



**République Algérienne Démocratique et
Populaire**

**Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique**

UNIVERSITÉ D'ELOUED

FACULTÉ DES SCIENCES ET DE TECHNOLOGIE

Mémoire de fin d'étude

LICENCE ACADEMIQUE

Domaine: Mathématiques et Informatique

Filière: Mathématiques

Spécialité: Modélisation mathématiques & simulation
numérique

Présenté par: Gori Khaled

Touil Ben Salem

Thème

Programmation Convexe

Sous l'encadrement de:

Miloudi Madjda MA(A)

Année universitaire : 2013 – 2014

Remerciements

La louange est à Allah, qui nous a facilité l'accomplissement de ce travail de recherche chose ne peut être qu'avec la volonté de Dieu -à lui la toute puissance et la Majesté- et que la louange initiale et finale appartient à Allah, Seigneur des mondes

Aussi, il nous fait plaisir que nous, au commencement de ce travail, présentons nos grands remerciements, estimations et reconnaissances à notre encadreur puissant "*MILOUDI Majda*" de nous avoir encourager moralement la durée de recherche et que ce travail est le fruit de ces encouragements

En fin, nous présentons nos veridiques remerciements à toute personne, du proche ou du loin, qui nous a donné un coup de main, à fin de terminer ce travail de recherche "Que la Grace et la paix soient sur notre prophète Muhammad ainsi que sur sa famille et ses compagnons".

Notations

$C(\mathbb{R}^N, \mathbb{R}) :$	l'ensemble des fonctions continues de $\mathbb{R}^N$ dans $\mathbb{R}$.
$C^1(E, F) :$	l'ensemble des fonctions continument différentiable d'ordre 1.
$C^2(E, F) :$	l'ensemble des fonctions continument différentiable d'ordre 2.
$\mathcal{L}(E, F) :$	l'ensemble des applications linéaires de E dans F .
$E, F :$	deux espaces vectoriels normés.
$T_x :$	la différentielle de f au point x .
$\nabla f(x) :$	le gradient de f .
$\partial f(x) :$	la dérivée partielle de f .
$H_f(x) :$	la matrice hessienne de f en x .
$J_f(x) :$	la matrice jacobienne de f en x .
$M_{p,N}(\mathbb{R}) :$	l'espace des matrices de p lignes et N colonnes.
$\ \cdot\ :$	la norme.
$ \cdot :$	la valeur absolue.
$\mathbb{k} :$	sous ensemble de $\mathbb{R}^N$.
$B(\bar{x}, \alpha) :$	la boule ouverte de centre $\bar{x}$ et de rayon α .
$\text{rang}(C) :$	rang de la matrice C .
$\text{card}(I^*(\bar{x})) :$	cardinal d'une partie $I^*(\bar{x})$.
$(\cdot, \cdot)_N :$	produit scalaire euclidien sur $\mathbb{R}^N$.
$\inf_{x \in K} f(x) :$	la borne inférieure de fonction f sur l'ensemble K .
$\sup_{x \in K} f(x) :$	la borne supérieure de fonction f sur l'ensemble K .
$\min :$	le minimum.
$\max :$	le maximum.

Table des matières

Introduction générale	1
1 Généralités sur optimisation	2
1.1 Définitions et rappels	2
1.1.1 Extremum d'une fonction	2
1.1.2 Applications différentiables	3
1.2 Convexité	6
1.2.1 Ensembles convexes	10
1.3 Optimisation sans contrainte	11
1.3.1 Définitions et conditions d'optimalité	11
1.3.2 Résultats d'existence et d'unicité	11
1.4 Optimisation avec contraintes	13
1.4.1 Existence, Unicité, et Conditions d'optimalité simple	14
1.4.2 Conditions d'optimalité dans le cas de contraintes égalité	15
1.4.3 Contraintes inégalités	17
2 Programmation convexe	19
2.1 Fonctions elliptiques, fonctions coercives	19
2.2 Application et exemples	21
2.2.1 Théorie générale du point selle	21
2.2.2 Applications de la théorie du point selle à l'optimisation	22
2.3 Le cas convexe	23

2.4	Algorithmes de minimisation avec contraintes	24
2.4.1	Méthodes de relaxation	25
2.4.2	Méthodes de projection	25
2.4.3	Méthodes de dualité	27
Bibliographie		31

Introduction générale

Nous étudions dans ce travail les problèmes de programmation convexe. Un problème de programmation convexe est un problème de minimisation d'une fonctionnelle convexe sur une partie convexe de $\mathbb{R}^N$.

D'une façon plus précise, dans le premier chapitre. On trouvera:

Généralités sur l'optimisation: La théorie qui concerne les problèmes d'optimisation sans et avec contraintes, les résultats d'existence et d'unicité de solutions de ces problèmes, caractérisation éventuelle d'une solution de ces problèmes, c'est-à-dire des conditions nécessaires et suffisantes dans certains cas.

Les conditions nécessaires font généralement intervenir la différentielle d'ordre 1 et 2. Pour la recherche des minimum locaux, on a des autres conditions nécessaires et suffisantes, on obtient aussi dans ce chapitre des conditions suffisantes dans les deux cas convenables, le cas des contraintes égalités, on a l'existence des multiplicateur de Lagrange, qui permet de déterminer le minimum d'une fonctionnelle sous ces contraintes. Aussi pour le cas des contraintes inégalités, On obtient une condition suffisante qui permet de déterminer le minimum d'une fonction sous ces contraintes (théorème de Karush-Tucker).

Dans le deuxième chapitre, nous considérons le problème de programmation non linéaire suivant:

$$(P) \begin{cases} \text{Trouver } \bar{x} \in K = \{x \in \mathbb{R}^N, g_i(x) \leq 0, 1 \leq i \leq p\} . \\ f(\bar{x}) = \inf_K f. \end{cases}$$

Tout d'abord, on donne quelques définitions et la théorie générale du point selle et l'application de la théorie du point selle à l'optimisation.

Pour commencer, nous établissons des conditions nécessaires et suffisantes en programmation convexe, pour qu'un élément de l'ensemble K soit solution du problème (P) : il s'agit des relations de Khun-Tucker (théorème 1.4.5), expriment l'existence de nombres λ_i vérifiant les conditions (1.4.1) et (1.4.5).

Au paragraphe 2.4.1, on obtient la méthode de relaxation qui permet d'un nombre fini des étapes. qui transforme le problème de minimisation d'une fonctionnelle sur un pavé de $\mathbb{R}^N$ à N problèmes de minimisation de cette fonctionnelle sur un intervalle de $\mathbb{R}$.

La construction effective d'algorithmes permettant d'approcher une solution $\bar{x}$ de problème (P) , c'est-à-dire, construction d'une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de l'ensemble K tels que $\lim_{n \rightarrow +\infty} x_n = \bar{x}$, le vecteur x_0 étant choisi arbitrairement, et on notera que, dans tous cas, la définition et l'étude de ces algorithmes utilisent de façon essentielle la caractérisation des solutions, que nous rappellerons à chacune de leurs utilisations.

Au paragraphe 2.4.2, on obtient des méthodes de projection sont des méthodes numériques qui utilisent l'opérateur de projection sur un convexe fermé. Ces algorithmes sont simples à décrire, mais ils soulèvent une question de calcul de la projection sur un convexe fermé.

Les problèmes sans contraintes "générales" sont beaucoup plus difficiles à traiter que les problèmes sans contraintes. D'ailleurs, on s'efforce souvent de les résoudre en les remplaçant par une suite de problèmes sans contraintes, ou avec contraintes faciles à prendre en compte (par exemple $K = \mathbb{R}_+^N$). Cette idée est à la base de méthode du dualité, étudiée au paragraphe 2.4.3 qui utilise la propriété du point selle de le Lagrangien et la solution du problème dual.

Bien que problème dual soit encore un problème d'optimisation avec contraintes, il est essentiel de noter que celle-ci sont faciles à prendre en compte numériquement, puisque l'opérateur de projection de $\mathbb{R}^N$ sur $\mathbb{R}_+^N$ est de forme très simple. Cette observation suggère l'application de la méthode du gradient à pas fixe avec projection ou problème dual. C'est le principe de la méthode d'Uzawa, qui existe au dernier paragraphe.

Chapitre 1

Généralités sur optimisation

L'objectif de ce chapitre est de rechercher des minima ou des maxima d'une fonction continue avec ou sans contrainte.

Pour commencer l'étude, on passe aux principales définitions et les résultats fondamentaux du calcul différentiel dans les espaces vectoriels normés.

Après avoir rappelé des conditions nécessaires et suffisantes d'extremum relatif ou local.

On donne aussi quelques résultats d'existence et d'unicité d'extremum.

1.1 Définitions et rappels

1.1.1 Extremum d'une fonction

Soit E un espace vectoriel et $f : E \rightarrow \mathbb{R}$. On dit que $\bar{x}$ est un minimum local de f s'il existe un voisinage V de $\bar{x}$ tel que:

$$f(\bar{x}) \leq f(x), \forall x \in V. \quad (1.1.1)$$

De même, on dit que $\bar{x}$ est un maximum local de f s'il existe un voisinage V de $\bar{x}$ tel que:

$$f(x) \leq f(\bar{x}), \forall x \in V.$$

On dit que $\bar{x}$ est un minimum global de f si:

$$f(x) \geq f(\bar{x}), \forall x \in E.$$

De même, on dit que $\bar{x}$ est un maximum global de f si:

$$f(x) \leq f(\bar{x}), \forall x \in E.$$

On dit que $\bar{x}$ est un extremum local (resp. global) de f si c'est un minimum local (resp. global) ou un maximum local (resp. global).

1.1.2 Applications différentiables

Définition 1.1.1 Soient E et F deux espaces vectoriels normés, f une application de E dans F et $x \in E$. On rappelle que f est différentiable en x s'il existe $T_x \in L(E, F)$ (où $L(E, F)$ est l'ensemble des applications linéaires de E dans F) telle que:

$$f(x+h) = f(x) + T_x(h) + \|h\| \varepsilon(h) \text{ avec } \varepsilon(h) \rightarrow 0 \text{ quand } \|h\| \rightarrow 0. \quad (1.1.2)$$

L'application T_x est alors unique et on note $Df(x) = T_x \in L(E, F)$ la différentielle de f au point x . Si f est différentiable en tout point de E , alors on appelle la différentielle de f l'application $Df : E \rightarrow L(E, F)$ qui à $x \in E$ associe l'application linéaire $Df(x)$ de E dans F .

