \forall X \in \mathbb{R}^n$: $X^T H_f(x)(f)X \geq 0$ et $H_f(x)$ est semi-définie négative $\Rightarrow \forall X \in \mathbb{R}^n$: $X^T H_f(x) X \leq 0$. Par ailleurs, une matrice symétrique vérifie $A = A^T$, ses valeurs propres sont réelles.*
De plus si A est définie positive alors elle est inversible et ses valeurs propres $\{\lambda_i\}_{i=1,n}$ sont positives.

— *Ces n-déterminants sont appelés les mineurs principaux dominants.*

Théorème 1.4.2. *[13] (conditions suffisantes d'optimalité) : Soit $f : \mathbb{R}^n \to \mathbb{R}$ une application de classe $\mathcal{C}^2$ sur $\mathbb{R}^n$. Si $X^* \in \mathbb{R}^n$ est un minimum local (respectivement un maximum local) de f alors :*

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

Condition d'optimalité du premier ordre et de plus,

$$H_{ij}(f(X^*)) = \nabla^2 f(X^*)$$

Condition d'optimalité du second ordre est définie positive (respectivement définie négative)
La condition du premier ordre sur le gradient de la fonction, $\nabla f(X^*) = 0$, permet d'identifier un point stationnaire sans que l'on sache s'il s'agit d'un minimum, d'un maximum ou d'un point de selle. La condition du second ordre portant sur le Hessien de f , $H_{ij}(f(X^*))$, détermine la nature de ce point critique.

Définition 1.4.2. *[13](Définition du point-selle)*
Soient $\mathcal{D}_1$ et $\mathcal{D}_2$ deux sous-ensembles et $f : \mathcal{D}_1 \times \mathcal{D}_2 \mapsto \mathbb{R}$. Nous disons que le point critique $(x_s,\ y_s) \in \mathcal{D}_1 \times \mathcal{D}_2$ est un point-selle de f sur $\mathcal{D}_1 \times \mathcal{D}_2$ si :

$$\forall\ (x,\ y) \in \mathcal{D}_1 \times \mathcal{D}_2 : f(x_s,\ y) \leq f(x_s,\ y_s) \leq f(x,\ y_s) \tag{1.4}$$

Dans ce cas $f(x_s,\ y_s)$ est appelée la valeur-selle de $f(x,\ y)$. Autrement dit, $y \mapsto f(x_s,\ y)$ atteint un maximum en y_s sur $\mathcal{D}_2$ et $x \mapsto f(x,\ y_s)$ atteint un minimum en x_s sur $\mathcal{D}_1$.
Comme exemple d'application, considérons :

$$f : \underbrace{[-2\ ;\ 2]}_{\mathcal{D}_1} \times \underbrace{[-2\ ;\ 2]}_{\mathcal{D}_2} \longmapsto \mathbb{R}$$

$$(x,\ y) \longmapsto f(x,y) = x^2 - y^2$$

Exemple 1.4.1. *f est le fonction définie sur $\mathbb{R}^2$ par :*

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

on recherche des points critiques alors on résoud le système : (Conditions nécessaire)

$$\begin{cases} \dfrac{\partial f(x,y)}{\partial x} = 0 \\ \dfrac{\partial f(x,y)}{\partial y} = 0 \end{cases} \Rightarrow \begin{cases} 3 - 3x^2 = 0 \\ -4y + 4y^3 = 0 \end{cases} \Rightarrow \begin{cases} x = \mp 1 \\ y = \mp 1 \ \textbf{\textit{ou}} \ \ y = 0 \end{cases}$$

donc les points critiques sont :$(1;0), (1;1), (1;-1), (-1;0), (-1;1), (-1;-1)$.

on étude la nature des points critiques : (Conditions suffisante)

la matrice Hessien est :

$$H(x,y) = \begin{pmatrix} \dfrac{\partial^2 f(x,y)}{\partial x^2} & \dfrac{\partial^2 f(x,y)}{\partial xy} \\ \dfrac{\partial^2 f(x,y)}{\partial yx} & \dfrac{\partial^2 f(x,y)}{\partial y^2} \end{pmatrix} \Rightarrow H(x,y) = \begin{pmatrix} -6x & 0 \\ 0 & -4 + 12y^2 \end{pmatrix}$$

- $(1;0) \to H(1;0) = \begin{pmatrix} -6 & 0 \\ 0 & -4 \end{pmatrix} \Rightarrow$ *det* $H_f(1;0) = 24 > 0$ *alors le point (1 ;0) est extrémum* $\to$ *tr* $H_f(1;0) = -10 < 0$ *donc maximum local.*
- $(1;1) \to H(1;1) = \begin{pmatrix} -6 & 0 \\ 0 & 8 \end{pmatrix} \Rightarrow$ *det* $H_f(1;1) = -48 < 0$ *alors le point (1 ;1) est point selle.*
- $(1;-1) \to H(1;-1) = \begin{pmatrix} -6 & 0 \\ 0 & 8 \end{pmatrix} \Rightarrow$ *det* $H_f(1;-1) = -48 < 0$ *alors le point (1 ;-1) est point selle.*
- $(-1;0) \to H(-1;0) = \begin{pmatrix} 6 & 0 \\ 0 & -4 \end{pmatrix} \Rightarrow$ *det* $H_f(-1;0) = -24 < 0$ *alors le point (-1 ;0) est point selle.*
- $(-1;1) \to H(-1;1) = \begin{pmatrix} 6 & 0 \\ 0 & 8 \end{pmatrix} \Rightarrow$ *det* $H_f(-1;1) = 48 > 0$ *alors le point (-1 ;1) est extrémum* $\to$ *tr* $H_f(-1;1) = 14 > 0$ *donc minimum local.*
- $(-1;-1) \to H(-1;-1) = \begin{pmatrix} 6 & 0 \\ 0 & 8 \end{pmatrix} \Rightarrow$ *det* $H_f(-1;-1) = 48 > 0$ *alors le point*

(-1 ;-1) est extrémum $\rightarrow$ *tr* $H_f(-1;-1) = 14 > 0$ *donc minimum local.*

Exemple 1.4.2. f *est le fonction définie sur* $\mathbb{R}^2$ *par :*

$$f(x, y) = 3x^3 - xy^2 + 2y^2 - 8$$

on recherche des points critiques alors on résoud le système :

$$\begin{cases} \dfrac{\partial f(x,y)}{\partial x} = 0 \\ \dfrac{\partial f(x,y)}{\partial y} = 0 \end{cases} \quad \textit{... Conditions nécessaire}$$

$$\begin{cases} 9x^2 - y^2 = 0 \\ 4y - 2xy = 0 \end{cases} \Rightarrow \begin{cases} 9x^2 - y^2 = 0 \\ y(4 - 2x) = 0 \end{cases} \Rightarrow \begin{cases} 9x^2 - y^2 = 0 \\ y = 0 \ \boldsymbol{ou} \ x = 2 \end{cases}$$

$$\Rightarrow \begin{cases} si \ y = 0 \Rightarrow 9x^2 - 0^2 = 0 \Rightarrow x = 0 \\ si \ x = 2 \Rightarrow 9 \times 2^2 - y^2 = 0 \Rightarrow y = \mp 6 \end{cases}$$

donc les points critiques sont : $\{(0;0), (2;6), (2;-6)\}$

la matrice Hessien est :(Conditions suffisante)

$$H(x, y) = \begin{pmatrix} \dfrac{\partial^2 f(x,y)}{\partial x^2} & \dfrac{\partial^2 f(x,y)}{\partial xy} \\ \dfrac{\partial^2 f(x,y)}{\partial yx} & \dfrac{\partial^2 f(x,y)}{\partial y^2} \end{pmatrix} \Rightarrow H(x, y) = \begin{pmatrix} 18x & -2y \\ -2y & 4 - 2x \end{pmatrix}$$

Etude de la nature de chacun des points critiques

- $(0;0) \rightarrow H(0;0) = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \Rightarrow$ *det* $H_f(0;0) = 0$ *on ne peut pas conclure, il faut développer les calculs jusqu'à l'ordre supérieur.*

- $(2;6) \rightarrow H(2;6) = \begin{pmatrix} 36 & -12 \\ -12 & 0 \end{pmatrix} \Rightarrow$ *det* $H_f(2;6) = -144 < 0$ *alors le point (2 ;6) est point selle (col).*

- $(2;-6) \rightarrow H(2;-6) = \begin{pmatrix} 36 & 12 \\ 12 & 0 \end{pmatrix} \Rightarrow$ *det* $H_f(2;-6) = -144 < 0$ *alors le point (2 ;-6) est point selle (col).*

Chapitre 2

Minimisation avec contraintes

On s'intéresse dans ce chapitre au problème d'optimisation avec contraintes,dont on va présenter sa forme générale, théorèmes d'existence et d'unicité et les conditions d'optimalité.

2.1 Formulation du problème

[3] Soit $f : \mathbb{R}^n \to \mathbb{R}$. On appelle problème de minimisation avec contraintes le problème suivant :

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

où C est l'ensemble des contraintes, C domaine non vide et fermé de $\mathbb{R}^n$.
Le problème (P) s'écrit aussi sous la forme générale suivante :

$$(P)\begin{cases}\min f(x),\\ h_i(x) = 0,\ \ i = 1, \ldots, p,\\ g_j(x) \leq 0,\ \ j = 1, \ldots, q,\end{cases}$$

- *Les fonctions $f, h_i, i = \overline{1,p}$ et $g_j, j = \overline{1,q}$ définies de C dans $\mathbb{R}$ sont de classe $\mathcal{C}^1$;*

- *Les conditions* $h_i(x) = 0,\ i = \overline{1,p}, g_j(x) \leq 0,\ j = \overline{1,q}$ *et* $x \in \mathbb{R}^n$ *sont appelées contraintes du problème (P).*

Définitions

[3]

- *On appelle solution réalisable ou admissible tout vecteur* x *vérifiant les contraintes du probléme (P).*

  L'ensemble $C = \{x \in : h_i(x) = 0, \forall i = \overline{1,p}$ *et* $g_j(x) \leq 0, \forall j = \overline{1,q}\}$ *est appelé ensemble des solutions réalisables ou admissibles.*
- *On appelle solution optimale du problème (P) (ou minimum global) une solution réalisable* x^* *qui minimise f(x) sur l'ensemble de toutes les solutions réalisables, c.à.d :*

$$\forall x \in C,\ f(x^*) \leq f(x).$$

- *On dit qu'un point* x^* *est un optimum (minimum) local de f si, et seulement si, il existe un voisinage* $V(x^*)$ *de* x^* *tel que* x^* *soit un min global de f sur* $V(x^*)$ *:*

$$\forall x \in V(x^*) \cap C,\ f(x^*) \leq f(x).$$

- *Un minimum est dit strict si les inégalités dans les définitions précédentes sont strictes.*
- *Si pour* $x \in C$ *et pour* $j \in \{1, 2, \ldots, q\}$ *on a* $g_j(x) = 0$*, on dit que la contrainte* g_j *est saturée ou active en* x*.*

