الجمــــهوريــــة الجـــزائريــــة الديــمقــراطيـــــــــة الشــعبيـــــــة
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعـلـيـم الـعـالـــي و الـبـحـث الـعـلـمـي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جـامـعـة عمّـــار ثـلـيــجــي بالأغـــواط
UNIVERSITE AMAR TELIDJI LAGHOUAT

كـلـيـــــة العلوم
FACULTE DES SCIENCES

DEPARTEMENT DE MATHEMATIQUES ET INFORMATIQUE

Mémoire de LICENCE

Domaine : Mathématiques et Informatique
Filière :Mathématiques
Option :Mathématiques

Par

Saadi Achour

THEME

Théorème de Karush-Kuhn-Tucker

Sous la direction du Pr. Mokhtari Abdelkader

Année universitaire 2014/2015

Dédicaces

Mes très chers parents ***Mr. Achour Mohamed*** *et* ***Mme. B.Ch.Ch*** *qui ont consenti d'énormes sacrifices pour mon éducation et mon bien être. Qu'****ALLAH*** *leur assure une longue vie et bien travail pour que je puisse veiller à leur bonheur !*

Toute ma famille.

touts mes frères.

Tout ce que j'aime.

Mon encadreur Pr. ***Mokhtari Abdelkader****.*

Tout Moslim.

Je dédie ce mémoire

SAADI_ACHOUR

REMERCIEMENTS

*Mes remerciements vont tout d'abord à **ALLAH** pour tous ce que m'a donné.*

*mes vifs remerciements à mon encadreur **Mr. Mokhtari Abdelkader**, Professeur en Mathématiques à l'université de Laghouat pour les moyens qu'il m'a procuré afin d'élaborer ce mémoire et pour tous ces efforts pendant les années d'étude.*

*Mes remerciements vont plus particulièrement à **Mr. Allaui Taher**, Magistère en Informatique, pour leur soutien pendant ces années d'étude.*

*Mes remerciements vont également à **Mr. Ouinten Youcef**, Maitre de conférence en Mathématiques à l'université de Laghouat pour ces inestimables conseils, ainsi à **Mr. Belabbaci Youcef**, Maitre de conférence en Mathématique à l'université de Laghouat et à **Mr. Messelmi Mohamed**, chef de département de Génie Informatique, ainsi que tout le personnel de la bibliothèque de l'université de Laghouat, qui ont tenu à notre disposition tous les moyens et livres ; sans oublier tout le personnel et les étudiants du département Maths et informatique de l'université de Laghouat.*

Un grand remerciement à toutes les personnes qui ont participé de prés ou de loin pour atteindre mon objectif.


بسم الله الرحمن الرحيم

Table des matières

Première partie

Chapitre 0 (Introduction générale)

0.1 La formulation des problèmes :

0.1.1 Réalisation

Une définition possible de la **R.O** (**R**echerche **O**pérationnele) pourrait être la suivante : il s'agit d'un ensemble de méthodes d'analyse scientifique (maths et info.) des phénomènes d'organisation qui traite de la maximisation d'un profit, d'une performance, d'un rendement ou bien de la minimisation d'un coût, d'une dépense. La **R.O**.est avant tout un outil d'**aide à la décision**.

Le schéma général suivi par ces méthodes est : Problème concret (de type **R.O**)→ modélisation→résolution par une méthode de **R.O**→ interprétation des résultats→prise de décision.

Tous les problèmes mentionnés ci-dessus et présentés dans ce cours peuvent se formaliser de la façon suivante. On cherche à maximiser une fonction f sur un ensemble X donné : $$max\{f(x), x \in X\}$$

∘ $f : X \to R$ est la fonction **objectif**. f peut être linéaire, quadratique, non linéaire...

∘ X est l'ensemble des solutions possibles dites **réalisables** (les contraintes). L'ensemble X est fini mais en général de trés grande taille.

On peut envisager de façon équivalente un problème de minimisation grâce à la relation $$minf = -max(-f).$$

L'optimisation est l'obtention du meilleure décision possible dans un ensemble de circonstances donnés.

Les application :

$$\left.\begin{array}{r} \checkmark Conception \\ \checkmark Construction \\ \checkmark Gestion\ d'unemaintenance. \end{array}\right\} \Longleftrightarrow$$ ✓minimiser l'effort ou maximiser unprofit.

Les méthodes d'optimisation sont déformées sous les formes :

✓Nombreuses

✓Diverses

✓Spécifique .

L'optimisation vise à résoudre des problèmes où l'on cherche à déterminer parmi un grand nombre de solutions candidates celle qui donne le meilleur rendement. Plus précisément, on cherche à trouver une solution satisfaisant un ensemble de contraintes qui minimise ou maximise une fonction donnée. L'application de l'optimisation est en expansion croissante et se retrouve dans plusieurs domaines. Les problèmes considères dans ce document s'écrivent sous la forme standard

$$\min_{x \in R^n} f(x)$$
$$\text{s.c. } x \in S$$

Où $x = (x_1, x_2, .., x_n)^\top$est un vecteur de R^n, $f : R^n \longrightarrow R$ est la fonction que l'on désire minimiser (appelée *fonction objectif*), $S \subseteq R^n$ est l'ensemble dans lequel les points doivent appartenir, et s.c. est l'abréviation de sous la ou les contraintes.

La formulation (1) signifie que l'on cherche à trouver une solution du domaine réalisable $x^* \in S$ dont la valeur de la fonction objectif est la plus petite.

0.1.2 Classification

✓Selon la nature des variables de décision :

∘ Optimisation discrète ou optimisation combinatoire.

✓Selon la nature des contraintes :

∘ Pas de contraintes ou contraintes faciles à satisfaire (un segment de la droite réelle) : optimisation sans contraintes.

∘ Optimisation sous contraintes : il est difficile de trouver un point satisfaisant les contraintes.

✓Propriétés spéciales des éléments du problème : linéarité, convexité.

0.2 Karush, Kuhn et Tucker

0.2.1 Le travail de Karush, Kuhn et Tucker

Les travaux de *Kuhn* et *Tucker* en 1951 sur les conditions *nécessaires* et *suffisants* pour les solutions des problèmes de *programmation mathématique* ont conduit à recherche sur problèmes des programmation *non linéaire*.

Formulation trouver $x^* = (x_1^*, x_2^*, .., x_n^*)^\top$ qui minimise $f(x)$ et qui vérifie les contraintes

$$g_j(x) \leq 0 \ , \ j = 1, .., m$$

et

$$h_j(x) = 0 \ , \ j = 1, .., p$$

On a alors le problème :

$$\begin{cases} \min f(x) \\ g_j(x) \leq 0, j = 1, .., m \quad x = (x_1, x_2, .., x_n)^\top / \ x \in R^n. \\ h_j(x) = 0, j = 1, .., p \end{cases}$$

0.2.2 Qui sont Karush, Kuhn et Tucker ?

Harold William Kuhn

(né le 29 juillet 1925) etait un mathématicien et économiste américain.

Professeur à l'Université Princeton, il publia plusieurs livres dontNonlinear programming avec Albert Tucker en 1951, puis contributions to the Theory of Games en 1976.

Albert William Tucker

(28 novembre 1905 - 25 janvier 1995) était un mathématicien américain d'origine canadienne qui a produit d'importantes contributions en topologie, théorie des jeux et programmation non-linéaire.

William William Karush

(01 mars 1917 – 22 février 1997) était professeur émérites de la California State

University de Northridge. Il étaitle premier à publier les conditions nécessaires pour les problèmes avec contraintes d'inégalité dans son mémoire de Master.

Notation le théorème de **Karush-Kuhn-Tucker** noté simplement Kuhn-Tucker ou **K.K.T**.

0.3 Plan du mémoire

Ce mémoire se décompose en introduction,trois chapitres et conclusion **:**

∘ **Introduction.**

∘ **Chapitre 01 :**

Dans ce chapitre, nous allons donner une vue générale sur l'espace euclidien et de *Hilbert* par la présentation de quelques notions et définitions, propriétés et des théorèmes fondamentaux avec les preuves (produit scalaire, théorème d'inégalité de Cauchy-Schwarz,..), des théorèmes de Riesz et de projection .

∘ **Chapitre 02 :**

Ce chapitre est le but du mémoire, nous allons présenter une étude fondamentale sur l'optimisation sans contraintes et sous contraintes, et comme début nous allons présenter les conditions d'optimalités avec et sans contraintes et étudier deux méthodes parmi plusieurs (substitition directe et les multiplicateurs de Lagrange) dans la cas dont les contraintes sont d'égalités, et pour le cas d'inégalités une étude fondamentale sur *les conditions de* $K.K.T$ avec dimonstration et les théorèmes qui sont fondamentaux dans cette étude.

∘ **Chapitre 03 :**

dans ce chapitre nous allons présenter deux applications d'optimisation avec contraintes d'égalité et d'inégalité avec les conditions de $K.K.T.$

∘ **Conclusion.**

Deuxième partie

Chapitre 1

Chapitre 1

Espace de Hilbert

1.1 Introduction

Le mathématicien *allemand David Hilbert* $(1862-1943)$, le premier à avoir introduit de manière systématique les espaces qui maintenant portent son nom, est connu pour les 23 fameux problèmes qu'il proposa au congrès international des mathématiciens en 1900. Ses idées ont profondément marqué l'ensemble des mathématiques jusqu'à l'heure actuelle, en particulier en introduisant des méthodes géométriques en analyse, donnant ainsi naissance à un domaine important des mathématiques : l'analyse fonctionnelle.

Les méthodes géométriques présentées dans ce mémoire sont basées sur la notion de produit scalaire et d'orthogonalité.

Dans ce paragraphe, nous introduisons l'espace de *Hilbert* avec ses propriétés élémentaires. Nous caractériserons la norme qui est déduite d'un produit scalaire. S'il n'y a pas de spécification le corps K est, comme auparavant, R ou C.

Dans cette partie K désigne R ou C et E est un *K-espace vectoriel.*

1.2 Généralités sur les formes bilinéaires et quadratiques

1.2.1 Formes bilinéaires symétriques et antisymétrique

Définition(1) *(les formes bilinéaires) On appelle forme bilinéaire sur $E \times E$ (E est un $K-espacevectoriel$) toute application $\varphi : E \times E \to K$ telle que :*

$$\forall \alpha \in K, \forall (x, x\prime, y) \in E^3, \varphi(\alpha x + x\prime, y) = \alpha\varphi(x, y) + \varphi(x\prime, y)$$

(φ est linéaire par rapport à la 1ère place)

$$\forall \beta \in K, \forall (x, y, y\prime) \in E^3, \varphi(x, \beta y + y\prime) = \beta\varphi(x, y) + \varphi(x, y\prime)$$

(φ est linéaire par rapport à la 2ème place) .

Notation

$*$Nous notons $\mathcal{L}(E, E; K)$ l'ensemble des formes bilinéaires sur $E \times E$.

$*$Il est immédiat que $\mathcal{L}$(E,E;K) est un K-espace vectoriel .

Proposision(1) *Soient ϕ une forme bilinéaire sur $E \times E$, $(n, p) \in (N^*)^2, \alpha_1, ..., \alpha_n, \beta_1, ..., \beta_p \in K, x_1, ..., x_n, y_1, ..., y_p \in E$. On a alors :*

$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i, \sum_{j=1}^{p} \beta_j y_j\right) = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}^{n} \alpha_i \beta_j \varphi(x_i, y_j).$$

Preuve On a, par récurrence immédiate sur n :

$$\forall Y \in E, \varphi\left(\sum_{i=1}^{n} \alpha_i x_i, Y\right) = \sum_{i=1}^{n} \alpha_i \varphi(x_i, Y),$$

d'où
$$\varphi\left(\sum_{i=1}^{n} \alpha_i x_i, \sum_{j=1}^{p} \beta_j y_j\right) = \sum_{i=1}^{n} \alpha_i \varphi\left(x_i, \sum_{j=1}^{p} \beta_j y_j\right) = \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{p} \beta_j \varphi(x_i, y_j)\right)$$
$$= \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}^{n} \alpha_i \beta_j \varphi(x_i, y_j).$$

Définition(2) *Une application $\varphi : E \times E \to K$ est dite symétriquesi et seulement si :*

$$\forall (x, y) \in E^2, \varphi(y, x) = \varphi(x, y).$$

On abrègera forme bilinéaire symétrique en : *fbs.*

Notation Nous notons $S(E;K)$ l'ensemble des formes bilinéaires symétriques sur $E \times E$. Il est immédiat que $S(E;K)$ est un $K-$espace vectoriel , sev de $\mathcal{L}(E,E;K)$.

La Proposition suivante est immédiate, et d'un usage commode.α

Proposition(2) *Pour qu'une application $\varphi : E \times E \longrightarrow K$ soit une fbs, il faut et il suffit que l'on ait :*

$\cdot\ \varphi$ *est symétrique*

$\cdot \varphi$ *est linéaire par rapport à la $2^{ème}$ place.*

1.2.2 Formes quadratiques

Définition(3) *Soit φ une fbs sur $E \times E$. On appelle forme quadratique associée à φ l'application, souvent notée ϕ, de E dans K définie par :*

$$\forall x \in E, \varphi(x) = \phi(x,x).$$

On abrègera forme quadratique en : *fq.*

Exemple(1) Le produit scalaire canonique sur R^n, défini par

$$R^n \times R^n \longrightarrow R$$
$$((x_1,...,x_n),(y_1,...,y_n)) \mapsto \sum_{k=1}^{n} x_k y_k \ ,$$

est une fbs et la fq associée est

$$R^n \longrightarrow R$$
$$(x_1,...,x_n) \mapsto \sum_{k=1}^{n} x_k^2 \qquad .$$

La proposition suivante est immédiate à partir de proposition1.

