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

CENTRE UNIVERSITAIRE DE MILA
INSTITUT DES SCIENCES ET DE LA TECHNOLOGIE

Réf. ***/11***

Mémoire de fin d'étude
Présenté pour l'obtention du diplôme de

Licence Académique

Domaine : **Mathématiques et Informatique**
Filière : **Mathématiques**
Spécialité : **Mathématiques Fondamentales**

Thème

La méthode du gradient réduit pour un problème d'optimisation avec contraintes

Présenté par :
- *Bouzerara Meriem*
- *Bouchama Nassima*

Dirigé par :
Dr. Boukaroura Khelladi Samia

Année universitaire 2010-2011

Table des matières

Introduction générale

Dans la vie courante, nous sommes fréquemment confrontés à des problèmes d'optimisation plus au moins complexes. Cela peut commencer au moment où l'on tente de ranger son bureau, de placer son mobilier, et aller jusqu'à un processus industriel, par exemple pour la planification des différentes taches. Ces problèmes peuvent être exprimés sous la forme générale d'un "problème d'optimisation". On définit alors une fonction objectif (fonction de coût ou fonction profit), que l'on cherche à optimiser (minimiser ou maximiser) par rapport à tous les paramètres (contraintes) concernés.

Il existe deux grandes catégories de problèmes celle des problèmes d'optimisation sans contraintes et celle des problèmes d'optimisation avec contraintes.

Dans ce mémoire on s'interesse aux problèmes d'optimisation avec contraintes. Dans le premier chapitre on donne les notions fondamentales pour l' optimisation avec contraintes en particulier les conditions d'optimalité. Dans le second chapitre, on va étudier la méthode du gradient projeté ensuite la méthode du gradient réduit et on termine ce chapitre en donnant la méthode du gradient réduit généralisé

Chapitre 1

Notions fondamentales pour l'optimisation avec contraintes

1.1 *Introduction*

Soit le problème d'optimisation avec contraintes :

$$(P)\begin{cases}\min f(x)\\ x \in D\end{cases}$$

où $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et D un sous ensemble de $\mathbb{R}^n$ $(D \subset \mathbb{R}^n)$ appelé ensemble des contraintes.

Ces contraintes sont données à l'aide de fonctions au moins continues (i.e. les fonctions $g_i, h_j \in C^1(\mathbb{R}^n)$), sous plusieurs formes :

1- Contraintes d'égalités :

L'ensemble des contraintes est donné par :

$$D = \{x \in \mathbb{R}^n \ / h_j(x) = 0, \forall j = 1, ..., l\}$$

2- Contraintes d'inégalités :

L'ensemble des contraintes est donné par :

$$D = \{x \in \mathbb{R}^n \ / g_i(x) \leq 0, \forall i = 1, ..., m\}$$

3- Contraintes d'égalités et d'inégalités :

L'ensemble des contraintes est donné par :

$$D = \{x \in \mathbb{R}^n \ / h_j(x) = 0, \forall j = 1, ..., l \text{ et } g_i(x) \leq 0, \forall i = 1, ..., m\}$$

Définition 1.1.1 *(minimum)*

1- *On dit que $x^\star \in D$ est un minimum local de f sur D si et seulement si*

$$\exists \varepsilon > 0 \ / \ f(x^\star) \leq f(x), \forall x \in B(x^\star, \varepsilon) \cap D.$$

2- *On dit que $x^\star \in D$ est un minimum global de f sur D si et seulement si*

$$f(x^\star) \leq f(x), \forall x \in D.$$

On rencontre par fois des classifications des problèmes d'optimisation, ce qui permet de choisir un algorithme de résolution adapté, par exemple on parle de :

Programmation linéaire :

Lorsque la fonction f est linéaire

$$f(x) = c^T x \text{ où } x \in \mathbb{R}^n \text{ et } c \in \mathbb{R}^n$$

et l'ensemble des contraintes D est donné par des fonctions affines

$$g_i(x) = c^T x + b \text{ où } c \in \mathbb{R}^n, x \in \mathbb{R}^n \text{ et } b \in \mathbb{R}.$$

En général, dans la programmation linéaire D est un polyèdre

$$D = \{x \in \mathbb{R}^n / Ax \leq b\}$$

où A est une $m \times n$ matrice et $b \in \mathbb{R}^n$.

Ces problèmes sont souvent résolus par la méthode du simplexe.

Programmation quadratique :

Lorsque f est quadratique, c'est-à-dire

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

où $b \in \mathbb{R}^n$ et A est une $n \times n$ matrice symétrique définie positive, et les fonctions g_i, h_j sont des fonctions affines.

En général D est un polyèdre convexe.

Programmation convexe :

Lorsque f et l'ensemble des contraintes D sont convexes.

Programmation différentiable :

Lorsque f et les fonctions g_i, h_j sont une ou deux fois différentiables.

1.2 *Résultats d'existence et d'unicité d'un minimum*

La plus part des théorèmes d'éxistence d'un minimum sont des variantes du théorème de Weirstrass.

Théorème 1.2.1 *(existence)*

Soit $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ *une fonction continue et soit* D *un sous ensemble de* $\mathbb{R}^n$, *alors* f *possède un minimum* $x^\star \in D$ $\left(i.e.\ \exists x^* \in D\ /\ f(x^*) = \min\limits_D f(x)\right)$ *dans l'un des deux cas suivants :*

1- *Si* D *est un compact (fermé et borné) de* $\mathbb{R}^n$.

2- *Si* D *est fermé et* f *coercive* $\left(\lim\limits_{\|x\|\to+\infty} f(x) = +\infty\right)$.

Théorème 1.2.2 *(unicité)*

Soit f *une fonction strictement convexe sur l'ensemble convexe* D, *alors il existe au plus un minimum globale* $x^* \in D$ *de* f.