Remarquons tout de suite que si f est une application linéaire de E dans F , alors f est différentiable, et $Df = f$. En effet si f est linéaire, $f(x+h) - f(x) = f(h)$, et donc l'égalité (1.1.2) est vérifiée avec $T_x = f$ et $\varepsilon = 0$. Voyons maintenant quelques cas particuliers d'espaces E et F .

Cas où $E = \mathbb{R}$ et $F = \mathbb{R}$. Si f est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, dire que f est différentiable en x revient à dire que f est dérivable en x . En effet, dire que f est dérivable en x revient à dire que: $\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$ existe, et $\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = f'(x)$, ce qui s'écrit encore:

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \varepsilon(h), \text{ avec } \varepsilon(h) \rightarrow 0 \text{ lorsque } h \rightarrow 0,$$

c'est-à-dire:

$$f(x+h) - f(x) = T_x(h) + h\varepsilon(h) \text{ avec } T_x(h) = f'(x)h$$

ce qui revient à dire que f est différentiable en x , et que sa différentielle en x est l'application linéaire $T_x: \mathbb{R} \rightarrow \mathbb{R}$, qui à h associe $f'(x)h$. On a ainsi vérifié que pour une fonction de $\mathbb{R}$ dans $\mathbb{R}$, la notion de différentielle coïncide avec celle de dérivée.

Exemple 1.1.1 Prenons $f: \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = \sin x$. Alors f est évidemment dérivable en tout point et sa dérivée vaut $f'(x) = \cos x$. La différentielle de f au point x est l'application linéaire $Df(x)$ qui à $h \in \mathbb{R}$ associe $Df(x)(h) = \cos x h$. La différentielle de f est l'application de $\mathbb{R}$ dans $L(\mathbb{R}, \mathbb{R})$, qui à x associe $Df(x)$.

Cas où $E = \mathbb{R}^N$ et $F = \mathbb{R}^P$. Soit $f: \mathbb{R}^N \rightarrow \mathbb{R}^P$, $x \in \mathbb{R}^N$ et supposons que f est différentiable en x ; alors $Df(x) \in L(\mathbb{R}^N, \mathbb{R}^P)$ et il existe une unique matrice $J_f(x) \in M_{P,N}(\mathbb{R})$ telle que:

$$Df(x)(y)_{\in \mathbb{R}^P} = J_f(x)y_{\in \mathbb{R}^P}, \forall y \in \mathbb{R}^N.$$

On confond alors l'application linéaire $Df(x) \in L(\mathbb{R}^N, \mathbb{R}^P)$ avec la matrice $J_f(x) \in M_{P,N}(\mathbb{R})$ qui la représente. On écrit donc:

$$J_f(x) = Df(x) = (a_{i,j})_{1 \leq i \leq P, 1 \leq j \leq N},$$

où $a_{i,j} = \partial_j f_i(x)$, ∂_j désignant la dérivée partielle par rapport à la j -ème variable. La matrice $J_f(x)$ s'appelle la matrice jacobienne de f au point x (d'où la notation J_f). Notons que si $N = P = 1$, la fonction f est de $\mathbb{R}$ dans $\mathbb{R}$ et la matrice jacobienne en x n'est autre que la dérivée en x : $J_f(x) = f'(x)$.

Exemple 1.1.2 Prenons $N = 3$ et $P = 2$; soit $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ définie par:

$$f(x) = \begin{pmatrix} x_1^2 + x_2^3 + x_3^4 \\ 2x_1 + x_2 \end{pmatrix}, \forall x = (x_1, x_2, x_3)^T \in \mathbb{R}^3.$$

Soit $h \in \mathbb{R}^3$ de composantes (h_1, h_2, h_3) . Pour calculer la différentielle de f (en x appliquée à h), on peut calculer:

$$\begin{aligned} f(x+h) - f(x) &= \begin{pmatrix} (x_1+h_1)^2 - x_1^2 + (x_2+h_2)^3 - x_2^3 + (x_3+h_3)^4 - x_3^4 \\ 2(x_1+h_1) - 2x_1 - 2(x_2+h_2) + 2x_2 \end{pmatrix} \\ &= \begin{pmatrix} 2x_1h_1 + h_1^2 + 3x_2^2h_2 + 3x_2h_2^2 + h_2^3 + 4x_3^3h_3 + 4x_3^2h_3^2 + h_3^4 \\ 2h_1 - 2h_2 \end{pmatrix}, \end{aligned}$$

et on peut ainsi vérifier l'égalité (1.1.2), avec:

$$Df(x)h = \begin{pmatrix} 2x_1h_1 + 3x_2^2h_2 + 4x_3^3h_3 \\ 2h_1 + h_2 \end{pmatrix},$$

et donc, avec les notations précédentes:

$$J_f(x) = \begin{pmatrix} 2x_1 & 3x_2^2 & 4x_3^3 \\ 2 & 1 & 0 \end{pmatrix}.$$

Cas où $E = \mathbb{R}^N$, $F = \mathbb{R}$. C'est en fait un sous-cas du paragraphe précédent, puisqu'on est ici dans le cas $p = 1$. Soit $x \in \mathbb{R}^N$ et f une fonction de E dans F différentiable en x ; on a donc $J_f(x) \in M_{1,N}(\mathbb{R})$, et on définit le gradient de f en x par $\nabla f(x) = (J_f(x))^T \in \mathbb{R}^N$. Pour $(x, y) \in (\mathbb{R}^N)^2$, on a donc:

$$Df(x)(y) = J_f(x)y = \sum_{j=1}^N \partial_j f(x) y_j = \nabla f(x) \cdot y,$$

où

$$\nabla f(x) = \begin{bmatrix} \partial_1 f(x) \\ \vdots \\ \partial_N f(x) \end{bmatrix} \in \mathbb{R}^N.$$

Voir [5].

Différentielle d'ordre 2, matrice hessienne

Revenons maintenant au cas général de deux espaces vectoriels normés E et F , supposons maintenant que $f \in C^2(E, F)$. Le fait que $f \in C^2(E, F)$ signifie que $Df \in C^1(E, L(E, F))$. Par définition, on a $D^2f(x) \in L(E, L(E, F))$ et donc pour $y \in E$, $D^2f(x)(y) \in L(E, F)$; en particulier, pour $z \in E$, $D^2f(x)(y)(z) \in F$. Considérons maintenant le cas particulier $E = \mathbb{R}^N$ et $F = \mathbb{R}$. On a:

$$f \in C^2(\mathbb{R}^N, \mathbb{R}) \Leftrightarrow [f \in C^1(\mathbb{R}^N, \mathbb{R})$$

et $\nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)]$. Soit $g = \nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)$, et $x \in \mathbb{R}^N$, alors, en confondant l'application linéaire $Dg(x)$ avec la matrice qui la représente, on a $Dg(x) \in M_N(\mathbb{R})$, et on peut définir la matrice hessienne de f , qu'on note H_f , par:

$$H_f(x) = Dg(x) = D(Df)(x) = (b_{i,j})_{i,j=1,\dots,N} \in M_N(\mathbb{R})$$

où $b_{i,j} = \partial_{i,j}^2 f(x)$ où $\partial_{i,j}^2$ désigne la dérivée partielle par rapport à la variable i de la dérivée partielle par rapport à la variable j . Notons que par définition (toujours avec l'abus de notation qui consiste à confondre l'application linéaire $Dg(x)$ avec la matrice qui la représente), $Dg(x)$ est la matrice jacobienne de $g = \nabla f$ en x .

Exemple 1.1.3 Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par:

$$f(x_1, x_2) = x_1^2 x_2^2 + x_1^2 + x_2^2 + 2ax_1 x_2 \quad (a \geq 0),$$

on a:

$$Df(x_1, x_2) = (\nabla f(x))^T = (2x_1 x_2^2 + 2x_1 + 2ax_2, 2x_1^2 x_2 + 2x_2 + 2ax_1),$$

alors:

$$D^2 f(x_1, x_2) = H_f(x_1, x_2) = \begin{pmatrix} 2x_2^2 + 2 & 4x_1 x_2 + 2a \\ 4x_1 x_2 + 2a & 2x_1^2 + 2 \end{pmatrix}.$$

Voir [7].

1.2 Convexité

Définition 1.2.1 (Convexité) Soit E un espace vectoriel et $f : E \rightarrow \mathbb{R}$. On dit que f est convexe si:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \text{ pour tout } (x, y) \in E^2 \text{ et } t \in [0, 1].$$

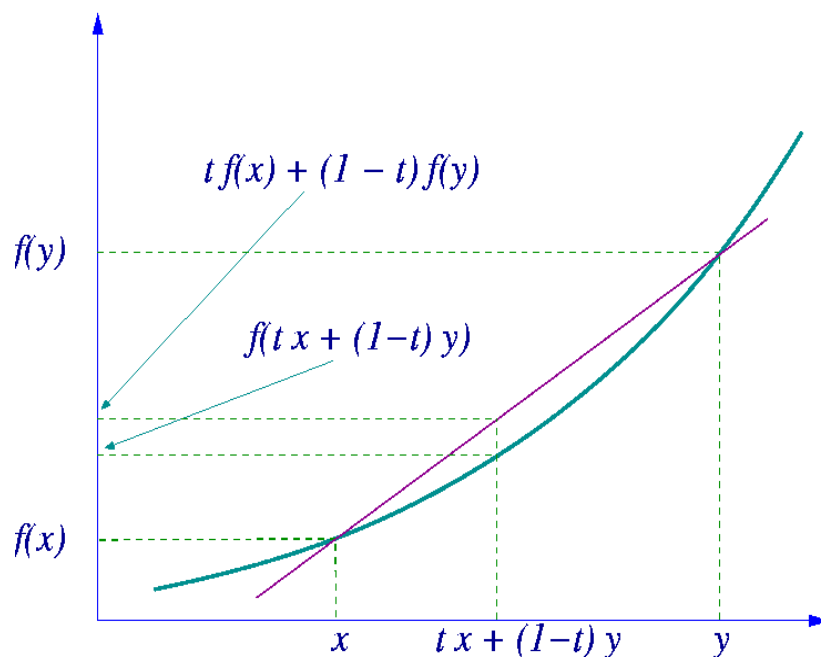
On dit que f est strictement convexe si:

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) \text{ pour tout } (x, y) \in E^2$$

t.q: $x \neq y$ et $t \in]0, 1[$.

