

République Algérienne Démocratique et Populaire.
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.
Université de Mouloud Mammeri, Tizi-Ouzou

Faculté des Sciences
Département des Mathématiques

Projet de Fin d'Etudes
En vue de l'obtention du Diplôme de Master
En Recherche Opérationnelle

Thème

Programmation mathématique et quelques applications

Présenté par:

Hamdidi Zakia
Bettahar Ghania

Devant le jury d'examen composé de :

<i>M^r</i> B. Oukacha	UMMTO	President
<i>M^r</i> M. Ouanes	UMMTO	Promoteur
<i>M^r</i> K. Merakeb	UMMTO	Examineur.

Sutenué le 06-10-2014

Remerciements

Nous remercions tout d'abord «DIEU» tout puissant d'avoir donné la santé et le courage pour effectuer ce projet de fin d'étude, dans des meilleures conditions comme nous tenons à adresser toutes les reconnaissances et notre gratitude à notre promoteur: *M^r* M.Ouanes pour ses précieux conseils et son suivi.

Nous remercions les membres du jury d'avoir accepté d'examiner notre travail.

Enfin, nous tenons à remercier tous ceux qui ont contribué de près et de loin à la réalisation de ce travail.

Dédicace

C'est avec de profonds sentiments de respect que je dédie ce mémoire :

♡ **A** mes chers parents qui m'ont tant donné , soutenu durant mon parcours universitaire et qui ont toujours été là pour moi ; ceux qui m'ont donné un honorable modèle de labeur et de persévérance , qu'ils trouvent ici l'expression de toute ma profond reconnaissance et ma parfaite considération.

♡ **A** mon promoteur *M^r* **ouanes** .

♡ **A** mon frère **Sid ali**

♡ **A** mes sœurs: **Nadjia , Malika , Hayat , Tiziri.**

♡ **A Arezki aklouche** qui ma donné le courage pendant mon parcours universitaire.

♡ **A** toute l'université **UMMTO.**

Dédicace

C'est avec de profonds sentiments de respect que je dédie ce mémoire :

♡ **A** mes chers parents qui m'ont tant donné , soutenu durant mon parcours universitaire et qui ont toujours été là pour moi ; ceux qui mon donné un honorable modèle de labeur et de persévérance , qu'ils trouvent ici l'expression de toute ma profond reconnaissance et ma parfait considération.

♡ **A** mon promoteur *M^r* **ouanes**.

♡ **A** mes frères **Wahab** , **Walid** et **youcef**.

♡ **A** mes sœurs **Nora** , **Hassina**.

♡ **A** mes amies .

♡ **A** toute l'université **UMMTO**.

Table des matières

Introduction générale :	5
1 Rappels	9
1.1 Introduction :	9
1.2 Espace euclidien :	9
1.3 La norme euclidienne :	10
1.3.1 Inégalité de Cauchy-Schwarz :	10
1.4 Notions topologiques :	11
1.4.1 Théorème de Weierstrass :	13
1.5 La convexité :	13
1.5.1 Ensemble convexe :	13
1.5.2 Fonction convexe :	13
1.6 Fonction d'appui :	16
1.7 Projection d'un point sur un ensemble convexe :	17
1.8 Programme convexe :	19
1.9 Les conditions nécessaires de Karush-Kuhn-Tucher :	22
1.9.1 Qualification des contraintes :	22
1.10 Extention à des problèmes avec contraintes d'égalités et d'inégalités :	
Condition de Lagrange :	24
1.11 Conclusion :	27
2 La dualité lagrangienne	28
2.1 Introduction :	28
2.2 Théorème faible de dualité :	30
2.3 Concavité de la fonction dual :	31
2.4 Théorème de dualité :	33

2.5	La dualité en programmation linéaire:	35
2.6	Saut de dualité:	36
2.7	Conclusion:	37
3	méthodes des moindres carrés	38
3.1	Introduction:	38
3.2	Notion de modèle et de régression linéaire multiple:	38
3.3	Critère des moindres carrés:	39
3.4	La forme standard:	40
3.5	Recherche d'une solution:	41
3.5.1	Solution géométrique:	41
3.5.2	Solution analytique:	42
3.5.3	Théorème:	42
3.5.4	Dérivation matricielle:	43
3.5.5	Proposition:	44
3.6	La forme quadratique:	45
3.7	Calcul de solution:	46
3.8	Théorème:	47
3.9	Conclusion:	48
4	La méthode SVM	49
4.1	Introduction:	49
4.2	Définition:	49
4.3	Hyperplan séparateur linéaire:	50
4.4	Rappel de géométrie:	50
4.4.1	Equation d'un hyperplan:	50
4.4.2	Distance à un hyperplan:	50
4.5	Hyperplan de marge maximal:	51
4.6	Problème primal:	51
4.7	Problème dual:	52
4.8	Avantages et inconvénients:	53
4.8.1	Avantages:	53
4.8.2	Inconvénients:	54
4.9	Conclusion:	54

Table des matières	5
Conclusion générale :	54
Bibliographie	56

Introduction générale:

La programmation mathématique se propose pour objet l'étude théorique des problèmes d'optimisation ainsi que la conception et la mise en œuvre des algorithmes de résolution. Ses applications sont extrêmement nombreuses et variées que se soit dans les sciences de l'ingénieur ou dans d'autres domaines des mathématiques appliquées, notamment en recherche opérationnelle, en analyse numérique, en automatique, en ingénierie, en économie mathématique etc...l'impact économique des méthodes et des outils (logiciel) issue de la programmation est aujourd'hui considérable. Des milliers d'entreprises les utilisent quotidiennement pour résoudre des problèmes liés à l'optimisation de leur rentabilité : des problèmes de localisation, de gestion de production, de logistique et de transport, de gestion de stocks, de tarification, d'optimisation de flux dans les réseaux ...

Dans la programmation mathématique, on peut indiquer deux tendances . La première, parfaitement constituée, est la programmation mathématique elle-même, elle traite des problèmes dans lesquels toute l'information initiale est complètement définie. La deuxième, dite programmation stochastique, concerne les problèmes dans lesquels l'information comporte des éléments indéterminés, ou bien les problèmes dont certains paramètres sont aléatoires mais définis par des caractéristiques probabilistes connues.

Tel est le cas, par exemple des programmes d'activité industrielle établis souvent dans des conditions d'une information incomplète sur la situation réelle de leur réalisation. Ou disons, celui d'un problème d'extrémum qui traduit le fonctionnement des instruments automatiques accompagnés de perturbation aléatoire. Notons que l'une des plus grandes difficultés de la

programmation stochastique réside dans la position de problème, surtout du fait que l'analyse de l'information initiale est très compliquée.

Voici les sections principales de la programmation mathématique devenues classiques.

Programmation linéaire: la fonction économique est linéaire; l'ensemble sur lequel on cherche l'extrémum de cette fonction et donné par un système linéaire d'égalités et d'inégalités.

La programmation linéaire comporte des méthodes spéciales pour leur résolution, bien plus avantageuses que celles relatives aux problèmes de forme générale. Ainsi on a vu apparaître dans la programmation linéaire les problèmes de transport.

Programmation non linéaire: la fonction économique et/ou les contraintes sont non linéaires. D'après l'usage, la programmation non linéaire compte les divisions suivantes :

Programmation convexe: la fonction économique (s'il s'agit de la minimisation) et l'ensemble sur lequel on cherche à résoudre le problème d'extrémum sont convexes.

programmation quadratique: la fonction économique est quadratique, les contraintes étant des égalités et des inégalités linéaires.