Théorème 1.2.3 *(existence et unicité)*

Soit f *une fonction strictement convexe sur* D, *et* D *un sous ensemble de* $\mathbb{R}^n$ *convexe et fermé, alors si* D *est borné ou* f *est coercive, il existe un unique minimum global* $x^* \in D$ *de* f, *c'est-à-dire*

$$\exists! x^* \in D\ /\ f(x^*) = \min_D f(x)$$

1.3 *Conditions d'optimalité*

Proposition 1.3.1 *(condition nécessaire)*

Soit $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, *une fonction de classe* $C^1(\mathbb{R}^n)$ *et* D *un ensemble de* $\mathbb{R}^n$ *et* $x^\star \in D$ *tels que* $f(x^\star) = \min\limits_D f(x)$, *alors :*

1- *si* $x^\star \in int(D)$, *alors* $\nabla f(x^\star) = 0$.

2- *si* D *est convexe, alors* $\langle \nabla f(x^\star), x - x^\star \rangle \geq 0$, $\forall x \in D$.

3- *si* f *est deux fois différentiable, alors* $H_f(x^\star)$ *la matrice hessienne de* f *au point* $x^\star$ *est définie positive.*

1.3.1 *Cas des contraintes d'inégalité*

On suppose dans cette partie que de l'ensemble D, sur le quel on veut minimiser f, est donné par des contraintes d'inégalité, i.e. on a le problème :

$$(P)\begin{cases} \min f(x) \\ x \in D \end{cases} \quad / \ f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

tel que

$$D = \{x \in \mathbb{R}^n \ / \ g_i(x) \leq 0, \forall i = 1, ..., m\},$$

où les g_i sont des fonctions de classe $C^1(\mathbb{R}^n)$.

Définition 1.3.2 *(contrainte saturée)*

Soit $g_i(x) \leq 0$ une contrainte d' inégalité et soit $x^\star \in D$.

Si $g_i(x^\star) < 0$, on dit que la contrainte est inactive en $x^\star$

Si $g_i(x^\star) = 0$, on dit que la contrainte est active ou saturée en $x^\star$.

Notons $I(x^\star)$ l'ensemble des indices des contraintes saturées en $x^\star$ c'est-à-dire :

$$I(x^\star) = \{i \in I = \{1, ..., m\} \ / \ g_i(x^\star) = 0\} \subset I$$

Définition 1.3.3 *(qualification des contraintes)*

Si les gradients $\{\nabla g_i(x^\star) \ / \ i \in I(x^\star)\}$ sont linéairement indépendants (libres), alors on dit que les contraintes du problème (P) sont qualifiées au point $x^\star$.

Dans ce cas $x^\star$ est dit point régulier du problème (P).

Remarque 1.3.4 .

***1**- Si les fonctions g_i sont affines $\forall i$, alors D est un polyèdre, c'est-à-dire*

$$D = \{x \in \mathbb{R}^n / Ax \leq b\}.$$

Dans ce cas les contraintes sont qualifiées en tout point $x^\star \in D$, i.e. $\forall x \in D$, x est regulier

***2**- D'une façon générale, si D est convexe alors la condition de qualification des contraintes est indépendante de $x^\star$, elle est exprimé par $intD \neq \phi$, c'est à dire*

$(\exists x^\circ \in D \ / g_i(x^\circ) < 0, \forall i = 1, ..., m) \implies$ l'ensemble D est qualifié en tout point $x \in D$.

Cette condition est connue par la condition de Slater.

Exemple 1.3.5 .

Dans $\mathbb{R}^2$, soit le problème :

$$(P) \begin{cases} \min -x_2 \\ x_1^2 + x_2^2 - 1 \leq 0 \end{cases}, \ \text{où } f(x) = f(x_1, x_2) = -x_2 \in C^\infty(\mathbb{R}^2)$$

On a l'ensemble des contraintes

$$D = \left\{x \in \mathbb{R}^2 / x_1^2 + x_2^2 \leq 1\right\} = B(0, 1),$$

on sait que $x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in D$, *on a* $g_1(x) = x_1^2 + x_2^2 - 1 \Rightarrow g_1(x^*) = g_1(0,1) = 0^2 + 1^2 - 1 = 0 \Rightarrow$ *la contrainte* $g_1(x) \leq 0$ *est saturée en* x^*

On a $\nabla g_1(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \Rightarrow \nabla g_1(x^*) = \nabla g_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \nabla g_1(x^*)$ *est libre*

Donc l'ensemble des contraintes D est qualifié au point $x^* \Rightarrow x^*$ *est un point régulier pour le probléme* (P)

Théorème 1.3.6 *(Kuhn –Tucker)*
Soient f *et* g_i, $\forall i \in I = \{1, ..., m\}$ *des fonctions de classe* $C^1(\mathbb{R}^n)$ *et soit l'ensemble des contraintes*

$$D = \{x \in \mathbb{R}^n \ / \ g_i(x) \leq 0, \forall i \in I\}$$

On suppose que l'ensemble D est qualifié au point $x^\star$ *(* $x^\star$*est régulier), alors :*
($x^\star$*est un minimum de f sur D)* $\Longrightarrow$ *il existe des nombres réels positifs* $\lambda_1, \lambda_2, ..., \lambda_n$ *appelés multiplicateurs de Kuhn-Tucker, tels que*

$$\begin{cases} \nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\star) = 0 \\ \lambda_i g_i(x^\star) = 0 \ , \ \forall i \in 1, ..., m \end{cases}$$

Remarque 1.3.7 .