Proposition(3) *Soient φ une fbs sur $E \times E$, ϕ la fq associée à φ. On a :*

1)$\forall n \in N^*, \alpha_1,...,\alpha_n \in K, \forall x_1,...,x_n \in E,$

$$\phi(\sum_{i=1}^{n} \alpha_i x_i) = \sum_{i=1}^{n} \alpha_i^2 \phi(x_i) + 2 \sum_{1 \le i < j \le n} \alpha_i \alpha_j \varphi(x_i, y_j)$$

2)$\forall (\alpha,\beta) \in K^2, \forall (x,y) \in E^2,\ \varphi(\alpha x + \beta y) = \alpha^2 \phi(x) + 2\alpha\beta\varphi(x,y) + \beta^2 \phi(y)$

3)$\forall (x,y) \in E^2, \phi(x+y) = \phi(x) + 2\varphi(x,y) + \phi(y)$

4)$\forall (x,y) \in E^2, \varphi(x,y) = 1/4(\phi(x+y) - \phi(x-y))$

5)$\forall (x,y) \in E^2, \phi(x+y) + \phi(x-y) = 2\phi(x) + \phi(y).$

Remarque(1) *Les formules 3)et 4)précédentes montrent que ϕ détermine entièrement φ ; φ est appelée la forme polaire de ϕ.*

Définition(4) Soit $\phi : E \to K$ une application. On dit que ϕ est une **forme quadratique** si et seulement s'il existe une fbs $\varphi : E \times E \to K$ telle que ϕ soit la fq associée à φ.

Notation Notons $Q(E)$ l'ensemble des fq sur E.

Rmarque(2) *Il est clair que l'application $U : S(E; K) \to Q(E)$ qui, à toute fbs φ sur $E \times E$ fait correspondre la fq associée à φ, et l'application $V : Q(E) \to S(E; K)$ qui, à toute fq ϕ sur E associe la forme polaire de ϕ, sont des bijections réciproques l'une de l'autre.*

1.3 Rappels sur les espaces euclidiens

1.3.1 Produit scalaire

Définition(1)*(produit scalaire) Soit E un espace vectoriel sur le corps K.*

Un produit scalaire sur E est une application $\varphi : E \times E \to K$ telle que pour tout $x, x_1, x_2, y \in E$, et $\lambda \in K$, on a :

1.$\varphi(x, x) \geq 0$ *et* $\varphi(x, x) = 0$ *si et seulement si* $x = 0$.

2.$\varphi(x, y) = \overline{(\varphi(y, x))}$

3.$\varphi(x_1 + x_2, y) = \varphi(x_1, y) + \varphi(x_2, y)$

4.$\varphi(\lambda x, y) = \lambda\varphi(x, y)$.

Remarque(1) *Les propriétés suivantes d'un produit scalaire induit par les axiomes* 1) *à* 4) *:*

i) $\langle 0, x\rangle = \langle x, 0\rangle = 0$, *pour tout* $x \in E$,

ii) $\langle \sum_{i=1}^{n} \lambda_i x_i, \sum_{j=1}^{m} \mu_j y_j\rangle = \langle \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_i \overline{\mu_j} \langle x_i, y_j\rangle$, *pour tout* $x_i, y_j \in E, \lambda_i, \mu_j \in K$.

Définition(2)*(autre définition de produit scalaire) On appelle produit scalaire sur*

E toute fbs φ sur $E \times E$ telle qu'en notant ϕ la fq associée à φ, on ait :

$$(i) \forall x \in E, \phi(x) \geq 0$$

$$(ii) \forall x \in E, (\phi(x) = 0 \Rightarrow x = 0)$$

Notation *Lorsque φ est un produit scalaire, on note souvent $(x|y)$ ou $<x,y>$ ou $x.y$ à la place de $\varphi(x, y)$.*

Définition(3) *On appelle espace euclidien tout couple (E, φ) où E est un $R-e.v$ de dimension finie et φ un produit scalaire sur E.*

On abrègera espace (vectoriel) euclidien en : eve.

Exemple(s)(2)

1)*Produit scalaire usuel sur $R^n, n \in N$*

L'application $\varphi : R^n \times R^n \mapsto R$ définie par :

$$\varphi((x_1, ..., x_n), (y_1, ..., y_n)) \to \sum_{k=1}^{n} x_k y_k.$$

2)*Produit scalaire canonique sur $M_{n.p}(R)$*

L'application $\varphi : (M_{n.p}(R))^2 \underset{(A,B) \mapsto tr(AB^t)}{\to} R$ est un produit scalaire sur $M_{n.p}(R)$, appelé produit scalaire canonique sur $M_{n.p}(R)$.

Le produit scalaire canonique sur $M_{n.1}(R)$ (ou $M_{1.n}(R)$) est, à la notation près, le produit scalaire usuel sur R^n.

Théorème(1)*(Inégalité de Cauchy-Schwarz) Soient (E, φ) un eve, ϕ la fq associée à φ. On a :*

Alors pour tout $x, y \in E$,

$$\varphi(x, y)^2 \leq \phi(x)\phi(y).$$

Ou

$$|\langle x, y \rangle| \leq \langle x, x \rangle^{1/2} \cdot \langle y, y \rangle^{1/2} .$$

Preuve On procède comme d'habitude, pour cette inégalité, on introduit un paramètre $\lambda \in R$ .De la positivité du produit scalaire, on déduit

$$\langle x - \lambda y, x - \lambda y \rangle \geq 0 \forall x, y \in E, \lambda \in R.$$

Mais, de la bilinéarité, on déduit :

$$\langle x-\lambda y, x-\lambda y\rangle = \langle x,x\rangle - 2\lambda\langle x,y\rangle + \lambda^2\langle y,y\rangle \geq 0, \forall x,y \in E, \lambda \in R.$$

Ce trinôme du second degré en λ reste toujours positif. Nécessairement, son discriminant est négatif ;

$$|\langle x,y\rangle|^2 \leq \langle x,x\rangle\cdot\langle y,y\rangle \,.$$

Ceci conduit à l'inégalité de Cauchy-Schwarz.

Proposition(1) *(Étude du cas d'égalité dans l'inégalité de Cauchy-Schwarz) Soient (E,φ)un eve, ϕ la fq associée à $\varphi, (x,y) \in E^2$;on a :*

$$\varphi(x,y)^2 = \phi(x)\phi(y) \iff (x,y) \text{ lié.}$$

Théorème(3)(Inégalité de Minkowski) Soient (E,φ) un eve, ϕ la fq associée à φ ; on a :

$$\forall (x,y) \in E^2, \phi(x+y)^{1/2} \leq \phi(x)^{1/2} + \phi(y)^{1/2}.$$

Proposition(2) *(Étude du cas d'égalité dans l'inégalité de Minkowski) Soient (E,φ) eve, ϕ la fq associée à φ, $(x,y) \in E^2$;On a :*

$$(\ \phi(x+y))^{1/2} = (\phi(x))^{1/2} + (\phi(y))^{1/2} \iff (x=0 \text{ ou } \exists \alpha \in R_+, y = \alpha x).$$

On traduit cette dernière condition par: (x,y) est positivement liée.

Proposition(3) *(et définition) Soient (E,φ) un eve, ϕ la fq associée à φ. L'application $\|.\| : E \underset{x\longrightarrow(\phi(x))^{1/2}}{\rightarrow} R$ est une norme sur E, appelée norme euclidienne associée à φ.*

Remarque(2) *Soient $\langle .,.\rangle$ un produit scalaire sur $E, ||.||$ la norme euclidienne associée. Les formules obtenues en 1.2 Prop. $3, 3), 4), 5)$ peuvent être récrites sous la forme suivante,pour tout (x,y)de E^2 :*

$$\cdot||x+y||^2 = ||x||^2 + 2\langle x,y\rangle + ||y||^2$$

$$\cdot||x+y||^2 - ||x-y||^2 = 4\langle x,y\rangle$$

$$\cdot||x+y||^2 + ||x-y||^2 = 2(||x||^2 + ||y||^2).$$

L'inégalité de Cauchy-Schwarz (Th. 1) peut être récrite :

$$\cdot(< x,y >)^2 \leq ||x||^2||y||^2.$$

Et l'inégalité de Minkowski (Th. 2) peut être récrite :

$\cdot||x+y|| \leq ||x|| + ||y||.$

1.3.2 Orthogonalités

Définition(4) *Soit* $(E, \langle\cdot,\cdot\rangle)$ *un espace vectoriel euclidien.*

1)*Soit* $(x, y) \in E^2$ *; on dit que* x *est orthogonal à* y*, et on note* $x \perp y$*, si et seulement si :*

$$\langle x, y\rangle = 0.$$

2) *Soient* $x \in E$,$A \in P(E)$ *; on dit que* x *est orthogonal à* A*, et on note* $x \perp A$*, si et seulement si :*

$$\forall a \in A, \langle x, a\rangle = 0.$$

3)*Pour toute partie* A *de* E*, on définit l'orthogonal de* A*, noté* $A^{\perp}$ *:*

$$A^{\perp} = \{x \in E; \forall a \in A, \langle x, a\rangle = 0\}.$$

4)*Une famille* $(x_i)_{i\in I}$ *d'éléments de* E *est dite orthogonale si et seulement si:*

$$\forall (i, j) \in I^2, (i \neq j \Rightarrow \langle x_i, x_j\rangle = 0).$$

5)*Une famille* $(xi)_{i\in I}$ *d'éléments de* E *est dite orthonormale si et seulement si :*

$\cdot(xi)_{i\in I}$ *est orthogonale* *et* $\cdot\forall i \in I, ||x_i|| = 1.$

Rmarque(3)

$1 \cdot 0 \perp x$ *pour tout* $x \in H$.

$2\cdot x \perp y = y \perp x$,*symétrie*

$3 \cdot x \perp x \Rightarrow x = 0$

$4 \cdot x \perp x_i$ *et* $\lambda_i \in K, i = 1, \ldots, n \Rightarrow x \perp \sum_{i=1}^{n} \lambda_i x_i$

$5 \cdot x \perp x_n$ *et* $x_n \xrightarrow{\|\cdot\|} x_0 \Rightarrow x \perp x_0$

$6\cdot$ *Soit* $\langle x_i, x_j\rangle = 0; i \neq j; i, j = 1, \ldots, n$ *alors*

$$\|x_1 + \cdots + x_n\|^2 = \|x_1\|^2 + \cdots + \|x_n\|^2.$$

Si on prend $n = 2$*, dans* (6) *on obtient*

$$\|x_1 + x_2\|^2 = \|x_1\|^2 + \|x_2\| \text{ si } x_1 \perp x_2,$$

Dans l'espace vectoriel R^2, c'est le théorème de Pythagore.

Proposition(4) *Soit $(E, \langle \cdot,\cdot \rangle)$ un eve.*

$1\cdot$*Pour toute partie A de E, $A^{\perp}$est un sev de E.*

$2 \cdot \forall (A, B) \in (P(E))^2, (A \subset B \Rightarrow A^{\perp} \supset B^{\perp})$.

$3 \cdot \forall A \in P(E), A^{\perp} = Vect(A)^{\perp}$ *(tel que $Vect(A)$ est l'ensemble engendré par A).*

$4\cdot$ *Pour tout sev F de $E : F \oplus F^{\perp} = E$, et donc :*

$$dim(F^{\perp}) = dim(E) - dim(F).$$

$5 \cdot \forall A \in P(E), A^{\perp\perp} = Vect(A)$.

$6 \cdot E^{\perp} = \{0\} et \{0\}^{\perp} = E$.

$7 \cdot \forall A \in P(E), A \cap A^{\perp} \subset \{0\}$.

$8\cdot$ *Pour tous sev F, G de E :*

$$(F + G)^{\perp} = F^{\perp} \cap G^{\perp} \text{ et } (F \cap G)^{\perp} = F^{\perp} + G^{\perp}.$$

Proposition(5) *Soient $(E, \langle \cdot,\cdot \rangle)$un eve, $(x_i)_{i\in I}$ une famille dans E*

$$Si \left\{ \begin{array}{c} (x_i)_{i\in I} \text{ est orthogonale} \\ \forall i \in I, x_i \neq 0 \end{array} \right|, \text{ alors } (x_i)_{i\in I} \text{ est libre.}$$

Proposition(6)*(Théorème de Pythagore)* *Soient $(E, \langle \cdot,\cdot \rangle)$ un eve, $(x, y) \in E^2$. On a :*

$$x \perp y \iff ||x + y||^2 = ||x||^2 + ||y||^2 \iff ||x - y||^2 = ||x||^2 + ||y||^2.$$

Théorème(4)*(Orthogonalisation de Schmidt)* *Soient $(E, \langle \cdot,\cdot \rangle)$ un eve, $p \in N^*, (e_1, ..., e_p)$ une famille libre dans E. Il existe $(V_1, ..., V_p) \in E^p$ tel que :*

$$\left\{ \begin{array}{c} V_1, ..., V_p \text{ sont deux à deux orthogonaux} \\ \forall k \in \{1, ..., p\}, Vect(V_1, ..., V_k) = Vect(e_1, ..., e_p) \end{array} \right\}$$

(et donc : $\forall k \in \{1, ..., p\}, V_k \neq 0$).