Problèmes à extrêmums multiples: on relève d'ordinaire dans ce domaine des classes spéciales des problèmes très fréquents dans les applications, par exemple, les problèmes relatifs à la minimisation des fonctions concaves sur un ensemble convexe.

Programmation en nombres entiers: dans laquelle les variables sont soumises à la contrainte d'intégrité.

Quel est le caractère particulier des problèmes de programmation mathématique?

En premier lieu, il est en général impossible d'appliquer à ces problèmes les méthodes de l'analyse classique portant sur la recherche des extrêmes liés; il en est ainsi du fait que même dans le cas le plus simple, celui des problèmes linéaires, ces problèmes admettent un extrême au point anguleux de la frontière de l'ensemble des contraintes, c'est-à-dire aux points où la dérivabilité est violée.

Le mémoire est organisé comme suit:

On commence avec une introduction générale dans laquelle on a parlé sur la programmation mathématique dont on a indiqué ces tendances et ces sections principales.

Le premier chapitre contient un survol rapide: espace euclidien, la norme euclidienne, inégalité de Cauchy-Schwarz, notion topologique, la convexité (ensemble convexe, fonction convexe), programme convexe, les conditions nécessaires de K-K-T, extension à des problèmes avec contraintes d'égalités et d'inégalité.

Le deuxième chapitre est réservé pour la présentation de la dualité lagrangienne.

Le troisième chapitre est réservé pour la méthode des moindres carrés.

Le quatrième chapitre est réservé pour la méthode SVM.

Chapitre 1

Rappels

1.1 Introduction :

Dans ce chapitre, on rappellera quelques notions de base sur la programmation mathématique, dont on a besoin dans les chapitres suivants.

1.2 Espace euclidien:

Une collection des points (vecteurs) $x^t = (x_1, x_2, \dots, x_n)^t$ espace à n coordonnées réelles x_1, x_2, \dots, x_n s'appelle **espace euclidien de dimension n** noté E_n si les conditions suivantes sont vérifiées :

Soient $x \in E_n$, $y \in E_n$ et α un nombre réel , alors :

1) $x + y = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$.

2) $\alpha x = (\alpha x_1, \alpha x_2, \dots, \alpha x_n)$.

3) $x \cdot y = \sum_{i=1}^n x_i y_i$.

1.3 La norme euclidienne:

Dans un espace euclidien on introduit la notion de norme euclidienne $\|x\| = \sqrt{x \cdot x}$, qui vérifié les relations suivantes:

$$1) \|x\| \geq 0, \text{ où l'inégalité } \|x\| = 0 \quad \text{Si } x = 0.$$

$$2) \|\alpha x\| = |\alpha| \cdot \|x\|.$$

$$3) \|x + y\| \leq \|x\| + \|y\|.$$

Exemple: Les trois normes usuelles pour les matrices sont :

$$\checkmark \|A\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}|$$

$$\checkmark \|A\|_2 = \sqrt{p(A^2)}, \text{ avec:}$$

$p(A^2)$: rayon spectral de A

$$\|A\|^2 = \max_{1 \leq i \leq n} (|\lambda_i|).$$

λ_i : valeur propre de A^2

$$\checkmark \|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$$

1.3.1 Inégalité de Cauchy-Schwarz:

Théorème:

soit $P = (x_1, x_2, \dots, x_n)$, $Q = (y_1, y_2, \dots, y_n)$, alors:

$$|P \cdot Q| \leq \|P\| \cdot \|Q\|.$$

1.4 Notions topologiques:

Définition 1:

$y \in S$ est un point intérieur de S s'il existe $\epsilon > 0$ tel que:

$$\|x - y\| < \epsilon \Rightarrow x \in S$$

autrement dit s'il existe une boule centrée en y et contenue dans S

L'ensemble des points intérieurs à S est appelé l'intérieur de S et est notés $\text{int}(S)$.

Exemple:

L'ensemble $A = \{x \in \mathbb{R}^n / \|x\| \leq 1\}$ a pour intérieur la boule unité $B_1(0) = \{x / \|x\| < 1\}$.

Définition 2:

une partie S est dite ouverte s'il coïncide avec son intérieur c'est à dire:

$$S = \text{int}(S)$$

Exemple:

Soit $r > 0$, l'ensemble des points $Br(A) = \{P \in \mathbb{R}^n / \|P - A\| < r\}$ est une boule de centre A et de rayon r .

Définition 3:

$S \subseteq \mathbb{R}^n$ est dit fermé si son complémentaire est ouvert.

Exemple:

L'ensemble $Sr(A) = \{P \in \mathbb{R}^n / \|P - A\| = r\}$ est appelé Sphère de centre A et de rayon r .

Exemple:

Dans \mathbb{R}^2 :

$Br(A)$ est un **disque** de centre A et de rayon r ,

$Sr(A)$ est un **cercle** de centre A et de rayon r .

Remarque :

$\|P - A\|$ est la distance de P à A .

1) \mathbb{R}^n et \emptyset sont ouverts et fermés en même temps.

2) On a des ensembles qui ne sont ni ouverts, ni fermés $[a,b[$.

3) $Br(A)$ est appelé **voisinage** de A .

Définition 4:

Un ensemble S dans \mathbb{R}^n est dit **compact** s'il est fermé et borné en même temps dans $[a,b]$.

1.4.1 Théorème de Weierstrass:

Toute fonction continue sur un compact est **bornée** de plus elle **atteint ses bornes**.

1.5 La convexité:**1.5.1 Ensemble convexe:**

Un ensemble $X \subset E_n$ est dit convexe si pour tout couple de points $x \in X$ et $y \in X$, le segment $[x,y]$ qui les relie lui appartient aussi.

La convexité de l'ensemble X signifie que:

$$\forall x \in X, \forall y \in X, \forall \lambda \in [0,1]:$$

$$\lambda x + (1 - \lambda)y \in X.$$

1.5.2 Fonction convexe:**Définition 1:**

On dit qu'une fonction scalaire f est convexe sur un ensemble convexe X si quels que soient $x, y \in X$, $\lambda \in [0,1]$, on a l'inégalité:

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

Par contre, si

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y),$$

la fonction f est dite concave.

Il est clair que si f est convexe, alors $-f$ est concave .

Quelque fois, pour simplifier les calculs des démonstration, on utilise la définition suivante:

Une fonction continue $f(x)$ est convexe sur un ensemble convexe X si, quelque que soient $x, y \in X$, on a la condition :

$$f\left(\frac{1}{2}x + \frac{1}{2}y\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y).$$

On voit sans peine que la somme des fonctions convexes (concaves) est une fonction convexe (concave).

Resultat:

1) f est de classe c^1 , f est convexe sur X si:

$$f(x^1) - f(x^2) \geq (x^1 - x^2) \nabla f(x^2), \quad \forall x^1, x^2 \in X.$$

Définition:

On définit la dérivée partielle première par rapport à x_i de la fonction $f(x_1, x_2, \dots, x_n)$:

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

elle est notée par $\frac{\partial f(x_1, \dots, x_n)}{\partial x_i}$ on définit le gradient par :

$$\nabla f(x_1, \dots, x_n) = \begin{pmatrix} \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix}$$

2) f est de classe C^2 , f est convexe sur \mathbb{R}^n si et seulement si :
 H_f est semi définie positive sur \mathbb{R}^n .