Si de plus f et les $g_i, \forall \ i = \{1, ..., m\}$ *sont convexes, alors on a une condition nécessaire et suffisante (i.e.* $x^\star$*est un minimum de f sur D)* $\Longleftrightarrow \exists \lambda_i \geq 0, \forall i = 1, ..., m$ *tels que*

$$\begin{cases} \nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\star) = 0 \\ \lambda_i g_i(x^\star) = 0 \ , \forall i = 1, ..., m \end{cases}$$

1.3.2 *Cas des contraintes d'égalité*

On suppose que l'ensemble des contraintes et du type égalité,

$$D = \{x \in \mathbb{R}^n \ / h_j(x) = 0, \forall j = 1, ..., l\}.$$

Puisque une contrainte d'égalité est équivalente à deux contraintes d'inégalité à savoir

$h_j(x) \leq 0$ et $h_j(x) \geq 0$,on a :

$$h_j(x) = 0 \iff \begin{cases} h_j(x) \leq 0 \\ -h_j(x) \leq 0 \end{cases}$$

Nous allons pouvoir se ramener au cas précédent, ce qu'il faut retenir dans ce cas est que :

- Les contraintes sont forcément saturées (évident).
- On a la même notion de contraintes qualifiées
- Lorsque l'on écrit la condition de Kuhn–Tucker pour les contraintes d'égalités an point x^* nous allons avoir :,

$$\nabla f(x^\star) + \sum_{i=1}^{l}(\alpha_i \nabla h_i(x^\star) - \beta_i \nabla h_i(x^\star)) = 0, \quad \alpha_i \geq 0, \beta_i \geq 0, \forall i = 1, ..., l$$
$$\Rightarrow \nabla f(x^\star) + \sum_{i=1}^{l}(\alpha_i - \beta_i)\nabla h_i(x^\star) = 0.$$

.On pose :

$$\mu_i = \alpha_i - \beta_i, \mu_i \in \mathbb{R}, \forall i = 1, ..., l.$$

Donc on obtient :

Si $x^\star$est un minimum de f sur D, alors $\exists \mu_i \in \mathbb{R}, \forall i = 1, ..., l$ tels que :

$$\nabla f(x^\star) + \sum_{i=1}^{l} \mu_i \nabla h_i(x^\star) = 0,$$

cette condition est dite condition de Lagrange.

Les multiplicateurs μ_i ne vérifient pas la condition de signe, ils s'appellent multiplicateurs de Lagrange.

1.3.3 *Cas des contraintes d'égalité et d'inégalité*

Dans ce cas général, on suppose qu'on a le problème suivant :

$$(P) \begin{cases} \min f(x) \\ x \in D \end{cases}$$

où

$$D = \{x \in \mathbb{R}^n \ / h_j(x) = 0, \forall j = 1, ..., l \text{ et } g_i(x) \leq 0, \forall i = 1, ..., m\}.$$

Théorème 1.3.8 *(Kuhn-Tucker-Lagrange)*
Soit f et les $g_i, \forall i = 1, ..., m$ et $h_j, \forall j = 1, ..., l$ des fonction de classe $C^1(\mathbb{R}^n)$, on veut

minimiser f sur D. On suppose que les contraintes sont qualifiées au point $x^\star \in D$, alors on a

($x^\star$ est minimum de f sur D) $\Rightarrow$ ($\exists \lambda_i \geq 0, i = 1, ..., m$ et $\exists \mu_j \in \mathbb{R}, \forall j = 1, ..., l$) tels que :

$$\begin{cases} \nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\star) + \sum_{j=1}^{l} \mu_j \nabla h_j = 0 \\ \lambda_i g_i(x^\star) = 0 \ , \forall i = 1, ..., m \end{cases}$$

1.3.4 *Conditions suffisantes d'optimalité*

Nons allons voir maintenant des conditions suffisantes d'optimalité pour des problèmes du type :

$$(P) \begin{cases} \min f(x) \\ g_i(x) \leq 0, \forall i = 1, ..., m \end{cases}$$

où f et les g_i sont des fonctions de classe $C^1(\mathbb{R}^n)$

Définition 1.3.9 *(fonction de Lagrange, point selle)*

***1**- La fonction*

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \text{ où } \lambda = (\lambda_1, \lambda_2, ..., \lambda_m)^T \in \mathbb{R}^m_+, \ (\lambda_i \geq 0), \text{ et } x \in \mathbb{R}^n$$

est dite fonction de Lagrange (ou Lagrangien) associée au problème (P).

***2**- On appelle point col (ou point selle) de la fonction de Lagrange $L(x, \lambda)$ sur l'ensemble $\mathbb{R}^n \times \mathbb{R}^m_+$, tout point $(x^\star, \lambda^\star) \in \mathbb{R}^n \times \mathbb{R}^m_+$ qui vérifie :*

$$L(x^\star, \lambda) \leq L(x^\star, \lambda^\star) \leq L(x, \lambda^\star), \forall \lambda \in \mathbb{R}^m_+, \forall x \in \mathbb{R}^n$$

On a $x^\star$ minimise $L(x, \lambda^*)$ sur $\mathbb{R}^n$ et λ^* minimise $L(x^\star, \lambda)$ sur $\mathbb{R}^m_+$
C'est-à-dire

$$L(x^\star, \lambda^\star) = \min_{\mathbb{R}^n} L(x, \lambda^\star) \text{ et } L(x^\star, \lambda^\star) = \max_{\lambda \in \mathbb{R}^m_+} L(x^\star, \lambda)$$

Remarque 1.3.10 .

