



## THESIS / THÈSE

### MASTER EN SCIENCES MATHÉMATIQUES

#### Sur la convergence d'algorithmes basés sur des méthodes de points intérieurs en programmation convexe

LELEUX, Frédéric

*Award date:*  
1993

*Awarding institution:*  
Universite de Namur

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX**

**NAMUR**

**FACULTE DES SCIENCES**

**Sur la convergence d'algorithmes basés sur des  
méthodes de points intérieurs en  
programmation convexe.**

**Promoteur : Monsieur le Professeur J-J STRODIOT**

**Frédéric Leleux**

Année académique 1992-1993.

Avant d'aller plus avant dans mon travail, je tiens à exprimer mes plus vifs remerciements à Monsieur STRODIOT pour avoir accepté d'être le promoteur de ce mémoire.

Je voudrais aussi exprimer toute ma gratitude à mes parents pour le soutien moral et logistique qu'ils m'ont apporté tout au long de mes études.

Enfin, je souhaite remercier de tout coeur les personnes qui, de près ou de loin, ont contribué à la réalisation de ce travail.

## TABLE DES MATIERES.

INTRODUCTION	4
NOTATIONS	6
PREMIERE PARTIE : Une méthode primale-duale appliquée à un programme convexe contraint linéairement.	7
1. Le problème et son dual.	8
2. La trajectoire centrale.	10
3. Algorithme conceptuel.	13
4. Implémentation numérique.	14
A. Voisinage de la trajectoire centrale.	14
B. Résolution approximative du système d'équations de K-T.	15
5. Méthode de Kortanek-Potra-Ye.	18
A. L'algorithme.	18
B. Convergence de l'algorithme.	18
6. Méthode de Zhu.	28
A. Algorithme de Zhu.	28
B. Classes de fonctions satisfaisant la condition SLC.	33
7. Complexité des méthodes dans le cas linéaire.	37
DEUXIEME PARTIE : Généralisation aux contraintes d'inégalités convexes.	39
Chapitre 1 : Algorithme à petits pas.	41
1. Présentation du problème.	41
A. Le problème.	41
B. La condition de self-concordance.	41
2. La méthode utilisée.	42
A. Hypothèses.	42
B. La méthode.	42
C. La méthode de Newton.	43
3. L'algorithme à pas courts.	44
4. Analyse de convergence.	45
Chapitre 2 : Algorithmes à pas longs.	53
1. Algorithme à pas longs se basant sur la fonction potentielle.	53
A. Le problème.	53
B. Description de l'algorithme.	54
C. Analyse de convergence.	56

2. Algorithme à pas longs se basant sur la fonction barrière logarithmique.	70
A. Généralités.	70
B. L'algorithme.	73
C. Analyse de convergence.	74
CONCLUSIONS.	79
ANNEXES.	80
BIBLIOGRAPHIE.	87

## INTRODUCTION.

Dans ce mémoire, nous généraliserons aux problèmes de programmation convexe les méthodes de points intérieurs développées pour résoudre les problèmes de programmation linéaire.

Dans une première partie, nous étudierons les problèmes de programmation convexe sous contraintes linéaires, c'est-à-dire les programmes

$$(LCCP) \quad \begin{cases} \min f(x) \\ \text{s.c. } Ax = b \\ x \geq 0 \end{cases}$$

où  $f(x)$  est une fonction convexe,  $A$  une matrice de dimension  $(m, n)$  avec  $n \geq m$  de rang plein et  $b \in R^m$ .

KORTANEK, POTRA et YE <sup>1</sup> ont développé un algorithme qui fournit une solution  $\varepsilon$ -optimale du problème (LCCP). Mais leur méthode primale-duale, basée sur la trajectoire centrale, présente un gros désavantage : à chaque itération, l'algorithme requiert la solution d'un système d'équations non-linéaires.

ZHU a remédié à cet état de fait en définissant la condition (SLC)<sup>2</sup>. Cette condition, imposée à la fonction objectif  $f(x)$ , permet, en effet, de présenter un algorithme où le système d'équations à résoudre est linéaire.

Nous clôturerons cette partie en étudiant la complexité de l'algorithme de ZHU dans le cadre de la programmation linéaire.