Remarque(4) *En imposant à $(V_1,...,V_p)$ la condition :*

$$\forall k \in \{1, ..., p\}, \langle V_k, e_k \rangle = 1.$$

Il y a alors unicité de $(V_1, ..., V_p)$, et la matrice de passage de $(e_1, ..., e_p)$ à $(V_1, ..., V_p)$ est triangulaire supérieure à termes diagonaux égaux à 1 :

$$\begin{pmatrix} 1 & & \cdots \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

On abrège base orthonormale (ou : orthonormée) en : b.o.n.

Corollaire(1) *(Théorème de la base orthonormée incomplète) Pour toute famille orthonormale $(e_1, ..., e_p)$ d'un eve E, il existe $e_{p+1}, ..., e_n \in E$ (où $n = dim(E)$) tels que $(e_1, ..., e_n)$ soit une b.o.n. de E .*

Corollaire(2) *Tout eve admet au moins une b.o.n.*

Définition(5) *Soient φ une fbs sur $E \times E$ et $B = (e_1, ..., e_n)$ une base de E. On appelle matrice de φ dans (ou : relativement à) B, et on note $Mat_B(\varphi)$, la matrice carrée d'ordre n, symétrique suivante :*

$$Mat_B(\varphi) = (\varphi(e_i, e_j))_{1 \leq i.j \leq n}.$$

Proposition(7) *Si $B = (e_1, ..., e_n)$ est une b.o.n. de $(E, \langle \cdot, \cdot \rangle)$ alors :*

$$\forall x \in E, x = \sum_{i=1}^{n} \langle e_i, x \rangle e_i.$$

Proposition(8) *Soient $(E, \langle \cdot, \cdot \rangle)$ un eve, B une base de E, $A = Mat_B(\langle \cdot, \cdot \rangle)$. Alors :*

$$\left| \begin{array}{c} B \text{ est orthogonale si et seulement si } A \in D_n(R) \\ B \text{ est orthonormale si et seulement si } A = I_n. \end{array} \right|$$

Proposition(9) *Si B est une b.o.n. de E, on a, pour tout (x, y) de E^2 et en notant $X = Mat_B(x)$, $Y = Mat_B(y)$:*

$$< x, y >= XY^t.$$

1.4 Espaces préhilbertiens et hilbertiens, théorèmes de projection et de reprèsantation Riesz

1.4.1 Généralités(Espaces préhilbertiens,espace de Hilbert)

Définition(1) *(espace préhilbertien) Un espace vectoriel muni d'un produit scalaire est appelé un espace préhilbertien.*

Théorème(1) *Dans un espace préhilbertien E, l'application $\|.\| : E \to R$, donnée par $\|x\| = \langle x, x\rangle^{1/2}$, pour tout $x \in E$, est une norme pour E.*

Preuve *Nous avons $\|x\| \geq 0$ et $\|x\| = 0$ si et seulement si $x = 0$,*

de plus

$$\|\lambda x\| = \langle \lambda x, \lambda x\rangle^{1/2} = (\lambda\overline{\lambda}\langle x, x\rangle)^{1/2} = (|\lambda|^2\langle x, x\rangle)^{1/2} = |\lambda|\langle x, x\rangle^{1/2} = |\lambda|\|x\|.$$

Enfin

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle$$
$$= \|x\|^2 + \langle x, y\rangle + \overline{\langle x, y\rangle} + \|y\|^2$$
$$= \|x\|^2 + 2Re\langle x, y\rangle + \|y\|^2$$
$$\leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2$$
$$\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2$$
$$= (\|x\| + \|y\|)^2,$$

D'où

$$\|x + y\| \leq \|x\| + \|y\|$$

La norme $\|.\|$ ainsi définie s'appelle la norme induite par le produit scalaire.

Remarque(1) *En utilisant la norme induite par le produit scalaire, les inégalités (1) et (2) deviennent respectivement*

$$|\langle x, y\rangle| \leq \|x\|\|y\| \qquad \text{et} \qquad |\langle x, y\rangle| \leq 1/2(\|x\|^2 + \|y\|^2)\ .$$

Théorème(2) *(Loi du parallélogramme) La norme induite par le produit scalaire satisfait l'égalité*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)\ .$$

Preuve

En effet,

$$\|x + y\|^2 + \|x - y\|^2 = \langle x + y, x + y\rangle + \langle x - y, x - y\rangle$$
$$= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle + \langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle$$
$$= 2(\|x\|^2 + \|y\|^2).$$

Les égalités

$$\|x + y\|^2 + \|x - y\|^2 = 2(\langle x, y\rangle + \langle y, x\rangle) = 4Re\langle x, y\rangle$$

$$\|x+iy\|^2 + \|x-iy\|^2 = 4Im\langle x,y\rangle$$

Nous amènent à

$$4\langle x,y\rangle = \|x+y\|^2 - \|x-y\|^2 + i(\|x+iy\|^2 - \|x-iy\|^2)$$

Cette égalité nous montre que deux produits scalaires différents sur E entraînent deux normes induites différentes.

Théorème(3) *Un produit scalaire est une fonction continue sur $E \times E$, par rapport à la norme induite.*

Preuve *Soient $x_0, y_0 \in E$ et les suites $\{x_n; n \in N\} \subset E$ et $\{y_n; n \in N\} \subset E$ telles que $x_n \xrightarrow{\|\cdot\|} x_0$ et $y_n \xrightarrow{\|\cdot\|} y_0$.*

Alors $|\langle x_n, y_n\rangle - \langle x_0, y_0\rangle| = |\langle x_n - x_0, y_n\rangle + \langle x_0, y_n - y_0\rangle|$

$$\leq |\langle x_n - x_0, y_n\rangle| + |\langle x_0, y_n - y_0\rangle|$$

$$\leq \|x_n - x_0\|\|y_n\| + \|x_0\|\|y_n - y_0\|$$

D'où

$$\lim_{n\longrightarrow+\infty} \langle x_n, y_n\rangle = \langle x_0, y_0\rangle \ .$$

Définition(2)*(Espace de Hilbert) Un espace de Hilbert (ou hilbertien) est un espace complet par rapport à la norme induite par un produit scalaire. En d'autres mots, un espace de Hilbert est un espace de Banach dont la norme est induite par un produit scalaire(généralement, le corps de base est R ou C).*

Dans toute la suite de la thèse, H sera un espace hilbertien de dimension infinie, et $(x,y) \in H^2 \longrightarrow \langle x,y\rangle$ un produit scalaire sur H.

1.4.2 Projection d'un point sur un convexe fermé(Théorème de projection)

Définition(3)*(la convexité) Soit E un e.v. réel et $K \subset E$,*

·On dit que K est convexe si

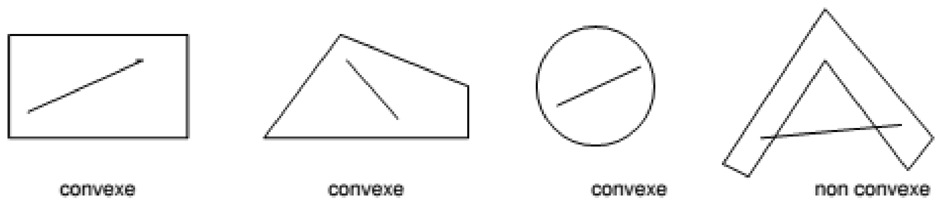
$$\forall (x,y) \in K^2, \forall t \in [0,1] : tx + (1-t)y \in K.$$

(on note encore $[x,y] \subset K$).

$\cdot$*On dit que K est dtrictement convexe si*

$$\forall (x,y) \in K^2, \forall t \in]0,1[: tx + (1-t)y \in K.$$

Ou Un ensemble K est convexe si, pour toute paire de points a, b de K, K contient aussi le segment ab.



Rappelle

$\cdot$ *On dit q'une suite (x_n) est de Cauchy si et seulement si :*

$$\forall \varepsilon > 0, \exists n_0 \in N / \forall p > q \geq n_0 \in N : | x_p - x_q | < \varepsilon.$$

$\cdot$*E est complet $\Leftrightarrow$ toute suit de Cauchy converge dans E .*

Remarque(2) *La notion d'« ensemble concave » n'existe pas.*

Théorème(3)*(Théoréme de projection) Soit $C \subset H$ une partie convexe fermée non vide. Pour tout $x \in H$, il existe un unique $y \in C$ tel que $d(x, C) = ||x - y||$, $y = x_C$ est appelé* **projeté orthogonal de** *x sur C.*

Preuve Pour tout $z \in C$ on a $||x - z|| \geq d(x, C)$ (qui existe, puisque c'est l'inf d'une partie non vide de R minorée par 0). Par définition, il existe $(x_n)_{n \in N} \in C^n$ tel que :

$$d(x, C) \leq ||x - x_n|| \leq d(x, C) + 1/n, \forall n \in N$$

Montrons que (x_n) est de **Cauchy** ; ainsi, H étant hilbertien c'est un espace complet, donc on pourra conclure quant à la convergence de la suite (x_n) :

Soit $p, q \in N$, et notons $\delta = d(x, C)$.

H étant un espace de Hilbert, l'identité du parallélogramme est vérifiée, donc

$$2(||x - x_p||^2 + ||x - x_q||^2) = ||(x - x_p) - (x - x_q)||^2 + ||(x - x_p) + (x - x_q)||^2$$

Donc on a

$$2(||x - x_p||^2 + ||x - x_q||^2) = ||x_p - x_q||^2 + 4||x - (x_p + x_q)/2||^2.$$

Or C est **convexe**, donc

$(x_p + x_q)/2 \in C$, donc $||x - (x_p + x_q)/2||^2 \geq \delta^2$.

En réorganisant, on a donc $||x_p - x_q||^2 \leq 2(||x - x_p||^2 - \delta^2 + ||x - x_q||^2 - \delta^2)$. Or il apparaît par comparaison que $\underset{p\to+\infty}{lim} ||x - x_p|| = \delta$, donc $\underset{p\to+\infty}{lim} ||x - x_p||^2 = \delta^2$. D'où pour tout $\varepsilon > 0$ ilexiste $n_0 \in N$ tel que pour tout $p \geq n_0$, $|(||x - x_p||^2 - \delta^2)| \leq \varepsilon^2/2$.

D'où pour $p, q \geq n_0$ on a $||x_p - x_q|| \leq \varepsilon$, donc (x_n) est bien une suite de Cauchy, donc elle converge vers $y \in \overline{C}$ car H est complet. Mais C est fermé, donc $C = \overline{C}$ et $y \in C$. On a $\underset{n\to+\infty}{lim} ||x - x_n|| = ||x - y||$ par continuité de l'application norme, donc par unicité de la limite,$||x - y|| = \delta = d(x, C)$.

Enfin, si on prend $y, z \in C$ vérifiant $||x - y|| = ||x - z|| = d(x, C)$: on construit une suite $(c_n)_{n\in N}$ telle que $c_n = y$ si n est pair, $c_n = z$ si n est impair. Pour tout $n \in N$ on a $||x - c_n|| = \delta$ donc avec le même raisonnement que ci-dessus, cela implique la convergence de la suite (c_n), et donc $y = z$. Il y a donc unicité du projeté orthogonal, ce que l'on souhaitait démontrer.

Proposition(1) *Soit C un convexe fermé non vide de $H, x \in H$ quelconque. Alors il existe un unique $y \in C$ vérifiant $\forall c \in C, \langle c - y, x - y\rangle \leq 0$, et $y = x_C$ projection orthogonale de x sur C (c'est une caractérisation de x_C)*

Preuve

·Si il existe $y \in C$ vérifiant l'hypothèse de la proposition, pour $z \in C$ on a :

$||z-x||^2 = ||z-y+y-x||^2 = ||z-y||^2 + ||y-x||^2 - 2\langle z-y, y-x\rangle$. Donc l'hypothèse assure que $||z - x||^2 \geq ||y - x||^2$ donc cela étant vrai pour tout $z \in C, ||y - x|| \leq d(x, C)$;mais $y \in C$ donc $||y - x|| \geq d(x, C)$ donc finalement $||x - y|| = d(x, C)$. Ayant unicité du projeté orthogonal d'aprés le *théorème01*, on a bien $y = x_C$.

·Vérifions maintenant que x_C vérifie bien l'hypothèse de l'énoncé.

On a pour tout $z \in C$,$||x - z|| \geq ||x - x_C||$ et $||x - z||^2 = ||x - x_C + x_C - z||^2 = ||x - x_C||^2 - 2\langle x - x_C, x_C - z\rangle + ||x_C - z||^2$.

Donc $2\langle z - x_C, x - x_C\rangle \leq ||z - x_C||^2$. Nous avons presque le résultat attendu, mais il nous faut se débarrasser du membre de droite. Pour cela, on paramètre le problème :

fixons $z_0 \in C$, ayant C convexe on a pour tout $\lambda \in [0, 1]$, $z = \lambda z_0 + (1 - \lambda)x_C \in C$.