  Une contrainte qui n'est pas active est dite inactive. On note $I(x)$ *l'ensemble des indices* j *correspondants aux contraintes actives en* x *:*

$$I(x) = \{j \in \{1, 2, \ldots, q\} : g_j(x) = 0\}.$$

- *Un vecteur d est une direction de descente s'il existe* τ *tel que :*

$$f(x + td) < f(x),\quad t \in [0, \tau].$$

- *Soit* $f : \mathbb{R}^n \to \mathbb{R},\ C \subseteq \mathbb{R}^n,\ x \in C$

  d est une direction admissible si $\exists \lambda' > 0$ *tel que :* $x + \lambda d \in C,\ \forall\ 0 \leq \lambda \leq \lambda'$

  Exemple 2.1.1. *Considérons le problème ci-dessous :*

$$(P_0) \begin{cases} \min\ f(x,y), \\ g_1(x,y) = (x+1)^2 + (y-2)^2 - 4 \leq 0, \\ g_2(x,y) = -y + \frac{3}{5}x + \frac{7}{5} \leq 0, \\ g_3(x,y) = y + x - 3 \leq 0, \\ (x,y) \in \mathbb{R}^2. \end{cases}$$

1. *considérons les directions* $d_0 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ *et* $d_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ *et le point* $x^0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

   (voir Figure 2.1)

2. *Quelles sont les contraintes actives en* x^0 *?*

   $g_1(x^0) = g_1(1,2) = (1+1)^2 + (2-2)^2 - 4 = 0,\ g_1$ *est active*

   $g_2(x^0) = g_2(1,2) = -2 + \frac{3}{5} \times 1 + \frac{7}{5} = 0,\ g_2$ *est active*

   $g_3(x^0) = g_3(1,2) = 2 + 1 - 3 = 0,\ g_3$ *est active*

   $I(x^0) = I(1,2) = \{1,2,3\}$

3. d_0 *est-elle une direction admissible pour* f *en* x^0 *?*

   Il faut avoir $\langle \nabla g_j(x^0), d_0 \rangle \leq 0, \forall j = \overline{1,3}$ *car* g_1, g_2 *et* g_3 *sont actives.*

   $\nabla g_1(x,y) = \begin{pmatrix} 2(x+1) \\ 2(y-2) \end{pmatrix} \Rightarrow \nabla g_1(1,2) = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$

   $\nabla g_2(x,y) = \begin{pmatrix} \frac{3}{5} \\ -1 \end{pmatrix} \Rightarrow \nabla g_2(1,2) = \begin{pmatrix} \frac{3}{5} \\ -1 \end{pmatrix}$

   $\nabla g_3(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \nabla g_3(1,2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

   $\langle \nabla g_1(1,2), d_0 \rangle = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \end{pmatrix} = -4 \leq 0$

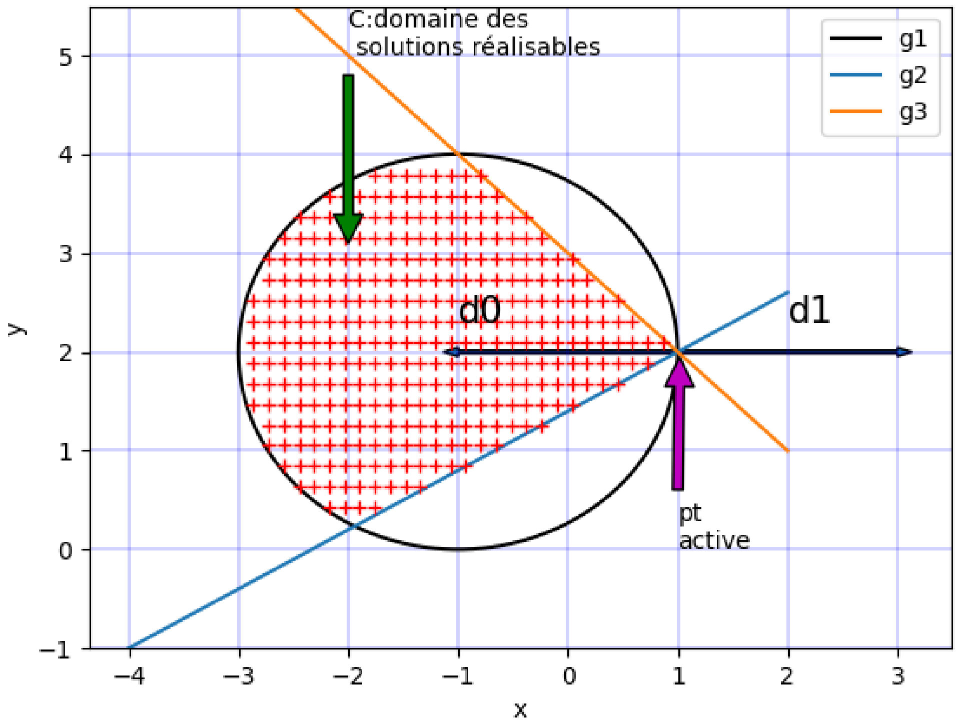


FIGURE 2.1 – Domaine des solutions réalisables

$$\langle \nabla g_2(1,2), d_0 \rangle = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{5} \\ -1 \end{pmatrix} = \frac{-3}{5} \leq 0$$

$$\langle \nabla g_3(1,2), d_0 \rangle = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -1 \leq 0$$

Donc d_0 est une direction admissible pour f en x^0.

4. *d_1 est-elle une direction admissible pour f en x^0 ?*

$$\langle \nabla g_1(1,2), d_1 \rangle = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \end{pmatrix} = 4 \geq 0$$

Donc d_1 n'est pas une direction admissible pour f en x^0

2.2 Résultats d'existence et d'unicité

Théorème 2.2.1. *[3](Existence)*

Supposons que f est continue, que C est un sous-ensemble fermé non vide de $\mathbb{R}^n$ et que l'une des conditions suivantes est satisfaite :

1. *Soit C est borné ;*
2. *Soit f est coercive.*

Alors le problème (P) admet au moins une solution.

Preuve 2.2.1. *[3] Soit $(x_n)_{n\in\mathbb{N}}$ une suite minimisante dans $f(C)$ i.e d'élément de C telle que $lim_{n\to+\infty} f(x_n) = inf f(C)$ comme C est fermé et borné, il existe une sous suite etraite $(x_\sigma(n))$ qui converge vers un $x^* \in C$ cette suite exraite vérifie : $x_{\sigma(n)} \to x^*$ et $f(x_{\sigma(n)}) \to inf_{y\in C} f(y)$ or f est continue, d'où par Unicité de la limite, il suit $f(x^*) = inf_{y\in C} f(y)$ avec $x^* \in C$ et f réalise son minimun sur C.*$\square$

Théorème 2.2.2. *(Existence et unicité)[3]*

Soit f une fonction continue et strictement convexe et soit C est un sous-ensemble non vide, convexe et fermé de $\mathbb{R}^n$.

Si C est borné ou si f est coercive, alors il existe un unique $x^* \in C$ solution de (P).

Preuve 2.2.2. *[3] supposons que f admette au moins un minimum m et soit $x_1 \neq x_2$ dans C réalisant ce minimum : $f(x_1) = f(x_2) = m$ par strict convexité de la fonction f on a alors :*

$$f\big(\frac{x_1 + x_2}{2}\big) < \frac{1}{2}\big(f(x_1) + f(x_2)\big) = m$$

ceci contredit le fait que m et le minimum donc $x_1 = x_2 = x^*$. $\square$

2.3 Conditions d'optimalité du 1er ordre

2.3.1 Conditions d'optimalité du 1er ordre générales

Théorème 2.3.1. *(Condition nécessaire du 1er ordre)[1]*
Si f est une fonction gâteaux différentiable et si C est un convexe et fermé de $\mathbb{R}^n$, alors toute solution x^* du problème (P) vérifie la propriété suivante :

$$\forall x \in C, \langle \nabla f(x^*), x - x^* \rangle \geq 0. \tag{2.1}$$

Preuve 2.3.1. *Soient x^* une solution du problème (P) et $x \in C$, alors :*

$$\forall x \in C, \ f(x^*) \leq f(x).$$

C est convexe, alors

$$x^* + t(x - x^*) \in C, \ \forall t \in [0, 1].$$

Donc

$$\forall x \in C, \ f(x^* + t(x - x^*)) - f(x^*) \geq 0.$$

D'où

$$\forall x \in C, \ \forall t \in [0, 1], \ \frac{f(x^* + t(x - x^*)) - f(x^*)}{t} \geq 0.$$

Alors

$$\forall x \in C, \ \forall t \in [0, 1], \ \lim_{t \to 0^+} \frac{f(x^* + t(x - x^*)) - f(x^*)}{t} \geq 0.$$

Par conséquent,

$$\forall x \in C, \ \langle \nabla f(x^*), \ x - x^* \rangle \geq 0.$$

Théorème 2.3.2. *[3](C.N.S. du 1er ordre dans le cas convexe)*
Soient f une fonction convexe et gâteaux différentiable et C est un convexe et fermé de $\mathbb{R}^n$. Soit x^* un élément quelconque de C.
La conditions (2.1) est nécessaire et suffisante pour que x^* soit solution du problème (P).

Preuve 2.3.2. *[3] La condition est nécessaire (voir Théorème 2.3.1). Il reste à montrer qu'elle est suffisante.*
Soit x^* un élément de C, comme f est convexe, alors :

$$\forall x \in C,\ f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle.$$

Puisque $\langle \nabla f(x^*), x - x^* \rangle \geq 0$, alors $\forall x \in C, f(x) \geq f(x^*)$. D'où x^* est une solution du problème (P).

2.3.2 Conditions d'optimalité du 1er ordre, cas en égalité

[8] Lorsqu'on a que des contraintes d'égalité, le problème (P) se réduit à la forme suivante :

$$(P_e) \begin{cases} \min \ f(x), \\ h_i(x) = 0, \ i = 1, \dots, p, \\ x \in \mathbb{R}^n. \end{cases}$$

Théorème 2.3.3. *[3](C.N. du 1er ordre - Contraintes en égalité)*
On suppose que :

- *f, h_i pour $i = \overline{1,p}$, sont de classe C^1 dans $\mathbb{R}^n$;*
- *Le problème (P_e) admet une solution x^* ;*
- *Les p vecteurs de $\mathbb{R}^n$: $\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_p(x^*)$sont linéairement indépendants (et donc $p \leq n$).*

Alors il existe p réels $\lambda_1^*,\ \lambda_2^*, \dots,\ \lambda_p^*$ tels que

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) = 0. \tag{2.2}$$

Preuve 2.3.3. *[3] Analogue à celle du théorème de Karush-Kuhn-Tucker qui sera démontré par la suite.*

Définition 2.3.1. *[3](Multiplicateurs de Lagrange)*
Les réels $\lambda_1^*,\ \lambda_2^*, \dots,\ \lambda_p^*$ du théorème 2.2.3 sont appelés multiplicateurs de Lagrange.