La deuxième partie aura pour objet l'étude des programmes convexes c'est-à-dire les programmes

$$\begin{cases} \min f(x) \\ \text{s.c. } g_i(x) \leq 0 \quad i = 1, \dots, m \end{cases}$$

---

<sup>1</sup> K.O. KORTANEK, F. POTRA, Y. YE, *On some efficient interior-point methods for nonlinear convex programming*. *Linear Algebra and Its Applications*, 152 (1991) pp169-189.

<sup>2</sup> J. ZHU, *A path-following algorithm for a class of convex programming problems*. *ZOR-Methods and Models of Operations Research*, 36 (1992) pp359-377.

où les fonctions  $f(x)$  et  $g_i(x)$ ,  $1 \leq i \leq m$ , sont supposées deux fois continûment différentiables .

Dans un premier chapitre, nous présenterons une méthode, développée par JARRE, basée sur la minimisation de la fonction potentielle <sup>1</sup>. Un algorithme à petits pas y est proposé, pour lequel la condition imposée à la fonction potentielle est la condition de self-concordance forte .

Nous consacrerons le second chapitre à l'étude d'algorithmes à grands pas, plus intéressants d'un point de vue pratique . Nous basant sur les travaux de DEN HERTOOG, ROOS et TERLAKY, nous fournirons la vitesse de convergence d'algorithmes basés sur la minimisation de la fonction potentielle <sup>2</sup> et de la fonction barrière logarithmique <sup>3</sup> .

---

<sup>1</sup> F. JARRE , *Interior-point methods for convex programming*. *Applied Mathematics and Optimization* (1992), pp 287-311.

<sup>2</sup> D. DEN HERTOOG, C. ROOS et T. TERLAKY, *A large-step analytic center method for a class of smooth convex programming problems*, *SIAM Journal on Optimization*, 2 (1992) pp 55-70 .

<sup>3</sup> D. DEN HERTOOG, C. ROOS et T. TERLAKY, *On the classical logarithmic barrier function method for a class of smooth convex programming problems*, *JOTA*, 73 (1992), pp 1-25

## NOTATIONS.

Nous dressons ci-dessous la liste des principaux symboles utilisés dans le texte :

.  $R$  : l'espace des réels.

.  $R^n$  : l'espace des vecteurs à  $n$  composantes dans  $R$ .

. Soit  $x \in R^n$

$$\text{Alors } X = \text{diag } x_i = \begin{pmatrix} x_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & x_n \end{pmatrix}$$

.  $e = (1, 1, \dots, 1)^t$

. Soit  $f: R^n \rightarrow R$ .

Alors  $\nabla f(x)$  : gradient de  $f$  au point  $x$ .

$\nabla^2 f(x)$  : matrice hessienne de  $f$  prise au point  $x$ .

.  $\varphi(x, \lambda)$  : fonction potentielle.

.  $\varphi_\mu(x)$  : fonction logarithmique.

.  $\lambda$  : paramètre de la fonction potentielle, borne supérieure du coût.

.  $\mu$  : paramètre de la fonction logarithmique, saut de dualité moyen.

$$\nabla^3 \varphi(x)[h, h, h] = \sum_{i,j,k=1}^n \frac{\partial^3 \varphi(x)}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k.$$

.  $\|x\|$  : norme euclidienne du vecteur  $x$ .

.  $\|x\|_H$  : H-norme du vecteur  $x$ .

. " $\triangleq$ " signifie " par définition ".

.  $\text{Ker}(A) = \{x \in R^n \text{ tel que } Ax = 0\} \triangleq$  noyau de  $A$ .

.  $R(A) = \{y \in R^m \text{ tel qu'il existe } x \in R^n \text{ pour lequel } Ax = y\}$ .

.  $\oplus$  : somme directe.

.  $\langle \cdot, \cdot \rangle$  désigne le produit scalaire usuel, c'est-à-dire  $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ .



## **PREMIERE PARTIE :**

**Une méthode primale - duale appliquée à un programme convexe contraint linéairement.**

## 1. Le problème et son dual.

Dans ce chapitre, nous considérons le problème primal suivant :

$$(P) \begin{cases} \text{minimiser } f(x) \\ \text{s.c. } Ax = b \\ x \geq 0 \end{cases}$$

où  $f : R^n \rightarrow R$  est une fonction convexe et différentiable

$A \in R^{m \times n}$  est une matrice de rang  $m$

$b \in R^m$ .

Le dual de Wolfe associé au problème primal (P) est :

$$(D) \begin{cases} \max f(x) - \nabla f(x)'x + b'y \\ \text{s.c. } s \equiv \nabla f(x) - A'y \geq 0 \end{cases}$$

où  $s \in R^n$  est le vecteur des variables d'écart  
et  $y \in R^m$ .