Si le problème (P) est donné sous la forme :

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

où f , g_i et h_j sont des fonctions de classe $C^1(\mathbb{R}^n)$, alors la fonction de Lagrange associée à (P) est :

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

où $\lambda = (\lambda_1, \lambda_2, ..., \lambda_m)^T \in \mathbb{R}_+^m$ et $\mu = \left(\mu_1, \mu_{2,...,}\mu_1\right)^T \in \mathbb{R}^l$, $x \in \mathbb{R}^n$

Théorème 1.3.11 *(propriété caractéristique des points selles)*

Le point $(x^\star, \lambda^\star) \in \mathbb{R}^n \times \mathbb{R}_+^m$ est un point selle de $L(x,\lambda)$ si et seulement si les conditions suivantes sont vérifiées :

1- $L(x^\star, \lambda^\star) = \min_{x \in \mathbb{R}^n} L(x, \lambda^\star)$

2- $g_i(x) \leq 0, \forall i = 1, ..., m$

3- $\lambda_i^\star g_i(x^\star) = 0$, $\forall i = 1, ..., m$

Théorème 1.3.12 *(suffisance de la condition du point selle)*
Soit le problème :

$$(P) \begin{cases} \min f(x) \\ x \in D \end{cases} \quad \text{où } D = \{x \in \mathbb{R}^n \ / g_i(x) \leq 0, \forall i = 1, ..., m\}$$

Si $(x^\star, \lambda^\star)$ est un point selle de $L(x,\lambda)$ alors $x^\star$ est un minimum global de f sur D.

Remarque 1.3.13 .

1- *Ce résultat est vrai pour un problème quelconque (convexe, non convexe, différentiable, non différentiable,...) mais l'éxistence des points selles n'est pas toujours sûr (en particulier en l'absence de la continuité).*

2- *Dans le cas d'un problème convexe (f et $g_i$convexes), la condition de Slater ($intD \neq \phi$) est nécessaire pour assurer l'existence des points selles et dans ce cas on a le théorème suivant :*

Théorème 1.3.14 .

Soientt f et $g_i, \forall i = 1, ..., m$ des fonctions convexes de classe C^1et soit l'ensemble des contraintes

$$D = \{x \in \mathbb{R}^n \ / g_i(x) \leq 0, \forall i = 1, ..., m\}$$

qualifié en $x^\star$. Alors

$$\begin{aligned} &(x^\star \text{ est minimum global de } f \text{ sur } D) \\ &\Leftrightarrow (\exists \lambda_i^\star \in \mathbb{R}_+^m \ / (x^\star, \lambda^\star) \text{ est un point selle de } L(x,\lambda) \text{ sur } \mathbb{R}^n \times \mathbb{R}_+^m) \\ &\Longleftrightarrow \begin{cases} \nabla_x L(x^\star, \lambda^\star) = 0 \\ \lambda_i^\star g_i(x^\star) = 0 \ , \forall i = 1, ..., m \end{cases} \end{aligned}$$

où

$$\nabla_x L(x^\star, \lambda^\star) = \frac{\partial L}{\partial x}(x^\star, \lambda^\star) = \nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla g_i(x^\star)$$

1.3.5 *Conditions d'optimalité nécessaires du second ordre*

Les résultats énoncés jusqu'à présent sont des conditions nécessaires pour résoudre le problème de minimisation avec contraintes, ils fournissent des candidats valables pour résoudre (P),

i.e. les points critiques de la fonction de Lagrange, où le problème est du type :

$$(P) \begin{cases} \min f(x) \\ g_i(x) \leq 0, \forall i = 1, ..., m \end{cases}$$

Toute fois il est nécessaire pour pouvoir conclure de savoir si la solution obtenue est effectivement un minimum.

Pour cela on peut grâce à une condition du second ordre restreindre le nombre de condidats, c'est l'objet du théorème suivant :

Théorème 1.3.15 .

On suppose que f et $g_i, \forall\ i = 1, ..., m$ sont de classe C^2 et que $x^\star$est un minimum (local) de f sur D et que $x^\star$ est régulier, alors $\exists \lambda_i \geq 0, \forall i = 1, .., m$ tels que les conditions de Kuhn-Tucker soient satisfaites et pour toute direction $d \in \mathbb{R}^n$ vérifiant :

$$\begin{cases} \langle \nabla g_i(x^\star), d \rangle = 0, i \in I_0^+(x^\star) \\ \langle \nabla g_i(x^\star), d \rangle \leq 0, i \in I_0(x^\star) \setminus I_0^+(x^\star) \end{cases}$$

où $I_0^+(x^\star) = \{1 \leq i \leq m \ / \ g_i(x^\star) = 0 \ et \ \lambda_i > 0\}$

on a

$$\langle \ \nabla^2_{xx} L(x^\star, \lambda^\star), d \ \rangle \geq 0$$

où $\nabla^2_{xx} L(x^\star, \lambda^\star) = \frac{\partial^2 L}{\partial x^2}(x^\star, \lambda^\star)$.

1.4 *Notion de dualité*

Définition 1.4.1 *(problème dual)*

Soit le problème de minimisation :

$$(P) \begin{cases} \min f(x) \\ x \in D \end{cases}$$

où $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ *de classe* C^1 *et* $D \in \mathbb{R}^n$ *l'ensemble des contraintes, et soit* $L(x, \lambda)$ *la fonction de Lagrange associée à* (P) *.*

Posons

$$q(\lambda) = \min L(x, \lambda) \ / \lambda \in \mathbb{R}_+^m \ \text{et} \ x \in D$$

alors le problème de maximisation

$$(C) \begin{cases} \max q(\lambda) \\ \lambda \in \mathbb{R}_+^m \end{cases} \Longleftrightarrow \begin{cases} \max \min L(x, \lambda) \\ \lambda \in \mathbb{R}_+^m, \ x \in D \end{cases}$$

est appelé problème "dual" de (P) *et* λ *est dite "variable duale " de* x *avec* (P) *problème primal.*

Remarque 1.4.2 .