Exemple 2.3.1. *cas* $\mathbb{R}^2$

On veut résourdre : $\begin{cases} \min\ f(x,y) = xy \\ h(x,y) = 4x^2 + y^2 - 4 = 0 \\ (x,y) \in \mathbb{R}^2 \end{cases}$

Le Lagrangien pour ce problèm est

$$\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda h(x,y), \quad \forall \lambda \in \mathbb{R}$$

$$\mathcal{L}(x,y,\lambda) = xy + \lambda(4x^2 + y^2 - 4)$$

Le système à résoudre pour trouver les points critiques est donc

$$\nabla\mathcal{L}(x,y,\lambda) = 0 \Rightarrow \begin{cases} \dfrac{\partial(x,y,\lambda)}{\partial x} = 0 \\ \dfrac{\partial(x,y,\lambda)}{\partial y} = 0 \\ \dfrac{\partial(x,y,\lambda)}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} y + 8\lambda x = 0 \\ x + 2\lambda y = 0 \\ 4x^2 + y^2 - 4 = 0 \end{cases}$$

$$\Rightarrow \begin{cases} y = -8\lambda x \\ x = -2\lambda y \\ 4x^2 + y^2 - 4 = 0 \end{cases} \Rightarrow \begin{cases} y = -8\lambda(-2\lambda y) = 16\lambda^2 y \\ x = -2\lambda y \\ 4x^2 + y^2 - 4 = 0 \end{cases} \Rightarrow \begin{cases} y(1 - 16\lambda^2) = 0 \\ x = -2\lambda y \\ 4x^2 + y^2 - 4 = 0 \end{cases}$$

- *Si* $y = 0$ *:*

$$\begin{cases} x = 0 \\ 4(0)^2 + (0)^2 = 0 \neq 4 \end{cases} \qquad \textit{alors pas de solution dans ce cas}$$

- *Si* $\lambda = \pm\frac{1}{4}$ *(car* $1 - 16\lambda^2 = 0$*) :*

$$\begin{cases} x = -2(\pm\frac{1}{4}y) = \pm\frac{1}{2}y \\ 4x^2 + y^2 = 4 \end{cases} \Rightarrow \begin{cases} x = \pm\frac{1}{2}y \\ 4(\pm\frac{1}{2}y)^2 + y^2 = 4 \end{cases}$$

$$\Rightarrow \begin{cases} x = \pm\frac{1}{2}y \\ y^2 = 2 \end{cases} \Rightarrow \begin{cases} x = \pm\frac{1}{2}y \\ y = \pm\sqrt{2} \end{cases}$$

dans ce cas on obtient 4 condidats $(\sqrt{2}/2, \sqrt{2})$; $(\sqrt{2}/2, -\sqrt{2})$; $(-\sqrt{2}/2, \sqrt{2})$; $(-\sqrt{2}/2, -\sqrt{2})$

Le tableau suivant permet de déterminer la nature des points critiques :

(x_0, y_0)	$(\frac{\sqrt{2}}{2}, \sqrt{2})$	$(\frac{\sqrt{2}}{2}, -\sqrt{2})$	$(\frac{-\sqrt{2}}{2}, \sqrt{2})$	$(\frac{-\sqrt{2}}{2}, -\sqrt{2})$
$f(x_0, y_0)$	*1*	*-1*	*-1*	*1*

les deux points critiques $(\frac{\sqrt{2}}{2}, -\sqrt{2})$, $(-\frac{\sqrt{2}}{2}, \sqrt{2})$ *sont des points minimaux.*

Exemple 2.3.2. *cas* $\mathbb{R}^3$

On veut résourdre :
$$\begin{cases} Min\ f(x_1, x_2, x_3) = 2x_1^2 + x_2^2 + 4x_3^2 \\ h_1(x_1, x_2, x_3) = x_1 + 2x_2 - x_3 = 6 \\ h_2(x_1, x_2, x_3) = 2x_1 - 2x_2 + 3x_3 = 12 \\ (x_1, x_2, x_3) \in \mathbb{R}^3 \end{cases}$$

les conditions de Lagrange :

$$\begin{cases} \nabla f(x_1, x_2, x_3) + \sum_{i=1}^{2} \lambda_i \nabla h_i(x_1, x_2, x_3) = 0. \\ \lambda_1, \lambda_2 \in \mathbb{R}. \end{cases}$$

$$\begin{cases} \begin{pmatrix} 4x_1 \\ 2x_2 \\ 8x_3 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ (x_1 + 2x_2 - x_3 - 6) = 0. \\ (2x_1 - 2x_2 + 3x_3 - 12) = 0. \\ \lambda_1, \lambda_2 \in \mathbb{R}. \end{cases}$$

— *Si* $\lambda_1 = 0$ *et* $\lambda_2 = 0$ *alors :* $\begin{cases} 4x_1 = 0 \\ 2x_2 = 0 \\ 8x_3 = 0 \end{cases} \Longrightarrow \begin{cases} x_1 = 0 \\ x_2 = 0 \\ x_3 = 0 \end{cases}$

$h_1(0,0,0) = 0 + 2 \times 0 - 0 = 0 \neq 6$ *non réalisable.*

— *Si* $\lambda_1 \neq 0$ *et* $\lambda_2 = 0$ *alors :* $\begin{cases} 4x_1 + \lambda_1 = 0 \\ 2x_2 + 2\lambda_1 = 0 \\ 8x_3 - \lambda_1 = 0 \\ x_1 + 2x_2 - x_3 - 6 = 0 \end{cases} \Longrightarrow \begin{cases} x_1 = \frac{6}{9.5} \\ x_2 = \frac{24}{9.5} \\ x_3 = \frac{-3}{9.5} \\ \lambda_1 = \frac{-24}{9.5} \end{cases}$

$h_1(\frac{6}{9.5}, \frac{24}{9.5}, \frac{-3}{9.5}) = \frac{6}{9.5} + 2 \times \frac{24}{9.5} - \frac{-3}{9.5} = \frac{57}{9.5} \neq 6$ *non réalisable.*

— *Si* $\lambda_1 = 0$ *et* $\lambda_2 \neq 0$ *alors :* $\begin{cases} 4x_1 + 2\lambda_2 = 0 \\ 2x_2 - 2\lambda_2 = 0 \\ 8x_3 - 3\lambda_2 = 0 \\ 2x_1 - 2x_2 + 3x_3 - 12 = 0 \end{cases} \Longrightarrow \begin{cases} x_1 = \frac{48}{25} \\ x_2 = \frac{-96}{25} \\ x_3 = \frac{36}{25} \\ \lambda_2 = \frac{-96}{25} \end{cases}$

$h_1(\frac{48}{25}, \frac{-96}{25}, \frac{36}{25}) = \frac{48}{25} + 2 \times \frac{-96}{25} - \frac{36}{25} = \frac{-180}{25} \neq 6$ *non réalisable.*

— $\lambda_1 \neq 0$ *et* $\lambda_2 \neq 0$ *alors On pose la matrice suivant :*

$$\begin{pmatrix} 1 & 2 & -1 & 0 & 0 \\ 2 & -2 & 3 & 0 & 0 \\ 4 & 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & 2 & -2 \\ 0 & 0 & 8 & -1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 12 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} x_1 = 5.044 \\ x_2 = 1.19 \\ x_3 = 1.43 \\ \lambda_1 = -7.52 \\ \lambda_2 = -6.32 \end{cases}$$

$h_1(5.044, 1.19, 1.43) = 5.044 + 2 \times 1.19 - 1.43 = 6$ *réalisable alors* h_1 *est active.*

$h_2(5.044, 1.19, 1.43) = 2 \times 5.044 - 2 \times 1.19 + 3 \times 1.43 = 12$ *réalisable alors* h_2 *est active.*

la valeur de la fonction f *au point* $(5.044, 1.19, 1.43)$ *est :*

$f(5.044, 1.19, 1.43) = 2 \times 5.044^2 + 1.19^2 + 4 \times 1.43^2 = 60.48$

2.3.3 Conditions d'optimalité du 1$^{\text{er}}$ ordre, cas en égalité et en inégalité

[3] Considérons un problème d'optimisation sous forme générale :

$$(P) \begin{cases} \min \ f(x), \\ h_i(x) = 0, \ i = 1, \dots, \ p \\ g_j(x) \leq 0, \ j = 1, \dots, \ q \\ x \in \mathbb{R}^n. \end{cases}$$

Théorème 2.3.4. *[3](Conditions d'optimalité non qualifiées)*
On suppose que f, h_i *pour* $i = \overline{1, p}$ *et* g_j *pour* $j = \overline{1, q}$ *sont de classe* C^1. *Soit* x^* *une solution de* (P).
Alors il existe $\lambda^* = (\lambda_1^*, \ \lambda_2^*, \dots, \ \lambda_p^*) \in \mathbb{R}^p$, $\mu^* = (\mu_1^*, \ \mu_2^*, \dots, \ \mu_q^*) \in \mathbb{R}_+^q$ *et* $\mu_0^* \in \mathbb{R}_+$

tels que :

- $\forall j \in \{0, 1, \ldots, q\}, : \quad \mu_j^* \geq 0;$ (2.3a)
- $h_i(x^*) = 0$ *pour* $i = \overline{1,p}$ *et* $g_j(x^*) \leq 0$ *pour* $j = \overline{1,q};$ (2.3b)
- $\forall j \in \{1, \ldots, q\} \; : \; \mu_j^* \; g_j(x^*) = 0;$ (2.3c)
- $\mu_0^* \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) + \sum_{i=1}^{q} \mu_j^* \nabla g_j(x^*) = 0.$ (2.3d)

Preuve 2.3.4. *La preuve consiste à remplacer le problème* (P) *avec contraintes par la suite de problèmes sans contraintes :*

$$(P_k) \begin{cases} \min \;\; f_k(x) = f(x) + k\alpha(x), \\ x \in \mathbb{R}^n \end{cases}$$

où $\alpha : \mathbb{R}^n \to \mathbb{R}$ *est une fonction de pénalisation des contraintes et* $k > 0$*.* α *est choisie de façon à avoir* (P) *et* (P_k) *équivalents (c.à.d. ayant les mêmes solutions).*

Pour tout entier k*, considérons le problème :*

$$(P_k) \begin{cases} \min \;\; f_k(x), \\ x \in B(x^*, \rho). \end{cases}$$

où

$$f_k(x) = f(x) + \frac{k}{2} \sum_{i=1}^{p} [h_i(x)]^2 + \frac{k}{2} \sum_{j=1}^{q} [g_j^+(x)]^2 + \|x - x^*\|^2, \tag{2.4}$$

avec $g_j^+(x) = max(0, g_j(x))$ *et* $B(x^*, \rho)$ *est une boule fermée (compact) centrée en* x^* *et de rayon* $\rho > 0$*.*

Remarque 2.3.1. *Les fonction* g_j *sont de classe* C^1*, alors* $(g_j^+)^2$ *est différentiable et*

$$\frac{d(g_j^+)^2}{dx}(x) = 2\frac{dg_i}{dx}(x) g_j^+(x).$$

- **Le problème** (P_k) **a au moins une solution** x_k

[3] En effet, f_k *est continue, elle atteint donc son minimum sur le compact* $B(x^*, \rho)$.