D'où en appliquant le petit calcul ci-dessus, ayant $z - x_C = \lambda(z_0 - x_C)$, nous avons $2\lambda\langle z_0 - x_C, x - x_C\rangle \leq \lambda^2||z_0 - x_C||^2$. Pour $\lambda \in]0;1]$ cela donne $2\langle z_0 - x_C, x - x_C\rangle \leq \lambda||z_0 - x_C||^2$.

Avec $\lambda \to 0+$ on a bien le résultat attendu.

1.4.3 Applications importantes du théorème de projection

Lemme(1) *Pour toute partie A de $H, A^{\perp}$est un sous-espace vectoriel fermé de H.*

Preuve Le fait que $A^{\perp}$soit un sous-espace est évident. Montrons qu'il est fermé :

Si on prend $y \in \overline{A^{\perp}}$ il existe une suite (y_n) d'éléments de $A^{\perp}$ qui converge vers y. La fonction $(x, y) \mapsto \langle x, y\rangle$ est continue d'aprés l'inégalité de **Cauchy-Schwarz** ; or $\forall x \in A, \forall n \in N$ on a $\langle x, y_n\rangle = 0$ donc avec $n \to +\infty$ on a $\langle x, y\rangle = 0$ par continuité, donc $y \in A^{\perp}$: $A^{\perp}$ est bien un fermé.

Nous avons maintenant deux petites propositions trés importantes pour la deuxiéme section de cet article, et qui généralisent des résultats déjà vus en dimension finie.

Proposition(2) *Soit F un sous-espace vectoriel fermé de H. On a*

$1 \cdot F \oplus F^{\perp} = H$

$2\cdot$ *Pour tout $x \in H$, on a $x_F = p_F(x)$ projection orthogonale sur F ; ainsi $x - x_F \in F^{\perp}$*

$3 \cdot F^{\perp^{\perp}} = F$.

Preuve

$1 + 2$: le fait que F et F$^{\perp}$sont en somme directe ne pose aucun problème. Montrons que pour $x \in H$ il existe $(y, z) \in F \times F^{\perp}$ tel que $x = y + z$.

F est un sous-espace vectoriel, c'est donc en particulier un convexe ;étant de plus fermé,d'aprés le *théorème01*, il existe un unique $x_F \in F$ tel que $d(x, F) = ||x - x_F||$. Notons $y = x - x_F$; d'aprés la *proposition1*, on a pour tout $f \in F\langle f - x_F, y\rangle \leq 0$. Soit $y_0 \in F$ fixé. Comme F est un espace vectoriel, $y_0 + x_F \in F$ et $-y_0 \in F$: cela implique $\langle y_0, y\rangle \leq 0$ et $\langle -y_0, y\rangle \leq 0$ donc $\langle y_0, y\rangle = 0$. Ceci étant vrai pour tout $y_0 \in F$, on a bien $y = x - x_F \in F^\perp$ donc les affirmations 1 et 2 sont démontrées.

3 : On a toujours $F \subset F^{\perp^\perp}$. Montrons l'inclusion inverse.

On a d'aprés l'affirmation 1 de cette proposition que $H = F \oplus F^\perp$. Donc pour $x \in F^{\perp^\perp}$il existe un unique couple $(x_F, x_{F^\perp}) \in F \times F^\perp$ tel que x = x$_F$ +x$_{F^\perp}$. Or l'on a $||x_{F^\perp}||^2 = ||x - x_F||^2$ donc $||x_{F^\perp}||^2 = \langle x - x_F, x - x_F\rangle = \langle x, x - x_F\rangle - \langle x_F, x - x_F\rangle$.

Ayant $x \in F^{\perp^\perp}$,$x - x_F \in F^\perp$ on a $\langle x, x - x_F\rangle = 0$; de même, comme $x_F \in F$ et $x - x_F \in F^\perp$, $\langle x_F, x - x_F\rangle = 0$. Donc $||x_{F^\perp}|| = 0$ donc $x_{F^\perp} = 0$: $x \in F$, ce qu'il fallait démontrer.

Proposition(3) *Soit F un sous-espace quelconque de H. Alors on a :*

$1 \cdot \overline{F} = F^{\perp^\perp}$

$2 \cdot F$ *dense dans* $H \iff F^\perp = \{0\}$.

Peuve

1· On a toujours $F \subset F^{\perp^\perp}$; donc $F \subset F^{\perp^\perp}$. Ainsi, le *lemme1* affirmant que $F^{\perp^\perp}$ est unfermé, on a $F \subset F^{\perp^\perp}$.

On a toujours $F \subset F$; par orthogonalité, on a donc $F^\perp \subset F^\perp$donc en recomposant, $F^{\perp^\perp} \subset F^{\perp^\perp}$. Or F est un fermé, donc l'affirmation 3 de la *proposition2* nous donne $F = F^{\perp^\perp}$. Ainsi, $F^{\perp^\perp} \subset F$. On a donc bien par double inclusion $F = F^{\perp^\perp}$.

2·On a F dense $\iff F = H \iff F^{\perp^\perp} = H$ d'aprés ci-dessus.

Or l'affirmation 1 de la proposition 2 nous donne $H = F^\perp \oplus F^{\perp^\perp}$donc F dense $\iff F^\perp = \{0\}$.

1.4.4 Théorème de représentation de Riesz, adjoint d'un endomorphisme en dimension finie

Théorème(4)*(de représentation de Riesz) Soit H' le* **dual topologique** *de H (c'est-à-dire l'ensemble des formes linéaires continues sur H).*

Soit $a \in H$, on note ϕ_a :$x \in H \mapsto < a, x > \in K$, $K = R$ ou C.

Alors $\phi : a \in H \mapsto \phi_a \in H'$ est bien définie, et c'est un isomorphisme de H sur

H'.

Preuve Tout d'abord, ϕ e st bien définie car l'inégalité de Cauchy-Schwarz assure la continuité de la fonction ϕ_a($a \in H$ fixé).

ϕ est injective : si $\phi_a = 0$ alors en particulier, $\phi_a(a) = \langle a, a\rangle = ||a||^2 = 0$ donc $a = 0$ et $ker\phi = \{0\}$.

Montrons que ϕ est surjective. Soit $L \in H\prime$. Si $L = 0L = \phi(0)$ et c'est bon ; sinon, notons $E = kerL$, E est un hyperplan de H.

Ayant $E = L^{\langle -1\rangle}(\{0\})$ et L continue, E est un fermé de H. D'aprés la *proposition02*, $E \oplus E^{\perp} = H$. Ayant L différente de l'application nulle, il existe $u \in E^{\perp}$tel que $L(u) \neq 0$. On a en fait $E = Vecta$: si on prend $x \in E^{\perp}$et si on note $w = x - \frac{L(x)}{L(a)}a$ on a $L(w) = L(x) - \frac{L(x)}{L(a)}L(a)$ par linéarité, donc $L(w) = L(x) - L(x) = 0 : w \in E$. Or $w \in E^{\perp}$ par construction, donc $w \in E \cap E \perp = \{0\}$ d'où $x = \frac{L(x)}{L(a)}a : x \in Vecta$ et par double inclusion (l'autre inclusion étant évidente) on a bien $E^{\perp} = Vecta$.

On a donc pour tout $x \in Hx = \frac{L(x)}{L(a)}a + x_E, x_E \in E$. Notons maintenant $b = \frac{L(x)}{\|a\|^2}a$.On pour tout $x \in H$

$\langle x, b\rangle = \langle \frac{L(x)}{L(a)}a + x_E, \frac{L(x)}{\|a\|^2}a\rangle = \frac{L(x)}{L(a)}\frac{L(x)}{\|a\|^2}\langle a, a\rangle = L(x)$ car $x_E \in E$ et $a \in E^{\perp}$. D'où l'on a $L = \phi_b$ et ϕ est bien surjective.

En conclusion, φ est bien un isomorphisme, il y a donc isomorphisme entre H et son dual topologique.

Ce théoréme, puissant, nous permet donc de pouvoir définir un adjoint dans un espace de *Hilbert* pour les endomorphismes continus, comme le montre la proposition suivante :

Proposition(4) *Soit $u \in \mathcal{L}_c(H)$ endomorphisme continu de H.*

Il existe un unique $v \in \mathcal{L}_c(H)$ tel que pour tout $(x, y) \in H^2$on ait $\langle u(x), y\rangle = \langle x, v(y)\rangle$

v est l'adjoint de u, noté u^*, et de plus on a $|||u||| = |||u^*|||$.

Peuve Soit $u \in \mathcal{L}_c(H)$. Fixons $y \in H$, et notons $\psi_y : x \in H \mapsto \langle u(x)|y\rangle$. C'est une forme linéaire(aisé à démontrer) et continue en tant que composée de fonctions continues (le produit scalaireest bien continu, u est continu par hypothèse). Donc $\psi_y \in H\prime$ le dual topologique de H : le théorème de **représentation de Riesz** permet d'affirmer l'existence d'un unique $v(y) \in H$ tel que $\langle u(x), y\rangle = \psi_y(x) = \langle v(y), x\rangle$.

v est bien une application linéaire : si on prend $y, z \in H$, on a $\psi_y + z(x) = \langle u(x), y + z\rangle = \langle u(x), y\rangle + \langle u(x), z\rangle$ donc $\psi_y + z(x) = \langle x, v(y)\rangle + \langle x, v(z)\rangle = \langle x, v(y) + v(z)\rangle$. Par unicité dans le *théorème2*, on a bien $v(y + z) = v(y) + v(z)$. De même, pour $\lambda \in K$ on aaussi $v(\lambda y) = \lambda v(y)$ et on a bien $v(0) = 0$.

Si on notait v' une application de $\mathcal{L}(H)$ vérifiant les hypothèses de la *proposition4*, à y fixé on aurait $\psi_y(x) = \langle x, v\prime(y)\rangle$ donc l'unicité imposée par le **théoréme de Riesz** nous donne $v\prime(y) = v(y)$, donc v est unique.

Enfin, v est bien continue : on a $\langle v(y), v(y)\rangle = \langle y, u(v(y))\rangle$, donc

$||v(y)||^2 \leq ||u(v(y))||.||y||$ d'aprés l'inégalité de **Cauchy-Schwarz**, et la continuité de u nous donne $||v(y)||^2 \leq |||u|||.||v(y)||.||y||$.

D'où pour $y \notin kerv$ on peut diviser par $||v(y)||$ et l'on a $||v(y)|| \leq |||u|||.||y||$, inégalité en core valable pour $v(y) = 0$. On a donc en particulier pour

$||y|| \leq 1||v(y)|| \leq |||u|||$ donc v est bornée sur la boule fermée unité : ayant v linéaire, on a donc v continue, et $|||v||| \leq |||u|||$.

On note maintenant $v = u^*$. On remarque que l'unicité imposée par la *proposition4* (unicité qui a été démontrée) impose $(u^*)^* = u$; ainsi, le raisonnement effectué ici avec u se refait avec u^* qui est bien élément de $\mathcal{L}_c(H)$, et l'on a $|||u^*||| \leq |||(u^*)^*|||$ c'est-à-dire $|||u^*||| \leq |||u|||$. On a donc bien par double inégalité $|||u||| = |||u^*|||$.

Rappel Soit $(E, \langle ., .\rangle)$ un espace euclidien, $u \in \mathcal{L}(E)$ un endomorphisme de E. Il existe un élément $u^* \in \mathcal{L}(E)$ unique, vérifiant : $\forall (x, y) \in E^2$:

$$\langle u(x), y\rangle = \langle x, u^*(y)\rangle \ .$$

On appelle u^* *l'endomorphisme adjoint* de u.

Troisième partie

Chapitre 2

Chapitre 2

Théorème de Karush, Kuhn et Tucker

2.1 Introduction à l'optimisation

2.1.1 Quelques notions d'optimisation

L'optimisation vise à résoudre des problèmes où l'on cherche à déterminer parmi un grand nombre de solutions candidates celle qui donne le meilleur rendement. Plus précisément, on cherche à trouver une solution satisfaisant un ensemble de contraintes qui minimise ou maximise une fonction donnée. L'application de l'optimisation est en expansion croissante et se retrouve dans plusieurs domaines. Les problèmes considères dans ce document s'écrivent sous la forme standardforme standard

$$\begin{aligned} &\min_{x \in R^n} f(x) \qquad (1) \\ &\text{s.c. } x \in S \end{aligned}$$

Où $x = (x_1, x_2, .., x_n)^\top$est un vecteur de R^n, $f : R^n \longrightarrow R$ est la fonction que l'on désire minimiser (appelée *fonction objectif*), $S \subseteq R^n$ est l'ensemble dans lequel les points doivent appartenir, et s.c. est l'abréviation de sous la ou les contraintes.

La formulation (1) signifie que l'on cherche à trouver une solution du domaine réalisable $x^* \in S$ dont la valeur de la fonction objectif est la plus petite.

Définition(1) *Une solution $x^* \in S$ est un minimum global de la fonction f sur le domaine S si*

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

La valeur optimale est $f(x^*)$.

Notez que le minimum global n'est pas nécessairement unique, mais la valeur optimale l'est. Par exemple, le problème d'optimisation suivant

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

posséde une infinité de minima globaux, soient $\{\frac{3\pi}{2} + 2k\pi : k \in Z\}$ mais une seule valeur optimale : -1.