Proposition 1.2.1 (1ère caractérisation de la convexité). Soit E un espace vectoriel normé (sur $\mathbb{R}$) et $f \in C^1(E, \mathbb{R})$ alors :

1. f convexe si et seulement si $f(y) \geq f(x) + Df(x)(y-x)$, pour tout couple $(x, y) \in E^2$.
2. f est strictement convexe si et seulement si $f(y) > f(x) + Df(x)(y-x)$ pour tout couple $(x, y) \in E^2$ telque $x \neq y$.



Preuve. 1. ($\Rightarrow$) Supposons que f est convexe: soit $(x, y) \in E^2$; on veut montrer que $f(y) \geq f(x) + Df(x)(y - x)$. Soit $t \in [0, 1]$, alors $f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x)$ grâce au fait que f est convexe. On a donc

$$f(x + t(y - x)) - f(x) \leq t(f(y) - f(x)) \quad (1.2.1)$$

Comme f est différentiable, $f(x + t(y - x)) = f(x) + Df(x)(t(y - x)) + t\varepsilon(t)$ où $\varepsilon(t)$ tend vers 0 lorsque t tend vers 0. Donc en reportant dans (1.2.1),

$$\varepsilon(t) + Df(x)(y - x) \leq f(y) - f(x), \forall t \in]0, 1[.$$

En faisant tendre t vers 0, on obtient alors:

$$f(y) \geq Df(x)(y - x) + f(x).$$

($\Leftarrow$) Montrons maintenant la réciproque: Soit $(x, y) \in E^2$, et $t \in]0, 1[$ (pour $t = 0$ ou $t = 1$ on n'a rien à démontrer). On veut montrer que $f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$:

On pose $z = tx + (1 - t)y$. On a alors par hypothèse: $f(y) \geq f(z) + Df(z)(y - z)$, et $f(x) \geq f(z) + Df(z)(x - z)$.

En multipliant la première inégalité par $1 - t$, la deuxième par t et en les additionnant, on obtient:

$$(1 - t)f(y) + tf(x) \geq f(z) + (1 - t)Df(z)(y - z) + tDf(z)(x - z).$$

$$(1 - t)f(y) + tf(x) \geq f(z) + Df(z)((1 - t)(y - z) + t(x - z)).$$

Et comme $(1 - t)(y - z) + t(x - z) = 0$, on a donc:

$$(1 - t)f(y) + tf(x) \geq f(z) = f(tx + (1 - t)y).$$

2. ($\Rightarrow$) On suppose que f est strictement convexe, on veut montrer que $f(y) > f(x) + Df(x)(y - x)$ si $y \neq x$. Soit donc $(x, y) \in E^2, x \neq y$. On pose $z = \frac{1}{2}(x + y)$, et comme f est convexe, on peut appliquer la partie 1 du théorème et écrire que $f(x + z) \geq f(x) + Df(x)(z)$. On a donc:

$$f(x) + Df(x)\left(\frac{y - x}{2}\right) \leq f\left(\frac{x + y}{2}\right).$$

Comme f est strictement convexe, ceci entraîne que $f(x) + Df(x)\left(\frac{y - x}{2}\right) < \frac{1}{2}(f(x) + f(y))$, d'où le résultat.

($\Leftarrow$) La méthode de démonstration est la même que pour le 1 Voir [7]. ■

Proposition 1.2.2 (2ème caractérisation de la convexité). Soit $E = \mathbb{R}^N$ et $f \in C^2(E, \mathbb{R})$. Soit $H_f(x)$ la hessienne de f au point x , i.e. $(H_f(x))_{i,j} = \partial_{i,j}^2 f(x)$. Alors:

1. f est convexe si et seulement si $H_f(x)$ est symétrique et positive pour tout $x \in E$ (c.à.d. $H_f(x)^T = H_f(x)$ et $H_f(x)y \cdot y \geq 0$ pour tout $y \in \mathbb{R}^N$).

2. f est strictement convexe si $H_f(x)$ est symétrique définie positive pour tout $x \in E$. (Attention la réciproque est fausse).

Preuve. ($\Rightarrow$) Soit f convexe, on veut montrer que $H_f(x)$ est symétrique positive. Il est clair que $H_f(x)$ est symétrique car $\partial_{i,j}^2 f = \partial_{j,i}^2 f$ car f est $\mathbb{C}^2$. Par définition, $H_f(x) = D(\nabla f(x))$ et $\nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)$: Soit $(x, y) \in E^2$, comme f est convexe et de classe $\mathbb{C}^1$, on a grâce à la proposition (1. 2.1):

$$f(y) \geq f(x) + \nabla f(x)(y - x). \quad (1.2.2)$$

Soit $\varphi \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$ définie par $\varphi(t) = f(x + t(y - x))$. Alors:

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt = [\varphi'(t)(t - 1)]_0^1 - \int_0^1 \varphi''(t)(t - 1) dt$$

c'est-à-dire:

$$f(y) - f(x) = \varphi'(0) + \int_0^1 \varphi''(t)(1 - t) dt.$$

Or

$$\varphi'(t) = \nabla f(x + t(y - x))(y - x),$$

et:

$$\varphi''(t) = D(\nabla f(x + t(y - x))(y - x)) = H_f(x + t(y - x))(y - x)(y - x).$$

On a donc:

$$f(y) - f(x) = \nabla f(x)(y - x) + \int_0^1 H_f(x + t(y - x))(y - x)(y - x)(1 - t) dt \quad (1.2.3)$$

Les inégalités (1.2.2) et (1.2.3) entraînent:

$$\int_0^1 H_f(x + t(y - x))(y - x)(y - x)(1 - t) dt \geq 0, \forall x, y \in E.$$

On a donc:

$$\int_0^1 H_f(x + tz)zz(1 - t) dt \geq 0, \forall x, \forall z \in E. \quad (1.2.4)$$

En fixant $x \in E$, on écrit (1.2.4) avec $z = \varepsilon y, \varepsilon > 0, y \in \mathbb{R}^N$. On obtient:

$$\varepsilon^2 \int_0^1 H_f(x + t\varepsilon y)yy(1 - t) dt \geq 0, \forall x, y \in E, \forall \varepsilon > 0;$$

et donc:

$$\int_0^1 H_f(x + t\varepsilon y)yy(1 - t) dt \geq 0, \forall \varepsilon > 0.$$

Pour $(x, y) \in E^2$ fixé, $H_f(x + t\varepsilon y)$ tend vers $H_f(x)$ uniformément lorsque $\varepsilon \rightarrow 0$, pour $t \in [0, 1]$. On a donc : $\int_0^1 H_f(x)yy(1 - t) dt \geq 0$, c.à.d. $\frac{1}{2}H_f(x)yy \geq 0$.

Donc pour tout $(x, y) \in (\mathbb{R}^N)^2$, $H_f(x)yy \geq 0$ donc $H_f(x)$ est positive.

($\Leftarrow$) Montrons maintenant la réciproque : On suppose que $H_f(x)$ est positive pour tout $x \in E$. On veut démontrer que f est convexe ; on va pour cela utiliser la proposition (1.2.1) et montrer que :

$$f(y) \geq f(x) + \nabla f(x)(y - x) \text{ pour tout } (x, y) \in E^2.$$

Grâce à (1.2.3), on a

$$f(y) - f(x) = \nabla f(x)(y - x) + \int_0^1 H_f(x + t(y - x))(y - x)(y - x)(1 - t)dt.$$

Or $H_f(x + t(y - x))(y - x)(y - x) \geq 0$ pour tout couple $(x, y) \in E^2$, et $1 - t \geq 0$ sur $[0, 1]$. On a donc : $f(y) \geq f(x) + \nabla f(x)(y - x)$ pour tout couple $(x, y) \in E^2$. La fonction f est donc bien convexe.

($\Leftarrow$) On suppose que $H_f(x)$ est strictement positive pour tout $x \in E$, et on veut montrer que f est strictement convexe.

On va encore utiliser la caractérisation de la proposition (1.2.1). Soit donc $(x, y) \in E^2$ tel que $y \neq x$. Alors :

$$f(y) = f(x) + \nabla f(x)(y - x) + \int_0^1 H_f(x + t(y - x))(y - x)(y - x)(1 - t)dt :$$

$$\int_0^1 H_f(x + t(y - x))(y - x)(y - x) > 0, \text{ si } x \neq y, 1 - t \neq 0, \text{ si } t \in]0, 1[.$$

Donc :

$$f(y) > f(x) + \nabla f(x)(y - x), \text{ si } x \neq y,$$

ce qui prouve que f est strictement convexe. Voir [7] ■

1.2.1 Ensembles convexes

Définition 1.2.2 Soit E un espace vectoriel sur $K = \mathbb{R}$, et soit $A \subset E$: On dit que A est convexe si :

$$\forall x \in A, \forall y \in A, \forall t \in [0, 1] : tx + (1 - t)y \in A.$$

Voir [6].

1.3 Optimisation sans contrainte

1.3.1 Définitions et conditions d'optimalité

Soit $E = \mathbb{R}^N$ et $f \in C(E, \mathbb{R})$. On s'intéresse à la recherche de $\bar{x} \in E$ tel que:

$$\begin{cases} \bar{x} \in E \\ f(\bar{x}) = \inf_E f \end{cases} \quad (1.3.1)$$

Ce problème est un problème d'optimisation sans contrainte.

Proposition 1.3.1 (*Condition nécessaire d'optimalité*).

Soit E un espace vectoriel normé, et soient $f \in C(E, \mathbb{R})$, et $x \in E$ tel que f est différentiable en $\bar{x}$. Si $\bar{x}$ est solution de (1.3.1) alors $Df(\bar{x}) = 0$.