- **La suite** (x_k) **converge vers** x^*

[3] La suite (x_k) *est dans le compact* $B(x^*, \rho)$ *et on peut en extraire une sous-suite, noté* (x_k)*, qui converge vers* $\tilde{x} \in B(x^*, \rho)$.
Comme $f_k(x_k) \leq f_k(x^*) = f(x^*) < +\infty$ *car* x_k *est le mimnimum de* f_k *sur* $B(x^*, \rho)$*, nous avons alors*

$$\sum_{i=1}^{p}[h_i(x_k)]^2 + \sum_{i=1}^{q}[g_j^+(x_k)]^2 \leq \frac{2}{k}\big[f(x^*) - f(x_k) - \|x_k - x^*\|^2\big]. \tag{2.5}$$

$[f(x^*) - f(x_k) - \|x_k - x^*\|^2]$ *étant borné, donc*

$$\forall i = \overline{1,p}, \lim_{k\to+\infty} h_i(x_k) = h_i(\tilde{x}) = 0,$$

et

$$\forall j = \overline{1,q}, \lim_{k\to+\infty} g_j^+(x_k) = g_j^+(\tilde{x}) = 0.$$

Alors $\tilde{x}$ *est une solution réalisable.*
D'autre part :

$$f(x_k) + \|x_k - x^*\|^2 \leq f_k(x_k) \leq f(x^*)$$

car $\frac{k}{2}[h_i(x_k)]^2 + \frac{k}{2}\sum_{j=1}^{q}[g_j^+(x_k)]^2 \geq 0$ *(c'est une somme de carrés).*
Comme x^* *est une solution du problème* (P)*, alors*

$$f(\tilde{x}) + \|\tilde{x} - x^*\|^2 \leq f(x^*) \leq f(\tilde{x}).$$

Donc $\|\tilde{x} - x^*\|^2 = 0$. *Par conséquent,* $\tilde{x} = x^*$.
Ce raisonnement peut être fait pour toute valeur d'adhérence de la suite (x_k).
Donc la suite (x_k) *converge vers* x^*.

- **Condition d'optimalité pour** (P_k)

[3] Comme (x_k) converge vers x^*, elle est dans $B(x^*, \rho)$ à partir d'un certain rang et donc x_k est un minimum local (sans contraintes) de f_k. Par conséquent, $\nabla f_k(x_k) = 0$ pour k assez grand, c.à.d.

$$\nabla f(x_k) + k \sum_{i=1}^{p} [h_i(x_k) \nabla h_i(x_k)] + k \sum_{j=1}^{q} [g_j^+(x_k) \nabla g_j(x_k)] + 2(x_k - x^*) = 0. \quad (2.6)$$

Posons

$$S_k = \left(1 + k^2 \sum_{i=1}^{p} [h_i(x_k)]^2 + k^2 \sum_{j=1}^{q} [g_j^+(x_k)]^2\right)^{\frac{1}{2}}$$

$\mu_0^k = \frac{1}{S_k}$, $\lambda_i^k = \frac{k h_i(x_k)}{S_k}$, *pour* $i = \overline{1,p}$ *et* $\mu_j^k = \frac{k g_j^+(x_k)}{Sk}$, *pour* $j = \overline{1,q}$.
La relation (2.6) devient :

$$\mu_0^k \nabla f(x_k) + \sum_{i=1}^{p} [\lambda_i^k \nabla h_i(x_k)] + \sum_{j=1}^{q} [\mu_j^k \nabla g_j(x_k)] + \frac{2}{S_k}(x_k - x^*) = 0. \quad (2.7)$$

Comme le vecteur $(\lambda_1^k, \ldots, \lambda_p^k, \mu_0^k, \mu_1^k, \ldots, \mu_q^k) \in \mathbb{R}_+^p \times \mathbb{R}_+ \times \mathbb{R}^q$ *est de norme 1, on peut en extraire une sous-suite convergente vers* $(\lambda_1^k, \ldots, \lambda_p^k, \mu_0^k, \mu_1^k, \ldots, \mu_q^k) \neq 0$ *et passer à la limite dans (2.7). Alors, on aura :*

$$\mu_0^* \nabla f(x^*) + \sum_{i=1}^{p} [\lambda_i^* \nabla h_i(x^*)] + \sum_{j=1}^{q} [\mu_j^* \nabla g_j(x^*)] = 0 \quad (2.8)$$

car toutes les fonctions considérées sont continues et $S_k \geq 1$. *Notons que* $\mu_j^* \geq 0, \forall j = \overline{0,q}$.

- **Relation (2.3c)**

[3] Si $g_j(x^*) < 0$, alors $g_j(x_k) < 0$ à partir d'un certain rang et $\mu_j^k = 0$. Par conséquent, en passant à la limite $\mu_j^* = 0$.

Remarque 2.3.2. *[3] Quelques remarques importantes :*

1. *Les réels λ_i^* et μ_j^* sont les multiplicateurs de Lagrange.*

2. *La relation (2.3b) est une relation de réalisabilité, c.à.d. que tout point x véerifiant (2.3b) est une solution réalisable.*
3. *La relation (2.3c) est une relation de complémentarité.*
4. *Les conditions du théorème (2.3d) sont dites non qualifiées car le réel μ_0^* peut être nul et on n'a pas de renseignements sur le minimum de f puisqu'elle n'apparaît nulle part dans la relation d'optimalité. Donc il est important de donner des conditions qui permettent d'assurer que μ_0^* soit non nul.*

   De telles conditions sont dites conditions de qualification ou de régularité. Lorsqu'elles sont vérifiées le problème est dit qualifié.

Définition 2.3.2. *[3](Point régulier ou condition de qualification 1)*
On dit qu'un élément $x^* \in \mathbb{R}^n$ est régulier pour les contraintes h_i pour $i = \overline{1,p}$ et g_j pour $j = \overline{1,q}$:

- *s'il est réalisable : $h_i(x^*) = 0$ pour $i = \overline{1,p}$ et $g_j(x^*) \leq 0$ pour $j = \overline{1,q}$*
- *si les vecteurs $\nabla h_i(x^*)$ pour $i = \overline{1,p}$ sont linéairement indépendants*
- *et si on peut trouver une direction $d \in \mathbb{R} \setminus \{0\}$ tel que*

$$\langle \nabla h_i(x^*), d \rangle = 0, \; i = \overline{1,p} \text{ et } \langle \nabla g_j(x^*), d \rangle < 0, \; \forall j = \overline{1,q}.$$

On dit aussi que x^* vérifie la condition de qualification de Mangasarian-Fromowitz que nous noterons (2.3.2).
La condition (2.3.3) suivante est plus forte que la condition (2.3.2).

Définition 2.3.3. *[3](Condition de qualification 2)*
On dit qu'un élément $x^* \in \mathbb{R}^n$ est régulier pour les contraintes h_i pour $i = \overline{1,p}$ et g_j pour $j = \overline{1,q}$:

- *s'il est réalisable,*
- *et si les vecteurs $\nabla h_i(x^*), \nabla g_j(x^*)$ pour $i = \overline{1,p}$ et $j = \overline{1,q}$ sont linéairement indépendants*

Exemple 2.3.3. *considérons le problème :*
$$\begin{cases} min\ f(x,y) = 4x^2+5y^2+6xy-6x-6y+3 \\ y-x+\frac{3}{11}=0 \\ x+y-1\leq 0 \\ x\geq 0 \\ y \geq 0 \\ (x,y)\in\mathbb{R}^2 \end{cases}$$

le point (0,0) n'est pas active

f est un fonction continue et strictement convexe.

$\nabla h(x,y) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $\nabla g_1(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\nabla g_2(x,y) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\nabla g_3(x,y) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

la famille $\{\nabla h(x,y),\ \nabla g_1(x,y),\ \nabla g_2(x,y),\ \nabla g_3(x,y)\}$ *sont libres.*

Théorème 2.3.5. *[3](Conditions de Karush-Kuhn-Tucker)*

On suppose que f, h_i pour $i=\overline{1,p}$ et g_j pour $j=\overline{1,q}$ sont de classe C^1. Soit x^* une solution de (P).

On suppose que x^* est un point régulier pour les contraintes $h_i,\ i=\overline{1,p}$ et $g_j,\ j=\overline{1,q}$.

Alors il existe $\lambda^*=(\lambda_1^*,\ \lambda_2^*,\dots,\ \lambda_p^*)\in\mathbb{R}^p$ et $\mu^*=(\mu_1^*,\ \mu_2^*,\dots,\ \mu_q^*)\in\mathbb{R}_+^q$ tels que :

- $\forall j\in\{1,\dots,q\},\ \mu_j^*\geq 0;$ (2.9a)
- $h_i(x^*)=0$ *pour* $i=\overline{1,p}$ *et* $g_j(x^*)\leq 0$ *pour* $j=\overline{1,q};$ (2.9b)
- $\forall j\in\{1,\dots,q\}: \mu_j^* g_j(x^*)=0;$ (2.9c)
- $\nabla f(x^*)+\sum_{i=1}^{p}\lambda_i^*\nabla h_i(x^*)+\sum_{j=1}^{q}\mu_j^*\nabla g_j(x^*)=0.$ (2.9d)

Preuve 2.3.5. *On montre que sous les hypothèses de régularité (2.3.2), le réel μ_0^* donné dans le théorème (2.3d) est non nul.*

On a $(\lambda_1^*,\ \lambda_2^*,\dots,\ \lambda_p^*,\ \mu_0^*,\ \mu_1^*,\ \mu_2^*,\dots,\ \mu_q^*)\neq 0$. On suppose que $\mu_0^*=0$ et supposons aussi que $\mu_j^*=0,\ \forall j\in I(x^*)$.

Alors le vecteur $\lambda^* = (\lambda_1^*,\ \lambda_2^*, \dots,\ \lambda_p^*) \neq 0$, *on a alors* $\sum\limits_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) = 0$. *Ce résultat contredit l'indépendances linéaire des* $\nabla h_i(x^*), i = \overline{1,p}$.

Donc il existe $j_0 \in I(x^*)$ *tel que* $\mu_{j_0}^* \neq 0$.

Nous avons, en prenant la direction d donnée par (2.3.2) :

$$0 = \sum_{i=1}^{p} \lambda_i^* \langle \nabla h_i(x^*), d\rangle + \sum_{j=1}^{q} \mu_j^* \langle \nabla g_j(x^*), d\rangle < \mu_{j_0}^* \langle \nabla g_{j_0}(x^*), d\rangle < 0,$$

d'où la contradiction.

L'ensemble des contraintes (2.9a)-(2.9d) sont appelées conditions de Karush-Kuhn-Tucker, notées K.K.T.

Définition 2.3.4. *(Lagrangien)*

On appelle lagrangien du problème (P) *la fonction définie de* $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ *dans* $\mathbb{R}$ *par :*

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

Remarque 2.3.3. *La relation (2.9d) s'écrit*

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0,$$

où ∇_x *désigne le gradient du lagrangien par rapport à la première variable.*

Dans le cas convexe, nous avons le résultat suivant :

Théorème 2.3.6. *(C.N.S. dans le cas convexe)*

On suppose que f, h_i *pour* $i = \overline{1,p}$ *et* g_j *pour* $j = \overline{1,q}$ *sont de classe* C^1 *et que* f *et* $g_j, j = \overline{1,q}$ *sont convexes et* $h_i, i = \overline{1,p}$ *sont affines. On suppose aussi que* x^* *est un point régulier. Alors,*

$$(x^* \text{ est solution de } (P)\ ) \Leftrightarrow (\text{les conditions } (2.9a)\text{-}(2.9d) \text{ sont satisfaites}).$$

Preuve 2.3.6. *[3] On montre que si les conditions (2.9a)-(2.9d) sont satisfaites alors* x^* *est une solution du problème* (P).