1*- Par fois le problème dual est plus facile à résoudre que le problème primal.*

2*- Si* $(x^\star, \lambda^\star)$ *est un point selle de* $L(x, \lambda)$*. Alors*

$$f(x^\star) = L(x^\star, \lambda^\star) = \max_{\lambda \in \mathbb{R}_+^m} \min_{x \in D} L(x, \lambda) = \min_{x \in D} \max_{\lambda \in \mathbb{R}_+^m} L(x, \lambda)$$

3*- Si le problème* (P) *est un problème convexe (i.e.* f *et* g_i *sont convexes* $\forall i$ *) et si* D *est qualifié (i.e.* $intD \neq \phi$ *car* D *est convexe), et* (P) *admet au moins une solution optimale* $x^\star$*, alors*

$$\min f(x) = \max_{\lambda \in \mathbb{R}_+^m} \min_{x \in D} L(x, \lambda) = \min_{x \in D} \max_{\lambda \in \mathbb{R}_+^m} L(x, \lambda) .$$

Chapitre 2

Méthode du gradient réduit

Dans ce chapitre on va étudier quelques méthodes pour résoudre des problèmes d'optimisation avec contraintes

La majorité des méthodes existentes pour ce genre de problème appartient à deux classes, celle des méthodes primales qui opérent directement sur le problème donné (problème primal), et celle des méthodes duales qui utilisent la notion de dualité, ces méthodes font appel à la fonction de Lagrange

Avant de donner la méthode du gradient projeté on a la résultat suivant :

2.1 *Projection sur un convexe fermé*

Proposition 2.1.1 .

Soit D un convexe fermé non vide de $\mathbb{R}^n$, alors pour tout $x \in \mathbb{R}^n$ il existe un unique $x^\circ \in D$ tel que$\|x - x^0\| \leq \|x - y\|, \forall y \in D$, c'est à dire

$$\forall x \in \mathbb{R}^n, \exists! x^0 \in D / \|x - x^0\| \leq \|x - y\|, \forall y \in D$$

Remarque 2.1.2 .

$$\begin{aligned} \forall x \quad &\in \quad \mathbb{R}^n, \exists! x^0 \in D / \|x - x^0\| \leq \|x - y\|, \forall y \in D \\ &\Leftrightarrow \quad \forall x \in \mathbb{R}^n, \exists! x^0 \in D \ / \ x^0 \text{ est solution optimale duproblème ave ccontraintes :} \\ &\quad \begin{cases} \min \|x - y\| \\ y \in D \end{cases} \end{aligned}$$

On note $x^0 = P_D(x)$ la projection orthogonale de x sur D.

On a également : $(x^0 = P_D(x)) \iff (\langle x - x^0, x^0 - y \rangle \geq 0, \forall y \in D)$.

P_D est appelé operateur de projection sur D

$$P_D : \mathbb{R}^n \to D \subset \mathbb{R}^n \quad x \to P_D(x) = x^\circ$$

Exemple 2.1.3 *(quelques operateurs de projection)*

***1**- Soit l'ensemble D donné par :*

$$D = \mathbb{R}^n_+ = \{x \in \mathbb{R}^n / \ x_i \geq 0, \forall i = 1, ..., m\},$$

D est un convexe fermé non vide de $\mathbb{R}^n$, alors $\forall y \in \mathbb{R}^n, \exists! y^0 \in D$, tel que

$$P_D(y) = y^0 = \left(y^0_{1,,} y^0_2, ..., y^n_n\right) \ \ où \ \ y^0_i = \max\left(y_{i,0}\right), \forall i = 1, ..., m$$

***2**- Soit l'ensemble D donné par :*

$$\begin{aligned} D &= \left\{x \in \mathbb{R}^n / \ \alpha_i \leq x_i \leq \beta_{i,} \ \forall i = 1, ..., m \ \ / \ \alpha_i < \beta_i\right\} \\ &= [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times ... \times [\alpha_n, \beta_n] \\ &= \prod_{i=1}^{n} [\alpha_i, \beta_i] \end{aligned}$$

D est un convexe fermé non vide de $\mathbb{R}^n$, alors $\forall y \in \mathbb{R}^n, \exists! y^\circ \in D$, tel que :

$$P_D(y) = y^0 = \left(y^0_{1,,} y^0_2, ..., y^0_n\right) \ où \ y^0_i = \max\left(\alpha_i, \ \min \ \left(y_i, \beta_i\right)\right), \forall i = 1, ..., m$$

2.2 *Méthode du gradient projeté*

Considérons maintenant le problème d'optimisation suivant :

$$(P) \begin{cases} \min f(x) \\ x \in D \end{cases}$$

où f est une fonction de classe $C^1\left(\mathbb{R}^n\right)$, et l'ensemble des contraintes D est un convexe fermé non vide de $\mathbb{R}^n$.

Rappelons que dans le cas d'optimisation sans contraintes l'algorithme du gradient, qui est une méthode de descente, s'écrit sous la forme générique suivante :

$$\begin{cases} x^0 \text{ donné (point de départ)} \\ x^{k+1} = x^k + \rho_k d^k \end{cases}$$

où $d^k = -\nabla f(x^k)$ la direction de descente et ρ_k le pas de deplacement, sont choisis

tels que

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

Lorsque l'on minimise sur un ensemble de contraintes D, il n'est pas sûr que x^k reste dans D, donc il est nécessaire de se ramener à D. On réalise cette derniere opération grâce à une projection sur l'ensemble D.

L'opérateur de projection associé est noté $P_D : \mathbb{R}^n \underset{x \longmapsto P_D(x)}{\longrightarrow} D$

Ce ci donne lieu alors, naturellement, à l'algorithme du gradient projeté, qui est identique à celui du gradient à la projection près.