Preuve. Supposons qu'il existe $\alpha > 0$ tel que $f(\bar{x}) \leq f(y)$ pour tout $y \in B(\bar{x}, \alpha)$. Soit $z \in E \setminus \{0\}$, alors si $|t| < \frac{\alpha}{\|z\|}$, on a $\bar{x} + tz \in B(\bar{x}, \alpha)$ (où $B(\bar{x}, \alpha)$ désigne la boule ouverte de centre $\bar{x}$ et de rayon α) et on a donc $f(\bar{x}) \leq f(\bar{x} + tz)$. Comme f est différentiable en $\bar{x}$, on a:

$$f(\bar{x} + tz) = f(\bar{x}) + Df(\bar{x})(tz) + |t| \varepsilon_z(t),$$

où $\varepsilon_z(t) \rightarrow 0$ lorsque $t \rightarrow 0$. On a donc $f(\bar{x}) + tDf(\bar{x})(z) + |t| \varepsilon_z(t) \geq f(\bar{x})$. Et pour $\frac{\alpha}{\|z\|} > t > 0$, on a $Df(\bar{x})(z) + \varepsilon_z(t) \geq 0$. En faisant tendre t vers 0, on obtient: $Df(\bar{x})(z) \geq 0, \forall z \in E$. On a aussi $Df(\bar{x})(-z) \geq 0, \forall z \in E$, et donc: $-Df(\bar{x})(z) \geq 0, \forall z \in E$. On en conclut que: $Df(\bar{x}) = 0$. Voir [6] ■

Remarque 1.3.1 *Attention, la proposition précédente donne une condition nécessaire mais non suffisante. En effet, $Df(\bar{x}) = 0$ n'entraîne pas que f atteigne un minimum (ou un maximum) même local en $\bar{x}$. Prendre par exemple $E = \mathbb{R}$, $\bar{x} = 0$ et la fonction f définie par: $f(x) = x^3$ pour s'en convaincre.*

1.3.2 Résultats d'existence et d'unicité

Théorème 1.3.1 (*Existence*) Soit $E = \mathbb{R}^N$ et $f : E \rightarrow \mathbb{R}$ une application telle que

(i) f est continue,

(ii) $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$.

Alors il existe $\bar{x} \in \mathbb{R}^N$ tel que $f(\bar{x}) \leq f(y)$ pour tout $y \in \mathbb{R}^N$.

Preuve. La condition (ii) peut encore s'écrire,

$$\forall A \in \mathbb{R}, \exists R \in \mathbb{R}, \|x\| \geq R \Rightarrow f(x) \geq A \quad (1.3.2)$$

On écrit (1.3.2) avec $A = f(0)$. On obtient alors:

$$\exists R \in \mathbb{R} \text{ telque } \|x\| \geq R \Rightarrow f(x) \geq f(0).$$

On en déduit que $\inf_{\mathbb{R}^N} f = \inf_{B_R} f$, où $B_R = \{x \in \mathbb{R}^N, |x| \leq R\}$. Or, B_R est un compact de $\mathbb{R}^N$ et f est continue donc il existe $x \in B_R$ tel que $f(\bar{x}) = \inf_{B_R} f$ et donc $f(\bar{x}) = \inf_{\mathbb{R}^N} f$. Voir [7] ■

1. Le théorème est faux si E est de dimension infinie (i.e. si E est espace de Banach au lieu de $E = \mathbb{R}^N$), car si E est de dimension infinie, B_R n'est pas compacte.
2. L'hypothèse (ii) du théorème peut être remplacée par

$$(ii)' \quad \exists b \in \mathbb{R}^N, \exists R > 0 \text{ telque } \|x\| \geq R \Rightarrow f(x) \geq f(b)$$

3. Sous les hypothèses du théorème il n'y a pas toujours unicité de $\bar{x}$ même dans le cas $N = 1$, prendre pour s'en convaincre la fonction f définie de $\mathbb{R}$ dans $\mathbb{R}$ par $f(x) = x^2(x-1)(x+1)$.

Théorème 1.3.2 (Condition suffisante d'unicité). Soit E un espace vectoriel normé et $f : E \rightarrow \mathbb{R}$ strictement convexe alors il existe au plus un $\bar{x} \in E$ tel que $f(\bar{x}) \leq f(y)$, $\forall y \in E$.

Preuve. Soit f strictement convexe, supposons qu'il existe $\bar{x}$ et $\bar{\bar{x}} \in E$ tels que: $f(\bar{x}) = f(\bar{\bar{x}}) = \inf_{\mathbb{R}^N} f$: Comme f est strictement convexe, si $\bar{x} \neq \bar{\bar{x}}$, alors:

$$f\left(\frac{1}{2}\bar{x} + \frac{1}{2}\bar{\bar{x}}\right) < \frac{1}{2}f(\bar{x}) + \frac{1}{2}f(\bar{\bar{x}}) = \inf_{\mathbb{R}^N} f;$$

ce qui est impossible; donc $\bar{x} = \bar{\bar{x}}$.

Ce théorème ne donne pas l'existence. Par exemple dans le cas $N = 1$ la fonction f définie par $f(x) = e^x$ n'atteint pas son minimum; en effet, $\inf_{\mathbb{R}} f = 0$ et $f(x) \neq 0$ pour tout $x \in \mathbb{R}$, et pourtant f est strictement convexe. Par contre, si on réunit les hypothèses des théorèmes 1. 3.1 et 1. 3.2, on obtient le résultat d'existence et d'unicité suivant: Voir [6] ■

Théorème 1.3.3 (*Existence et unicité*). Soit $E = \mathbb{R}^N$, et soit $f : E \rightarrow \mathbb{R}$. On suppose que:

- (i) f continue,
- (ii) $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$,
- (iii) f est strictement convexe ;

alors il existe un unique $\bar{x} \in \mathbb{R}^N$ tel que $f(\bar{x}) = \inf_{\mathbb{R}^N} f$. Voir [1]

Théorème 1.3.4 (*Caractérisation des points tels que $f(\bar{x}) = \inf_E f$*). Soit E espace vectoriel normé et f une fonction de E dans $\mathbb{R}$. On suppose que $f \in \mathcal{C}^1(E, \mathbb{R})$ et que f est convexe. Soit $\bar{x} \in E$. Alors:

$$f(\bar{x}) = \inf_E f \Leftrightarrow \nabla f(\bar{x}) = 0.$$

Preuve. ($\Rightarrow$) Supposons que $f(\bar{x}) = \inf_E f$ alors on sait (voir Proposition 1.3.1) que $Df(\bar{x}) = 0$ (la convexité est inutile).

($\Leftarrow$) Si f est convexe et différentiable, d'après la proposition 1.2.1, on a: $f(y) \geq f(\bar{x}) + Df(\bar{x})(y - \bar{x})$ pour tout $y \in E$ et comme par hypothèse $Df(\bar{x}) = 0$, on en déduit que $f(y) \geq f(\bar{x})$ pour tout $y \in E$. Donc $f(\bar{x}) = \inf_E f$. Voir [7] ■

1.4 Optimisation avec contraintes

Définition 1.4.1 Soit $E = \mathbb{R}^N$, soient $f \in C(E, \mathbb{R})$ et K un sous ensemble de E . On s'intéresse à la recherche de $\bar{x} \in K$ tel que:

$$\begin{cases} \bar{x} \in K. \\ f(\bar{x}) = \inf_K f \end{cases} \quad (1.4.1)$$

Ce problème est un problème de minimisation avec contrainte (ou "sous contrainte") au sens où l'on cherche $\bar{x}$ qui minimise f en astreignant $\bar{x}$ à être dans K . Voyons quelques exemples de ces contraintes (définies par l'ensemble K), qu'on va expliciter à l'aide des p fonctions continues, $g_i \in C(E, \mathbb{R})$ $i = 1, \dots, p$.

Contraintes égalités:

On pose $K = \{x \in E, g_i(x) = 0, i = 1, \dots, p\}$. On verra plus loin que le problème de minimisation de f peut alors être résolu grâce au théorème des multiplicateurs de Lagrange.

Contraintes inégalités:

On pose $K = \{x \in E, g_i(x) \leq 0, i = 1, \dots, p\}$. On verra plus loin que le problème de minimisation de f peut alors être résolu grâce au théorème de Kuhn–Tucker.

Programmation convexe:

Dans le cas où f est convexe et K est convexe, on dit qu'on a affaire un problème de "programmation convexe".

1.4.1 Existence, Unicité, et Conditions d'optimalité simple

Théorème 1.4.1 (*Existence*). Soit $E = \mathbb{R}^N$ et $f \in C(E, \mathbb{R})$.

1. Si K est un sous-ensemble fermé borné de E , alors il existe $\bar{x} \in K$ tel que $f(\bar{x}) = \inf_K f$.
2. Si K est un sous-ensemble fermé de E , et si f est croissante à l'infini, c'est-à-dire que $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$, alors $\exists \bar{x} \in K$ tel que $f(\bar{x}) = \inf_K f$.

Preuve. 1. Si K est un sous-ensemble fermé borné de E , comme f est continue, elle atteint ses bornes sur K , d'où l'existence de $\bar{x}$.

2. Si f est croissante à l'infini, alors il existe $R > 0$ tel que si $\|x\| > R$ alors $f(x) > f(0)$; donc $\inf_K f = \inf_{K \cap B_R} f$ où B_R désigne la boule de centre 0 et de rayon R . L'ensemble $K \cap B_R$ est compact, car intersection d'un fermé et d'un compact. Donc, par ce qui précède, il existe $\bar{x}$ tel que $f(\bar{x}) = \inf_{K \cap B_R} f = \inf_{B_R} f$. Voir [1] ■

Théorème 1.4.2 (*Unicité*). Soient $E = \mathbb{R}^N$ et $f \in C(E, \mathbb{R})$. On suppose que f est strictement convexe et que K est convexe. Alors il existe au plus un élément $\bar{x}$ de K tel que $f(\bar{x}) = \inf_K f$.

Preuve. Supposons que $\bar{x}$ et $\bar{\bar{x}}$ soient deux solutions du problème (1.4.1), avec $\bar{x} \neq \bar{\bar{x}}$. Alors:

$$f\left(\frac{1}{2}\bar{x} + \frac{1}{2}\bar{\bar{x}}\right) < \frac{1}{2}f(\bar{x}) + \frac{1}{2}f(\bar{\bar{x}}) = \inf_K f.$$

On aboutit donc à une contradiction.

Des théorèmes d'existence 1.4.1 et d'unicité 1.4.2, on déduit immédiatement le théorème d'existence et d'unicité suivant: ■

Théorème 1.4.3 (*Existence et unicité*). Soient $E = \mathbb{R}^N$, $f \in C(E, \mathbb{R})$ une fonction strictement convexe et K un sous ensemble convexe fermé de E . Si K est borné ou si f est croissante à l'infini, c'est-à-dire si $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$, alors il existe un unique élément $\bar{x}$ de K solution du problème de minimisation (1.4.1), i.e. tel que $f(\bar{x}) = \inf_K f$.