De (2.9b), on déduit que x^* est réalisable. Comme toutes les fonctions sont convexes, alors le lagrangien est convexe par rapport à la variable x et la condition (2.9d) est équivalente à dire que x^* est un minimum de $\mathcal{L}(x, \lambda^*, \mu^*)$. On obtient donc :

$$\forall x \in \mathbb{R}^n : \mathcal{L}(x^*, \lambda^*, \mu^*) \leq \mathcal{L}(x, \lambda^*, \mu^*)$$

Si $x \in C$, alors $h_i(x) = 0$ et $g_j(x) \leq 0$ de sorte que

$$\sum_{i=1}^{p} \lambda_i^* h_i(x) + \sum_{j=1}^{q} \mu_j^* g_j(x) = \sum_{j=1}^{q} \mu_j^* g_j(x) \leq 0,$$

puisque $\mu_j^* \geq 0$ d'après (2.9a).
De plus avec la relation de complémentarité (2.9c), on voit que

$$\mathcal{L}(x^*, \lambda^*, \mu^*) = f(x^*).$$

Finalement, on obtient

$$f(x^*) = \mathcal{L}(x^*, \lambda^*, \mu^*) \leq \mathcal{L}(x, \lambda^*, \mu^*) = f(x^*) \leq f(x).$$

Ce qui preuve que x^* est solution du problème (P).

Exemple 2.3.4. *On veut résourdre :* $\begin{cases} min\ f(x,y) = 4x^2 + 5y^2 + 6xy - 6x - 6y + 3 \\ x + y \leq 1 \\ x \geq 0 \\ y \geq 0 \\ (x,y) \in \mathbb{R}^2 \end{cases}$

On écrit les contraintes sous la forme générale : $\begin{cases} min\ f(x,y) = 4x^2 + 5y^2 + 6xy - 6x - 6y + 3 \\ x + y - 1 \leq 0 \\ x \geq 0 \\ y \geq 0 \\ (x,y) \in \mathbb{R}^2 \end{cases}$

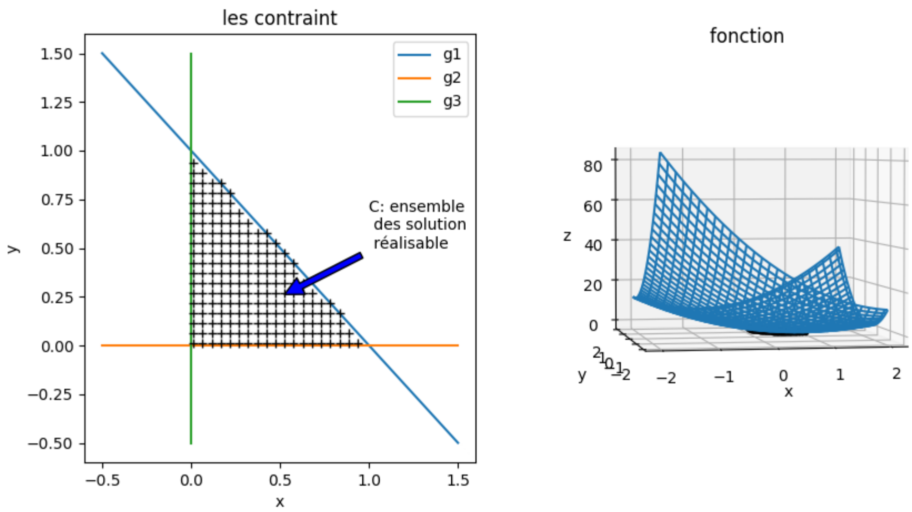


FIGURE 2.2 – la fonction f et les contraintes g_i

les condition de KKT

on a

$$\begin{cases} \nabla f(x,y) + \sum\limits_{i=1}^{3} u_i \nabla g_1(x,y) = 0 \\ u_1, u_2, u_3 \geq 0 \\ u_i g_i(x,y) = 0, \quad i = \overline{1,3} \end{cases}$$

$$\Rightarrow \begin{cases} \begin{pmatrix} 8x+6y-6 \\ 10y+6x-6 \end{pmatrix} + u_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + u_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0 \\ u_1, u_2, u_3 \geq 0 \\ u_1(x+y-1) = 0 \\ u_2(-x) = 0 \\ u_3(-y) = 0 \end{cases}$$

- *si* $u_1 = 0$ *et* $u_2 = 0$ *et* $u_3 = 0$

$$\begin{cases} 8x + 6y = 0 \\ 10y + 6x - 6 = 0 \end{cases} \Rightarrow \begin{cases} x = \frac{6}{11} \\ y = \frac{3}{11} \end{cases} \quad \textit{réalisable}$$

d'après le théorème de Weierstrass le pt$(\frac{6}{11}, \frac{3}{11})$ *est minimum global.*

- $u_1 = u_3 = 0$ *et* $u_2 > 0 \Rightarrow x = 0$ *et* $y = \frac{6}{11}$ *et* $u_2 = \frac{-24}{10} < 0$ *refuse.*
- *si* $u_1 = u_2 = 0$ *et* $u_3 > 0 \Rightarrow y = 0$ *et* $x = \frac{6}{8}$ *et* $u_3 = \frac{-3}{2} < 0$ *refuse.*
- $u_2 = u_3 = 0$ *et* $u_1 > 0 \Rightarrow x = \frac{2}{3}$ *et* $y = \frac{1}{3}$ *et*$u_1 = \frac{-4}{3} < 0$ *refuse.*
- *si* $u_1 > 0$ *et* $u_2 > 0$*et* $u_3 = 0 \Rightarrow y = 1$ *et* $x = 0$ *et* $u_1 = -4 < 0$ *et* $u_2 = u_1 = -4 < 0$ *refuse.*
- *si* $u_1 > 0$ *et* $u_3 > 0$*et* $u_2 = 0 \Rightarrow x = 1$ *et* $y = 0$ *et* $u_1 = u_3 = -2 < 0$ *refuse.*
- *si* $u_1 = 0$ *et* $u_2 > 0$*et* $u_3 > 0 \Rightarrow y = 0$ *et* $x = 0$ *et* $u_2 = u_3 = -6 < 0$ *et* $u_2 = u_1 = -4 < 0$ *refuse.*
- *si* $u_1 > 0$ *et* $u_2 > 0$*et* $u_3 > 0 \Rightarrow u_2 = -u_3$ *et* $u_1 = 6$ *et* $x = y = 0$ *et refuse.*

2.4 Conditions d'optimalité du second ordre

Les conditions d'optimalité du premier ordre permettent de déterminer les bons condidats à la solution de (P). *Les Conditions d'optimalité du second ordre vont permettre dans un premiers temps de restreindre encore le nombre de candidats.*

Théorème 2.4.1. *[9](Condition nécessaire du second ordre)*
On suppose que f, h_i, $i = \overline{1,p}$ *et* g_j, $j = \overline{1,q}$ *sont de classe* $\mathcal{C}^2$, *que* x^* *est un minimum de* f *sur* C *et que la condition (CQ1) est vérifiée.*
Alors il existe $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_p^*) \in \mathbb{R}^p$ *et* $\mu^* = (\mu_1^*, \mu_2^*, \ldots, \mu_q^*) \in \mathbb{R}^q$ *tels que :*

- *Les relations (2.9a)-(2.9d) de K.K.T. sont satisfaites*
- *Pour toute direction* $d \in \mathbb{R}^n \setminus \{0\}$ *vérifiant :*

$$\begin{cases} \langle \nabla h_i(x^*), d \rangle = 0, & \textit{pour } i = \overline{1,p}; \\ \langle \nabla g_j(x^*), d \rangle = 0, & \textit{pour } j \in I^+(x^*); \\ \langle \nabla g_j(x^*), d \rangle \leq 0, & \textit{pour } j \in I(x^*) \setminus I^+(x^*); \end{cases} \tag{2.10}$$

où $I^+(x^*) = \{j \in \{1, 2, \ldots, q\} : g_j(x^*) = 0$ *et* $\mu_j^* > 0\}$, *on a*

$$\langle H_{xx}\mathcal{L}(x^*,\ \lambda^*,\ \mu^*)d, d\rangle \geq 0, \tag{2.11}$$

avec

$$H_{xx}\mathcal{L}(x^*,\ \lambda^*,\ \mu^*) = H_f(x) + \sum_{i=1}^{p} \lambda_i^* H_{h_i}(x^*) + \sum_{j=1}^{q} \mu_j^* H_{g_j}(x^*)$$

désigne la seconde dérivée de $\mathcal{L}$ *au point* $(x^*,\ \lambda^*,\ \mu^*)$.

Définition 2.4.1. *[3](Contraintes fortement actives)*
L'ensemble $I^+(x^*)$ *est l'ensemble des contraintes fortement actives.*
Lorsque $I^+(x^*) = I(x^*)$, *c.à.d.* $g_j(x^*) = 0 \Leftrightarrow \mu_j^* > 0$ *on dit qu'il y'a stricte complémentarité.*
Le résultat suivant donne une condition suffisante du second ordre.

Théorème 2.4.2. *[3](Condition suffisante du second ordre)*
On suppose que $f, h_i,\ i = \overline{1,p}$ *et* $g_j,\ j = \overline{1,q}$ *sont de classe* $\mathcal{C}^2$. *Soit* $x^* \in \mathbb{R}^n$ *vérifiant les conditions de K.K.T. avec les multiplicateurs de Lagrange* $\lambda^*,\ \mu^*$.
Si la matrice hessienne du lagrangien au point $(x^*,\ \lambda^*,\ \mu^*)$, $\nabla^2_{xx}\mathcal{L}(x^*,\ \lambda^*,\ \mu^*)$, *est définie positive sur le sous-espace*

$$\tau = \{d \in \mathbb{R}^n \setminus \{0\} : < \nabla h_i(x^*), d\rangle = 0,\ i = \overline{1,p} \text{ et } < \nabla g_j(x^*), d\rangle = 0, j \in I^+(x^*)\}$$

alors x^* *est un minimum strict de* f *sur* C.

Chapitre 3

Algorithmes

Dans ce chapitre nous allons introduire une classe importante d'algorithmes de résolution des problèmes d'optimisation sous contraintes. Bien entendu, nous ne pouvons pas être exhaustifs ; nous présentons les méthodes "de base" les plus classique. et avec cela, la plut part de ces algorithmes exploitent les conditions d'optimalité dont on a vu qu'elles permettaient (au mieux) de déterminer des minima locaux. La question de la détermination de minima globaux est difficile et dépasse le cadre que nous nous sommes fixés. Néanmoins, nous décrirons dans la section suivante, un algorithme probabiliste permettant de "déterminer" x un minimum global.