Définition:

La matrice Hessianne de $f(x_1, \dots, x_n)$ est de la forme:

$$\mathbf{H}_{f(x_1, \dots, x_n)} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Définition 2:

On appelle **sous gradient de f** au point x^o , tout vecteur $\varphi \in \mathbb{R}^n$ vérifiant:

$$f(x) \geq f(x^o) + \varphi^T(x - x^o) \quad \forall x \in \mathbb{R}^n.$$

Définition 3:

L'ensemble des sous gradients de f en x^o est appelé

Sous différentiel noté $\partial f(x^o)$ tel que:

$$\partial f(x^o) = \text{conv}\{\nabla f_i(x^o)\}.$$

Définition:

On appelle enveloppe convexe d'un ensemble S , le plus petit ensemble convexe contenant S et on noté **conv(S)**.

1.6 Fonction d'appui:**Définition:**

On appelle **fonction d'appui** pour la fonction f au point x^o :

La fonction $f_{app(x^o)}(x) = f(x^o) + \varphi^T(x - x^o)$.

Tel que :

1) $f_{app(x^o)}(x^o) = f(x^o)$.

2) $f_{app(x^o)}(x) \leq f(x), \forall x \in \mathbb{R}^n$.

3) $f_{app(x^o)}(x)$ est linéaire.

Remarque:

Si f est différentiable en x^o alors $\partial f(x^o) = \{\nabla f(x^o)\}$.

1.7 Projection d'un point sur un ensemble convexe :

Définition :

On appelle projection d'un point v sur un ensemble convexe X , un point $p = p(v)$ de X tel que :

$$\|p - v\| = \inf_{x \in X} \|x - v\| = \delta$$

On dit que δ est la distance de v à p .

Théorème 1 :

Quels que soient l'ensemble convexe fermé X et le point v , il existe un point unique $p \in X$, projection de v sur X .

Théorème 2 :

La projection, p_{x^o} de x^o sur un convexe X fermé est **unique**, tel que $v = x^o$.

Proposition 1 :

Pour que p_{x^o} , soit la projection de x^o sur X convexe fermé, il faut et il suffit que :

$$(x - p_{x^o})(x^o - p_{x^o}) \leq 0, \forall x \in X.$$

Définition :

On définit un **hyperplan** par :

$$H = \{x \in \mathbb{R}^n / cx = \alpha, c \in \mathbb{R}^n, \alpha \in \mathbb{R}\} .$$

Remarque :

1) Dans \mathbb{R}^2 :

$(c_1, c_2)(x_1, x_2) = \alpha, c_1x_1 + c_2x_2 = \alpha$ est une **droite**.

2) Dans \mathbb{R}^3 c'est un **plan**.

Définition :

On définit un demi espace par $E = \{x \in \mathbb{R}^2 / cx \leq \alpha\}$.

Proposition 2 :

Soit $x^o \notin X$ convexe fermé alors il existe un hyperplan :

$H = \{x \in \mathbb{R}^n / cx = \alpha\}$ Tel que :

$$cx_o = \alpha \text{ et } cx < \alpha \quad \forall x \in X.$$

Définition :

H est appelé **hyperplan d'appui** au convexe X en un points x_o si toutes les points de X sont de même côté de H .

Définition :

On appelle **epigraphe** de $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ L'ensemble $epi(f) = \{(x,r) \in X \times \mathbb{R} / f(x) \leq r\}$.

Proposition :

Une fonction f est convexe si et seulement si son $epi(f)$ est convexe.

1.8 Programme convexe :

Définition 1 :

Un problème de programmation mathématique est convexe s'il consiste à minimiser une fonction convexe (resp maximiser une fonction concave) sur un domaine convexe fermé.

Soit :

$$(P) \left\{ \begin{array}{l} \min f(x) \\ g_i(x) \leq 0, \quad i = 1, \dots, m \\ x \in S \subset \mathbb{R}^n, \quad f, g_i \text{ convexes} \end{array} \right.$$

(P) est un problème de programmation convexe (ou simplement:un programme convexe).

La propriété fondamentale de programme convexe apparaitre alors dans le résultat suivant :

Théorème :

Pour un programme convexe, tout optimum local est un optimum global.

Démonstration :

On peut toujours considérer un programme convexe de la forme :

$$\left\{ \begin{array}{l} \text{minimiser } f(x) \\ \text{sous la contrainte :} \\ x \in S \end{array} \right.$$

Ou f est une fonction convexe et S est un ensemble convexe.
Soit x^o un optimum local, pour montrer que x^o est un optimum global. Considérons $y \in S (y \neq x^o)$ quelconque et montrons que nécessairement: $f(x^o) \leq f(y)$.

Raisonnons par l'absurde :

En supposant que $f(x^o) > f(y)$.

En vertu de la convexité de f , ceci implique que:

$$\forall \theta \in]0,1], f(x^o + \theta(y - x^o)) \leq (1 - \theta)f(x^o) + \theta f(y) < f(x^o)$$

D'où il résulte une contradiction avec le fait que x^o est un optimum local.

Donc pour tout $y \in S$, on a bien: $f(x^o) \leq f(y)$ ce qui assure que x^o est un optimum global.

Condition nécessaire d'optimalité :

On s'intéresse ici au problème suivant:

$$(P') \left\{ \begin{array}{l} \text{Minimiser } f(x) \\ \text{sous les contraintes:} \\ g_i(x) \leq 0 \quad i \in I = \{1, 2, \dots, m\} \\ x \in \mathbb{R}^n \end{array} \right.$$

Qui n'est autre que le problème (P) avec la condition $S \equiv \mathbb{R}^n$.

Toutes les fonctions f et $g_i(x)$ ($i \in I$) sont supposées continues et différentiables.

On notera X l'ensemble des solutions de (P') c'est à dire :

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

Nous supposerons que X est non vide, par contre X peut très bien avoir un intérieur vide, c'est le cas, en particulier si l'on impose que certaines contraintes soient vérifiées l'égalité de Lagrange suivante :

$$\left\{ \begin{array}{l} \nabla f(x) + \sum_{i=1}^n \lambda_i \nabla g_i(x) = 0 \\ \lambda_i g_i(x) = 0 \quad i = 1 \dots m. \end{array} \right.$$

1.9 Les conditions nécessaires de Karush-Kuhn-Tucher :

1.9.1 Qualification des contraintes :

Lemme1 :

* Pour que (QC) soit vérifiée en tout point $x \in X$ il suffit que l'une des conditions (a) ou (b) soit réalisée :

a) Toutes les fonctions g_i sont linéaires ou affines (**Karlin 1959**) .

b) Toutes les fonctions g_i sont convexes différentiables et il existe $\bar{x} \in \mathbb{R}^n$ vérifiant :

$$g_i(\bar{x}) < 0, \forall i \in I(\text{Slater 1950}) .$$

* Pour (QC) soit vérifiée en un point $x^o \in X$, il suffit que l'on ait :

c) Les gradients $\nabla g_i(x^o)$ ($i \in I^o$) des contraintes saturées en x^o sont linéairement indépendants (Fiacco et McCormick 1968) .

* La référence suivante est fondamentale et donne, sous l'hypothèse de qualification des contraintes, une condition nécessaire d'optimalité local pour un problèmes d'optimisation avec contraintes du type P' .