Remarque 1.4.1 On peut remplacer $E = \mathbb{R}^N$ par E espace de Hilbert de dimension infinie dans le dernier théorème, mais on a besoin dans ce cas de l'hypothèse de convexité de f pour assurer l'existence de la solution.

Proposition 1.4.1 (*Condition simple d'optimalité*). Soient $E = \mathbb{R}^N$, $f \in C(E, \mathbb{R})$ et $\bar{x} \in K$ tel que $f(\bar{x}) = \inf_K f$. On suppose que f est différentiable en $\bar{x}$

1. Si $\bar{x} \in \overset{\circ}{K}$, alors $\nabla f(\bar{x}) = 0$.
2. Si K est convexe, alors $\nabla f(\bar{x})(x - \bar{x}) \geq 0$ pour tout $x \in K$.

Preuve. 1. Si $\bar{x} \in \overset{\circ}{K}$, alors il existe $\varepsilon > 0$ tel que $B(\bar{x}, \varepsilon) \subset K$ et $f(\bar{x}) \geq f(x)$ $\forall x \in B(\bar{x}, \varepsilon)$. Alors on a déjà vu (voir preuve de la Proposition 1.3.1) que ceci implique $\nabla f(\bar{x}) = 0$.

2. Soit $x \in K$. Comme $\bar{x}$ réalise le minimum de f sur K , on a: $f(\bar{x} + t(x - \bar{x})) = f(tx + (1-t)\bar{x}) \geq f(\bar{x})$ pour tout $t \in]0, 1]$, par convexité de K . On en déduit que: $\frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t} \geq 0$ pour tout $t \in]0, 1]$. En passant à la limite lorsque t tend vers 0 dans cette dernière inégalité, on obtient: $\nabla f(\bar{x})(x - \bar{x}) \geq 0$. Voir [7] ■

1.4.2 Conditions d'optimalité dans le cas de contraintes égalité

Dans tout ce paragraphe, on considèrera les hypothèses et notations suivantes:

$$\begin{aligned} f &\in C(\mathbb{R}^N, \mathbb{R}), g_i \in C^1(\mathbb{R}^N, \mathbb{R}), i = 1, \dots, p; \\ K &= \{x \in \mathbb{R}^N, g_i(x) = 0, \forall i = 1, \dots, p\}; \\ g &= (g_1, \dots, g_p)^T \in C^1(\mathbb{R}^N, \mathbb{R}^p). \end{aligned}$$

Remarque 1.4.2 (*Quelques rappels de calcul différentiel*). Comme $g \in C^1(\mathbb{R}^N, \mathbb{R}^p)$, si $x \in \mathbb{R}^N$; alors $Dg(x) \in L(\mathbb{R}^N, \mathbb{R}^p)$, ce qui revient à dire, en confondant l'application linéaire $Dg(x)$ avec sa matrice, que $Dg(x) \in M_{p,N}(\mathbb{R})$. Par définition, $\text{Im}(Dg(x)) =$

$\{Dg(x)z, z \in \mathbb{R}^N\} \subset \mathbb{R}^p$, et $\text{rang}(Dg(x)) = \dim(\text{Im}(Dg(x))) \leq p$. On rappelle de plus que:

$$Dg(x) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial x_1} & \cdots & \frac{\partial g_p}{\partial x_N} \end{pmatrix}$$

et que $\text{rang}(Dg(x)) \leq \min(N, p)$. De plus, si $\text{rang}(Dg(x)) = p$, alors les vecteurs $(Dg_i(x))_{i=1, \dots, p}$ sont linéairement indépendants dans $\mathbb{R}^N$.

Théorème 1.4.4 (Multiplicateurs de Lagrange). Soit $\bar{x} \in K$ tel que $f(\bar{x}) = \inf_K f$. On suppose que f est différentiable en $\bar{x}$ et $\dim(\text{Im}(Dg(\bar{x}))) = p$ (ou $\text{rang}(Dg(\bar{x})) = p$), alors: il existe $(\lambda_1, \dots, \lambda_p)^T \in \mathbb{R}^p$ tels que:

$$\nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0.$$

(Cette dernière égalité a lieu dans $\mathbb{R}^N$).

Remarque 1.4.3 (Utilisation pratique du théorème de Lagrange). Soient $f \in C^1(\mathbb{R}^N, \mathbb{R})$, $g = (g_1, \dots, g_p)^T$ avec $g_i \in C(\mathbb{R}^N, \mathbb{R})$ pour $i = 1, \dots, p$; et soit $K = \{x \in \mathbb{R}^N, g(x) = 0, i = 1, \dots, p\}$. Le problème qu'on cherche à résoudre est le problème de minimisation suivant:

$$\begin{cases} \bar{x} \in K \\ f(\bar{x}) = \inf_K f \end{cases} \quad (1.4.2)$$

D'après le théorème des multiplicateurs de Lagrange, si u est solution de (1.4.2) et $\text{rang}(Dg(\bar{x})) = p$, alors il existe $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ tel que $\bar{x}$ est solution du problème

$$\begin{cases} \nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0 \\ g_i(\bar{x}) = 0, i = 1, \dots, p, \end{cases} \quad (1.4.3)$$

Le système (1.4.3) est un système non linéaire de $(N + p)$ équations et à $(N + p)$ inconnues $(\bar{x}_1, \dots, \bar{x}_N, \lambda_1, \dots, \lambda_p)$.

Ce système sera résolu par une méthode de résolution de système non linéaire (Newton par exemple).

Remarque 1.4.4 On vient de montrer que si $\bar{x}$ solution de (1.4.2) alors $\bar{x}$ solution de (1.4.3). Par contre, si $\bar{x}$ est solution de (1.4.3), ceci n'entraîne pas que $\bar{x}$ est solution de (1.4.2).

1.4.3 Contraintes inégalités

Soient $f \in C(\mathbb{R}^N, \mathbb{R})$ et $g_i \in C^1(\mathbb{R}^N, \mathbb{R})$ $i = 1, \dots, p$, on considère maintenant un ensemble K de la forme: $K = \{x \in \mathbb{R}^N, g_i(x) \leq 0, i = 1 \dots p\}$ et on cherche à résoudre le problème de minimisation (1.4.1) qui s'écrit:

$$\begin{cases} \bar{x} \in K \\ f(\bar{x}) \leq f(x) \quad \forall x \in K \end{cases}$$

Remarque 1.4.5 Soit $\bar{x}$ une solution de (1.4.2) et supposons que $g_i(x) < 0$; pour tout $i = 1, \dots, p$. Il existe alors $\varepsilon > 0$ tel que si $x \in B(\bar{x}, \varepsilon)$ alors $g_i(x) < 0$ pour tout $i = 1, \dots, p$. On a donc $f(\bar{x}) \leq f(x) \quad \forall x \in B(\bar{x}, \varepsilon)$. On est alors ramené à un problème de minimisation sans contrainte, et si f est différentiable en $\bar{x}$, on a donc $\nabla f(\bar{x}) = 0$.

Théorème 1.4.5 (Kuhn–Tucker). Soit $f \in C(\mathbb{R}^N, \mathbb{R})$, soit $g_i \in C^1(\mathbb{R}^N, \mathbb{R})$; pour $i = 1, \dots, p$, et soit $K = \{x \in \mathbb{R}^N, g_i(x) \leq 0, i = 1, \dots, p\}$. On suppose qu'il existe $\bar{x}$ solution de (1.4.2), et on pose $I(\bar{x}) = \{i \in 1, \dots, p, g_i(\bar{x}) = 0\}$. On suppose que f est différentiable en $\bar{x}$ et que la famille (de $\mathbb{R}^N$) $\{\nabla g_i(\bar{x}), i \in I(\bar{x})\}$ est libre. Alors il existe une famille $(\lambda_i)_{i \in I(\bar{x})} \subset \mathbb{R}_+$ telle que:

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0.$$

Voir [7].

Remarque 1.4.6 1- Le théorème de Kuhn-Tucker s'applique pour des ensembles de contrainte de type inégalité. Si on a une contrainte de type égalité, on peut évidemment se ramener à deux contraintes de type inégalité: $K = \{x \in \mathbb{R}^N, g(x) = 0\} = \{x \in \mathbb{R}^N, g_1(x) \leq 0 \text{ et } g_2(x) \geq 0\}$ $g_1 = g, g_2 = -g$ on remarque que la famille $\{\nabla g_1(\bar{x}), \nabla g_2(\bar{x})\} = \{\nabla g(\bar{x}), -\nabla g(\bar{x})\}$ n'est pas libre. On ne peut donc pas appliquer le théorème de Kuhn-Tucker sous la forme donnée précédemment dans ce cas.

2- Dans la pratique, on a intérêt à écrire la conclusion du théorème de Kuhn-Tucker i.e. l'existence de la famille $(\lambda_i)_{i \in I(\bar{x})}$ sous la forme du système de $N + p$ équations et $2p$ inéquations à résoudre suivant:

$$\nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0 \quad (1.4.4)$$

$$\lambda_i g_i(\bar{x}) = 0, \quad i = 1, \dots, p \quad (1.4.5)$$

$$g_i(\bar{x}) \leq 0, \quad i = 1, \dots, p$$

$$\lambda_i \geq 0, \quad i = 1, \dots, p.$$

Voir [7].

Chapitre 2

Programmation convexe

On s'intéresse dans ce chapitre, aux problèmes de programmation convexe.

On étudie l'existence et l'unicité de solutions de ces problèmes et on donne aussi quelques méthodes numériques de résolution de ce type des problèmes.

2.1 Fonctions elliptiques, fonctions coercives

Définition 2.1.1 Soit $f : \mathbb{R}^N \rightarrow \mathbb{R}$. On dit que f est une fonction elliptique si elle est de classe C^1 et si $\exists \alpha > 0$ telle que $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2 ; \forall x, y \in \mathbb{R}^N$.

Définition 2.1.2 Soit $\Omega \subset \mathbb{R}^N$ un ensemble non borné et $f : \Omega \rightarrow \mathbb{R}$. On dit que f est coercive sur Ω si on a $\lim_{x \in \Omega, \|x\| \rightarrow +\infty} f(x) = +\infty$.