3.1 Méthode du gradient projeté

[6] Dans le cas d'optimisation sans contraintes, l'algorithme du gradient s'énonce comme suit :

$$\begin{cases} x^{(0)} \in \mathbb{R}^n \text{ un point initial donné,} \\ x^{(k+1)} = x^{(k)} + p^{(k)} d^{(k)}, d^{(k)} \in \mathbb{R}^n \backslash \{0\}, p^{(k)} \in \mathbb{R}_+^*, \end{cases}$$

où $p^{(k)}$ et $d^{(k)}$ sont choisis de telle sorte que :

$$f(x^{(k+1)}) \leq f(x^{(k)}).$$

Dans le cas avec contraintes, on ne peut pas être sur que $x^{(k)}$ reste sur l'ensemble des contraintes C. Il est donc nécessaire de se ramener sur C, et ceci grâce à une projection sur C, l'opérateur associé étant noté Π_C où C est convexe fermé et non vide :

$$\Pi_C : \mathbb{R}^n \to C$$
$$x \to \Pi_C(x).$$

On obtient ainsi l'algorithme du gradient projeté, qui est identique à celui du gradient à la projection près :

Initialisation :

$k = 0$, *choix de* $x^{(0)}$ *et* $\varepsilon > 0$.

Itération k :

Tant que le critère d'arrèt est non satisfait :

$$\hat{x}^{(k+1)} = x^{(k)} - p^{(k)} \nabla f(x^{(k)})$$
$$x^{(k+1)} = \Pi_C \hat{x}^{(k+1)}$$
$$k = k + 1$$

Fin tant que

Théorème 3.1.1. *[6](Convergence).*
Soit f une fonction de classe C^1 définie de $\mathbb{R}^n$ dans $\mathbb{R}$. On suppose que f est elliptique de dérivée lipschitzienne. Alors, si on choisit le pas $p^{(k)}$ dans un intervalle $[\beta_1, \beta_2]$ tel que $0 < \beta_1 < \beta_2 < \frac{2\alpha}{M^2}$ (où α et M sont respectivement les constantes d'elliplicité

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \alpha \|x - y\|^2$$

et d'inégalité lipschitzienne

$$(\|f(x) - f(y)\| \leq M \|x - y\|).$$

La suite $(x^{(k)})_k$ définie par la mèthode du gradient projeté converge vers la solution du problème

Exemple 3.1.1. *[6] Pour des contraintes de type $x \geq 0$, l'opérateur de projection est*

$$\Pi_C(x) = \max(x_i, 0)_{i=1,\ldots,m}.$$

Pour des contraintes de type $a \leq x \leq b$, l'opérateur de projection est

$$\Pi_C(x) = \begin{cases} b, & si \;\; x > b, \\ x, & si \;\; a \leq x \leq b, \\ a, & si \;\; x < a. \end{cases}$$

3.2 Méthode Lagrange-Newton pour des contraintes égalité

Cas d'un problème quadratique avec des contraintes affines égalités

[6] Considérons un problème quadratique avec contraintes affines de type égalité :

$$\begin{cases} \min \frac{1}{2} x^T Q x - c^T x, \\ Ax = b \end{cases}$$

où Q est une matrice carrée de $\mathbb{R}$ d'ordre $n \times n$, c un vecteur de $\mathbb{R}^n$, A une matrice de $\mathbb{R}$ d'ordre $m \times n$ et b un vecteur de $\mathbb{R}^m$.
*Les conditions de **KKT** de premier ordre nous donnent*

$$\begin{cases} \nabla_x (\frac{1}{2} x^T Q x - c^T x + \lambda^T (Ax - b)) = 0, \\ Ax = b, \end{cases}$$

où $\lambda \in \mathbb{R}^m$ est le vecteur multiplicateur de Lagrange associé aux contraintes $Ax = b$.

Par conséquent, le couple optimal (x^*, λ^*) *est la solution du système :*

$$\begin{cases} Qx^* - c + A^T\lambda^* = 0, \\ Ax^* = b. \end{cases}$$

Ce système peut s'écrire comme suit :

$$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} c \\ b \end{pmatrix}$$

Si la matrice $\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix}$ *est inversible, le système admet une solution que l'on peut calculer par une des méthodes de résolution des systèmes linéaires.*

Cas d'un problème non quadratique

[6] Considérons un problème non quadratique suivant :

$$\begin{cases} \min f(x), \\ g(x) = 0 \end{cases}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ *et* $g : \mathbb{R}^n \to \mathbb{R}^m$ *sont de classe* C^2.

Les conditions de premier ordre de ***KKT*** *s'écrivent comme suit :*

$$\begin{cases} \nabla_x L(x, \lambda) = 0, \\ g(x) = 0, \end{cases}$$

où $\lambda \in \mathbb{R}^m$ *et* $L(x, \lambda) = f(x) + \lambda^T g(x)$ *est le lagrangien du problème.On résoud ce système d'équations non linéaires par la méthode de Newton. C'est l'algorithme de Lagrange-Newton énoncé comme suit :*

- *Initialisation :* $k = 1$*, Choix de* $(x^{(0)}, \lambda^{(0)}) \in \mathbb{R}^n \times \mathbb{R}^m$.
- *Itération* k

— *Tant que le critère d'arrêt est non satisfait, faire*

Résoudre le système linéaire suivant :

$$\begin{pmatrix} H_{x,x}L(x^{(k)},\lambda^{(k)}) & \nabla_x g(x^{(k)})^T \\ \nabla_x g(x^{(k)}) & 0 \end{pmatrix} \begin{pmatrix} d^{(k)} \\ y^{(k)} \end{pmatrix} = - \begin{pmatrix} \nabla_x L(x^{(k)},\lambda^{(k)})^T \\ g(x^{(k)}) \end{pmatrix}$$

où $H_{x,x}L(x^{(k)},\lambda^{(k)})$ *est le hessien de* L *par rapport à* x.

Poser

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + d^{(k)} \\ \lambda^{(k+1)} &= \lambda^{(k)} + y^{(k)} \\ k &= k+1. \end{aligned}$$

— *Fin tant que*

3.3 Méthode de Newton projetée

Cas de contraintes de type $x \geq 0$

[6] Considérons le problème suivant :

$$\begin{cases} \min f(x) \\ x \geq 0 \end{cases} \tag{3.1}$$

où f *est de classe* C^2 *sur* $\mathbb{R}^n$ *. Si* $H^{(k)}$ *designe la matrice hessienne de* f *en* $x^{(k)}$, *une itération de la méthode de Newton est de la forme suivante :*

$$x^{(k+1)} = x^{(k)} - [H^{(k)}]^{-1}(\nabla f(x^{(k)}).$$

On peut affiner en introduisant un pas $\alpha^{(k)} > 0$, *ce qui donne :*

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)}[H^{(k)}]^{-1}(\nabla f(x^{(k)}).$$

Si on projette sur l'ensemble des contraintes, on obtient :

$$[x^{(k+1)} = x^{(k)} - \alpha^{(k)}[H^{(k)}]^{-1}(\nabla f(x^{(k)})]^+$$

où $x^+ = \max(x_i, 0)_{1 \leq i \leq n}$. *On a le théorème de convergence suivant :*

Théorème 3.3.1. *[6]On suppose que* f *est convexe et est de classe* C^2 *et que le problème (3.1) a une solution unique* x^* *vérifiant* $\frac{\partial f}{\partial x_i}(x^*) > 0, \quad \forall i \in I_a(x^*)$, *(*$I_a(x^*)$ *est l'ensemble des indices actifs). On suppose de plus que l'on peut trouver* m_1 *et* m_2 *deux réels strictement positifs tels que :*

$$m_1 \|z\|^2 \leq z^T \nabla^2 f(x) z \leq m_2 \|z\|^2,$$

dans chacun des cas suivants :

- *pour tout* $z \in \{x / f(x) \leq f(x^{(0)})\}$, *d'une part ;*
- x *dans une boule centrée en* x^* *et* $z \neq 0$ *tel que* $z_i = 0$ *pour* $i \in I_a(x^*)$,*d'autre part.*

Alors la suite $x^{(k)}$ *engendrée par l'algorithme converge vers* x^* *et le taux de convergence est superlineaire. On peut alors généraliser l'algorithme précédent au problème à contraintes bornées.*

Cas de contraintes de bornes

[6] Dans ce cas, le problème se présente comme suit :

$$\begin{cases} \min \ f(x) \\ a \leq x \leq b \end{cases}$$

Nous appliquons les itérés de la méthode de Newton projetée comme dans le cas du problème (3.1), en utilisant la projection suivante :

$$[x^{(k+1)} = x^{(k)} - \alpha^{(k)} [H^{(k)}]^{-1} (\nabla f(x^{(k)})]^{Proj}$$

où pour tout $z \in \mathbb{R}^n, z^{Proj} = \Pi_C(z)$ *désigne le vecteur :*

$$[z]_i^{Proj} = \begin{cases} b_i & si \ b_i \leq z_i, \\ z_i & si \ a_i \leq z_i \leq b_i, \\ a_i & si \ z_i \leq a_i. \end{cases}$$

3.4 Méthodes de pénalisation

Principe de la méthode

[6] Le principe est de remplacer la fonction objectif originale du problème par une fonction de pénalité. La fonction de pénalité est composée de :

- *la fonction objectif originale du problème d'optimisation avec contraintes, et*
- *un terme supplémentaire pour chaque contrainte.*
- *Ce terme est positif si le point courant x viole les contraintes et nul dans le cas contraire.*

Donc on remplace le problème avec contraintes suivant :

$$\begin{cases} f(x) \to min, \\ x \in C \subset \mathbb{R}^n, \end{cases} \tag{3.2}$$

par un problème sans contraintes suivant :

$$\begin{cases} f(x) + \frac{1}{\mu}\alpha(x) \to min, \\ x \in \mathbb{R}^n, \end{cases} \tag{3.3}$$

où $\alpha : \mathbb{R}^n \to \mathbb{R}$ est une fonction de pénalisation des contraintes et $\mu > 0$. Le but est de trouver des fonctions α telles que les problèmes (3.2) et (3.3) soient équivalents.

Nous supposons que α vérifie les propriétés suivantes :

1. *La fonction α est continue sur $\mathbb{R}^n$.*
2. $\forall x \in \mathbb{R}^n, \alpha(x) \geq 0.$
3. $\alpha(x) = 0 \Leftrightarrow x \in C.$

Nous donnons quelques exemples de pénalisation pour différentes contraintes :

1. *Pour une contrainte $x \geq 0$, on a la fonction $\alpha(x) = \|x^-\|^2$ '*
2. *Pour une contrainte $h(x) = 0$, on a la fonction $\alpha(x) = \|h(x)\|^2$*
3. *Pour une contrainte $g(x) \geq 0$, on a la fonction $\alpha(x) = \|g(x)^-\|^2$*

où $\| - \|$ est la norme euclidienne de $\mathbb{R}^n$ et $x^- = \max(-x, 0)$.

La méthode de pénalisation dans le cas de contraintes égalités

Dans le cas de problème d'optimisation avec contraintes de type égalité,la fonction pénalité utilisée est la fonction quadratique.
Le problème à résoudre est le suivant :

$$\begin{cases} f(x) \to min, \\ g_i(x) = 0, \qquad 1 \leq i \leq m. \end{cases} \tag{3.4}$$

La fonction quadratique de pénalisation $q(x, \mu)$ pour (3.4)est :

$$q(x, \mu) = f(x) + \frac{1}{2\mu} \sum_{i=1}^{m} g_i^2(x)$$

où $\mu > 0$ est le paramètre de pénalisation.
En tendant vers zéro, nous pénalisons les violations des contraintes.
Le principe est donc : on considère une séquence de valeurs de $\{\mu_k\}$ avec $\mu_k \to 0, k \to +\infty$ et on calcule le minimum x_k de $q(x, \mu_k)$ pour chaque k.