Référence 1 : (Karush, Kuhn et Tucker 1951) :

On suppose que les fonctions f et g_i ($i \in I$) sont continûment différentiables et que l'hypothèse de qualification des contraintes est vérifiée en $x^o \in X$ avec :

$$X = \{x \in \mathbb{R}^n / g_i \leq 0, \forall i \in I\}.$$

Alors une condition nécessaire pour que x^o soit un optimum local de P' et qu'il existe des nombres $\lambda_i \geq 0$ ($i \in I$) appelé :

Multiplicateurs de Karush-Kuhn-Tucker tel que :

$$KKT \left\{ \begin{array}{l} \nabla f(x) + \sum_{i \in I} \lambda_i \nabla g_i(x^o) = 0 \\ \text{et:} \\ \lambda_i g_i(x^o) = 0 \quad (\forall i \in I) \end{array} \right.$$

1.10 Extention à des problèmes avec contraintes d'égalités et d'inégalités : Condition de Lagrange:

Les conditions de Karush-Kuhn-Tucker s'étendent sans difficulté à des problèmes comportant à la fois des contraintes d'égalités et des contraintes d'inégalités de la forme :

$$(P') \left\{ \begin{array}{l} \text{Min } f(x) \\ \text{Sous contraintes:} \\ g_i(x) \leq 0 \quad i \in I = \{1, 2, \dots, m\} \\ h_l(x) = 0 \quad l \in L = \{1, 2, \dots, p\} \\ x \in \mathbb{R}^n \end{array} \right.$$

Reference 2 :

On suppose que les fonctions f, g_i ($i \in I$) et h_l ($l \in L$) sont continûment différentiables et que (QC) est vérifiée en x^o solution de P'_1 .

Alors une condition nécessaire pour que x^o soit un optimum local de P'_1 est qu'il existe des nombres $\lambda_i \geq 0$ ($i \in I$) et μ_l ($l \in L$) (μ_l non contraintes en signe) tel que :

$$KKT \left\{ \begin{array}{l} \nabla f(x^o) + \sum_{i \in I} \lambda_i \nabla g_i(x^o) + \sum_{l \in L} \mu_l \nabla h_l(x^o) = 0 \\ \text{et:} \\ \lambda_i g_i(x^o) = 0 \quad (\forall i \in I) \\ h_l(x^o) = 0. \quad \forall l \in L. \end{array} \right.$$

9) Conditions suffisantes d'optimalité :

« Points – cols » **et Fonction de Lagrange :**

Nous allons étudier maintenant des conditions suffisantes d'optimalité

pour des problèmes du type :

$$(P) \left\{ \begin{array}{l} \text{Min} f(x) \\ g_i(x) \leq 0 \quad (i \in I) \\ x \in S \subset \mathbb{R}^n \end{array} \right.$$

Remarquons que quand $S = \mathbb{R}^n$ on retrouve problème (P') .

Néanmoins ce qui va dans ce paragraphe s'applique plus généralement aux problèmes du type (P) . S pourra être par exemple un ensemble de points à coordonnées entières (programmation en nombres entières).

Associons à chaque contraintes $i \in I$ un nombre réel $\lambda_i \geq 0$ appelé **multiplificateur de Lagrange**. La fonction de Lagrange associée au problème (P') , est par définition la fonction :

$$L(x, \lambda) = f(x) + \sum_{i \in I} \lambda_i g_i(x).$$

Dans toute la suite nous supposons que les fonctions f , g_i , ainsi que l'ensemble S ont toutes les propriétés requises pour que le problème $\min_{x \in S} \{L(x, \lambda)\}$ ait une solution optimale, $\forall \lambda \geq 0$.

Définition :

Soit $\bar{x} \in S$ et $\bar{\lambda} \geq 0$.

On dit que $(\bar{x}, \bar{\lambda})$ est un **point-col** pour la fonction de Lagrange si :

$$L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \quad \forall x \in S.$$

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \quad \forall \lambda \geq 0 .$$

Reference 1 :(propriété caractéristique des points-cols)

Soit $\bar{x} \in S$ et $\bar{\lambda} \geq 0$; $(\bar{x}, \bar{\lambda})$ est un point-col pour $L(x, \lambda)$ si et seulement si :

a) $L(\bar{x}, \bar{\lambda}) = \min_{x \in S} L(x, \bar{\lambda})$;

b) $g_i(\bar{x}) \leq 0 \quad (\forall i \in I)$;

c) $\bar{\lambda}_i g_i(\bar{x}) = 0 \quad (\forall i \in I)$.

Reference 3: (suffisance de la condition de point-col)

Si $(\bar{x}, \bar{\lambda})$ et $\bar{\lambda} \geq 0$ est un point-col de $L(x, \lambda)$ alors: \bar{x} est un optimum global de (p') .

1.11 Conclusion:

Ce chapitre constitue un exposé rapide et général des notions de base qui sont indispensable pour la définition d'un problème de programmation mathématique .

Chapitre 2

La dualité lagrangienne

2.1 Introduction :

Considérons un problème de programmation mathématique de type :

$$(P) = \begin{cases} \text{minimiser } f(x) & \text{sous les contraintes :} \\ g_i(x) \leq 0 & i \in I = \{1, \dots, m\} \\ x \in S \subset R^n \end{cases}$$

En associant à chaque contrainte $g_i(x) \leq 0$ un multiplicateur de lagrange $\lambda_i \geq 0$, la fonction de lagrange $L(x, \lambda)$, s'écrit :

$$L(x, \lambda) = f(x) + \sum_{i \in I} \lambda_i g_i(x)$$

le problème (P) peut être résolu si l'on sait déterminer un point-col de la fonction de lagrange c'est-à-dire un couple $(\bar{x}, \bar{\lambda})$ vérifiant :

$$(P) \begin{cases} L(\bar{x}, \bar{\lambda}) = \min_{x \in S} L(x, \bar{\lambda}) \\ g(\bar{x}) \leq 0 \\ \bar{\lambda}_i \cdot g_i(\bar{x}) = 0 \quad (i=1, \dots, m) \end{cases}$$

Considérons maintenant la fonction W définie, pour tout $\lambda \geq 0$, par :

$$W(\lambda) = \inf_{x \in S} \{L(x, \lambda)\}$$

Dans la suite, nous supposons que les fonctions f et g_i et l'ensemble S sont tels que, en tout point $\bar{\lambda}$ où $W(\bar{\lambda})$ a une valeur finie, il existe $\bar{x} \in S$ tel que :

$W(\bar{\lambda}) = L(\bar{x}, \bar{\lambda})$. on pourra donc écrire :

$$W(\bar{\lambda}) = \min_{x \in S} \{L(x, \bar{\lambda})\}$$

Remarquons que cette condition est satisfaite, en particulier, si les fonctions f et g_i sont continues sur S , et que S est compact (Théorème de Weierstrass).

Un autre cas intéressant est celui où S est un ensemble discret de cardinalité finie.

Nous allons voir que la recherche d'un *point – col*, lorsqu'il existe, peut se faire précisément en résolvant le problème :

$$(D) \begin{cases} \max_{\lambda} W(\lambda) = \max_{\lambda} \{ \min_{x \in S} L(x, \lambda) \} \\ \lambda \in \mathbb{R}_+^m \end{cases}$$

(D) est appelé le problème dual de (P). Par opposition (P) appelé le problème primal. W est la fonction dual.

Notons, et c'est un point important, que la fonction dual W et le problème dual (D) sont définis même lorsqu'il n'existe pas de point-col.

2.2 Théorème faible de dualité

Dans ce théorème nous n'avons pas besoin d'hypothèses particulière sur x , ni sur les fonctions f et g_i , $i = 1, \dots, m$.

Pour tout $\lambda \in \mathbb{R}_+^m$, la valeur de la fonction dual $W(\lambda)$ est un minorant de l'optimum global $f(x^*)$ de (P), autrement dit, si $W(\lambda^*)$ est la valeur optimal du problème dual:

$$\forall \lambda \geq 0 : \quad W(\lambda) \leq W(\lambda^*) \leq f(x^*).$$

Démonstration: considérons $\lambda \geq 0$ quelconque.