2.2.1 *Algorithme du gradient projeté*

Algorithm 2.2.1 *(gradient projeté)*

1- Initialisation :

Poser $k = 0$, *choisir* $x^0 \in D$, *donner* $\varepsilon > 0$ *(précision demandée), donner* ρ_0 *le pas de déplacement.*

2- Itération k *:*

Calculer

$$\begin{aligned} d^k &= -\nabla f(x^k) \\ \hat{x}^{k+1} &= x^k + \rho_k d^k \\ x^{k+1} &= P_D\left(\hat{x}^{k+1}\right) \end{aligned}$$

3- Critère d'arrêt :

Si $\|\nabla f(x^{k+1} - \nabla f(x^k))\| < \varepsilon$, *stop.*

Sinon, poser $k = k + 1$ *et retourener à l'étape* 2.

Remarque 2.2.2 .

1- *Le pas de déplacement* ρ_k *peut être choisit de plusieurs manières (pas fixe, pas variable ou pas optimal)*

2- *On peut choisir un autre critère d'arrêt, tel que*

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

3- *L'appelation la plus convenable pour cette méthode est la méthode du gradient avec projection car le gradient n'est pas projeté, mais c' est le point obtenu à chaque itération qui est projeté sur l'ensemble* D.

2.2.2 *Convergence de la méthode*

La convergence de la méthode du gradient projeté est assurée à partir du théorème suivant :

Théorème 2.2.3 *(convergence)*

Soit le problème

$$(P)\begin{cases} \min f(x) \\ x \in D \end{cases}$$

où f est une fonction du classe $C^1(\mathbb{R}^n)$ et D un ensemble convexe fermé non vide de $\mathbb{R}^n$. On suppose que :

1- $\exists \alpha > 0$ / $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|, \forall x, y \in \mathbb{R}^n$ *(i.e. f est α-élliptique).*

2- $\exists M > 0$ / $\|\langle \nabla f(x) - \nabla f(y) \rangle\| \leq M \|x - y\|, \forall x, y \in \mathbb{R}^n$ *(i.e. ∇ f est un fonction lipschitzienne).*

Alors :

i- *il existe un unique élément $x^\star \in D$ solution optimal du problème (P) (i.e. $x^\star$ minimum de f sur D).*

ii- *si $0 < \rho_k < \frac{2\alpha}{M}$, alors la suite $\left(x^k\right)$ générée par l'algorithme du gradient projeté converge vers $x^\star$.*

Remarque 2.2.4 .

Remarque 2.2.5 ***1***- *Si f est $\alpha-$elliptique, alors f est strictement convexe et coércive*

2- *L'algorithme du gradient projeté est simple à écrire mais il souléve deux questions*

- *Comment calculer la projection ?*
- *Si D n' est pas convexe, que faire ?*

Pour répendre à la deuxième question on utilise une autre méthode.

Pour répendre à la premier question, on peut donner des projections dans les cas simples tels que $D = \mathbb{R}^n_+$ ou

$$D = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times ... \times [\alpha_n, \beta_n] = \prod_{i=1}^{n} [\alpha_i, \beta_i], \alpha_{1i} < \beta_i$$

comme on l'a vu dans l'exemple 2.1.4

Si on a pas une projection sur D predéterminée, on va essayer de résoudre à chaque itération k le problème de minimisation avec contraintes

$$\begin{cases} \min \|x - x^k\| \\ x \in D \end{cases}$$

Ce qui est presque impossible à faire .

2.3 *Méthode du gradient réduit*

une autre méthode, applicable au cas des contraintes linéaires, est la méthode du gradient réduit due à *P. Wolf* (1962).

bien que très proche, au niveau du principe, des méthodes de gradient projeté , elle en diffère assez sensiblement au niveau de la mise en oeuvre et elle est généralement reconnue comme plus efficace .

Ceci s'explique, sans aucun doute, par le fait qu'elle constitue une extension directe de la méthode du simplexe.

Le problème à résoudre est supposé mis sous forme standard, c' est-à-dire avec des contraintes d'égalité seulement :

$$(P)\begin{cases} \min f(x) \\ Ax = b \\ x \geq 0 \end{cases}$$

où A est une $(m \times n)$ matrice, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$.

On suppose $m < n$ et $rang\,(A) = m$.

Soit B une base, c'est-à-dire une sous matrice carrée $(m \times m)$ réguliere extraite de A,

En posant $A = [B, N]$, et $x = [x_B,\, x_N]$, les contraintes $Ax = b$ peuvent s' écrire :

$$Bx_B + Nx_N = b$$

Ce qui permet d'exprimer les variables de base (x_B) en fonction des variables hors base (x_N) :

$$x_B = \overline{b} - \overline{N}x_N, (\text{ avec } \overline{b} = B^{-1}b, \overline{N} = B^{-1}N)$$

On suppose dans la suite que la base B est non dégénérée, c'est-à-dire telle que $\overline{b} > 0$ (et cette condition doit être vérifié de tout au long de la procédure).

La variation df de f pour un placement dx compatible avec les contraintes s'écrit :

$$df = \frac{\partial f}{\partial x_B}dx_B + \frac{\partial f}{\partial x_N}dx_N$$

avec $dx_B = -\,\overline{N}dx_N$

d'où :

$$\begin{aligned} df &= \left(\frac{\partial f}{\partial x_N} - \frac{\partial f}{\partial x_B}\overline{N}\right)dx_N \\ &= u_N.dx_N \end{aligned}$$

en posant :

$$u_N = \left(\frac{\partial f}{\partial x_N} - \frac{\partial f}{\partial x_B} \overline{N} \right)$$

Par définition u_N est appelé **gradient réduit** relativement à la base B.

A l'étape courante de l'algorithme, soit $x = [x_B,\, x_N]$ la solution réalisable obtenue.