Remarque : Cette formulation du dual est équivalente en programmation convexe à la formulation lagrangienne. En effet,

$$\text{soit le problème (P)} \begin{cases} \min f(x) \\ \text{s.c. } g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

$$\text{Le dual lagrangien s'écrit : (DL)} \begin{cases} \max \min_x L(x, y, z) \\ \text{s.c. } y \geq 0 \end{cases}$$

où  $L(x, y, z) = f(x) + y'g(x) + z'h(x)$ .

$$\text{Ce programme (DL) est équivalent à } \begin{cases} \max L(x, y, z) \\ \text{s.c. } y \geq 0 \\ x \text{ solution de } \min_x L(x, y, z) \end{cases}$$

Puisque les fonctions  $f$  et  $g$  sont convexes et puisque  $h$  est affine, ce problème

$$\text{est équivalent au problème dual de Wolfe } \begin{cases} \max L(x, y, z) \\ \text{s.c. } y \geq 0 \\ \nabla_x L(x, y, z) = 0 \end{cases}$$

Dans le cas qui nous intéresse ici, puisque  $L(x, y, z) = f(x) + y'(b - Ax) - s'x$ , nous obtenons comme formulation du dual de Wolfe :

$$\begin{cases} \max f(x) + y'(b - Ax) - s'x \\ \text{s.c. } s \geq 0 \\ \nabla f(x) - A'y - s = 0 \end{cases}$$

$$\text{ie } \begin{cases} \max f(x) + y'b - x'(A'y + s) \\ \text{s.c. } s \geq 0 \\ s = \nabla f(x) - A'y \end{cases}$$

$$\text{ie } \begin{cases} \max f(x) - \nabla f(x)'x + y'b \\ \text{s.c. } s = \nabla f(x) - A'y \geq 0 \end{cases}$$

Notons  $\mathcal{P} = \{x \in R^n \text{ tel que } Ax = b, x \geq 0\}$  l'ensemble primal admissible et

$\mathcal{D} = \{(y, s) \in R^m \times R^n \text{ tel que } A'y + s = \nabla f(x), s \geq 0\}$  l'ensemble dual admissible.

Par le théorème de dualité faible, nous savons que le saut de dualité est :

$$\nabla f(x)'x - b'y = x's \geq 0$$

De plus, si  $\bar{x} \in \mathcal{P}$  et  $(\bar{y}, \bar{s}) \in \mathcal{D}$  sont tels que  $\bar{x}'\bar{s} = 0$ , alors  $\bar{x}$  est la solution optimale de (P) et  $(\bar{y}, \bar{s})$  est la solution optimale de (D).

Par conséquent, dans une méthode primale-duale, une condition suffisante pour atteindre l'optimalité est de considérer la solution du système :

$$\begin{cases} Ax = b \\ A'y + s = \nabla f(x) \\ x's = 0 \\ x \geq 0, s \geq 0 \end{cases}$$

Nous allons pour cela rechercher une suite  $(x^k, s^k)$  de points admissibles telle que  $(x^k)'s^k$  tende vers 0 lorsque k tend vers  $\infty$ .

Pour simplifier les notations, nous écrirons désormais, puisque la matrice  $A$  est de rang m et donc qu'il existe une bijection entre  $y$  et  $s$ ,  $s \in \mathcal{D}$  pour  $(y, s) \in \mathcal{D}$ .

## 2. La trajectoire centrale.

Soient  $\mathcal{P}^\circ$  et  $\mathcal{D}^\circ$  les intérieurs relatifs de  $\mathcal{P}$  et  $\mathcal{D}$  respectivement,

$$\text{ie } \mathcal{P}^\circ = \{x \in R^n \text{ tel que } Ax = b, x > 0\}$$

$$\mathcal{D}^\circ = \{(y, s) \in R^m \times R^n \text{ tel que } A'y + s = \nabla f(x), s > 0\}.$$

Puisque nous nous intéressons à une méthode de points intérieurs, nous ferons l'hypothèse que  $\mathcal{P}^\circ \neq \emptyset$  et  $\mathcal{D}^\circ \neq \emptyset$ .