Par définition:

$$x \in S \implies W(\lambda) \leq f(x) + \sum_{i \in I} \lambda_i g_i(x).$$

pour x solution de (P) on a: $g(x) \leq 0$ donc $\lambda.g(x) \leq 0$ et par suit:

$$W(\lambda) \leq f(x) \quad \forall \lambda \geq 0.$$

En particulier pour x^* solution optimal de (P) on a:

$$W(x) \leq W(x^*) \leq f^*.$$

2.3 Concavité de la fonction dual:

la fonction dual W est une fonction concave de λ .

Démonstration:

Prenons λ^1 et λ^2 quelconques et pour $\theta \in [0,1]$ quelconque définissons $\lambda = \theta\lambda^1 + (1 - \theta)\lambda^2$

Alors, il existe $\bar{x} \in S$ tel que:

$$W(\lambda) = f(\bar{x}) + \sum_{i=1}^m \lambda_i g_i(\bar{x})$$

On a alors par définition de $W(\lambda^1)$, $W(\lambda^2)$:

$$W(\lambda^1) \leq f(\bar{x}) + \sum_{i=1}^m \lambda_i^1 g_i(\bar{x})$$

$$W(\lambda^2) \leq f(\bar{x}) + \sum_i \lambda_i^2 g_i(\bar{x});$$

multiplions la première inéquation par $\theta \geq 0$, la seconde par $(1 - \theta) \geq 0$ et additionnons, il vient:

$$\begin{aligned}\theta W(\lambda^1) + (1 - \theta)W(\lambda^2) &\leq f(\bar{x}) + \sum_i [\theta \lambda_i^1 + (1 - \theta) \lambda_i^2] g_i(\bar{x}) \\ &= f(\bar{x}) + \sum_i \lambda_i g_i(\bar{x}) \\ &= W(\lambda)\end{aligned}$$

Cette dernière propriété est absolument générale et ne suppose rien sur la convexité des fonctions f et g_i , ni sur la convexité de l'ensemble S .

Définition:

une paire (x^*, y^*) de points satisfait les conditions d'optimalité pour le problème primal si:

- ✓ x^* est un minimum global de: $f(x) + \sum_{i=1}^m \lambda_i^* f_i(x)$ sur x .
- ✓ $\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0$.
- ✓ $f_i(x^*) \leq 0, i = 1, \dots, m$.
- ✓ $\lambda_i^* \geq 0, i = 1, \dots, m$.

Pour les prochaines théorèmes nous avons besoin de l'hypothèse de convexité des fonctions f et g_i , pour $i = 1, \dots, m$, sur x convexe.

2.4 Théorème de dualité

(a) Si le problème (P) admet un point-col (x^*, λ^*) alors, on a:

$$\max(D) = W(\lambda^*) = f(x^*) = \min(P).$$

Autrement dit, la valeur optimal de problème primal (P) est égale à la valeur optimal de problème dual (D). (b) Réciproquement, s'il existe x^* solution de (P) et $\lambda^* \geq 0$ tels que:

$$W(\lambda^*) = f(x^*).$$

alors (P) admet un point-col et (x^*, λ^*) est un tel point-col.

Démonstration :

a) Comme (x^*, λ^*) est un point-col; on a:

$$L(x^*, \lambda^*) = f(x^*) + \lambda^* g(x^*) = f(x^*)$$

$$= \min_{x \in S} \{L(x, \lambda^*)\} = W(\lambda^*).$$

D'autre part en vertu de la propriété (01) (théorème faible de dualité) on a, pour tout $\lambda \geq 0$:

$$W(\lambda) \leq f(x^*).$$

On en déduit que:

$$f(x^*) = W(\lambda^*) = \max_{\lambda \geq 0} W(\lambda)$$

(b) Réciproquement, supposons qu'il existe x^* solution de (P) et $\lambda^* \geq 0$ tels que:

$$W(\lambda^*) = f(x^*)$$

Par définition de $W(\lambda^*)$ on a:

$$\forall x \in S : W(\lambda^*) \leq f(x) + \lambda^* g(x)$$

En particulier, en prenant $x = x^*$ on a:

$$W(\lambda^*) = f(x^*) \leq f(x^*) + \lambda^* g(x^*)$$

d'où on déduit $\lambda^* g(x^*) \geq 0$.

comme λ^* et $g(x^*) \leq 0$ on a aussi $\lambda^*g(x^*) \leq 0$ et par suite:

$$\lambda^*g(x^*) = 0$$

Comme $\lambda^*.g(x^*)$ est une somme de termes tous négatifs ou nuls, on en déduit que :

$$\forall i : \lambda^*g_i(x^*) = 0$$

et par suite (x^*, λ^*) est un point-col.

2.5 La dualité en programmation linéaire:

Dans ce cas, f et g_i étant convexes, il existe un point-col et les conditions de k-k-T sont nécessaires et suffisantes.

Considérons le problème suivant:

$$(PL) \begin{cases} \min c^t . x \\ b - Ax \leq 0 \\ x \in \mathbb{R}^n \end{cases}$$

$$c = (c_1, c_2, \dots, c_n)^t .$$

$$b = (b_1, b_2, \dots, b_n)^t .$$

$$A_{(m \times n)} .$$

La fonction de lagrange s'écrit:

$$L(x, \lambda) = c^t \cdot x + \lambda^t (b - Ax)$$

$$W(\lambda) = \min_x \{ (c^t - \lambda^t A)x + \lambda^t \cdot b \}$$

On a donc : $W(\lambda) = \lambda^t \cdot b$ pour tout λ tel que : $c^t - \lambda^t \cdot A = 0$,
 $W(\lambda) = -\infty$ pour les autres valeurs de λ .

Le problème dual:

$$\begin{cases} \max W(\lambda) \\ \lambda \in \mathbb{R}_+^m \end{cases}$$

est alors le programme linéaire:

$$(DPL) \begin{cases} \max \lambda^t b & \text{sous les contraintes:} \\ c^t - \lambda^t A = 0 \\ \lambda \geq 0 \end{cases}$$

Cette propriété est absolument générale et ne suppose rien sur la convexité des fonctions f et g_i ni sur la convexité de l'ensemble S .

2.6 Saut de dualité :

Pour les programmes de type (P) qui n'admettent pas de point-col (c'est le cas général, entre autre, pour les programmes discrets où S est un en-

semble discret) On a:

$$W(\lambda^*) < f(x^*)$$

et la différence $f(x^*) - W(\lambda^*)$ est appelée le saut de dualité.

D'autre part les conditions de k-k-T qui sont ici nécessaires et suffisantes montrent qu'à l'optimum:

$$\lambda^t (b_i - A_i x) = 0 \quad (\forall i = 1, \dots, m).$$

2.7 Conclusion:

La notion de dualité est un concept fondamental en programmation linéaire et conduit à un résultat important d'un point de vue théorique et pratique.

Chapitre 3

méthodes des moindres carrés

3.1 Introduction :

L'étude d'un phénomène peut , le plus souvent être schématisé de la manière suivante:

on s'intéresse à une grandeur b , que nous appellerons par la suite réponse ou variable expliquée, qui dépend d'un certain nombre de variables v_1, v_2, \dots, v_n que nous appellerons facteurs ou variables explicatives.

3.2 Notion de modèle et de régression linéaire multiple:

On cherche à mettre en évidence la liaison (relation fonctionnelle) pouvant exister entre la variable expliquée b et les variables explicatives v_1, v_2, \dots, v_n . On s'intéresse aux modèles dits linéaires, i.e aux modèles du type:

$$b = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = \sum_{j=1}^n \alpha_j v_j.$$

Les α_j sont des réels appelés coefficients du modèle.