Pour améliorer cette solution, en se déplace dans la direction $d = [d_B,\, d_N]$, où d_N est définie par :

$$\begin{cases} d_j = 0 \quad \text{si } u_j > 0 \text{ et } x_j = 0 \\ d_j = -u_j \quad \text{sinon} \end{cases} \quad \forall j$$

pour tous les indices j correspondants à des variables hors base et :

$$d_B = -\overline{N} d_N$$

Autrement dit , dans l'espace des variables hors base (variables indépendantes) on se déplace dans la direction opposé au gradient réduit et, si ce déplacement tend à rendre une variable x_j négative, on impose $d_j = 0$ (ceci n'est autre qu'une projection sur l'orthant positive dans l'espace des variables hors base).

Si $d_N = 0$, nous montrerons ci-dessous que les conditions de Kuhn-Tucker sont satisfaites.Si f est convexe cela signifie que l'optimum est atteint. Si f n' est pas convexe, on a un optimum local.

Si $d_N \neq 0$, on cherche $\rho^\star$ minimisant la fonction φ de ρ :

$$\varphi(\rho) = f(x + \rho d)$$

Comme les contraintes des positivité des variables doivent toujours être satisfaites, on doit avoir :

$$x_j + \rho_{\max} \geq 0, \ \forall j.$$

ce qui impose :

$$\rho \leq \rho_{\max} = \min_{d_j < 0} \left\{ \frac{-x_j}{d_j} \right\}$$

On cherchera donc $\rho^\star$ minimisant $\varphi(\rho)$ sur $[0, \rho_{\max}]$ (minimisation unidimensionnelle)

Deux situations peuvent alors se présenter :

1- $\rho^\star < \rho_{\max}$

Dans ce cas, aucune variable ne s'annule. La procédure ce poursuit à partir du nouveau point $x' = x + \rho^\star d$.

On recalculera u_N en utilisant la même base B.

2- $\rho^\star = \rho_{\max}$

Dans ce cas, une variable s'annule, soit x_s.

Si x_s est une variable hors bas, alors on poursuit la procédure sans changer de base. Par contre, si x_s est une variable de base, on effectue un changement de base : n'importe quelle variable hors base $x_t > 0$ peut venir remplacer x_s à condition que l'élément pivot $\overline{A}_{st}$ soit strictement positif.

La procédure ce poursuit alors avec cette nouvelle base.

Pour terminer, vériririons que la condition $d_N = 0$, implique les conditions de Kuhn-Tucker (puisque les contraintes sont linéaires, l'hypothèse de qualification est toujours vérifiée).

En associant aux contraintes $Ax = b$, des multiplicateurs $\mu \in \mathbb{R}^m$ (de signe quelconque) et aux contraintes $x \geq 0$, des multiplicateurs $\lambda \geq 0$ ($\lambda \in \mathbb{R}^n_+$), les conditions de Kuhn-Tucker.s'écrivent :

$$\begin{cases} -\frac{\partial f}{\partial x} + \mu A + \lambda I = 0 \ldots\ldots (1) \\ \lambda x = 0 \ldots\ldots\ldots\ldots\ldots (2) \end{cases}$$

où I est la matrice unité $(n \times n)$.

En posant :

$$\mu = \frac{\partial f}{\partial x_B} B^{-1}, \lambda = [\lambda_B \ , \ \lambda_N] \text{ avec}$$
$$\lambda_B = 0 \text{ et } \lambda_N = \left(\frac{\partial f}{\partial x_N} - \frac{\partial f}{\partial x_B} . \overline{N} \right) = u_N$$

la relation (1) est vérifiée.

Remarquons alors que $d_N = 0$, implique $u_N \geq 0$ d'où $\lambda_N \geq 0$ et par suite $\lambda \geq 0$.

D'autre part, comme pour tout j hors base : $d_j = 0$ on a nécessairement : $u_j > 0 \implies x_j = 0$, et on en déduit que les conditions de complémentairité (2) sont satisfaites.

Les conditions de Kuhn-Tucker.sont donc bien vérifiées lorsque $d_N = 0$.

La méthode du gradient réduit, sous sa forme primitive exposée ci- dessus, ne converge pas globalement.

Il est facile de voir, en effet, que la direction de déplacement d peut subir des discontinuités brusques par exemple lorsqu'une contrainte cesse d'être active, néanmoins, il est possible d'obtenir la convergence globale en modifiant la méthode et en définissant la direction de déplacement d_N par :

$$\begin{cases} d_j = -u_j & \text{si } u_j < 0 \\ d_j = -x_j . u_j & \text{si } u_j > 0 \end{cases} , \forall j$$

pour tous les indices j correspondant à des variables hors base.

2.4 *La méthode du gradient réduit généralisé (GRG)*

La méthode du gradient réduit généralisé (*Abadie* et *Carpentier* 1962 , *Abadie* et *Giugou* 1970) combine les idées des méthodes de linéarisation, et celles du gradient réduit de *Wolfe*.

La méthode a lengtemps été considérée comme une des plus efficaces pour traiter le cas général où les contraintes, comme la fonction à optimiser, sont non linéaires (on fera cependant l'hypothèse qu'elles sont continûment différentiable).

On peut toujours supposer le problème mis sous forme standard :

$$(P)\begin{cases} \min f(x) \\ g(x) = (g_1(x), ...g_m(x))^{\intercal} = 0 \\ x \geq 0 \end{cases}$$

on note $J = \{1, ..., m\}$ l'ensemble des indices des variables.

Soit x^0 une solution réalisable (solution courante) , c'est-à-dire vérifiant

$$g\left(x^0\right) = 0 \text{ et } x \geq 0$$

Un sous ensemble de variable $B \subset J$, où $|B| = m$, est appelé une base si et seulement si la matrice (jacobien) :

$$\frac{\partial g}{\partial x_B}\left(x^0\right) = \left[\frac{\partial g_i}{\partial x_j}\left(x^0\right)\right]_{\substack{i=1,....,m \\ j \in B}}$$

est inversible.