Définissons maintenant la notion d'optimalité dans un voisinage restreint. Pour introduire cette idée, on va définir $B_\varepsilon(x^*)$ comme étant l'ensemble des points de R^ndont la distance à x^* est inférieure à ε, un scalaire positif donné. Cet ensemble est communément appelé une boule de rayon ε centrée en x^* , et s'écrit formellement :

$B\varepsilon(x^*) = \{x \in R^n :\parallel x - x^* \parallel < \varepsilon\}$.

Définition(2) *Une solution $x^* \in S$ est un minimum local de la fonction f sur le domaine S si*

$$f(x^*) \leq f(x) \quad , \forall x \in S \cap B\varepsilon(x^*).$$

Les maxima sont définis de façon similaire, il suffit de remplacer les inégalités $(\leq)$ aux définitions 1 et 2 par $(\geq)$. La *figure 1* illustre le cas d'une fonction d'une seule variable possédant trois minimums locaux,dont un minimum globale

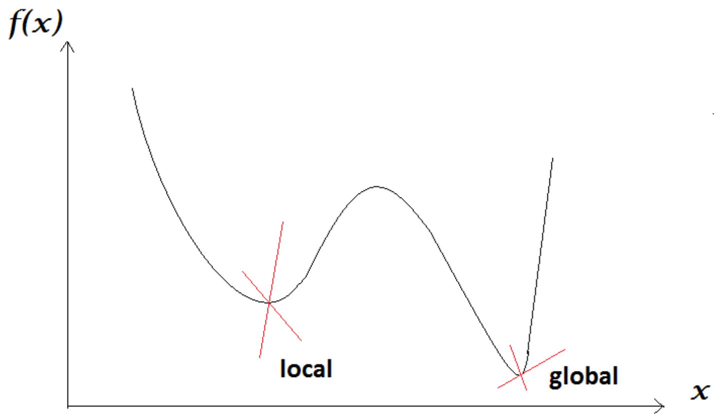


figure1 la represantation de deux minimums local et global

2.1.2 Rappels

Le gradient d'une fonction $f : R^n \to R$ différentiable. évalué au point $x \in R^n$est un vecteur de R^n s'écrivant

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, ..., \frac{\partial f(x)}{\partial x_n}\right)^T.$$

La dérivée directionnelle de f en $x \in R^n$ dans la direction unitaire $d \in R^n$est

$$f'(x) = \lim_{t \to 0^+} \frac{f(x+td)-f(x)}{t} = d \cdot \nabla f(x).$$

De plus, si les dérivées secondes de f existent et sont continues, alors la matrice Hessienne s'écrit

$$\nabla^2 f(x) = \begin{bmatrix} f''_{x_1x_1} & f''_{x_1x_2} & -- & f''_{x_1x_n} \\ f''_{x_2x_1} & f"{}_{x_2x_2} & -- & f''_{x_2x_n} \\ | & | & \diagdown & | \\ f''_{x_nx_1} & f''_{x_nx_2} & -- & f''_{x_nx_n} \end{bmatrix}.$$

Exemple(1) Le gradient et la matrice *Hessienne* de la fonction

$f(x) = -(x_1 - 2x_2) - e^{x_1}$ sont

$\nabla f(x) = (-2(x_1 - 2x_2) - e^{x_1}, 4(x_1 - 2x_2))^T, \nabla^2 f(x) = \begin{bmatrix} -2 - e^{x_1} & 4 \\ 4 & -8 \end{bmatrix}.$

Des notions d'algébre serviront au classement des optima.

Définition(3) *Une matrice symétrique A de dimension $n \times n$ est dite*

$\cdot$ *Semi-définie positive si* $\langle y, Ay \rangle = y^T Ay \geq 0, \forall y \in R^n$.

$\cdot$ *Définie positive si* $\langle y, Ay \rangle = y^T Ay > 0, \forall y \neq 0 \in R^n$.

$\cdot$ *Semi-définie négative si* $\langle y, Ay \rangle = y^T Ay \leq 0, \forall y \in R^n$.

$\cdot$ *Définie négative si* $\langle y, Ay \rangle = y^T Ay < 0, \forall y \neq 0 \in R^n$.

Si aucune des propriétés ci-dessus n'est satisfaite, la matrice A est dite indéfinie.

Une façon simple de vérifier si une matrice est définie positive est de considérer les *déterminants* des n sous matrices suivantes de A

$$d_1 = \det\left([a_{11}]\right),$$
$$d_2 = \det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right),$$
$$d_3 = \det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{22} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right), ..., d_n = \det\left(\begin{bmatrix} f''_{x_1x_1} & f''_{x_1x_2} & \cdots & f''_{x_1x_n} \\ f''_{x_2x_1} & f''{}_{x_2x_2} & \cdots & f''_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f''_{x_nx_1} & f''_{x_nx_2} & \cdots & f''_{x_nx_n} \end{bmatrix}\right).$$

Ces n déterminants sont appelés les mineurs principaux dominants. Dans le cas la matrice est inversible, et donc lorsque $d_n \neq 0$, il est toujours facile de déterminer si une matrice est définie positive, définie négative ou bien indéfinie. En effet, il suffit d'analyser le signe de chacun des mineurs principaux dominants.

Classement d'une matrice inversible

Soit $d_1, d_2, ...d_n$ les mineurs principaux dominants d'une matrice A symétrique inversible de dimension $n \times n$ et $e_1, e_2, ..., e_n$ les mineurs principaux dominants de la matrice $-A$.

$\cdot$ Si $d_i > 0$ pour chaque $i \in \{1, 2..., n\}$ alors A est définie positive ;

$\cdot$ Si $e_j > 0$ pour chaque $j \in \{1, 2..., n\}$ alors A est définie négative ;

$\cdot$ Sinon A est indéfinie.

La vérification qu'une matrice singuliére est semi-définie positive, semi-définie négative ou indéfinie requiert beaucoup plus de calculs. Par exemple, pour montrer qu'une matrice est semi-définie positive, il ne suffit pas de vérifier que les mineurs principaux dominants sont tous positifs ou nuls. Une possibilité consiste à analyser toutes valeurs propres de la matrice. Ce sujet sera abordé dans des cours plus avancés.

Exemple(Suit de *l'exemple1*) Les déterminants des sous-matrices principales dominantes de

$$\nabla^2 f(x) = \begin{bmatrix} -2 - e^{x_1} & 4 \\ 4 & -8 \end{bmatrix}.$$

sont respectivement

$d_1 = \det([-2 - e^{x_1}]) = -2 - e^{x_1} < 0$ et $d_2 = det \begin{bmatrix} -2 - e^{x_1} & 4 \\ 4 & -8 \end{bmatrix} = (-2 - e^{x_1})(-8) - (4)(4) = 8e^{x_1} > 0$.

Leurs signes sont indépendant de la valeur de x. La matrice $\nabla^2 f(x)$ n'est donc pas définie positive ni même semi-définie positive.

Cependant, on remarque que les mineurs principaux dominants de la matrice $-\nabla^2 f(x)$ sont

$\mathrm{e}_1 = \det([2 + e^{x_1}]) = 2 + e^{x_1} > 0$ et $e_2 = det \begin{bmatrix} 2 + e^{x_1} & -4 \\ -4 & 8 \end{bmatrix} = (2+e^{x_1})(8)(4)(4) = 8e^{x_1} > 0$.

On peut alors conclure que la matrice $-\nabla^2 f(x)$ est définie positive, et donc la matrice $\nabla^2 f(x)$ est définie négative, et ce pour toutes les valeurs de x.

2.2 Optimisation à plusieurs variables sans contraintes

Dans cette section nous considérons le cas particulier du problème d'optimisation (1) sans contraintes où $S = R^n$, c'est-à-dire ;

$$\min_{x \in R^n} f(x) \qquad (2)$$

où $f \in C^2$(cette notation signifie f est différentiable aux moins deux fois, et que ses dérivées sont continues).

2.2.1 Conditions d'optimalité

Soit x un minimum local de f sur $S = R^n$. La *définition2* assure qu'il existe un scalaire $\varepsilon > 0$ tel que

$$f(x^*) \leq f(x) \qquad\qquad \forall x \in B\varepsilon(x^*),$$

en particulier pour n'importe quelle direction unitaire $d \in R^n$ et scalaire $t > 0$ suffisamment petit, le point $x = x^* + td$ appartient à la boule $B\varepsilon(x^*)$ et donc

$$f(x^*) \leq f(x^* + td).$$

Ceci implique que la dérivée directionnelle de f en x dans n'importe la direction unitaire $d \in R^n$ satisfait

$$f_d'(x^*) = \lim_{t \to 0^+} \frac{f(x^*+td)-f(x^*)}{t} \geq 0.$$

Le résultat précédent est valide pour n'importe quelle direction unitaire d , et donc il est valide en particulier pour $-\frac{\nabla f(x^*)}{\|\nabla f(x^*)\|}$. Il s'ensuit que

$$0 \leq \ f_d'(x^*) = d \cdot \nabla f(x^*) = -\frac{\nabla f(x^*) \cdot \nabla f(x^*)}{\|\nabla f(x^*)\|} = -\|\nabla f(x^*)\|,$$

et donc, $\nabla f(x^*) = 0$ (car on a montré que $0 \leq -\|\nabla f(x^*)\|$, or $\|\nabla f(x^*)\|$ est toujours positif). Nous avons donc montré la condition d'optimalité suivante.

Condition nécessaire de premier ordre

Si x est un minimum local de la fonction f sur R^n, alors $\nabla f(x^*) = 0$ (c'est-à-dire, x est un point critique).

Cette définition donne une condition nécessaire mais pas suffisante. En effet, il est possible que le gradient soit nul en un point et que ce point ne soit pas un minimum local (par exemple, ce point peut être un maximum local ou un *point de selle*).

Un point critique $x \in R^n$ est un point de selle si pour n'importe quel $\varepsilon > 0$, il existe deux points $a \in B\varepsilon(x^*)$ *et* $b \in B\varepsilon(x^*)$ qui soient tels que $f(a) < f(x^*) < f(b)$. Un point de selle n'est donc ni un minimum local ni un maximum local.

Les conditions de second ordre nous permettent de distinguer ces cas. Considèrons encore une fois x*un minimum local de f sur $S = R^n$. A l'aide du développement de Taylor d'ordre 2 de f autour du minimum local x^*on obtient pour toute direction unitaire $d \in R^n$et pour $t \in R$ suffisament petit

$$f(x^*) \leq f(x^* + td) \approx f(x^*) + td \cdot \nabla f(x^*) + \tfrac{t^2}{2} d^T \nabla^2 f(x^*) d$$
$$= f(x^*) + \tfrac{t^2}{2} d^T \nabla^2 f(x^*) d.$$

Pour passer de la première à la deuxième ligne, on a utilisé le fait que $\nabla f(x^*) = 0$ en un minimum local. En simplifiant cette dernière expression, et en divisant par $t^2 > 0$ on obtient

$$\tfrac{1}{2} d^T \nabla^2 f(x^*) d \geq 0$$

une expression indépendante de t . On a alors la condition suivante.

Condition nécessaire de second ordre

Si x^* est un minimum local de la fonction f sur R^n, alors $\nabla f(x^*) = 0$ et

$y^T \nabla^2 f(x^*) y \geq 0$ pour tout $y \in R^n$ (c'est-à-dire, la matrice Hessienne $\nabla^2 f(x^*)$ est semi-définie positive).

La contraposée de cette condition assure que si la matrice des dérivés secondes $\nabla^2 f(x^*)$ n'est pas semi-définie positive, alors x^* n'est pas un minimum local de f . De façon similaire, si la matrice n'est pas semi-définie négative, alors x^* n'est pas un maximum local de f . Et donc, si la matrice est indéfinie en un point critique x^* de la fonction f , alors x^* est un point selle. En effet, si la matrice est indéfinie alors elle n'est ni semi-définie positive ni semi-définie négative, et cons équemment x^* n'est ni un minimum ni un maximum local de f .

De plus, si la condition de second ordre est strictement satisfaite, on obtient la condition suffisante suivante.

Condition suffisante de second ordre

Soit x$^* \in R^*$.Si $\nabla f(x^*) = 0$ et $y^T \nabla^2 f(x^*) y \geq 0$ pour tout $y \in R^n \backslash \{0\}$ (c'est-à-dire, la matrice Hessienne$\nabla^2 f(x^*)$est définie positive), alors x^* est un minimum local de la fonction f sur R^n.

Nous pouvons spécialiser les conditions d'optimalités de deuxième ordre aux fonctions de deux variables.

Pour s'attaquer à un problème d'optimisation, on procédera de la façon suivante. En un premier temps, on utilisera les conditions nécessaires du premier ordre pour identifier tous les points critiques de f. C'est-à-dire,on trouvera tous les points où le gradient s'annule. Ensuite, pour chacun de ces points, on évaluera la matrice des dérivées secondes, et on utilisera les conditions d'optimalité de deuxième ordre pour classer les points. On ne pourra conclure seulement lorsque cette matrice sera inversible. Et dans ce cas, si elle est définie positive il s'agira d'un minimum local, si elle est définie négative il s'agira d'un maximum local, et autrement (c'est-à-dire si la matrice est indéfinie) il s'agira d'un point selle.