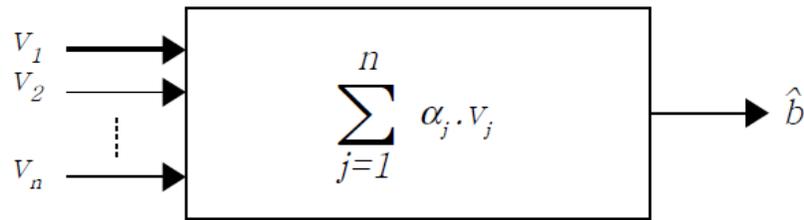


FIG. 3.1

On cherche à expliquer la variable b par n autres variables v_1, v_2, \dots, v_n mais on n'est pas certain que b ne dépend que de ces variables, dans l'idéal $\hat{b} = b$, mais le plus souvent $\hat{b} \approx b$ avec $b \neq \hat{b}$.

3.3 Critère des moindres carrés:

On cherche donc un modèle qui nous permet d'obtenir un \hat{b} le plus « proche » possible de b pour cela, on effectue m mesures ($m > n$) des variables v_1, v_2, \dots, v_n et de b on cherche alors $\alpha_1, \alpha_2, \dots, \alpha_n$ tel que pour $i=1 \dots m$,

$$\hat{b}_i = \alpha_1 v_{i,1} + \alpha_2 v_{i,2} + \dots + \alpha_n v_{i,n} = \sum_{j=1}^n \alpha_j v_{i,j}.$$

soit le plus proche de b_i .

En utilisant les notations matricielles, le système:

$$\left\{ \begin{array}{l} \hat{b}_1 = \alpha_1 v_{1,1} + \alpha_2 v_{1,2} + \dots + \alpha_n v_{1,n} \\ \hat{b}_2 = \alpha_1 v_{2,1} + \alpha_2 v_{2,2} + \dots + \alpha_n v_{2,n} \\ \cdot \\ \cdot \\ \cdot \\ \hat{b}_m = \alpha_1 v_{m,1} + \alpha_2 v_{m,2} + \dots + \alpha_n v_{m,n} \end{array} \right.$$

S'écrit:

$$\hat{\mathbf{b}} = \underbrace{\begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,n} \\ v_{2,1} & v_{2,2} & \dots & v_{2,n} \\ \vdots & \vdots & \dots & \vdots \\ v_{m,1} & v_{m,2} & \dots & v_{m,n} \end{pmatrix}}_A = \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}}_x$$

Ainsi on cherche $x = (\alpha_1, \alpha_2, \dots, \alpha_n)^t$ tel que Ax soit le plus « proche » possible de b , on comprend alors que la notion de distance apparaît. On rappelle que la distance euclidienne usuelle est définie comme suit:

$$\forall x, y \in \mathbb{R}^m, \quad d(x, y) = \sqrt{\|x - y\|^2}.$$

Où $\|\cdot\|$ est la norme euclidienne .
on souhaite que $d(\hat{b} = Ax, b)$ soit minimale, ce qui s'écrit:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2.$$

3.4 La forme standard:

On appelle forme standard d'un problème de moindres carrés la donnée de:

$$\mathbf{A} = \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ v_{m,1} & v_{m,2} & \cdots & v_{m,n} \end{pmatrix} \in \mu_{m,n}$$

A : c'est la matrice des données.

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^n$$

b : vecteur réponse.

$$\mathbf{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} \in \mathbb{R}^n$$

réalisant: $\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$.

x : l'expression du critère.

3.5 Recherche d'une solution:

On suppose que les variables explicatives sont linéairement indépendantes (i.e $\text{rang}(A)=n$).

3.5.1 Solution géométrique:

On cherche à exprimer un vecteur comme combinaison linéaire de n vecteurs indépendants. Cette combinaison linéaire appartient, par définition

d'un espace vectoriel, à l'espace vectoriel engendré par ces variables explicatives.

$$\text{vect} = (v_1, v_2, \dots, v_n) = \text{Im}(A)$$

est un vecteur de \mathbb{R}^m ($m > n$). On cherche $\hat{b} \in \text{Im}(A)$ Tel que: $\|\hat{b} - b\|^2$ soit minimal: c'est la définition de la projection orthogonale de b sur $\text{Im}(A)$.

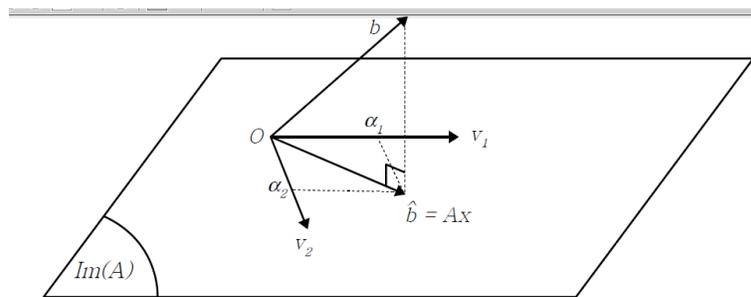


FIG. 3.2

3.5.2 Solution analytique:

Notation : Soit $E(x) = \|Ax - b\|^2$ la fonction erreur, on sait que:

$$E(x) \Rightarrow \dot{E}(x) = 0$$

3.5.3 Théorème:

Si E est strictement convexe, alors:

$$E(x) \Leftrightarrow \dot{E}(x) = 0$$

on cherche donc $x \in \mathbb{R}^n$ tel que $\acute{E}(x) = 0$.

3.5.4 Dérivation matricielle:

On considère une forme linéaire $f : \mathbb{R}^k \rightarrow \mathbb{R}$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{R}^k$$

Définition 1:

On appelle dérivée de f en x et on note $\frac{\partial f}{\partial \mathbf{x}}$ ou $\nabla f(x)$ ou encore $\acute{f}(x)$ le vecteur colonne des dérivées partielles de f par rapport aux x_i :

$$\frac{\nabla \mathbf{f}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Définition 2:

On appelle dérivée directionnelle en x dans la direction d et on note $Df(x,d)$, la limite quand elle existe:

$$Df(x,d) = \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon d) - f(x)}{\varepsilon}$$

3.5.5 Proposition :

Si f est dérivable en x alors quelque soit la direction $d \in \mathbb{R}^k$:

$$Df(x,d) = (f'(x) \setminus d) = f'(x)^t d.$$

Remarque:

On retrouve ce resultat via les dérivées partielles: soit d une direction quelconque:

$$\begin{aligned} Df(x,d) &= \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon d) - f(x)}{\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{(x + \varepsilon d \setminus \alpha) - (x \setminus \alpha)}{\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{(x \setminus \alpha) + \varepsilon(d \setminus \alpha) - (x \setminus \alpha)}{\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} (d \setminus \alpha) \\ &= (d \setminus \alpha) = (\alpha \setminus d) = \alpha^t d \\ &= f'(x) = \alpha. \end{aligned}$$

3.6 La forme quadratique :

Définition :

une forme quadratique est un polynôme homogène de degré 2 avec un nombre quelconque de variables:

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

$$x = (x_1, x_2, \dots, x_n) \mapsto \sum_{i,j=1}^k \alpha_{i,j} x_i x_j.$$

En notant $A = (\alpha_{i,j}) \in \mu_{k,k}(\mathbb{R})$, f s'écrit:

$$f : \mathbb{R}^k \rightarrow \mathbb{R}$$

$$x \rightarrow x^t A x = (x \setminus A x) = (A x \setminus x).$$

Calcul de dérivée :

$$\begin{aligned}
Df(x,d) &= \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon d) - f(x)}{\varepsilon} \\
&= \lim_{\varepsilon \rightarrow 0} \frac{A(x + \varepsilon d) + \varepsilon(Ax \setminus d) + \varepsilon(Ad \setminus x)}{\varepsilon} + \varepsilon^2(Ad \setminus d) - (Ax \setminus x) \\
&= (Ax \setminus d) + (Ad \setminus x) = (Ax \setminus d) + (A^t d \setminus x) \\
&= ((A + A^t)x \setminus d) \\
&\implies f'(x) = (A + A^t)x
\end{aligned}$$

Si A est symétrique, si $A^T = A$ on a:

$$\frac{\partial(x^T Ax)}{\partial x} = 2Ax.$$

3.7 Calcul de solution:

On a:

$$\begin{aligned}
 E(x) &= \| Ax - b \|^2 \\
 &= \| Ax \|^2 - 2(Ax \cdot b) + \| b \|^2 \\
 &= x^T A^T Ax - 2x^T A^T b + b^T b
 \end{aligned}$$

Ainsi:

$$\begin{aligned}
 \acute{E} &= \frac{\partial x^T A^T Ax}{\partial x} - 2 \frac{\partial x^T A^T b}{\partial x} + \frac{\partial b^T b}{\partial x} \\
 &= 2A^T Ax - 2A^T b + 0
 \end{aligned}$$

donc :

$$\begin{aligned}
 \acute{E} = 0 &\Leftrightarrow 2A^T Ax - 2A^T b = 0 \\
 &\Leftrightarrow A^T Ax = A^T b.
 \end{aligned}$$

3.8 Théorème :

Soit $A \in \mu_{m \times n}$ avec $m > n$, et $b \in \mathbb{R}^m$, une condition nécessaire et suffisante pour que $x \in \mathbb{R}^n$ réalise le minimum de $E(x) = \| Ax - b \|^2$ est que:

$$A^T Ax = A^T b \quad (*)$$

Les équations (*) sont appelées **équations normales** ce système admet toujours, au moins une solution. Si la matrice $A^T A$ est **régulière** i.e, si $\text{rang} A = n$ alors la solution est **unique**.

3.9 Conclusion :

La méthode de moindres carrés est utilisée pour résoudre plusieurs problèmes dans différents domaines (statistique ,biologie ...) .

La résolution d'un problème de moindres carrés via les équations normales possède deux inconvénients majeures. d'une part, la perturbation due aux erreurs d'arrondi lorsque l'on passe par les équations normales peut être importante. En effet, si la matrice des données A est légèrement perturbé: $\hat{A} = A + \delta A$ le passage aux équations normales va amplifier la perturbation:
$$(A + \delta A)^t(A + \delta A) = A^t A + \delta A^t A + A^t \delta A + \delta A^t A...$$

Alors qu'en passant par d'autres méthodes de résolution (on peut passer par la méthode de la décomposition LU pour résoudre le système), La perturbation des données sera moindre.

Chapitre 4

La méthode SVM

4.1 Introduction:

L'apprentissage machine basé sur la notion de généralisation à partir d'un grand nombre de données couvre des domaines. Tels que la reconnaissance de forme et de la régression ne cesse d'avoir un développement dans ses méthodes et techniques, ceux-ci ne les a pas empêché de dévoiler des limites qui réduisent leur efficacité face a la complexité des problèmes du domaine. En même temps d'autre méthodes ou étaient misent en œuvre et dès leur première apparition elles ont surpassé les méthodes existantes au paravant. Les SVMs sont une des nouvelles méthodes largement utilisés recemment. Dans ce qui suit on va présenter ce paradigme .

4.2 Définition :

Les "Support Vector Machins" appelés aussi « maximum margin classifier » (en français machine à vecteur de support ou séparateur à vaste marge) sont des techniques d'apprentissages supervisé basé sur la théorie de l'apprentissage statique (généralement considérés comme la première réalisation pratique de cette théorie[6]) et respectant les principes du (SRM) « Structural risk minimisation » (trouver un séparateur qui minimise la somme de l'erreur de l'apprentissage [1]), un SVM repose sur les 2 notion de vaste marge et fonction Kernel.

Les SVMs sont considéré comme l'un des modèles les plus important parmi la famille des méthodes a Kernel. Ils ont gagné une forte popularité grâce a leur succès dans la reconnaissance des chiffres manuscrits avec un taux d'erreur de 1.1 en phase de test (le même temps marqué par un réseaux de neurone soigneusement construit)[2].

4.3 Hyperplan séparateur linéaire :

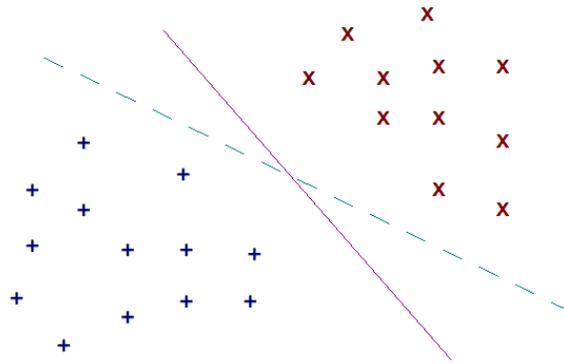


FIG. 4.1 – *hyperplan séparateur linéaire*

4.4 Rappel de géométrie :

4.4.1 Equation d'un hyperplan :

$$a_0 + a_1x_1 + \dots + a_dx_d = 0, \text{ ou encore } a_0 + a^T x = 0.$$

4.4.2 Distance à un hyperplan :

$$h(x) / \|a\|, \text{ où } h(x) = a_0 + a_1x_1 + \dots + a_dx_d$$

$$\text{et } \|a\|^2 = \sum_{i=1}^d a_i^2.$$

4.5 Hyperplan de marge maximal :

On impose une marge en $h(x) = 1$ et $h(x) = -1$ (après le réglage des paramètres a_0, \dots, a_d) soit une marge de $2/\|a\|$.

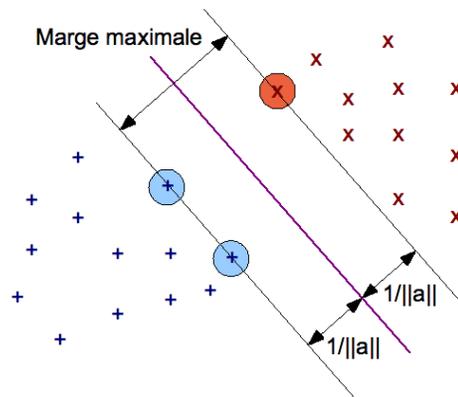


FIG. 4.2 – *Hyperplan de marge maximale*

4.6 Problème primal:

Etant donné un ensemble d'apprentissage $\{(x_1, u_1), \dots, (x_m, u_m)\}$, $u_i \in \{1, -1\}$ représentant la classe de chaque objet $x_i \in \mathbb{R}^d$, on cherche à :

$$\begin{cases} \text{minimiser } \frac{1}{2}\|a\|^2 \\ \text{Sous les contraintes: } u_i h_i(x) \geq 1 & \text{Si } i = 1 \dots m \end{cases}$$

La résolution de ce problème à $d + 1$ paramètres est possible par programmation quadratique. Si d est petit ($\leq 10^3$), mais défficile au delà de quelques contraintes. Et si $d \geq 10^5$ on utilise la forme dual. Mais entre 10^3 et 10^5 en utilise le programme quadratique.