Le vecteure x peut alors se partitionner en $x = [x_B,\ x_N]$ où x_B sont les variables de base (variables dépendantes), et x_N sont les variables hors- base (variables dépendantes).

A partir de x^0, on va rechercher une direction de déplacement permettant de diminuer la fonction f tout en satisfaisant les contraintes.

La variation df de f, pour un déplacement infinitésimal $dx = [dx_B,\ dx_N]$ compatible avec les contraintes s'écrit :

$$\begin{aligned} df &= \left[\frac{\partial f}{\partial x_N} - \left(\frac{\partial f}{\partial x_B}\right)\left(\frac{\partial g}{\partial x_B}\right)^{-1}\left(\frac{\partial g}{\partial x_N}\right)\right] dxN \\ &= u_N.dx_N \end{aligned}$$

où le vecteur :

$$u_N = \frac{\partial f}{\partial x_N} - \left(\frac{\partial f}{\partial x_B}\right)\left(\frac{\partial g}{\partial x_B}\right)^{-1}\left(\frac{\partial g}{\partial x_N}\right)$$

est appelé le **gradient reduit généralisé** .

Tous les jacobiens sont calculés au point courant x^0, et ceci restera valable dans la suite.

On définit alors la direction de déplacement d_N relative aux variables indépendantes par :

$$\begin{cases} d_j = 0 \text{ si } \ u_j \text{ et } x_j = 0 \ (j \in \mathbb{N}) \\ d_j = -u_j \text{ sinon } (j \in \mathbb{N}) \end{cases}$$

La dérection de déplacement d_B suivant les variables de base se déduit alors de d_N par :

$$d_B = -\left(\frac{\partial g}{\partial x_B}\right)^{-1}\frac{\partial g}{\partial x_N}d_N$$

Si $d_N = 0$, on démontre que les conditions de Khun-Tucker sont satisfaites au point courant x^0.

Si $d_N \neq 0$, on cherche $\rho \geq 0$ minimisant la fonction ψ de ρ

$$\psi(\rho) = f(x^\circ + \rho d)$$

Comme les conditions de positivité des variables doivent toujours être satisaites, on doit avoir :

$$x_j + \rho d \geq 0 \quad (\forall j)$$

ce qui impose

$$\rho \leq \rho_{\max} = \min_{d_j<0}\left\{-\frac{x_j}{d_j}\right\}$$

On cherchera donc $\rho^\star$ minimisant $\psi(\rho)$ sur $[0, \rho_{\max}]$ (minimisation unidimentionnelle).

ρ étant ainsi déterminé, le point

$$\hat{x} = x^\circ + \rho^\star d$$

satisfait évidemment les contraintes de positivité $(\hat{x} \geq 0)$.

Mais comme le déplacement d_B des variables de base a été obtenu, à partir de d_N, par linéarisation des contraintes, $\hat{x}$ n'a en principe aucune raison de satisfaire les contraintes $g(x) = 0$.

Pour obtenir un point x^1 réalisable, à partir de $\hat{x}$, on applique la méthode de Newton

à la résolution du système non linéaire (on variable x_B) :

$$g\left(x_B,\hat{x}_N\right)=0.$$

où

$$\hat{x}_N = x_N^0 + \rho^\star d_N$$

reste fixé, et où

$$\hat{x}_B = x_B^0 + \rho^\star d_B$$

est le point de départ des itérations,

Ceci conduit à la formule itérative suivante :

$$x_B^{k+1} = x_B^k - \left[\frac{\partial g}{\partial x_B}(x^0)\right]^{-1} g\left(x_B^k, \hat{x}_N\right)$$

Notons qu'en principe, l'inverse du jacobien $\left[\frac{\partial g}{\partial x_B}\right]^{-1}$ devrait être évolué, à chaque itération k au point x_B^k.

Cependant, dans la plupart des cas, il suffira d'effectuer toujours ce calcul avec l'inverse du jacobien en x_B^0, il en résulte un gain considérable en temps de calcul.

La convergence de la méthode vers un point satisfaisant les conditions de Kuhn-Tucker, est très difficile à établir.

En fin, le principal avantage de la méthode GRG semble résider dans le fait qu'elle peut exploiter efficacement la faible densité des matrices (jacobiens) et permettre ainsi la résolution de certains problèmes de taille relativement importante (quelques dizaines de variables et de contraintes).

Conclusion générale

En conclusion, on peut dire que la méthode de graduit réduit et celle de graduit réduit généralisé sont parmis les méthodes les plus performantes pour la résolution des problèmes d'optimisation avec contraintes linéaires, mais elles sont un peu compliquées comparées à la méthode du gradient projeté.

La méthode du gradient réduit comporte des opérations analogues à celles de la méthode du simplexe, en particulier utilisation des bases et de changement de base, de ce fait elle peut être sujet à des phénomènes de dégénérescence, surtout que sa convergence (théorique) est très difficile à établir et il n'existe pas beaucoup de tests numériques sur cette méthode.

Bibliographie

[1] **P. G. Ciarlet**, *Introduction à l'analyse numérique matricielle et à l'optimisation,* Masson, Paris (1988).

[2] **H. M. Kharoubi,** *Sur quelques méthodes de gradient réduit sous contraintes linéaires,* RAIRO-Analyse numérique, Tome 13 N°2 (1979).

[3] **M. Minoux**, *Programmation mathématique, théorie et algorithme,*Tome 1, Dunod, Paris (1984).

[4] **A. Rondepierre, S. Tordeux**, *Introduction à l'optimisation numérique,* INSA Toulouse (2010).