Nous pouvons introduire pour le problème primal une fonction barrière, c'est-à-dire une fonction  $\varphi_\mu : \mathcal{P}^\circ \rightarrow R : x \rightsquigarrow \varphi_\mu(x) = f(x) - \mu \sum_{i=1}^n \ln x_i$  avec  $\mu > 0$ . Au problème primal seront alors associés les problèmes :

$$(P\mu) \begin{cases} \text{minimiser } f(x) - \mu \sum_{i=1}^n \ln x_i \\ \text{s.c. } Ax = b \\ x > 0 \end{cases}$$

Le théorème suivant nous fournira une caractérisation des solutions des problèmes  $(P\mu)$ .

**Théorème 1 :** soit  $\mu > 0$ .

$x$  est solution optimale de  $(P\mu)$

si et seulement si

$$\exists y \in R^m, \exists s \in R^n \text{ tels que } \begin{cases} Ax = b \\ A'y + s = \nabla f(x) \\ Xs = \mu e \\ x > 0 \end{cases}$$

Preuve :

Puisque  $f$  est convexe,  $\varphi_\mu$  est strictement convexe sur  $\mathcal{P}^\circ$ . Par conséquent, les conditions de KUHN-TUCKER du problème  $(P\mu)$  sont nécessaires et suffisantes.

Le langrangien s'écrit :  $L(x, y) = f(x) - \mu \sum_{i=1}^n \ln x_i + y'(b - Ax)$

Le système des conditions de K-T s'écrit donc :  
il existe  $y \in R^m$  tel que

$$\begin{cases} \nabla f(x) - \mu X^{-1}e - A'y = 0 \\ Ax = b \\ x > 0 \end{cases}$$

En posant  $s = \mu X^{-1}e$ , ce dernier système s'écrit :

$$\begin{cases} A'y + s = \nabla f(x) \\ Ax = b \\ Xs = \mu e \\ x > 0 \end{cases}$$

□

Remarque : on a également  $s > 0$  puisque  $Xs = \mu e$ .

Définition 1 : Soit  $C = \left\{ (x(\mu), s(\mu)) \in R^n \times R^n \text{ tels que } \begin{array}{l} x(\mu) \text{ solution de (P}\mu) \\ s(\mu) \text{ solution de (D}\mu) \end{array}, \mu > 0 \right\}$ .

On appelle C la trajectoire centrale primale-duale.

Par le théorème précédent, on a :

$$C = \left\{ (x, s) \in R^n \times R^n \text{ tels que } \begin{array}{l} Ax = b \\ A'y + s = \nabla f(x) \\ Xs = \mu e \end{array}, \mu > 0 \right\}.$$

Proposition : L'ensemble des points  $(x(\mu), s(\mu))$  de la trajectoire centrale C où  $0 < \mu \leq \bar{\mu}$  est borné.

Preuve :

soit  $x^0 \in \mathcal{P}^0$  et  $s^0 \in \mathcal{D}^0$ .

Par définition, on a :  $Ax^0 = b$   
et  $Ax(\mu) = b$ .

Donc,  $x^0 - x(\mu) \in \text{Ker } A$ .

De même,  $s^0 \in c - R(A')$   
et  $s(\mu) \in c - R(A')$ .

Par conséquent,  $s^0 - s(\mu) \in R(A')$ .

Puisque A est de rang plein,  $R^n = \text{Ker } A \oplus R(A')$ .

Donc,  $\langle x^0 - x(\mu), s^0 - s(\mu) \rangle = 0$

et  $(x^0)'s(\mu) + (s^0)'x(\mu) = (x^0)'s^0 + x(\mu)'s(\mu)$ .

On obtient alors,  $\forall \mu \in ]0, \bar{\mu}]$  et  $1 \leq i \leq n$  :

$$\begin{aligned}
0 < s_i^0 x_i(\mu) &\leq (s^0)'x(\mu) && \text{puisque } s^0 > 0 \text{ et } x(\mu) > 0. \\
&\leq (s^0)'x(\mu) + (x^0)'s(\mu) \\
&= (x^0)'s^0 + x(\mu)'s(\mu) \\
&\leq (x^0)'s^0 + n\bar{\mu} && \text{par définition de C.}
\end{aligned}$$

Donc,  $\forall i \in [1, n]$ , on a :  $0 < x_i(\mu) \leq \frac{(x^0)'s^0 + n\bar{\mu}}{s_i^0}$

et  $x(\mu)$  est borné.

Le même raisonnement s'applique à  $s(\mu)$ .