Exemple 3.4.1. *[6] On considère le problème suivant :*

$$\begin{cases} \min \ x_1 + x_2 \\ x_1^2 + x_2^2 = 2. \end{cases}$$

La fonction de pénalisation quadratique associée est :

$$q(x, \mu) = x_1 + x_2 + \frac{1}{2\mu}(x_1^2 + x_2^2 - 2)^2.$$

Pour $\mu = 1$, on trouve le minimum de q est $(-1.1; -1.1)$;
Pour $\mu = 0.1$, on trouve le minimum de q est $(-1; -1)$;
La solution du problème est donc $(-1; -1)$.

La méthode de pénalisation dans le cas de contraintes mixtes

[6] Considérons le problème d'optimisation avec contraintes mixtes suivant :

$$\begin{cases} f(x) \to min, \\ g_i(x) = 0, & 1 \leq i \leq m, \\ h_j(x) \geq 0, & 1 \leq j \leq l. \end{cases}$$

On définit la fonction q de pénalisation comme suit :

$$q(x, \mu) = f(x) + \frac{1}{2\mu} \sum_{i=1}^{m} g_i^2(x) + \frac{1}{2\mu} (h_j^-(x))^2,$$

où $y^- = max(-y, 0)$.

Algorithme de la méthode de pénalisation quadratique

[6] Soient $\mu_0 > 0, \varepsilon_0 > 0$, et le point initial x_0^s donnés.

Tant que $\|\nabla q(x, \mu_k)\| > \varepsilon_k$ faire

— *Trouver le minimum approximatif de $q(., \mu_k)$, en partant de x_k^s ; soit le minimum x_k ;*

— *Choisir le nouveau paramètre de pénalisation $\mu_{k+1} \in [0, \mu_k]$.*

— *Choisir le nouveau point x_{k+1}^s*

Fin Tant que

Théorème 3.4.1. *[6]*

Supposons que chaque solution x_k est un minimum global exact de $q(x, \mu_k)$ dans l'algorithme précédent et que $\mu_k \to 0$. Alors, chaquen point x^* limite de la séquence $\{x_k\}$ est une solution du problème (3.4).

preuve. *Soit x une solution globale du problème (3.4), alors*

$$f(\overline{x}) \leq f(x), \forall x : g_i(x) = 0, 1 \leq i \leq m.$$

Puisque $x_k minimise q(., \mu_k)$ pour chaque k, alors on a

$$q(x_k, \mu_k) \leq q(\overline{x}, \mu_k),$$

d'où

$$f(x_k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} g_i^2(x_k) \le f(\overline{x}) + \frac{1}{2\mu_k}\sum_{i=1}^{m} g_i^2(\overline{x}) = f(\overline{x}) \tag{3.5}$$

alors

$$\sum_{i=1}^{m} g_i^2(x_k) \le 2\mu_k(f(\overline{x}) - f(x_k)). \tag{3.6}$$

Supposons que x^* est un point limite de $\{x_k\}$, donc il existe une sous-séquence $\mathcal{K}$ telle que

$$\lim_{k\in\mathcal{K}, k\to+\infty} x_k = x^*.$$

En passant aux limites dans (3.6), $k \to +\infty, k \in \mathcal{K}$, on obtient

$$\sum_{i=1}^{m} g_i^2(x^*) = \lim_{k\in\mathcal{K}}\sum_{i=1}^{m} g_i^2(x_k) \le \lim_{k\in\mathcal{K}} 2\mu_k(f(\overline{x}) - f(x_k)) = 0;$$

d'où, $g_i(x^*) = 0,\ 1 \le i \le m$. x^* est alors réalisable.

D'un autre coté, ayant $k \to \infty, k \in \mathcal{K}$ dans l'equation (3.5), μ_k non négative et $g_i^2(x_k),\ 1 \le i \le m$ non négatives, on a

$$f(x^*) \le f(x^*) + \lim_{k\in\mathcal{K}} \frac{1}{2\mu_k}\sum_{i=1}^{m} g_i^2(x_k) \le f(\overline{x}).$$

Ayant x^* réalisable et $f(x^*) \le f(\overline{x})$, alors on conclut que x^* est un minimum global.

Théorème 3.4.2. *[6]*
Si les ε_k de l'algorithme précédent vérifient

$$\lim_{k\to\infty} \varepsilon_k = 0$$

et les paramètres de pénalisation vérifient $\mu_k \to 0$, alors tous les points limites x^* de la séquence $\{x_k\}$ pour lesquelles les gradients $\nabla g_i(x^*)$ sont linéairement indépendants, on a x^* est un point de Karush-Kuhn-Tucker du problème (3.4). Pour ces points, on a une infinité de sous-séquence $\mathcal{K}$ tel que

$$\lim_{k\in\mathcal{K}} x_k = x^*,$$

$$\lim_{k\in\mathcal{K}} -g_i(x_k)/\mu_k = \lambda_i^*, \quad 1 \leq i \leq m, \tag{3.7}$$

où λ^* est le vecteur multiplicateur qui satisfait les conditions de KKT.

preuve.

Soit $q(x_k, \mu_k) = f(x_k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} g_i^2(x_k)$, d'où

$$\nabla_x q(x_k, \mu_k) = \nabla f(x_k) + \sum_{i=1}^{m} \frac{g_i(x_k)}{\mu_k}\nabla g_i(x_k).$$

En appliquant le critère d'arrêt de l'algorithme, on obtient :

$$\|\nabla f(x_k) + \sum_{i=1}^{m} \frac{g_i(x_k)}{\mu_k}\nabla g_i(x_k)\| \leq \varepsilon_k \tag{3.8}$$

d'où

$$\|\sum_{i=1}^{m} g_i(x_k)\nabla g_i(x_k)\| \leq \mu_k(\varepsilon_k + \|\nabla f(x_k)\|)$$

Quand $k \to \infty$ pour $k \in \mathcal{K}$, on a

$$\varepsilon_k + \|\nabla f(x_k)\| \to \|\nabla f(x^*)\|$$

d'où, comme $\mu_k \to 0$, on a

$$\mu_k(\varepsilon_k + \|\nabla f(x_k)\|) \to 0.$$

On obtient alors

$$\sum_{i=1}^{m} g_i(x^*)\nabla g_i(x^*) = 0.$$

Par hypothèse, on a $\nabla g_i(x^*)$, $\quad 1 \leq i \leq m$ linéairement indépendants et $g_i(x^*) = 0, \quad 1 \leq i \leq m$, d'où x^* est réalisable. On va maintenant démontrer la condition de KKT et montrer que la limite (3.7) du théorème 3.4.2 se produit.
On dénote la matrice des gradients des contraintes par $A(x)$,

$$A(x) = (\nabla g_i(x))_{1\leq i\leq m}$$

et

$$\lambda_k = -\frac{g(x_k)}{\mu_k}$$

on a alors comme dans (3.8) :

$$A(x_k)^T\lambda_k = (\nabla g_i(x))\left(\frac{-g_i(x_k)}{\mu_k}\right)_{1\leq i\leq m} = \nabla f(x_k) - \nabla_x q(x_k,\mu) \tag{3.9}$$

$$\|\nabla_x q(x_k,\mu)\| \leq \varepsilon_k.$$

Si on multiplie l'équation (3.9) par $A(x_k)$, on a

$$\lambda_k = (A(x_k)A^T(x_k))^{-1}A(x_k)(\nabla f(x_k) - \nabla_x q(x_k,\mu)).$$

D'où

$$\lim_{k\in\mathcal{K}} \lambda_k = \lambda^* = (A(x^*)A^T(x^*))^{-1}A(x^*)\nabla f(x^*).$$

Conformément à l'équation (3.13), on a

$$\nabla f(x^*) - A^T(x^*)\lambda^* = 0,$$

d'où λ^* satisfait les conditions de KKT. Donc x^* est un point de KKT.

Exemple 3.4.2. *[6]Considérons le problème de minimisation suivant :*

$$\begin{cases} min\ x, \\ x^2 \geq 0, \\ x+1 \geq 0, \end{cases}$$

pour lequel $x^* = -1$ est la solution. Ecrire la fonction de pénalisation $q(x,\mu)$ associée à ce problème et trouver les minimums locaux
Montrer que pour chaque séquence $\{\mu_k\}_k$ telle que $\mu_k \to 0$, il existe une séquence de minimums locaux $x(\mu_k)$ qui convergent vers -1.

Solution :

On définit la fonction de pénalisation $q(x,\mu)$ comme suit :

$$q(x,\mu) = x + \frac{1}{2\mu}((x^2)^-)^2 + \frac{1}{2\mu}((x+1)^-)^2$$

avec

$$(x^2)^- = \max(0,-x^2) = 0, \forall x,$$

$$(x+1)^- = \max(0,-1-x)$$

— *Si* $x \geq -1$, *alors* $(x+1)^- = 0$. *On a alors* $q(x,\mu) = x$.

— *Si* $x < -1$, *alors* $(x+1)^- = -x-1$. *On a alors* $q(x,\mu) = x + \frac{1}{2\mu}(x+1)^2$.

Dans les deux cas, on trouve que min $q(x,\mu)$ *est atteint pour* $x \to -1$.

3.5 Méthode de programmation quadratique successive (S.Q.P)

Contraintes d'égalité

[12] Considérons un problème d'optimisation (différentiable) avec contraintes d'égalité :

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

$$s.c. : \; h_i(x) = 0, i = 1, \ldots, p.$$

où $f : \mathbb{R}^n \to \mathbb{R}$ *et* $h_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, p$, *sont supposées au moins différentiables.*

D'après ce qui a été vu au chapitre 2, les conditions d'optimalité de Lagrange s'écrivent :

$$\nabla L(x;\lambda) = 0 \; soit : \; \begin{pmatrix} \nabla f(x) + \sum\limits_{i=1}^{p} \lambda_i \nabla h_i(x) \\ h(x) \end{pmatrix} = 0$$

Pour résoudre ce système d'équations, utilisons la méthode de Newton dont une itération s'écrit ici :

$$H[L](x^k;\lambda^k) \begin{pmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = -\nabla L(x^k;\lambda^k), \qquad (3.10)$$

soit :

$$\begin{pmatrix} H_x[L](x^k;\lambda^k) & \nabla h(x^k)^T \\ \nabla h(x^k) & 0 \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = - \begin{pmatrix} \nabla_x L(x^k;\lambda^k) \\ h(x^k) \end{pmatrix}$$

où $\nabla h(x)$ désigne la matrice jacobienne de l'application $h : \mathbb{R}^n \to \mathbb{R}^p$ définie par :

$$\nabla h(x)^T = [\nabla h_1(x)| \dots |\nabla h_p(x)].$$

Posons :$H_k = H_x[L](x_k; \lambda_k), d = x^{k+1} - x^k$ et $\mu = \lambda^{k+1}$. L'itération (3.10) s'écrit donc :

$$\begin{pmatrix} H_k & \nabla h(x^k)^T \\ \nabla h(x^k) & 0 \end{pmatrix} \begin{pmatrix} d \\ \mu - \lambda^k \end{pmatrix} = - \begin{pmatrix} \nabla_x L(x^k; \lambda^k) \\ h(x^k) \end{pmatrix}$$

et est bien définie à condition que la matrice $H[L](x^k; \lambda^k)$ soit inversible. Ce sera le cas si :