Exemple(1) Considèrons la fonction de trois variables $f(x) = 6x_1^2 + x_2^3 + 6x_1x_2 + 3x_2^2 + \frac{1}{4}x_3^4 - \frac{1}{3}x_3^3$.Nous allons identifier tous les points critiques de cette fonction, et allons utiliser les conditions d'optimalité *afin* de déterminer leur nature (minimum, maximum ou point de selle). Le gradient et la matrice Hessienne de cette fonction sont

$$\nabla f(x^*) = \begin{bmatrix} 12x_1 + 6x_2 \\ 3x_2^2 + 6x_1 + 6x_2 \\ x_3^3 - x_3^2 \end{bmatrix} \text{ et } \nabla^2 f(x^*) = \begin{bmatrix} 12 & 6 & 0 \\ 6 & 6x_2 + 6 & 0 \\ 0 & 0 & 3x_3^2 - 2x_3 \end{bmatrix}.$$

○ **Etape 1 : Identification des points critiques**

La condition nécessaire de premier ordre $\nabla f(x_1, x_2, x_3) = (0,0,0)^T$nous permet d'identifier quatre points critiques : $x_A = (0,0,0)^T, x_B = (0,0,1)^T, x_C = (\frac{1}{2}, -1, 0)^T$ et $x_D = (\frac{1}{2}, -1, 1)^T$.

○ **Etape 2 : Nature des points critiques**

Les conditions de deuxième ordre nous permettent de classer certains de ces points critiques.

Point critique	$\nabla^2 f(x^*)$ d_1	d_2	d_3	$-\nabla^2 f(x^*)$ e_1	e_2	e_3	Conclusion
$x_A = (0,0,0)^T$	12	36	**0**	-12	36	0	**matrice singulière** $\Longrightarrow$ indéterminé
$x_B = (0,0,1)^T$	**12**	**36**	**36**	-12	36	36	$\mathbf{d}_i > \mathbf{0}$ $\Longrightarrow$ minimum local
$x_C = (\frac{1}{2}, -1, 0)^T$	12	-36	**0**	-12	-36	0	**matrice singulière** $\Longrightarrow$ indéterminé
$x_D = (\frac{1}{2}, -1, 1)^T$	12	$\mathbf{-36}$	$\mathbf{-36}$	$\mathbf{-12}$	$\mathbf{-36}$	36	**matrice inversible et** $\mathbf{d}_i \not> \mathbf{0}, \mathbf{e}_i \not> \mathbf{0}$ $\Longrightarrow$ point selle

Cette analyse permet d'identifier un minimum local x_B et un point selle x_D. Une analyse plus approfondie serait nécessaire pour classer les deux points critiques x_A et x_C.

2.3 Optimisation à plusieurs variables avec contraintes d'égalités

Dans la plupart des problèmes d'optimisation, les variables ne sont pas libres de prendre n'importe quelle valeur.

Elles sont habituellement restreintes à un domaine. Dans cette section, nous nous penchons sur les problèmes sous la forme générale (1). Le résultat suivant, cité sans preuve, sera fréquemment utilisé.

Théorème(1) *Si l'ensemble S est fermé et borné, et si la fonction f est continue sur S, alors il existe un minimum global atteint en un point de S et un maximum global atteint en un point de S.*

On a un autre forme équivalent

$$\begin{cases} Min\ f(x) \\ g_j(x) = 0, j = 1, 2, .., m \end{cases} ; x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Généralement on a $m < n$

$\cdot$ Si $m > n \rightarrow \begin{cases} redondance\ dans\ les\ contraintes \\ trop\ de\ contraintes. \end{cases}$,

$\cdot$ Un nombre excessif de contrainte $\underset{Conduit\ à}{\rightarrow}$ problème sans solution.

Il existe plusieurs méthodes pour ce type de problèmes. Parmi elles, on peut citer :

$\circ$ Méthode de substitution directe

$\circ$ Méthode des multiplicateurs de Lagrange.

2.3.1 Méthode de substitution directe

Définition(1)

$$\left.\begin{array}{c} dx_1, dx_2, .., dx_n \\ une\ variation\ \textbf{admissible} \\ au\ tour\ du\ point \\ x^* = (x_1^*, .., x_n^*)^t \end{array}\right\} \rightleftarrows \begin{cases} g_i(x_1^* + dx_1, .., x_n^* + dx_n) = 0 \\ \forall i = 1, .., m \end{cases}.$$

Problème sans contrainte

$\cdot$ n variables indépendantes

$\cdot$ fonction objectif peut être évaluée pour tout ensemble de n nombres

Problème avec contrainte

au moins une variable indépendante devient liée aux autres variables avec l'ajout de chaque équation contrainte

$$\left.\begin{array}{c} n\ variables \\ m\ contraintes \end{array}\right\} \rightarrow n - m \text{ variables indépendantes}$$

Si on choisit les valeurs de $n - m$ variables, les valeurs des variables restantes sont déterminées par les équations contraintes

En théorie

m variables quelconques peuvent être exprimées en fonction

des autres $n-m$ variables en utilisant les m contraintes →On remplace les m variables dans la fonction objectif→On obtient une nouvelle fonction qui ne dépend que de n-m variables →La nouvelle fonction n'est liée à aucune contrainte →On obtient alors un problème d'optimisation sans contraintes.

Cette méthode de substitution directe paraît simple en théorie mais, elle peut s'avérer difficile en pratique.

souvent les contraintes ne sont pas linéaires ce qui ne permet pas d'exprimer m variables quelconques en fonction des autres $n-m$ variables

Mais dans des cas simples, la méthode peut s'avérer très utile.

Exemple(1)

Trouver les dimensions d'une boite de volume maximum qui peut être circonscrite par une sphère de rayon 1.

Le problème peut être formulé comme suit :

$$\begin{cases} Min\ f(x) = 8x_1x_2x_3 \\ x_1^2+x_2^2+x_3^2=1 \end{cases} ;\ x_1>0, x_2>0, x_3>0.$$

On a $x_1^2+x_2^2+x_3^2=1 \implies x_3=(1-x_1^2-x_2^2)^{\frac{1}{2}}$

$$f(x)=8x_1x_2(1-x_1^2-x_2^2)^{\frac{1}{2}}$$

$$\begin{cases} \frac{\partial f(x)}{\partial x_1} = 8x_2\left((1-x_1^2-x_2^2)^{1/2} - \frac{x_1^2}{(1-x_1^2-x_2^2)^{1/2}}\right)=0 \\ \frac{\partial f(x)}{\partial x_2} = 8x_1\left((1-x_1^2-x_2^2)^{1/2} - \frac{x_2^2}{(1-x_1^2-x_2^2)^{1/2}}\right)=0 \end{cases} \implies \begin{cases} x_1=1/\sqrt{3} \\ x_2=1/\sqrt{3} \end{cases}.$$

La solution est $x^*=\begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$ et donne un volume maximum de $f(x^*)=\frac{8}{3\sqrt{3}}$,

La matrice Hessienne au point x^*

$$H=\begin{bmatrix} -32/\sqrt{3} & -16/\sqrt{3} \\ -16/\sqrt{3} & -32/\sqrt{3} \end{bmatrix}$$

H est définie négative→ x^* correspond à un maximum.

2.3.2 Méthode des multiplicateurs de Lagrange

On introduit une variable additionnelle pour chaque contrainte.

Pour un problème à n variables et m contraintes, on ajoute donc m variables pour avoir en tout $n+m$ variables

Pour illustrer la méthode on prend un exemple simple où $n=2$ et $m=1$

Condition nécessaire pour le cas général

n variables et m contraintes sous forme d'égalités

Théorème(2) *Une condition nécessaire pour que la fonction f sous les contraintes $g_j(x)=0, i=1,2,..,m$ admette un minimum relatif au point x^*est que les dérivées partielles de la fonction de Lagrange*

$$L=L(x_1,..,x_n,\lambda_1,..,\lambda_m)$$

Par rapport à chacun de ces arguments soient nulles au point x^*.

$$x^* \text{ est un minimum relatif de } f \implies \begin{cases} \frac{\partial L}{\partial x_i}(x^*,\lambda)=0, i=1,2,..,n \\ \frac{\partial L}{\partial \lambda_i}(x^*,\lambda)=0, i=1,2,..,m \end{cases}.$$

Remarque(1)

On a m contraintes, on peut choisir n-m variables indépendantes

$x_1,x_2,..,x_m$ *les variables dépendantes*

$x_{m+1},x_{m+2},..,x_n$ *les variables indépendantes*

On utilise les m premiers coefficients de dL pour définir les $\lambda_i, i=1,..,m$

$$\frac{\partial f}{\partial x_i}(x^*)+\sum_{j=1}^{m}\lambda_j\frac{\partial g_j}{\partial x_i}(x^*)=0, i=1,..,n.$$

Les conditions suffisantes pour le cas général

Théorème(3) *Une condition suffisante pour que la fonction f sous les contraintes $g_i(x)=0, i=1,2,..,m$ admette un minimum relatif au point x^* est que la forme quadratique*

$$F=\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\partial^2 L}{\partial x_i \partial x_j}dx_i dx_j$$

évaluée au point $x = x^*$ soit positive pour toutes les valeurs de dx pour lesquelles les contraintes sont satisfaites.

En pratique, comment montrer que la forme quadratique

$$F = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\partial^2 L}{\partial x_i \partial x_j} dx_i dx_j$$

est positive ou négative pour toute variation admissible dx c.à.d. définie positive ou définie négative ?

On utilise le polynôme défini par

$$P(z) = \begin{vmatrix} H_L - I_n z & J\left(\frac{g_1,..,g_m}{x_1,..,x_n}\right)^t \\ J\left(\frac{g_1,..,g_m}{x_1,..,x_n}\right) & \begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{matrix} \end{vmatrix}$$

Il a été montré $(*)$ que :

$$\left.\begin{matrix} \text{les racines du polynôme} \\ P(z) \\ \text{sont toutes positives} \end{matrix}\right\} \Longrightarrow \quad F = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\partial^2 L}{\partial x_i \partial x_j} dx_i dx_j > 0, \text{ pour toute}$$

variation admissible dx.

$$P(z) = \begin{vmatrix} H_L - I_n z & (\nabla g_1 \ \cdots \ \nabla g_m) \\ \begin{pmatrix} (\nabla g_1)^t \\ \vdots \\ (\nabla g_m)^t \end{pmatrix} & \begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{matrix} \end{vmatrix} \qquad H_L = \left(\frac{\partial^2 L}{\partial x_i \partial x_j}\right)_{i,j=1,..,n}.$$

Exemple(2) Trouver les dimensions d'une cannette cylindrique (le haut et bas compris) fabriquée à l'aide d'une feuille d'aluminium de volume maximum et telle que la surface totale est de 24π

$$\begin{cases} Max\ f(x_1, x_2) = \pi x_1^2 x_2 \\ 2\pi x_1^2 + 2\pi x_1 x_2 = 24\pi \end{cases} ;\ x_1 > 0, x_2 > 0.$$

$$\begin{aligned} L(x_1, x_2, \lambda) &= f(x_1, x_2) + \lambda g(x_1, x_2) \\ &= \pi x_1^2 x_2 + \lambda(2\pi x_1^2 + 2\pi x_1 x_2 - 24\pi) \end{aligned}$$

$$\begin{cases} \frac{\partial L}{\partial x_1} = 2\pi x_1 x_2 + 4\pi\lambda x_1 + 2\pi\lambda x_2 = 0 \\ \frac{\partial L}{\partial x_2} = \pi x_1^2 + 2\pi\lambda x_2 = 0 \\ \frac{\partial L}{\partial \lambda} = 2\pi x_1^2 + 2\pi x_1 x_2 - 24\pi = 0 \end{cases} \Longrightarrow \begin{cases} x_1^* = 2 \\ x_2^* = 4 \\ \lambda^* = -1 \end{cases}$$

On évalue ces expressions au point $x^* = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$, $\lambda^* = -1$, on obtient alors

$$P(z) = \begin{vmatrix} 4\pi - z & 2\pi & 16\pi \\ 2\pi & -z & 4\pi \\ 16\pi & 4\pi & 0 \end{vmatrix} = 272\pi^2 z + 192\pi^3$$

$P(z) = 272\pi^2 z + 192\pi^3 = 0 \implies z = -\frac{12}{17}\pi < 0$. alors x$^* = (2,4)^t$ est maximum local de f s.c g.

2.4 Optimisation à plusieurs variables avec contraintes d'inégalités

Soit le problème de la forme

$$\begin{cases} Min\ f(x) \\ g_j(x) \leq 0, j = 1, 2, .., m \end{cases} ; x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Définition(1) *Soit une contrainte inégalité $g_i(x) \leq 0$ et $x_0 \in R^n$*

· *Si x_0 satisfait $g_i(x_0) < 0$ on dit que la contrainte est inactive en x_0*

· *Si x_0 satisfait $g_i(x_0) = 0$ on dit que la contrainte est active ou saturée en x_0.*

On divise l'ensemble des indices en deux sous ensembles

$J_1 = \{j / g_j(x^*) = 0\}$

$J_2 = \{j / g_j(x^*) < 0\}$

Proposition(1) *Si h est une variation admissible autour du point x^* alors nous avons l'inégalité suivante sur les contraintes saturées*

$$\forall j \in J_1, \nabla g_j(x^*) \cdot h \leq 0.$$

De plus, nous avons la proposition suivante :

Proposition(2) *Si x^* est un minimum de la fonction f alors on a :*

$$\nabla f(x^*) \cdot h \geq 0$$

pour toute variation admissible h,

Cette inégalité porte aussi le nom d'inégalité d'Euler.