□

Le théorème suivant montre l'intérêt de la trajectoire centrale.

**Théorème 2 :** si  $\mu \rightarrow 0$ ,

alors chaque point d'accumulation  $(x^*, s^*)$  de la suite  $(x(\mu), s(\mu))$  est une solution primale-duale.

Preuve :

Comme  $Xs = \mu e$ , on a  $x_i s_i = \mu \quad \forall i$  c'est-à-dire  $\mu = \frac{x' s}{n}$ .

Donc, si  $\mu \rightarrow 0$ ,  $x' s \rightarrow 0$  et, par le théorème de dualité faible,  $(x^*, s^*)$  est une solution primale-duale.

□

Ainsi donc, en variant le paramètre  $\mu$  et en le faisant tendre vers 0, nous allons rechercher les solutions optimales des problèmes (P) et (D) en nous déplaçant le long de la trajectoire centrale.

Cet aspect est l'idée de base de l'algorithme conceptuel suivant.

### 3. Algorithme conceptuel.

Initialisation: soit  $\varepsilon > 0$  le paramètre de précision de l'algorithme

soit  $\mu^0 > 0$  et  $\sigma$  choisi de sorte que  $0 < \sigma < 1$

posons  $k = 0$

Pas 1 : si  $\mu^k > \varepsilon$

alors calculer  $(x^k, s^k) = (x(\mu^k), s(\mu^k))$  en résolvant le système

$$\begin{cases} Ax^k - b = 0 \\ A'y^k + s^k - \nabla f(x^k) = 0 \\ X^k s^k - \mu^k e = 0 \end{cases}$$

Pas 2 :  $\mu^{k+1} = \sigma \mu^k$

$k = k + 1$

revenir au pas 1.

Cet algorithme est théorique. A chaque itération, le système d'équations à résoudre est non linéaire. En outre, les points obtenus sont situés exactement sur la trajectoire centrale. En pratique, on imposera la proximité plutôt que la position exacte le long de la trajectoire centrale.

#### 4. Implémentation numérique.

Pratiquement, on va suivre approximativement la trajectoire centrale. On va, pour cela, définir une notion de voisinage.

##### A. Voisinage de la trajectoire centrale :

La trajectoire centrale est définie comme l'ensemble

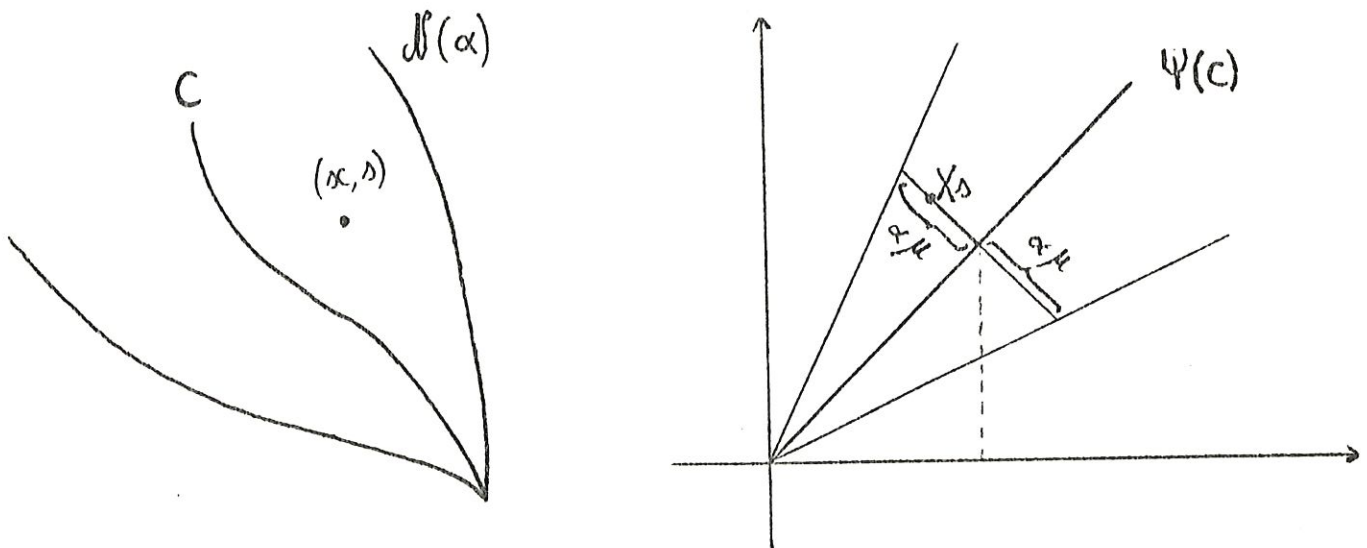
$$C = \left\{ (x, s) \in R^n \times R^n \text{ tels que } \begin{array}{l} (x, s) \text{ soit strictement admissible} \\ Xs = \mu e \end{array}, \mu > 0 \right\}$$