Proposition 2.1.1 Soit $f : \mathbb{R}^N \rightarrow \mathbb{R}$ une fonction elliptique. Alors elle est strictement convexe et coercive. Elle vérifie en plus l'inégalité:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2} \|y - x\|^2 ; \forall x, y \in \mathbb{R}^N$$

Proposition 2.1.2 (Caractérisation de l'ellipticité avec " ∇^2 "). Soit f une fonction de classe C^2 . Alors f est elliptique si et seulement si $\exists \beta > 0$ tel que:

$$\langle \nabla^2 f(x) h, h \rangle \geq \beta \|h\|^2, \forall x, h \in \mathbb{R}^N$$

Proposition 2.1.3 Soit $p \in \mathbb{N}^*$, $f_1, f_2, \dots, f_p: \mathbb{R}^N \rightarrow \mathbb{R}$ des fonctions de classe C^1 et convexes et $\lambda_1, \lambda_2, \dots, \lambda_N$ des constantes strictement positives. Si au moins une des fonctions $f_1, \dots, f_p$ est elliptique, alors $f \equiv \lambda_1 f_1 + \lambda_2 f_2 + \dots + \lambda_p f_p$ est elliptique.

Preuve. Soit $i \in \{1, 2, \dots, p\}$ tel que f_i soit elliptique. Alors il existe $\alpha > 0$ tel que

$$\langle \nabla f_i(y) - \nabla f_i(x), y - x \rangle > \alpha \|x - y\|^2, \forall x, y \in \mathbb{R}^N.$$

D'autre part, comme tous les f_k sont convexes on a:

$$\langle \nabla f_k(y) - \nabla f_k(x), y - x \rangle \geq 0, \forall x, y \in \mathbb{R}^N; \forall k \neq i$$

En multipliant la première inégalité par λ_i et les autres par λ_k et en sommant en k , on obtient immédiatement le résultat. Voir [1] ■

Exemple 2.1.1 (Fonctions elliptiques):

1) Si $N = 1$: Toute fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ de classe C^2 et satisfaisant: $\exists \alpha > 0, f''(x) \geq \alpha, \forall x \in \mathbb{R}$ est une fonction elliptique (c'est une conséquence immédiate de la proposition 2.1.2).

a) $f(x) = ax^2 + bx + c$ avec $a > 0$,

b) $f(x) = x^2 + \sin(x)$ (car $f''(x) = 2 - \sin(x) \geq 1; \forall x \in \mathbb{R}$).

2) **Le cas général** ($N \in \mathbb{N}^*$) Soit $f: \mathbb{R}^N \rightarrow \mathbb{R}$ donnée par $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle; \forall x \in \mathbb{R}^N$ avec $A \in M_N(\mathbb{R})$ une matrice carrée réelle et symétrique de taille n , avec $b \in \mathbb{R}^N$. On calcule facilement: $\nabla f(x) = Ax - b$, $\nabla^2 f(x) = A$ (donc la hessienne de f est constante). Comme A est symétrique alors: on sait que toutes les valeurs propres de A sont réelles et que $\langle Ax, x \rangle \geq \lambda_{\min} \|x\|^2; \forall x \in \mathbb{R}^N$, où $\lambda_{\min} \in \mathbb{R}$ est la plus petite valeur propre de A .

Rémarquons que l'inégalité précédente devient égalité si x est un vecteur propre associé à la valeur propre $\lambda_{\min}$.

Rappelons aussi que A est une matrice définie positive si et seulement si $\lambda_{\min} > 0$. (par définition une matrice carrée réelle $B \in M_N(\mathbb{R})$ est définie positive si $\langle Bx, x \rangle > 0, \forall x \in \mathbb{R}^N, x \neq 0$).

En utilisant la Proposition 2.1.2 on déduit que f est une fonction elliptique si et seulement si A est une matrice définie positive.

2.2 Application et exemples

On rappelle le résultat suivant de projection sur un convexe fermé:

Proposition 2.2.1 (*Projection sur un convexe fermé*) Soit E un espace de Hilbert, muni d'une norme $\|\cdot\|_E$ induite par un produit scalaire $\langle \cdot, \cdot \rangle$ et soit K un convexe fermé non vide de E . Alors, pour tout $x \in E$, il existe un unique $x_0 \in K$ tel que $\|x - x_0\| \leq \|x - y\|, \forall y \in K$. On note $x_0 = p_K(x)$ la projection orthogonale de x sur K . On a également: $x_0 = p_K(x)$ si et seulement si $\langle x - x_0, x_0 - y \rangle \geq 0, \forall y \in K$.

Proposition 2.2.2 (*Contraction de la projection orthogonale*)

$$\|P_K(x) - P_K(y)\| \leq \|x - y\|, \forall x, y \in E.$$

2.2.1 Théorie générale du point selle

Soit K et P deux ensembles quelconques et l'application $L : K \times P \rightarrow \mathbb{R}$.

Définition 2.2.1 On dit que le couple $(\bar{x}, \mu) \in K \times P$ est un point selle de L si on a:

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \mu) \leq L(x, \mu), \forall x \in K, \forall \lambda \in P$$

(Autrement dit: $\bar{x}$ est un point de minimum de la fonction $x \in K \rightarrow L(x, \lambda) \in \mathbb{R}$ et λ est un point de maximum de la fonction $\lambda \in P \rightarrow L(x, \lambda) \in \mathbb{R}$).

Exemple 2.2.1 Prendre $K = P = \mathbb{R}$ et $L(x, \lambda) = x^2 - \lambda^2$. Il est facile de voir que $(0, 0)$ est un point selle de L . On introduit les fonctions suivantes: $G : K \rightarrow \mathbb{R}$ et $H : P \rightarrow \mathbb{R}$ données par: $G(x) = \sup_{q \in P} L(x, q)$ et respectivement par $H(\lambda) = \inf_{y \in K} L(y, \lambda)$.

Remarque 2.2.1 $(\bar{x}, \mu)$ est un point selle de L si et seulement si $G(\bar{x}) = L(\bar{x}, \mu) = H(\mu)$. On a le résultats suivant:

Proposition 2.2.3

$$\sup_{\lambda \in P} \inf_{x \in K} L(x, \lambda) \leq \inf_{x \in K} \sup_{\lambda \in P} L(x, \lambda)$$

(autrement dit: $\sup_{\lambda \in P} H(\lambda) \leq \inf_{x \in K} G(x)$).

Théorème 2.2.1 Si $(\bar{x}, \mu)$ est un point selle de L sur $K \times P$ alors:

$$\sup_{\lambda \in P} \inf_{x \in K} L(x, \lambda) = L(\bar{x}, \mu) = \inf_{x \in K} \sup_{\lambda \in P} L(x, \lambda)$$

(autrement dit : $\sup_{\lambda \in P} H(\lambda) = L(\bar{x}, \mu) = \inf_{x \in K} G(x)$). Voir [1]

2.2.2 Applications de la théorie du point selle à l'optimisation

On utilise ici de nouveau les notations du début de la section 1.1.1, spécifiques aux problèmes d'optimisation. Nous introduisons la fonction $L : \mathbb{R}^N \times \mathbb{R}_+^p \rightarrow \mathbb{R}$ donnée par

$$L(x, \lambda) = f(x) + \sum_{i=1}^p \lambda_i g_i(x) \equiv f(x) + \langle \lambda, g(x) \rangle$$

(où on voit λ comme un vecteur colonne de composantes $\lambda_1, \lambda_2, \dots, \lambda_p$). On a le résultat suivant:

Théorème 2.2.2 Si $(\bar{x}, \mu)$ est un point selle de L sur $\mathbb{R}^N \times \mathbb{R}_+^p$ alors $\bar{x}$ est un point de minimum de f sur K . En plus $(\bar{x}, \mu)$ satisfait les conditions de Kuhn-Tucker, c'est à dire:

$$\begin{cases} \nabla f(\bar{x}) + \sum_{i=1}^p \mu_i \nabla g_i(\bar{x}) = 0 \\ \mu_i g_i(\bar{x}) = 0, \forall i = 1, \dots, p \end{cases} \quad (2.2.1)$$

Preuve. Nous avons $L(\bar{x}, \mu) \geq L(\bar{x}, \lambda), \forall \lambda \in \mathbb{R}_+^p$ ce qui donne:

$$\sum_{i=1}^p \mu_i g_i(\bar{x}) \geq \sum_{i=1}^p \lambda_i g_i(\bar{x}), \forall \lambda \in \mathbb{R}_+^p. \quad (2.2.2)$$

En prenant respectivement $\lambda = 0$ et $\lambda = 2\mu$ dans l'inégalité précédente on obtient:

$$\langle \mu, g(\bar{x}) \rangle = 0. \quad (2.2.3)$$

Maintenant pour un $j \in \{1, 2, \dots, p\}$ fixé prenons dans (2.2.2) $\lambda = \mu + e_j$. On déduit alors $g_j(\bar{x}) \leq 0$. Comme ceci est vrai pour tout j , on déduit $\bar{x} \in K$.

D'autre part, on se sert de l'inégalité $L(\bar{x}, \mu) \leq L(x, \mu), \forall x \in \mathbb{R}^N$ c'est à dire

$$f(\bar{x}) + \langle \mu, g(\bar{x}) \rangle \leq f(x) + \langle \mu, g(\bar{x}) \rangle \quad \forall x \in \mathbb{R}^N. \quad (2.2.4)$$

En utilisant l'égalité (2.2.3) et le fait que $\langle \mu, g(x) \rangle \leq 0$ pour $x \in K$, on déduit $f(\bar{x}) \leq f(x), \forall x \in K$.

Ceci nous dit que $\bar{x}$ est un point de minimum de f sur K .