Définition(2) *Nous dirons que les contraintes sont qualifiées au point x^* si*

∘ *ou bien les fonctions g_j sont affines.*

∘ *ou bien les vecteurs $\nabla g_j(x^*)$ sont linéairement indépendants.*

Nous avons alors la caractérisation d'une variationou direction admissible,

Proposition(3) *Si les contraintes sont qualifiée au point x^*,alors h est une variation ou direction admissible si et seulement si*

$$\nabla g_j(x^*) \cdot h \leq 0, \forall j \in J_1.$$

Dans le cas général si l'ensemble des contraintes actives n'est pas connu. Alors on peut présenter une étude approfondie sur les conditions de **KKT.**

2.5 Théorème de Karush, Kuhn et Tucker

On définit un autre forme des problèmes

Problème à résoudre On considère f et $(\varphi_i)_{i\in I}$ fonctions de classe C^1 de R^nsur R.

Soit P le problème "minimiser f sur $X = \{x \in R^n; \varphi_i(x) \leq 0, \forall i \in I\}$".

2.5.1 Résultats auxiliaires

Théorème(1)*(Théorème de projection sur R^n) Soit C un convexe non vide fermé de R^n.*

Pour tout $b \in R^n$, il existe un et un seul élément $p(b)$ de R^n tel que $p(b) \in C$ et $\|b - p(b)\| = \inf_{c\in C} \|b - c\|$. Cet élément vérifie $\langle p(b) - b, p(b) - c\rangle \leq 0$ pour tout $c \in C$ et réciproquement, si $\langle u - b, u - c\rangle \leq 0$ pour tout $c \in C$, alors $u = p(b)$.

Preuve (chapitre1).

Théorème(2) *(lemme de Farkas-Minkowski) Soit I un ensemble fini,$(a_i)_{i\in I}$ et b des éléments de R^n. Alors*

$$\{w \in R^n; \langle a_i, w\rangle \geq 0 \text{ pour tout } i \in I\} \subset \{w \in R^n; \langle b, w\rangle \geq 0\}$$

si et seulement si il existe une famille $(\lambda_i)_{i\in I}$ de réels positifs, tels que $b = \sum_{i\in I} \lambda_i a_i$.

Preuve

· Si $b = \sum_{i\in I} \lambda_i a_i$ et si $\langle b, w\rangle \geq 0$ pour tout $i \in I$ alors $\lambda_i\langle a_i, w\rangle \geq 0$ pour tout $i \in I$ car $\lambda_i \geq 0$ *et* $\sum_{i\in I} \lambda_i\langle a_i, w\rangle = \langle\sum_{i\in I} \lambda_i a_i, w\rangle = \langle b, w\rangle \geq 0$.

· La réciproque est beaucoup moins évidente et se fait en plusieurs étapes.

Soit $C = \{\sum_{i\in I} \lambda_i a_i; \lambda_i \geq 0 \text{ pour tout } i \in I\}$.

$i)$ Montrons que C est un cône convexe fermé de sommet 0.

$\rightarrow$ Si $x \in C, x = \sum_{i\in I} \lambda_i a_i$ et $\lambda x = \sum_{i\in I}(\lambda\lambda_i)a_i$ avec $\lambda\lambda_i \geq 0$ si $\lambda \geq 0$, donc si $x \in C$ et si $\lambda \geq 0$,alors $\lambda x \in C$: C est bien un cône de sommet 0.

Si $(x, y) \in C^2$ et si $\theta \in [0, 1], x = \sum_{i\in I} \lambda_i a_i, y = \sum_{i\in I} \mu_i a_i; \lambda_i, \mu_i \geq 0$ et

$\theta x+(1-\theta)y = \sum_{i\in I}(\theta\lambda_i+(1-\theta)\mu_i)a_i$ avec $\theta\lambda_i+(1-\theta)\mu_i \geq 0$ donc $\theta x+(1-\theta)y \in C$ et C est convexe.

Si les a_i sont linéairement indépendants, C est isomorphe à $(R_+)^{cardI}$et c'est donc un fermé.

$\rightarrow$ C fermé est plus difficile lorsque les a_i nesont pas linéairement indépendants. Il existe alors des réels μ_i tels que $\sum_{i\in I} \mu_i a_i = 0$ avec $J = \{i, \mu_i < 0\} \neq \emptyset$. Dans ces conditions, tout vecteur $\sum_{i\in I} \lambda i a_i$ s'ecrit aussi $v = \sum_{i\in I}(\lambda_i+t\mu_i)a_i$ où t est quelconque.

Posonst $t_0 = \min\limits_{j\in J}(-\frac{\lambda_j}{\mu_j})$. Alors l'un au moins des $\lambda_i + t_0\mu_i$ est nul et, pour tout $i \in I, \lambda_i + t_0\mu_i \geq 0$.

Donc $C = \bigcup\limits_{j\in J}\{\sum_{i\in I\setminus\{j\}} \lambda_i a_i, \lambda_i \geq 0 \text{ pour tout } i \in I\setminus\{j\}\}$.

On recommence ce raisonnement jusqu'à ce que C soit décrit comme une réunion finie de cônes associés à des vecteurs linéairement indépendants et donc C est bien fermé.

$ii)$ On va appliquer le *Th1* à C : on va montrer que si $b \notin C$, alors l'inclusion est fausse, c'est-à-dire qu'il existe $u \in R^n$ tel que $\langle a_i, u_i\rangle \geq 0$ pour tout $i \in I$ et $\langle b, u\rangle < 0$.

D'aprés le $Th1$, on a, pour tout $v \in C$, $\langle p(b) - b, p(b) - v\rangle \leq 0$.

Par suite, $\langle v, p(b) - b\rangle \geq \langle p(b) - b, p(b)\rangle = \|p(b) - b\|^2 + \langle p(b) - b, b\rangle$.

On a $\|p(b) - b\|^2 > 0$ car $p(b) \neq b (b \notin C)$.

Soit $\alpha \in]\langle p(b) - b, b\rangle, \langle p(b) - p, p(b)\rangle[$et $u = p(b) - b$. On a donc $\langle v, u\rangle > \alpha$ pour tout $v \in C$ et $\langle b, u\rangle < \alpha$.

En particulier, comme $0 \in C, \alpha < 0$ et $\langle b, u\rangle < 0$. D'autre part, pour tout $i \in C$ et pour tout $\lambda > 0, \lambda\alpha_i \in C$ $\langle \alpha_i, u\rangle > \frac{\alpha}{\lambda}$ donc et en faisant $\lambda \to +\infty$, on obtient $\langle \alpha_i, u\rangle \geq 0$. D'où le résultat.

2.5.2 Les relations de Kuhn-Tucker

Définition(1) *Soit $X = \{x \in R^n; \varphi_i(x) \leq 0$ pour tout $i \in I\}$. Un arc de courbe Γ défini par $\gamma : R_+ \to R^n$de classe C^1est dit admissible en $\overline{x} \in X$ si $\gamma(0) = \overline{x}$ et si $\gamma(\theta) \in X$ pour θ assez petit. On appelle alors **direction admissible** en $\overline{x}$ le vecteur $v = \gamma\prime(0)$.On note $C(\overline{x})$ la fermeture du cône formé par l'ensemble des directions admissibles en $\overline{x}$.*

Théorème(3) *Si f admet en $\overline{x}$ un minimum local sur X, alors $\langle \nabla f(\overline{x}), w\rangle \geq 0$ pour tout $w \in C(\overline{x})$.*

Preuve Si w est direction admissible en $\overline{x}$, il existe $\gamma : R_+ \to R^n$ de classe C^1 telle que $\gamma(\theta) = \overline{x} + \theta w + \theta\varepsilon(\theta)$ et $f(\gamma(\theta)) = f(\gamma(0)) + (f \circ \gamma)\prime(0)\theta + \theta\varepsilon(\theta)$ avec $f(\gamma(\theta)) = f(\gamma_1(\theta),\cdots,\gamma_n(\theta))$ d'oû $(f \circ \gamma)\prime(\theta) = \sum_{i=1}^{n} D_i f(\gamma(\theta))\gamma_i'(\theta)$ donc, pour $\theta = 0, (f \circ \gamma)\prime(0) = \langle \nabla f(\gamma(0)), \gamma\prime(0)\rangle = \langle \nabla f(\overline{x}), w\rangle$.

Par suite, $f(\gamma(\theta)) = f(\overline{x}) + \theta(\langle \nabla f(\overline{x}), w\rangle + \varepsilon(\theta))$.

Comme $\gamma(\theta) \in X$ et que $\overline{x}$ est minimum local sur X, on a $f(\gamma(\theta)) - f(\overline{x}) \geq 0$ et donc $\langle \nabla f(\overline{x}), w\rangle + \varepsilon(\theta) \geq 0$ pour tout θ assez petit. En faisant $\theta \to 0$, on obtient $\langle \nabla f(\overline{x}), w\rangle \geq 0$.

Si w n'est pas direction admissible, $w = \lim_n w_n$ avec w_n direction admissible, donc $\langle \nabla f(x), w\rangle = \lim_n \langle \nabla f(x), w_n\rangle \geq 0$.

Définition(2) *Si $x \in X$, on note $I(x) = \{i \in I; \varphi_i(x) = 0\}$ et on appelle $I(x)$ l'ensemble des contraintes saturées en x.*

On pose $C^*(x) = \{w \in R^n; \langle \nabla\varphi_i(x), w\rangle \leq 0$ pour tout $i \in I(x)\}$.

Théorème(4) *On a $C(x) \subset C^*(x)$ mais la réciproque est fausse en général.*

Preuve En remplaçant f par φ_i dans le *Th3*, on a :

$$\varphi_i(\gamma(\theta)) = \varphi_i(x) + \theta(\langle \nabla\varphi_i(x), w\rangle + \varepsilon(\theta)).$$

Si $i \in I(x)$, alors $\varphi_i(x) = 0$ et $\varphi_i(\gamma(\theta)) \leq 0$ car $\gamma(\theta) \in X$.

En faisant $\theta \to 0$, on obtient $\langle \nabla\varphi_i(x), w\rangle \leq 0$.

(Ceci pour w direction admissible, sinon, $w = \lim_n w_n$ et on procède de même que précédemment.)

Donc, si $w \in C(x)$, alors $w \in C^*(x)$.La réciproque n'est pas toujours vraie.

Exemple(1)(sur R^2) :$\varphi_1(x) = -x_1, \varphi_2(x) = -x_2, \varphi_3(x) = -(1-x_1)^3 + x^2$

$$X = \{x \in R^2; x_1 \geq 0, x_2 \geq 0, x_2 \leq (1-x_1)^3\}$$

Pour $x = (1,0), I(x) = \{2,3\}, \nabla\varphi_2(x) : \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \nabla\varphi_2(x) : \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

On a donc $C^*(x) = \{w = (w_1, 0); w_1 \in R\}$: $(1,0) \in C^*(x)$ et $(1,0) \notin C(x)$ car sinon on aurait $(1+\theta, 0) \in X$, ce qui est faux.

Définition(3) *On dit que les contraintes sont qualifiées en x si $C(x) = C^*(x)$.*

La condition précédente étant trés peu aisée à vérifier, on donnera ici des conditions suffisantes (mais non nécessaires) de qualification des contraintes *(QC)* :

Lemme admis *Pour que (QC) soit vérifiée en tout point x de X, il suffit que les φ_i soient linéaires, ou convexes si X est d'intèrieur non vide.*

Pour que (QC) soit vérifiée en $\overline{x}$, il suffit que la famille $(\nabla\varphi_i(\overline{x}))_{i \in I(\overline{x})}$ soit libre.

Remarque(1) *Les conditions du lemme précédent peuvent être combinées. Ainsi, par exemple,*

1) *(QC) est vérifiée en tout $x \in X$ si les φ_i sont linéaires, soit non linéaires convexes et s'*il*existe $\overline{x} \in X$ tel que $\varphi_i(\overline{x}) < 0$ pour tout φ_i non linéaire.*

2)*(QC) est vérifiée en $\overline{x}$ si les φ_i sont, soit linéaires, soit non linéaires avec*

$$(\nabla\varphi_i(\overline{x}))_{i \in I(\overline{x}) \setminus \{j; \varphi_j \text{ linéaire}\}} \textit{libre.}$$

Thérème(5) *(Kuhn-Tucker) Si f admet un minimum en*

$\overline{x} \in X = \{x; \varphi_i(x) \leq 0$ *pour tout* $i \in I\}$, *et si les contraintes sont qualifiées en* $\overline{x}$, *alors il existe une famille* $(\lambda_i(\overline{x}))_{i \in I(\overline{x})}$ *de réels positifs ou nuls tels que*

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

Preuve D'aprés le *Th3*,$\langle \nabla f(\overline{x}), w \rangle \geq 0$ pour tout $w \in C(\overline{x})$. Or

$C(\overline{x}) = C^*(\overline{x}) = \{w; \langle \nabla \varphi_i(\overline{x}), w \rangle \leq 0, \forall i \in I(\overline{x})\} = \{w; \langle -\nabla \varphi_i(\overline{x}), w \rangle \geq 0, \forall i \in I(\overline{x})\}$

On a donc $\{w; \langle -\nabla \varphi_i(\overline{x}), w \rangle \geq 0$ pour tout $i \in I(\overline{x})\} \subset \{w; \langle \nabla f(\overline{x}), w \rangle \geq 0\}$ donc, d'aprés le *Th2*, il existe une famille $(\lambda_i(\overline{x}))_{i \in I(\overline{x})}$ de réels positifs ou nuls tels que $\nabla f(x) = -\sum_{i \in I(\overline{x})} \lambda_i(\overline{x}) \nabla \varphi_i(\overline{x})$.