4.7 Problème dual :

Lagrangien $L(a, a_0, \alpha) = \frac{1}{2} \|a\|^2 - \sum_{i=1}^m \alpha_i (u_i (ax_i + a_0) - 1)$, où les $\alpha_i \geq 0$

sont appelés multiplicateurs de Lagrange ou variables duales.

Le théorème de Kuhn-Tucker montre que le problème primal et sa formulation duale ont la même solution , qui correspond à un point-selle du Lagrange (il faut le minimiser par variables primaires a et a_0 et le maximiser par rapport aux variables duales α_i).

Au point -selle , la dérivée du lagrangien par rapport aux variables primaires s'annule :

$$\partial L(a, a_0, \alpha) / \partial a = 0 \quad \text{et} \quad \partial L(a, a_0, \alpha) / \partial a_0 = 0$$

$$\text{On obtient : } \sum_{i=1}^m \alpha_i u_i = 0 \quad \text{et} \quad a = \sum_{i=1}^m \alpha_i u_i x_i$$

On peut ainsi éliminer les variables primaires et le problème d'optimisation s'écrit alors sous la forme duale : trouver α maximisant

$$\left\{ \begin{array}{l} \max \sum_{i=1}^m \alpha - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j u_i u_j x_i x_j \\ \text{sous contraintes} \quad \alpha_i \geq 0 \quad \text{pour} \quad i = 1 \dots m \\ \text{et} \quad \sum_{i=1}^m \alpha_i u_i = 0 \end{array} \right.$$

Le problème ne depend plus de la dimension d de l'espace d'entrée mais de la taille m de l'ensemble d'apprentissage, et même d'un nombre très inférieur à m .

En effet le théorème de Karush-kuhn-Tucker montre que les multiplicateurs α_i sont non nuls uniquement pour les points qui sont sur les hyperplans frontière ($h(x) = \pm 1$). Ces points sont appelés Vecteur Support.

On retrouve formellement ce que l'on observe intuitivement: l'hyperplan solution ne dépend que des vecteurs supports.

Les méthodes d'optimisation quadratiques standards suffisent donc en pratique: Une fois que l'on dispose des vecteurs des variables duales α_i , on

obtient le vecteur normal de l'hyperplan $a = \sum_{i=1}^m \alpha_i u_i x_i$ puis $a_0 = u_i - ax_i$

à partir d'un des vecteurs supports x_i .

L'hyperplan séparateur a donc pour équation

$$h(x) = ax + a_0 = \sum_{i=1}^m \alpha_i u_i x_i x + a_0.$$

Cas réel : (Exemple : bruités, exceptions)

Lorsque les exemples ne peuvent pas être séparés par un hyperplan, on peut rechercher l'hyperplan faisant le moins d'erreurs possible à l'aide de variables ressort. On impose alors : $u_i h(x_i) \geq 1 - \xi_i$.

Et on minimise : $\frac{1}{2} \|a\|^2 + C \sum_{i=1}^m \xi_i$ où C est une constante, nouveau

paramètre du système, réglant le compromis entre la marge entre les exemples et les erreurs, au sens de leur distance au "bon" côté de l'hyperplan.

Le problème dual revient alors à maximiser :

$$\left\{ \begin{array}{l} \text{maximiser } \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j u_i u_j x_i x_j \\ 0 \leq C \text{ pour } i = 1, \dots, m \\ \text{et } \sum_{i=1}^m \alpha_i u_i = 0 \end{array} \right.$$

Limité aux problèmes linéairement séparables.

4.8 Avantages et inconvénients :

4.8.1 Avantages :

- Absence d'optimum local .
- Contrôle explicite du compromis entre la complexité du classifieur et de l'erreur.
- Possibilité d'utilisation de structure de données comme les chaînes de caractères et arbres comme des entrées .
- Traitement des données a grandes dimensions .

4.8.2 Inconvénients :

- Demande des données négatives et positive en même temps .
- Besoin d'une bonne fonction **Kernel** .
- Problème de stabilité des calculs dans la resolution de certains programme quadratique a contraintes .

4.9 Conclusion :

Les SVMs présentent un alternatif utile aux différentes méthodes de classification classique, leurs principes de vaste marge et fonction Kernel les permettent des taux de classification et de minimisation très importants.

Conclusion générale

Ce mémoire contient quatre chapitres : dans le premier chapitre on a rappelé quelques notions de base .

Dans le deuxième chapitre , on a étudié **la dualité lagrangienne**. et dans le troisième chapitre, on a étudié **La méthode des moindres carrés**. L'objet de cette méthode est de fournir un outil d'interprétation de données. Plus précisément, lorsqu'on dispose de données dépendant de deux paramètres x et y , on peut les représenter dans le plan muni d'un repère, en marquant x en abscisse et y en ordonnée ; si le "nuage de points" qu'on obtient a l'allure d'une droite, on veut savoir quelle est l'équation de cette droite , c'est-à-dire quelle loi relie les deux paramètres de la mesure. La méthode des moindres carrés permet d'obtenir cette droite.

Dans le quatrième chapitre, on a traité **la méthode SVM**, et on a donné une vision générale et une vision parument mathématique des SVMs. Cette méthode de classification est basée sur la recherche d'un meilleur hyperplan qui permet de séparer au mieux des ensembles de données. On a exposé le cas linéaire (hyperplan séparateur linéaire et l'hyperplan de marge maximal) qui nécessitent l'utilisation d'une bonne fonction **Kernel** pour changer l'espace.

Cette méthode est applicable pour des tâches de classification à deux classes, mais il existe des extensions pour la classification multi-classes.

Les méthodes SVMs représentent une méthode d'apprentissage statistique caractérisée par un background théorique solide qui leur a permis l'ex-

tension vers plusieurs variantes. Les SVMs sont marquées par une grande capacité de généralisation et une convergence assurée qui les placent au premiers rang des outils d'analyse en data mining.

Le data mining, en pleine évolution, connaît encore des difficultés pour la manipulation de bases de données enregistrées dans ces bases, posent des difficultés pour beaucoup de ces algorithmes.

Bibliographie

- [1] : Vojislav Kecman. "Learning and soft computing support vector machines, neural net works, and fuzzy logic models" the MIT press 2001.
- [2] : L . Botton et al. "Comparison of classifier méthodes: a case study in handwritten digit " recognition . Proceeding of the 12 th IAPR International conference on pattern recognition, vol 2.
- [3] : Michel Minoux. Programmation mathématique, théories et algorithmes, 2^e eddition , lavoisier 2008 .
- [4] : K . Karmanov. Programmation mathématique, édition Mir.Moscou 1977 .
- [5] : Tony Bourdier. ESIAL-Mathématiques numériques 2007-2008 .
- [6] : Bernhard Scholkopf, Alexander J . Smola . "Learning with kernels , support vector machine , régularisation,optimisation, and Beyoud", the MIT press 2002 .
- [7] : Martin Law "A simple introduction to support vector machines", lecteur foe CSE 802. Departement of computer Science and Engineering. Michigan State University 2011.
- [8] : History of support Vector Machiness [en ligne] .<< <http://www.SVms.org/history.html>>> (9/11/2012) .

[9] : Colon Campbell, Yiming Ying "Learning with support vector machines".

SYNTHESIS LECTURES ON ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING 10", Morgan Claypool publisher 2011.

[10] : Bernhard Scholkopf, Alexander j. Smola "Learning with Kernels support vector Machines Regularization, optimization, and Beyond ", the MIT press 2002 .