1. *Les colonnes $\nabla h_1(x), \dots, \nabla h_p(x)$ sont linéairement indépendants : c'est l'hypothèse de qualification des contraintes.*
2. *Quel que soit $d \neq 0$ tel que $(x^k)d = 0,\ d^T H_k d > 0$: c'est la condition suffisante d'optimalité du second ordre dans le cas de contraintes d'égalité.*

Revenons à l'itération (3.10). Elle s'écrit encore :

$$\begin{cases} H_k d + \sum\limits_{i=1}^{p} (\mu_i - \lambda_i^k) \nabla h_i(x^k) = -\nabla_x L(x^k; \lambda^k) \\ \nabla h_i(x^k)^T d + h_i(x^k) = 0, & i = 1, \dots, p \end{cases}$$

Or : $\nabla_x L(x^k; \lambda^k) = \nabla f(x_k) + \sum\limits_{i=1}^{p} \lambda_i^k \nabla h_i(x^k)$, d'où :

$$\begin{cases} H_k d + \sum\limits_{i=1}^{p} \mu_i \nabla h_i(x^k) = -\nabla f(x^k) \\ \nabla h_i(x^k)^T d + h_i(x^k) = 0, & i = 1, \dots, p \end{cases}$$

On reconnait dans le système ci-dessus les conditions d'optimalité de Lagrange du problème quadratique suivant :

$$(QP_k) \begin{cases} \min_{d \in R^n} \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d \\ s.c. \quad h_i(x^k) + \nabla h_i(x^k)^T d = 0. \end{cases}$$

Le problème (QP_k) peut être vu comme la minimisation d'une approximation quadratique du Lagrangien de (P_E) avec une approximation linéaire des contraintes.

Comme son nom l'indique, la méthode SQP (Sequential Quadratic Programming) consiste à remplacer le problème initial par une suite de problèmes quadratiques sous contraintes linéaires plus faciles à résoudre. L'algorithme est le suivant :

Algorithme SQP avec contraintes d'égalité.*[12]*

Données : $f : \mathbb{R}^n \to R$, $h : \mathbb{R}^n \to \mathbb{R}^p$ *différentiables,* x_0 *point initial,* $\lambda_0 \to \mathbb{R}^p$ *multiplicateur initial,* $\epsilon > 0$ *précision demandée.*

Sortie : une approximation x_l *de la solution.*

1. $k := 0$;
2. *Tant que* $\|\nabla L(x^k; \lambda_k)\| > \epsilon$,
   (a) *Résoudre le sous-problème quadratique :*

$$(QP_k) \begin{cases} \min_{d \in R^n} \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d \\ s.t. \quad h_i(x^k) + \nabla h_i(x^k)^t d = 0. \end{cases}$$

   et obtenir la solution primale d_k *et le multiplicateur* λ' *associé à la contrainte d'égalité*

   (b) $x^{k+1} = x_k + d_k;\ \lambda_{k+1} = \lambda';\ k = k + 1$;
3. *Retourner* x_k.

Contraintes d'inégalité

[12] Intéressons nous maintenant aux problèmes avec contraintes d'égalité et d'inégalité :

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

$$s.c. : g_j(x) \leq 0,\ j = 1, \dots, q,$$

$$h_i(x) = 0,\ i = 1, \dots, p.$$

Selon le même principe qu'avec contraintes d'égalité seules, on linéarise les contraintes et on utilise une approximation quadratique du lagrangien :

$$L(x; \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x), \quad avec: \quad \lambda j \geq 0, j = 1, \ldots, q.$$

Algorithme SQP avec contraintes d'égalité et d'inégalité.

Données : $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^q, h : \mathbb{R}^n \to \mathbb{R}^p$ *différentiables,* x_0 *point initial,* $\lambda_0 \in \mathbb{R}^q_+$ *et* $\mu_0 \in \mathbb{R}^p$ *multiplicateurs initiaux,* $\epsilon > 0$ *précision demandée.*

Sortie : une approximation x^* *de la solution.*

1. $k := 0$;
2. *Tant que* $\|\nabla L(x^k; \lambda^k)\| > \epsilon$,
   (a) *Résoudre le sous-problème quadratique :*

$$(QP_k) \begin{cases} \min_{d \in R^n} \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d \\ s.t. \quad g_j(x^k) + \nabla g_j(x^k)^T d = 0, i = 1, \ldots, q, \\ h_i(x^k) + \nabla h_i(x^k)^T d = 0, i = 1, \ldots, p. \end{cases}$$

   et obtenir la solution primale d_k *et les multiplicateurs* λ' *et* μ' *associés aux contraintes d'inégalité et d'égalité respectivement.*

   (b) $x^{k+1} = x_k + d_k; \lambda^{k+1} = \lambda'; \mu^{k+1} = \mu'; k = k + 1$;
3. *Retourner* x_k.

3.6 Méthode d'Uzawa

[6] L'idée générale de la méthode est de considérer le Lagrangien L au lieu de la fonction f. Ce choix est motivé par, au moins, deux raisons :

— *La fonction de Lagrange L englobe la fonction f et les cotraintes g et h ; donc L représente bien le problème ;*

— *Ensuite, nous avons vu qu'une condition necessaire du premier ordre pour que* x^* *soit un minimum de f sous les contraintes est que* x^* *(associé aux multiplicateurs de Lagrange) soit un point critique de L.*

Notre problème à considérer est :

$$\begin{cases} \min f(x) \\ h_j(x) = 0 \quad 1 \leq j \leq l, \\ g_i(x) \leq 0 \quad 1 \leq i \leq m. \end{cases}$$

Le Lagrangien du problème est :

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

Définition 3.6.1. *On appelle point-selle de L sur $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$ tout triplet $(x^*, \mu^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$ vérifiant l'équation suivante :*

$$L(x^*, \mu, \lambda) \leq L(x^*, \mu^*, \lambda^*) \leq L(x, \mu^*, \lambda^*), \forall (x, \mu, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}.$$

On a alors le résultat suivant :

Théorème 3.6.1. *[6]*
Supposons que f, h et g soient des fonctions de classe C^1 et que le triplet $(x^*, \mu^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$ soit un point selle de L sur $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$. Alors, ce triplet vérifie les conditions de ***KKT****.*

Théorème 3.6.2. *[6]*
Supposons que f, h et g soient des fonctions convexes et de classe C^1. Alors, le triplet $(x^*, \mu^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$ est un point selle de L sur $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$ si et seulement s'il vérifie les conditions de KKT. L'algorithme d'Uzawa se base sur ces théorèmes. Il consiste à chercher un triplet $(x^*, \mu^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$ vérifiant les conditions de KKT de la façon suivante :

1. *Pour (μ^*, λ^*) fixé dans $\mathbb{R}^l \times \mathbb{R}^{+m}$, nous allons chercher le minimum sur $\mathbb{R}^n$ de la fonction définie ainsi :*

$$x \to L(x, \mu^*, \lambda^*).$$

2. *Pour x^* fixé sur $\mathbb{R}^n$, on cherche le maximum sur $\mathbb{R}^l \times \mathbb{R}^{+m}$ de la fonction définie ainsi :*
$$(\mu, \lambda) \to L(x^*, \mu, \lambda).$$

En faisant ces deux calculs simultanément, on obtient l'algorithme d'Uzawa suivant :

Initialisation – Poser $k = 0$;

– Choisir $\mu^{(0)} \in \mathbb{R}^l$ et $\lambda^{(0)} \in \mathbb{R}^{+m}$

Itération k : Tant que le critère d'arrèt n'est pas satisfait, faire

a) Calculer $x^{(k)} \in \mathbb{R}^n$ solution du problème :
$$\begin{cases} \min \ L(x, \mu^{(k)}, \lambda^{(k)}) \\ x \in \mathbb{R}^n. \end{cases}$$

b) Calculer $\mu^{(k+1)}$ et $\lambda^{(k+1)}$ comme suit :
$$\begin{cases} \mu_j^{(k+1)} = \mu_j^{(k)} + ph_j(x^{(k)}), & 1 \leq j \leq l, \\ \lambda_i^{(k+1)} = \max(0, \lambda_i^{(k)} + pg_i(x^{(k)})), & 1 \leq i \leq m \end{cases}$$
où $p > 0$ est un réel fixé.

Théorème 3.6.3. *[6](Convergence)*

On suppose que f est de classe C^1 et elliptique, que h est affine, g est convexe de classe C^1 et que g et h sont lipschitziennes. On suppose de plus que la fonction de Lagrange L possède un point selle (x^*, μ^*, λ^*) sur $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^{+m}$. Alors, il existe p_1, p_2 avec $0 < p_1 < p_2$ tels que $\forall p \in [p_1, p_2]$ la suite $(x^{(k)})_{k>0}$ générée par l'algorithme converge vers x^*.

Bibliographie

[1] *M. Bierlaire, Introduction à l'optimisation différentiel le. Presses polytechniques et universitaires romandes, 2006.*

[2] *O.Boudjaatat, N.Nesrouche, R.Bouderbane, S.Dekkiche, M.Azi, optimisation non linéaire avec contrainte, Mémoire préparé En vue de l'obtention du diplôme de licence Mathématiques fondamentales,Centre Universitaire de Mila, 2014.*

[3] *M. Bouraine, Cours - Optimisation non linéaire en dimension finie avec contraintes. Université A. Mira - Béjaia, 2019-2020.*

[4] *M.Chebah, R.Arezki, La programmation mathématique Avec les méthodes des points intérieurs, mémoire de fin d'études en vue l'obtention du diplôme de master en mathématiques, recherche operationnelle,université mouloud mammeri de Tizi-Ouzou, 08 .07.2013.*

[5] *I.S. Ciuperca, Optimisation, Cours en Master M1 SITN, Université Lyon 1, 19.02.2020.*

[6] *S. Douar-Radjef, optimisation -Cours et exercices. Université des sciences et de la Technologie d'Oran. Mohamed Boudiaf, 2019.*

[7] *J. Gauvin, Leçons de programmation mathématique. Ecole Polytechnique de Montreal, 1995.*

[8] *J.B. Hiriart-Urruty, Les mathématiques du mieux faire : Premiers pas en optimisation. Volume 1, Paris : Ellipses, 2007.*

[9] J.B. Hiriart-Urruty, Optimisation et analyse convexe - exercices et problèmes corrigés, avec rappels de cours. EDP Sciences, 2009.

[10] A.Intissar et J.K.Intissar, Calcul Differentiel fondements Applications, cours exercices avec solutions 2016-2017.

[11] Y. Privat et D. Smets, Optimisation linéaire convexité, Cours de Licence 3 - LM239, Université Pierre et Marie Curie – Paris 6 et CNRS 2017-2018.

[12] A. Rondepierre, Méthodes numériques pour l'optimisation non linéaire déterministe. Insa Toulouse 2017.

[13] S. Kenouche, cours : Méthodes Mathématiques et Algorithmes pour la Physique Version corrigée. Département des Sciences de la Matière - UMKB, 2020.