Remarque(2)

1) En convenant que $\lambda_i(\overline{x}) = 0$ *si* $i \notin I(\overline{x})$, *le vecteur* $\lambda(\overline{x})$ *est appelé vecteur de Kuhn-Tucker en* $\overline{x}$.

2) Si $(\nabla \varphi_i(\overline{x}))_{i \in I(\overline{x})}$ *est libre, alors* $\lambda(\overline{x})$ *est unique.*

2.5.3 Extension à des problèmes avec des contraintes d'égalités et d'inégalités

Problème Minimiser f sous les contraintes :

$\rightarrow \varphi_i(x) \leq 0$ pour tout $i \in I$;

$\rightarrow \psi_j(x) = 0$ pour tout $j \in J$;

les fonctions f, φ_i, ψ_j étant de classe C^1.

On note $X = \{x \in R^n; \varphi_i(x) \leq 0$ pour tout $i \in I$ et $\psi_j(x) = 0$ pour tout $j \in J\}$ et on pose $I(\overline{x}) = \{i \in I; \varphi_i(\overline{x}) = 0\}$ et

$C^*(x) = \{w \in R^n; \langle \nabla \varphi_i(\overline{x}), w \rangle \leq 0, \forall i \in I$ et $\langle \nabla \psi j(\overline{x}), w \rangle = 0, \forall j \in J\}$.

On définit de même les directions admissibles et $C(\overline{x})$ et on a l'analogue du *théorème de Kuhn-Tucker* :

Thérème(6) *Si f admet sur X un minimum en $\overline{x}$ en le uel les contraintes sont* *qualifiées, alors il existe des nombres positifs ou nuls* $\lambda_i(\overline{x})$ *pour* $i \in I$ *et des nombres* $\mu_j(\overline{x})$ *(non nécessairement positifs) pour* $j \in J$ *tels que :*

$\nabla f(\overline{x}) + \sum_{i \in I} \lambda_i(\overline{x}) \nabla \varphi_i(\overline{x}) + \sum_{j \in J} \mu_j(\overline{x}) \nabla \psi_j(\overline{x}) = 0$ *et* $\lambda_i(\overline{x})\varphi_i(\overline{x}) = 0$ *pour tout* $i \in I$.

Autre problème Minimiser f sous les contraintes $\psi_j(x) = 0$ pour tout $j \in J$; les fonctions f, ψ_j étant de classe C^1(contraintes égalités seulement).

On note $X = \{x \in R^n; \psi_j(x) = 0 \text{ pour tout } j \in J\}$.

Théorème(7) *Si* f *admet sur* X *un minimum en* $\overline{x}$ *en lequel les contraintes sont qualifiées, alors il existe une famille de nombres* $(\mu_j(\overline{x}))_{j \in J}$*(non nécessairement positifs) telle que :*

$$\nabla f(\overline{x}) + \sum_{j \in J} \mu_j(\overline{x}) \nabla \psi_j(\overline{x}) = 0.$$

Remarque(3) *on retrouve les multiplicateurs de Lagrange et le théoréme des extrema liés.*

2.5.4 Cas des fonctions convexes

Définition(4)*(fonction concave)* f *est concave sur* X *si et seulement si,* $\forall (x, y) \in X^2$ *et* $\forall \lambda \in [0, 1]$, *on a :*

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

f *est strictement concave sur* X *si et seulement si :*

$$f((1-\lambda)x + \lambda y) > (1-\lambda)f(x) + \lambda f(y).$$

Définition(5)*(fonction convexe)* f *est convexe sur* X *si et seulement si,* $\forall (x, y) \in X^2$ *et* $\forall \lambda \in [0, 1]$, *on a :*

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

f *est strictement convexe sur* X *si et seulement si :*

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y).$$

Propriétés(1)*(importantes)*

$\circ$ f *concave* $\iff$ $-f$ *convexe*

$\circ$ *Si f et g sont des fonctions concaves (resp.convexes), alors* $\forall(a,b) \in R_+^2, (a.f + b.g)$ *est une fonction concave (resp.convexe).*

$\circ$ *Si f est une fonction concave et g est une fonction croissante et concave, alors la fonction $g(f(x))$ est concave.*

$\circ$ *Si f est une fonction convexe et g est une fonction croissante et convexe, alors la fonction $g(f(x))$ est convexe.*

$\circ$ *Une fonction affine est à la fois concave et convexe.*

Théorème(8) *Si X est convexe, si f et les φ_i sont convexes et s'il existe des nombres λ_i tels que les relations de Kuhn-Tucker soient satisfaites en $\overline{x}$, alors f admet sur X un minimum en $\overline{x}$.*

Preuve Pour tout $x \in X, \varphi_i(x) \leq 0$ et $\lambda_i \geq 0$ pour $i \in I(\overline{x})$ donc $\lambda_i \varphi_i(x) \leq 0$ et

$$f(\overline{x}) \leq f(\overline{x}) - \sum_{i \in I(\overline{x})} \lambda_i \varphi_i(x) = f(\overline{x}) - \sum_{i \in I(\overline{x})} \lambda_i(\varphi_i(x) - \varphi_i(\overline{x}))$$

car $\varphi_i(\overline{x}) = 0$ pour $i \in I(\overline{x})$. Comme les φ_i sont convexes, $\varphi_i(x) - \varphi_i(\overline{x}) \geq \langle \nabla \varphi_i(\overline{x}), x - \overline{x} \rangle$ donc $f(\overline{x}) \leq f(\overline{x}) - \langle \sum_{i \in I(\overline{x})} \lambda_i \nabla \varphi_i(\overline{x}), x - \overline{x} \rangle = f(\overline{x}) + \langle \nabla f(\overline{x}), x - \overline{x} \rangle$

d'aprés le *Th5*(*Kuhn-Tucker*). Mais f étant convexe, $\langle \nabla f(\overline{x}), x - \overline{x} \rangle \leq f(x) - f(\overline{x})$. D'oû $f(\overline{x}) \leq f(x)$ si $x \in X$ et $\overline{x}$ est bien minimum de f sur$X$.

Quatrième partie

Chapitre 3

Chapitre 3

Applications(des conditions de KKT)

3.1 Exemple1

$\circ$ Extrémums de $f : (x, y) \mapsto ax+by+c$ sous la contrainte $\varphi(x, y) = x^2+y^2-1 = 0$.

$$\nabla f(x,y) : \begin{pmatrix} a \\ b \end{pmatrix} \qquad ; \qquad \nabla\varphi(x,y) : \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

Condition nécessaire :

$$\nabla f(\overline{x}, \overline{y}) + \mu\nabla\varphi(\overline{x}, \overline{y}) = 0 \text{ soit } \begin{cases} a + 2\mu\overline{x} = 0 \\ b + 2\mu\overline{y} = 0 \end{cases},$$

On trouve $\overline{x} = -\frac{a}{2\mu}$ *et* $\overline{y} = -\frac{b}{2\mu}$.

On a aussi $\varphi(x, y) = 0$ donc $a^2 + b^2 = 4\mu^2$ *et* $\mu = \pm\frac{\sqrt{a^2+b^2}}{2}$.

$\cdot$ Si $\mu = \frac{\sqrt{a^2+b^2}}{2}$,alors $\overline{x} = -\frac{a}{\sqrt{a^2+b^2}}, \overline{y} = -\frac{b}{\sqrt{a^2+b^2}}$ et $f(\overline{x}, \overline{y}) = c - \sqrt{a^2 + b^2}$.

$\cdot$ Si $\mu = -\frac{\sqrt{a^2+b^2}}{2}$,alors $\overline{x} = \frac{a}{\sqrt{a^2+b^2}}, \overline{y} = \frac{b}{\sqrt{a^2+b^2}}$ et $f(\overline{x}, \overline{y}) = c + \sqrt{a^2 + b^2}$.

φ étant convexe, la contrainte est qualifiée. De plus, X est fermé borné (cercle unité) et f est continue donc les extrémums existent.

On a donc $\min\limits_X f = c - \sqrt{a^2 + b^2}$, $\max\limits_X f = c + \sqrt{a^2 + b^2}$, avec $\mathrm{Arg}_X \min f = \left\{\left(-\frac{a}{\sqrt{a^2+b^2}}, -\frac{b}{\sqrt{a^2+b^2}}\right)\right\}$ et $\mathrm{Arg}_X \max f = \left\{\left(\frac{a}{\sqrt{a^2+b^2}}, \frac{b}{\sqrt{a^2+b^2}}\right)\right\}$.

3.2 Exemple2

∘ Extrémums de $f : (x, y, z) \mapsto x + y + z$ sous les contraintes $\varphi_1(x, y, z) = x^2 + z^2 - 1 \leq 0$ et $\varphi_2(x, y, z) = y^2 + z^2 - 1 \leq 0$.

fest continue, $X = \varphi_1^{-1}(]-\infty, 0]) \cap \varphi_2^{-1}(]-\infty, 0])$ est fermé comme intersection de 2 fermés (images réciproques de $]-\infty, 0]$ fermé par les φ_i qui sont continues), et X est borné, car si $(x, y, z) \in X, x^2 + y^2 + z^2 \leq 2$.

On a donc existence des extrémums de f sur X.

$$\nabla f(x, y, z) : \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} ; \nabla \varphi_1(x, y, z) : \begin{pmatrix} 2x \\ 0 \\ 2z \end{pmatrix} ; \nabla \varphi_2(x, y, z) : \begin{pmatrix} 0 \\ 2y \\ 2x \end{pmatrix}.$$

$$\text{Conditions de Kuhn-Tucker :} \begin{cases} 1 + 2\lambda_1 \overline{x} = 0 \\ 1 + 2\lambda_1 \overline{y} = 0 \\ 1 + 2(\lambda_1 + \lambda_2)\overline{z} = 0 \\ \lambda_1(\overline{x}^2 + \overline{z}^2 - 1) = 0 \\ \lambda_2(\overline{y}^2 + \overline{z}^2 - 1) = 0 \end{cases}.$$

On trouve $\overline{x} = -\frac{1}{2\lambda_1}$ et $\overline{y} = -\frac{1}{2\lambda_2}$.

On a nécessairement $\lambda_1 \neq 0$ et $\lambda_2 \neq 0$ donc$\varphi_1(x, y, z) = 0$ et $\varphi_2(x, y, z) = 0$ soit

$$\begin{cases} x^2 + z^2 - 1 = 0 \\ y^2 + z^2 - 1 = 0 \end{cases}.$$

En faisant la différence, on obtient $\overline{x}^2 = \overline{y}^2$ et, comme $\overline{x}$ et $\overline{y}$ sont de même signe, on a alors $\overline{x} = \overline{y}$, donc $\lambda_1 = \lambda_2 = \lambda$, puis $\overline{x} = \overline{y} = -\frac{1}{2\lambda}$ et $\overline{z} = -\frac{1}{4\lambda} = \overline{x}/2$. Enfin $\overline{x}^2(1 + 1/4) - 1 = 0$ d'oú $\overline{x} = \pm\frac{2}{\sqrt{5}}$.

On a danc $\min_X f = -\sqrt{5}$; $\max_X f = \sqrt{5}$ avec $\mathrm{Arg}_X \min f = \left\{\left(-\frac{2}{\sqrt{5}}, -\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right)\right\}$ et $\mathrm{Arg}_X \max f = \left\{\left(\frac{2}{\sqrt{5}}, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)\right\}$.

Cinquième partie

Conclusion

Le théorème de K.K.T donne une condition nécessaire pour qu'un point $\overline{x}$ soit une solution optimale. Dans le cas où f est convexe, cette condition devient suffisante.

Notons qu'on n'a pas un algorithme de recherche de la méthode de manière générale.

Sixième partie

Bibliographie

Bibliographie

[1]Algèbre et Géometrie PC-PSI-PT Jean-Marie Monier, Dunod.

[2]Grapheset Recherche Opérationnelle, Notes de cours de J.-F. Scheid,2010-2011.

[3] Introduction à l'analyse numérique matricielle et à l'optimisation, PHILIPPE G.CIARLET, Dunod.

[4] Le théorème de représentation de Riesz Adjoint d'un endomorphisme en dimension infinie, Julien Baglio, Algèbre linéaire.

[5] MATHÉMATIQUES Algèbre et géométrie GUY AULIAC-GUY AULIAC - JEAN DELCOURT - RÉMY GOBLOT, Dunod.

[6] Notes de cours Optimisation, Licence MIA, L3, UFR DE MATHEMATIQUES ET INFORMATIQUE,Paris Descartes.

[7] Optimisation avec contrainte, Chapitre 5.

[8] Recherche Opérationnelle, Paul Feautrier, 20 mai 2004.

[9] Techniques d'optimisation classiques Y. Ouinten, Université de Laghouat, 2007-2010.

[10] Vade mecum : Optimisation statique.Lagrangien et conditions de Kuhn et Tucker, d'Économie – Julien Grenet, École Normale Supérieure, Année 2007.

[11] WALER HENGARTNER, MARCEL LAMBERT et CORINA REISCHER, Introduction à l'analyse fonctionnelle, Les Presses de l'Université du Québec (1981).