Définition 2 : On appelle  $\alpha$ -voisinage de la trajectoire centrale, l'ensemble défini par :

$$\mathcal{N}(\alpha) = \left\{ (x, s) \in R^n \times R^n \text{ tels que } \begin{array}{l} (x, s) \text{ soit strictement admissible} \\ \|Xs - \mu e\| \leq \alpha \mu \text{ où } \mu = \frac{x's}{n} \end{array} \right\}$$

Pour interpréter géométriquement la notion de voisinage, considérons l'application  $\psi : R^n \times R^n \rightarrow R^n \rightsquigarrow Xs = (x_i s_i)_{i=1, \dots, n}$

Comme tous les points de la trajectoire centrale  $C$  sont tels que  $x_i s_i = \mu$  pour  $i = 1, \dots, n$ , l'image par  $\psi$  de  $\mathcal{N}(\alpha)$  est un cône sphérique de  $R_+^n$ , symétrique par rapport à  $\psi(C)$ . Ceci apparaît clairement grâce au schéma suivant :



Voyons à présent comment on résoud en pratique le système d'équations de K-T.



**B. Résolution approximative du système d'équations de K-T.**

Afin de bien nous fixer les idées, voyons d'abord le cas de la programmation linéaire.

• En programmation linéaire :

$$\text{Soit le programme linéaire (P) } \begin{cases} \min & c'x \\ \text{s.c.} & Ax = b \\ & x \geq 0 \end{cases}$$

La fonction barrière logarithmique associée au problème (P) est la fonction  $\varphi_\mu : R^n \rightarrow R : x \rightsquigarrow \varphi_\mu(x) = c'x - \mu \sum_{i=1}^n \ln x_i$  et le système de K-T à résoudre est :

$$(*) \begin{cases} Ax = b & (i) \\ A'y + s - c = 0 & (ii) \\ Xs - \mu e = 0 & (iii) \end{cases}$$

Nous voulons que les solutions soient strictement admissibles. Autrement dit, nous imposons aux équations (i) et (ii) d'être exactement vérifiées. Par contre, la troisième équation peut être résolue approximativement.

Effectuons dès lors un pas de la méthode Newton sur le système (\*). Nous obtenons :

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A' & I \\ S & 0 & X \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ Xs - \mu e \end{pmatrix}$$

c'est-à-dire :

$$(*') \begin{cases} Ad_x = 0 \\ A'd_y + d_s = 0 \\ Sd_x + Xd_s = \mu e - Xs \end{cases}$$

Après plusieurs opérations matricielles élémentaires, nous obtenons comme solutions :

$$\begin{aligned} d_x &= S^{-1} \left( XA'(AS^{-1}XA')^{-1} AS^{-1} - I \right) (Xs - \mu e) \\ d_y &= (AS^{-1}XA')^{-1} AS^{-1} (Xs - \mu e) \\ d_s &= -A'(AS^{-1}XA')^{-1} AS^{-1} (Xs - \mu e) \end{aligned}$$

• *Généralisation au cas non linéaire.*

$$\text{Le système d'équations K-T à résoudre est : (**)} \begin{cases} Ax = b & (i) \\ A'y + s = \nabla f(x) & (ii) \\ Xs - \mu e = 0 & (iii) \end{cases}$$

L'équation (ii), qui correspond au caractère admissible des solutions du problème dual, n'est plus linéaire. Deux possibilités s'offrent donc à nous pour résoudre le système (\*\*):

- ✓ soit on linéarise le système (\*\*) en effectuant un pas de la méthode Newton
- ✓ soit on impose au nouveau point d'être dual-admissible.

a) Pas de Newton sur le système.

$$\text{le système (**)} \text{ s'écrit : } \begin{cases} Ax - b = 0 \\ -\nabla f(x) + A'y + s = 0 \\ Xs - \mu e = 0 \end{cases}$$