D'autre part, de (2.2.4), on déduit que $\bar{x}$ est un point de minimum de la fonction $x \in \mathbb{R}^N \rightarrow f(x) + \sum_{i=1}^p \mu_i g_i(x)$. Ceci nous donne, en utilisant le Théorème 2.17, la première

égalité de (2.2.1). Finalement, de (2.2.2) et (??) on a: $\sum_{j=1}^p \mu_j g_j(\bar{x}) = 0$ et $\mu_i g_i(\bar{x}) \leq 0$ pour tout $i = \{1, \dots, p\}$ ce qui nous donne la deuxième partie de (2.2.1). Ceci finit la preuve du Théorème. Voir [1] ■

Remarque 2.2.2 *Ce résultat nous donne une condition suffisante pour avoir un minimum de f sur K . On verra aussi une réciproque, sous des hypothèses supplémentaires. On cherchera alors à résoudre le problème appelé le problème dual: $\sup_{\lambda \in \mathbb{R}_+^p} \inf_{x \in \mathbb{R}^N} L(x, \lambda)$ (justifié par le Théorème 2.2.1) plus simple à résoudre que le problème de minimisation de f sur K (qu'on appellera problème primal).*

2.3 Le cas convexe

On verra ici des réciproques des Théorèmes 1.4.5 et 2.2.2 dans le cas où f et g_j sont convexes.

Théorème 2.3.1 *(Réciproque du Théorème 1.4.5). Supposons que toutes les fonctions $f, g_1, g_2, \dots, g_p$ sont convexes. Supposons aussi qu'il existe $\mu_1, \mu_2, \dots, \mu_p \in [0, +\infty[$ et $\bar{x} \in K$ tels que les conditions de Kuhn-Tucker du Théorème 1.4.5 soient satisfaites. Alors*

a) $\bar{x}$ est un point de minimum de f sur K .

b) $(\bar{x}, \mu)$ est un point selle de L .

(Rappel: $L : \mathbb{R}^N \times \mathbb{R}_+^p \rightarrow \mathbb{R}$, $L(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle$).

Preuve. a) Nous avons $f(\bar{x}) \leq f(y) - \sum_{i=1}^p \mu_i g_i(y), \forall y \in K$ (car $g_i(y) \leq 0, \mu_i \geq 0$). Ensuite de (1.4.5) et de la convexité de g_i on déduit:

$$\mu_i g_i(y) = \mu_i [g_i(y) - g_i(\bar{x})] \geq \mu_i \langle \nabla g_i(\bar{x}), y - \bar{x} \rangle.$$

On obtient alors des inégalités précédentes et de (1.4.4) $f(\bar{x}) \leq f(\bar{x}) + \langle \nabla f(\bar{x}), y - \bar{x} \rangle, \forall y \in K$. En utilisant cette fois-ci la convexité de f nous obtenons $f(\bar{x}) \leq f(y), \forall y \in K$. ce qui finit la preuve de a).

b) Nous avons $L(\bar{x}, \lambda) = f(\bar{x}) + \langle \lambda, g(\bar{x}) \rangle \leq f(\bar{x}) + \langle \mu, g(\bar{x}) \rangle = L(\bar{x}, \mu), \forall \lambda \in \mathbb{R}_+^p$ car $\langle \lambda, g(\bar{x}) \rangle \leq 0$ et $\langle \lambda, g(\bar{x}) \rangle = 0$.

D'autre part, comme les fonctions $f, g_1, g_2, \dots, g_p$ sont convexes, alors l'application $x \in \mathbb{R}^N \rightarrow L(x, \mu)$ est convexe. Mais l'égalité (1.4.4) nous dit que le gradient de cette application s'annule en $\bar{x}$. On déduit alors que $\bar{x}$ est un point de minimum de cette application et ceci finit la preuve. On finit cette section par les corollaires suivants: Voir [1] ■

Corollaire 2.3.1 (*Lien entre point selle et conditions de Kuhn-Tucker*) Supposons que les fonctions $f, g_1, \dots, g_p$ sont convexes. Alors $(\bar{x}, \mu)$ est un point selle de L si et seulement si $\bar{x} \in K$ et les conditions (1.4.4) et (1.4.5) sont satisfaites.

Corollaire 2.3.2 Supposons que les fonctions $f, g_1, \dots, g_p$ sont convexes. Soit $\bar{x}$ un point de minimum de f sur K tel que les contraintes de K sont qualifiées en $\bar{x}$. Alors $(\bar{x}, \mu)$ est un point selle de L .

2.4 Algorithmes de minimisation avec contraintes

Le but de cette section est de donner des algorithmes numériques pour la résolution du problème de minimisation avec contraintes: $\min_{x \in K} f(x)$ avec $K \subset \mathbb{R}^N$ un ensemble de contraintes.

On rappelle les deux cas de contraintes que nous considérons dans ce cours:

1. K est un pavé, c'est à dire, il est de la forme $K = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_N, b_N]$ (avec la convention qu'on peut avoir $-\infty$ à la place de a_i ou $+\infty$ à la place de b_i , donc le pavé peut être de volume infini).

2. K est de la forme $K = \{x \in \mathbb{R}^N, g_i(x) \leq 0, i = 1, 2, \dots, p\}$ avec $p \in \mathbb{N}^*$ et $g_1, g_2, \dots, g_p$ des fonctions de $\mathbb{R}^N$ dans $\mathbb{R}$ données (K est donné par des contraintes inégalités larges).

2.4.1 Méthodes de relaxation

On suppose ici

$$K = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_N, b_N] = \prod_{i=1}^N [a_i, b_i]. \quad (2.4.1)$$

L'algorithme dans ce cas rassemble à l'algorithme de relaxation pour minimisation sans contraintes.

Algorithme:

On se donne $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_N^{(k)})^T$ et on va calculer $x^{(k+1)} = (x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_N^{(k+1)})^T$ en N pas successifs. On a:

$$\begin{aligned} f(x_1^{(k+1)}, x_2^{(k)}, \dots, x_N^{(k)}) &= \min_{y \in [a_1, b_1]} f(y, x_2^{(k)}, \dots, x_N^{(k)}), \\ f(x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k)}, \dots, x_N^{(k)}) &= \min_{y \in [a_2, b_2]} f(x_1^{(k+1)}, y, x_3^{(k)}, \dots, x_N^{(k)}), \\ f(x_1^{(k+1)}, \dots, x_N^{(k+1)}) &= \min_{y \in [a_N, b_N]} f(x_1^{(k+1)}, \dots, x_{N-1}^{(k+1)}, y). \end{aligned}$$

Comme dans le cas sans contraintes, on a le théorème de convergence suivant:

Théorème 2.4.1 *Si $f : \mathbb{R}^N \rightarrow \mathbb{R}$ est une fonction elliptique et l'ensemble K est donné par (2.4.1), alors la méthode de relaxation pour la minimisation de f sur K est bien définie et converge. Voir [1]*

2.4.2 Méthodes de projection

Méthodes de gradient avec projection

Dans le cadre des algorithmes de minimisation avec contraintes que nous allons développer maintenant, nous considérerons $E = \mathbb{R}^N$; $f \in C^1(\mathbb{R}^N, \mathbb{R})$ une fonction convexe, et K fermé convexe non vide. On cherche à calculer une solution approchée de $\bar{x}$, solution du problème:

$$\begin{cases} \bar{x} \in K \\ f(\bar{x}) = \inf_K f \end{cases}. \quad (2.4.2)$$

Algorithme du gradient a pas fixe avec projection sur K (GPFK)

Soit $\rho > 0$ donne, on considère l'algorithme suivant:

Algorithme (GPFK)

Initialisation: $x_0 \in K$

Itération: x_n connu $x_{n+1} = p_K(x_n - \rho f(x_n))$

ou p_K est la projection sur K .

Lemme 2.4.1 Soit $(x_n)_n$ construite par l'algorithme(GPFK). On suppose que $x_n \rightarrow x$ quand $n \rightarrow +\infty$. Alors x est solution de (2.4.2).

Théorème 2.4.2 (Convergence de l'algorithme GPFK) Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$, et K convexe fermé non vide. On suppose que:

1. Il existe $\alpha > 0$, tel que $\langle \nabla f(x) - \nabla f(y)(x - y) \rangle \geq \alpha \|x - y\|^2$, pour tout $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$.
2. Il existe $M > 0$, tel que $\|\nabla f(x) - \nabla f(y)(x - y)\| \leq M \|x - y\|$ pour tout $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$, alors:
 - a. Il existe un unique élément $\bar{x} \in K$ solution de (2.4.2),
 - b. Si $0 < \rho < \frac{2\alpha}{M^2}$, la suite $(x_n)_n$ définie par l'algorithme (GPFK) converge vers $\bar{x}$ lorsque $n \rightarrow +\infty$. Voir [7]

Algorithme du gradient a pas optimal avec projection sur K (GPOK)

L'algorithme du gradient à pas optimal avec projection sur K s'écrit:

Initialisation $x_0 \in K$

Itération x_n connu

$$w_n = -\nabla f(x_n); \text{ calculer } \rho_n \text{ optimal dans la direction } w_n$$

$$x_{n+1} = p_K(x_n + \rho_n w_n).$$

Remarque 2.4.1 On pourrait aussi utiliser un algorithme de type Quasi-Newton avec projection sur K . Les algorithmes de projection sont simples à décrire, mais ils soulèvent deux questions:

1. Comment calculent on p_K ?
2. Que faire si K n'est pas convexe?

2.4.3 Méthodes de dualité

Supposons que les hypothèses suivantes sont:

$$\left\{ \begin{array}{l} f \in C^1(\mathbb{R}^N, \mathbb{R}). \\ g_i \in C^1(\mathbb{R}^N, \mathbb{R}), i = 1, \dots, p. \\ K = \{x \in \mathbb{R}^N, g_i(x) \leq 0, i = 1, \dots, p\}. \end{array} \right.$$

On définit un problème "primal" comme étant le problème de minimisation d'origine.
C-à-d:

$$\left\{ \begin{array}{l} \bar{x} \in K \\ f(\bar{x}) \leq f(x) \text{ pour tout } x \in K \end{array} \right.$$

On définit le "Lagrangien" comme étant la fonction L définie de $\mathbb{R}^N \times \mathbb{R}^p$ dans $\mathbb{R}$ par:

$$L(x, \lambda) = f(x) + \lambda \cdot g(x) = f(x) + \sum_{i=1}^p \lambda_i g_i(x),$$

avec $g(x) = (g_1(x), \dots, g_p(x))^T$ et $\lambda = (\lambda_1, \dots, \lambda_p)^T$: On note $\mathbb{R}_+^p$ ensemble défini par:

$$\mathbb{R}_+^p = \{\lambda \in \mathbb{R}^p, \lambda = (\lambda_1, \dots, \lambda_p), \lambda_i \geq 0 \text{ pour tout } i = 1, \dots, p\}.$$

Remarque 2.4.2 Le théorème de Kuhn-Tucker entraîne que si $\bar{x}$ est solution du problème primal alors il existe $\lambda \in \mathbb{R}_+^p$ tel que $D_1 L(\bar{x}, \lambda) = 0$ (c'est-à-dire $Df(\bar{x}) + \lambda \cdot Dg(\bar{x}) = 0$) et $g(\bar{x}) = 0$. On définit alors l'application M de $\mathbb{R}^p$ dans $\mathbb{R}$ par: $M(\lambda) = \inf_{x \in \mathbb{R}^N} L(x, \lambda)$, pour tout $\lambda \in \mathbb{R}^p$. On peut donc remarquer que $M(\lambda)$ réalise le minimum (en x) du problème sans contrainte, suivant

$$\left\{ \begin{array}{l} x \in \mathbb{R}^N \\ L(x, \lambda) = \inf_{y \in \mathbb{R}^N} L(y, \lambda). \end{array} \right. \quad (2.4.3)$$

Lemme 2.4.2 L'application M de $\mathbb{R}^p$ dans $\mathbb{R}$ définie précédent est concave (l'application $-M$ est convexe), c'est-à-dire que pour tout $\lambda, \mu \in \mathbb{R}^p$ et pour tout $t \in]0, 1[$ on a

$$M(t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu).$$

Preuve. Soit $\lambda, \mu \in \mathbb{R}^p$ et $t \in]0, 1[$, on veut montrer que $M(t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu)$. Soit $x \in \mathbb{R}^N$, alors:

$$L(x, t\lambda + (1-t)\mu) = f(x) + (t\lambda + (1-t)\mu)g(x) = tf(x) + (1-t)f(x) + (t\lambda + (1-t)\mu)g(x).$$

On a donc

$$L(x, t\lambda + (1-t)\mu) = tL(x, \lambda) + (1-t)L(x, \mu).$$

Par définition de M , on en déduit que pour tout $x \in \mathbb{R}^N$,

$$L(x, t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu).$$

Or, toujours par définition de M :

$$M(t\lambda + (1-t)\mu) = \inf_{x \in \mathbb{R}^N} L(x, t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu).$$

On considère maintenant le problème d'optimisation dit "dual" suivant:

$$\begin{cases} \mu \in \mathbb{R}_+^p. \\ M(\mu) \geq M(\lambda), \forall \lambda \in \mathbb{R}_+^p. \end{cases} \quad (2.4.4)$$

Voir [1]. ■

Proposition 2.4.1 *Sous les hypothèses précédentes, soit L définie par $L(x, \lambda) = f(x) + \lambda g(x)$ et $(\bar{x}, \mu) \in \mathbb{R}^N \times \mathbb{R}_+^p$ un point selle de L sur $\mathbb{R}^N \times \mathbb{R}_+^p$, alors:*

- 1) $\bar{x}$ est solution du problème primal,
- 2) μ est solution du problème dual,
- 3) $\bar{x}$ est solution du problème (de minimisation sans contrainte) avec $\lambda = \mu$.

On admettra cette proposition.

Réciproquement, on peut montrer que (sous des hypothèses convenables sur f et g), si μ est solution de dual, et si x solution de (de minimisation sans contrainte) avec $\lambda = \mu$, alors $(\bar{x}, \mu)$ est un point selle de L , et donc $\bar{x}$ est solution de primal.

De ces résultats découle l'idée de base des méthodes de dualité: on cherche μ solution de dual. On obtient ensuite une solution $\bar{x}$ du problème primal, en cherchant x comme solution du problème de minimisation sans contrainte) avec $\lambda = \mu$. La recherche de la solution μ du problème dual peut se faire par exemple par l'algorithme très classique d'Uzawa, que nous décrivons maintenant

Algorithme d'Uzawa

L'algorithme d'Uzawa consiste à utiliser l'algorithme du gradient à pas fixe avec projection (**GPFK**) pour résoudre de manière itérative le problème dual (2.4.4). On cherche donc $\mu \in \mathbb{R}_+^p$ tel que $M(\mu) \geq M(\lambda)$ pour tout $\lambda \in \mathbb{R}_+^p$. On se donne $\rho > 0$; et on note $P_{\mathbb{R}_+^p}$ la projection sur le convexe $\mathbb{R}_+^p$. L'algorithme (**GPFK**) pour la recherche de μ s'écrit donc :

Initialisation: $\mu_0 \in \mathbb{R}_+^p$.

Itération : $\mu_{n+1} = P_{\mathbb{R}_+^p}(\mu_n + \rho \nabla M(\mu_n))$.

Pour définir complètement l'algorithme d'Uzawa, il reste à préciser les points suivants:

1. Calcul de $\nabla M(\mu_n)$.
2. calcul de $P_{\mathbb{R}_+^p}(\lambda)$, λ dans $\mathbb{R}^p$.

La réponse au point 2: est simple pour $\lambda \in \mathbb{R}^p$, on calcule $P_{\mathbb{R}_+^p}(\lambda) = \sigma$ avec $\sigma = (\sigma_1, \dots, \sigma_p)$ en posant $\sigma_i = \max(0, \lambda_i)$, pour $i = 1, \dots, p$, où $\lambda = (\lambda_1, \dots, \lambda_p)^T$. La réponse au point 1: est une conséquence de la proposition suivante.

On suppose que pour tout $\lambda \in \mathbb{R}^p$, le problème (2.4.3) admet une solution unique, notée x_λ et on suppose que l'application définie de $\mathbb{R}^p \rightarrow \mathbb{R}^N$ définie par $\lambda \rightarrow x_\lambda$ est différentiable.

Alors: $M(\lambda) = L(x_\lambda, \lambda)$, M est différentiable en λ pour tout $\lambda \in \mathbb{R}^p$ et $\nabla M(\lambda) = g(x_\lambda)$.

En conséquence, pour calculer $\nabla M(\lambda)$, on est ramené à chercher x_λ solution du problème de minimisation sans contrainte (2.4.3). On peut donc maintenant donner le détail de l'itération générale de l'algorithme d'Uzawa :

Itération de l'algorithme d'Uzawa. Soit $\mu_n \in \mathbb{R}_+^p$ connu;

1. On cherche $x_n \in \mathbb{R}^N$ solution de:

$$\begin{cases} x_n \in \mathbb{R}^N; \\ L(x_n, \mu_n) \leq L(x, \mu_n), \forall x \in \mathbb{R}^N \end{cases}$$

(On a donc $x_n = x_{\mu_n}$)

2. On calcule $\nabla M(\mu_n) = g(x_n)$

$$3. \bar{\mu}_{n+1} = \mu_n + \rho \nabla M(\mu_n) = \mu_n + \rho g(\mu_n) = ((\bar{\mu}_{n+1})_1, \dots, (\bar{\mu}_{n+1})_p)^T$$

$$4. \mu_{n+1} = P_{\mathbb{R}_+^p}(\bar{\mu}_{n+1}), \text{ c'est-à-dire } \mu_{n+1} = ((\mu_{n+1})_1, \dots, (\mu_{n+1})_p)^T, \text{ avec } (\mu_{n+1})_i = \max(0, (\bar{\mu}_{n+1})_i)$$

pour tout $i = 1, \dots, p$.

On a alors le résultat suivant de convergence de l'algorithme

1. Il existe $\alpha > 0$, tel que $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2$, pour tout $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$.

2. Il existe $M_f > 0$ tel que $\|\nabla f(x) - \nabla f(y)(x - y)\| \leq M_f \|x - y\|$ pour tout $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$

3. Pour tout $\lambda \in \mathbb{R}_+^p$, il existe un unique élément $x_\lambda \in \mathbb{R}^N$ tel que $L(x_\lambda, \lambda) \leq L(x, \lambda)$ pour tout $x \in \mathbb{R}^N$.

Alors si $0 < \rho < \frac{2\alpha}{M^2}$, la suite $((x_n, \mu_n))_n \in \mathbb{R}^N \times \mathbb{R}_+^p$ donnée par l'algorithme d'Uzawa vérifie:

1. $x_n \rightarrow \bar{x}$ quand $n \rightarrow +\infty$ où $\bar{x}$ est la solution du problème primal,
2. $(\mu_n)_{n \in \mathbb{N}}$ est bornée. Voir [7]

Bibliographie

- [1] **Ciuperca Ionel Sorin**; Cours d'optimisation; Cours à l'ISFA, en M1SAF, Paris. 1984.
- [2] **D. Azé, J.B. Hiriart-Urruty**; Analyse variationnelle et optimisation; Cépaduès, 2010.
- [3] **M. Bergounioux**; Optimisation et contrôle des systèmes linéaires; Dunod, Paris, 2001.
- [4] **P.G. Ciarlet**; Introduction à l'analyse numérique matricielle et à l'optimisation; Masson, Paris. 1982.
- [5] **P.G. Ciarlet, B. Miara, J.M. Thomas**; Exercices l'analyse numérique matricielle et d'optimisation; Masson, Paris. 1987.
- [6] **Raphaële Herbin**; Cours d'Analyse numérique; Université Aix Marseille 1 Licence de mathématiques, 2003-2004, 8 juillet 2004.
- [7] **Raphaële Herbin**; Cours d'Analyse numérique; Université Aix Marseille Licence de mathématique, 18 novembre 2013.

Résumé

L'objet essentiel de ce mémoire est de donner une description et une analyse relativement complètes des méthodes les plus couramment utilisées en optimisation sous contraintes dans le cas convexe. Particulièrement les méthodes de projection et la méthode de dualité.

La convexité joue un rôle très important à l'existence et à l'unicité de la solution et aussi à la convergence de ces méthodes.

Mots clés: Optimisation, Convexité, Projection, Programmation, convexe, Dualité.

ملخص:

قدمنا في هذا العمل الطرق الأكثر استعمالاً لحل مسائل optimisation sous contraintes في حالة المحدبة. خاصة طرق الإسقاط وطريقة dualité. خاصية التحدب مهمة في إثبات وجود وحدانية حل هذه المسائل وإثبات تقارب الطرق المذكورة.

Abstract:

The main purpose of this paper is to provide a relatively complete description and analysis of the methods most commonly used in constrained optimization in the convex case. Particular projection methods and the method of duality.

Convexity plays a very important role in the existence and uniqueness of the solution and also the convergence of these methods.

Keywords: Optimization, Convexity, Projection, convex Programming, Duality.