Effectuons un pas de la méthode Newton sur ce système et recherchons  $(d_x, d_y, d_s)$  solution de :

$$\begin{pmatrix} A & 0 & 0 \\ -\nabla^2 f(x) & A' & I \\ S & 0 & X \end{pmatrix} \begin{pmatrix} d_x \\ d_y \\ d_s \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ Xs - \mu e \end{pmatrix}$$

$$\text{c'est-à-dire : } \begin{cases} Ad_x = 0 \\ -\nabla^2 f(x)d_x + A'd_y + d_s = 0 \\ Sd_x + Xd_s = \mu e - Xs \end{cases}$$

En éliminant  $d_s$ , nous recherchons finalement  $(d_x, d_y)$  solution de

$$(1) \begin{cases} Ad_x = 0 \\ [\nabla^2 f(x) + X^{-1}S]d_x - A'd_y = \mu X^{-1}e - s \end{cases}$$

b) Caractère dual-admissible du nouveau point :

Soit  $(x, y, s)$  un point dual-admissible, c'est-à-dire un point vérifiant

$$A'y + s - \nabla f(x) = 0. \quad (i)$$

Nous allons imposer au nouveau point d'être également dual-admissible; nous recherchons donc  $(d_x, d_y, d_s)$  de telle sorte que

$$A'(y + d_y) + (s + d_s) - \nabla f(x + d_x) = 0 \quad (ii)$$

Effectuant (ii)-(i), nous obtenons immédiatement

$$d_s + A'd_y - \nabla f(x + d_x) + \nabla f(x) = 0$$

ou encore  $d_s = \nabla f(x + d_x) - \nabla f(x) - A'd_y$  (iii)

Nous basant sur le résultat obtenu en programmation linéaire, nous remplaçons la deuxième équation du système (\*) par l'équation (iii). Finalement, le système à résoudre sera le suivant :

$$\begin{cases} Ad_x = 0 \\ d_s = \nabla f(x + d_x) - \nabla f(x) - A'd_y \\ Sd_x + Xd_s = \mu e - Xs \end{cases}$$

En éliminant  $d_s$ , nous recherchons  $(d_x, d_y)$  solution de :

$$(2) \begin{cases} Ad_x = 0 \\ X[\nabla f(x + d_x) - \nabla f(x) - A'd_y] + Sd_x + Xs - \mu e = 0 \end{cases}$$

KORTANEK, POTRA et YE ont proposé ensemble, sur base de cette méthode, un algorithme résolvant simultanément les problèmes primal et dual. Le désavantage le plus important de cette méthode est, qu'à chaque itération, l'algorithme impose de résoudre approximativement le système d'équations non-linéaires (2). Afin de remédier à ce problème, ZHU a proposé un deuxième algorithme, basé sur celui de KORTANEK- POTRA -YE, pour lequel le système d'équations à résoudre est le système linéaire (1).

L'étude de ces deux algorithmes et de leur convergence feront l'objet des deux paragraphes suivants.

## 5. Méthode de KORTANEK-POTRA-YE.

### A. L'algorithme.

Initialisation : soit  $\varepsilon > 0$  le paramètre de précision de l'algorithme.  
soit  $0 < \alpha < 0.1$   
soit  $(x^0, s^0)$  un point de départ tel que  $(x^0, s^0) \in \mathcal{N}(\alpha)$   
 $k = 0$

Pas 1 : si  $\mu^k = \frac{(x^k) \cdot s^k}{n} \leq \varepsilon$  alors stop.  
sinon effectuer le pas 2.

Pas 2 : poser  $\sigma = 1 - \frac{\alpha}{\sqrt{n}}$   
résoudre approximativement en  $(d_x, d_y)$  le système  $\begin{cases} A \cdot d_x = 0 \\ r(d_x, d_y) = 0 \end{cases}$   
où  $r(d_x, d_y) = X^k [\nabla f(x^k + d_x) - \nabla f(x^k) - A^t y^k] + S^k \cdot d_x + X^k s^k - \sigma \mu^k e$   
c'est-à-dire calculer  $(d_x, d_y)$  solution du système  $\begin{cases} A d_x = 0 \\ \|r(d_x, d_y)\| \leq \frac{1}{2} \cdot \alpha \mu^k \end{cases}$   
mettre à jour  $\begin{aligned} x^{k+1} &= x^k + d_x \\ y^{k+1} &= y^k + d_y \\ s^{k+1} &= \nabla f(x^{k+1}) - A^t y^{k+1} \\ k &= k + 1 \end{aligned}$

revenir au pas 1.

remarque : la norme utilisée est la norme  $\|\cdot\|_2$  définie par  $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ .

### B. Convergence de l'algorithme.

L'algorithme de KORTANEK-POTRA-YE est directement motivé par le théorème suivant :

**Théorème 3 :** pour tout  $k$ , nous avons :

- 1)  $\|(X^k)^{-1} d_x\|^2 + \|(s^k)^{-1} d_s\|^2 < 1$
- 2)  $\|X^{k+1} s^{k+1} - \mu^{k+1} e\| \leq \alpha \cdot \mu^{k+1}$
- 3)  $\mu^{k+1} \leq \left(1 - \frac{0.253 \cdot \alpha}{\sqrt{n}}\right) \cdot \mu^k$

remarques : ▪ la partie 1) du théorème entraîne que  $x^{k+1} > 0$  et  $s^{k+1} > 0$ .  
 En effet, si 1) est vérifié, on a :

$$\|(X^k)^{-1} d_x\|^2 < 1 \text{ c'est-à-dire } \sum_{i=1}^n \left( \frac{d_{x_i}}{x_i^k} \right)^2 < 1$$

$$\begin{aligned} \text{Donc, } \forall i = 1, \dots, n, \text{ on a: } & \left( \frac{d_{x_i}}{x_i^k} \right)^2 < 1 \\ & \Leftrightarrow (d_{x_i})^2 < (x_i^k)^2 \\ & \Leftrightarrow -x_i^k < d_{x_i} < x_i^k. \end{aligned}$$

Par conséquent,  $-x_i^k - d_{x_i} < 0 \quad \forall i = 1 \dots n$   
 ou encore  $x_i^k + d_{x_i} > 0 \quad \forall i = 1 \dots n$   
 et donc  $x^k + d_x > 0$   
 c'est-à-dire  $x^{k+1} > 0$ .

Le même raisonnement s'applique à  $s^{k+1}$ .

▪ La partie 2) implique que  $(x^{k+1}, s^{k+1})$  se trouve dans le  $\alpha$ -voisinage de la trajectoire centrale.

▪ La partie 3) établit la convergence linéaire globale de l'algorithme.

Avant de fournir une preuve du théorème de convergence, nous allons démontrer plusieurs lemmes techniques.

Afin d'alléger les notations, nous noterons  $x^k$  par  $x$ ,  $x^{k+1}$  par  $\bar{x}$ . Ce même procédé sera appliqué aux variables  $y, s$  et  $\mu$ . Nous supposerons en outre que l'itération courante est l'itération  $k$ .

Lemme 1 :  $\|Xs - \sigma\mu e\|^2 \leq 2(\alpha\mu)^2$

preuve :

soit  $(x, s)$  la solution courante de l'algorithme. Les vecteurs  $Xs - \mu e$  et  $\mu e$  sont orthogonaux. En effet,  $\langle Xs - \mu e, \mu e \rangle = (Xs)' \mu e - \mu^2 e' e$

$$= \mu(Xs)' e - n\mu^2$$

On sait, par définition, que  $\mu = \frac{x' s}{n} = \frac{\sum_{i=1}^n x_i s_i}{n}$ .

Par conséquent,  $\langle Xs - \mu e, \mu e \rangle = \frac{\sum_{i=1}^n x_i s_i}{n} \cdot \sum_{i=1}^n x_i s_i - \frac{\left( \sum_{i=1}^n x_i s_i \right)^2}{n} = 0$

et donc  $Xs - \mu e$  et  $\mu e$  sont bien orthogonaux.

D'autre part,  $(x, s)$  se trouve dans le  $\alpha$ -voisinage de la trajectoire centrale,

c'est-à-dire  $\|Xs - \mu e\| \leq \alpha \cdot \mu$ .

Regroupant ces deux résultats, nous obtenons :

$$\begin{aligned} \|Xs - \sigma \mu e\|^2 &= \left\| Xs - \left(1 - \frac{\alpha}{\sqrt{n}}\right) \mu e \right\|^2 && \text{par définition de } \sigma \\ &= \|Xs - \mu e\|^2 + \left\| \frac{\alpha}{\sqrt{n}} \mu e \right\|^2 && \text{car } Xs - \mu e \perp \mu e \\ &= (\alpha \mu)^2 + (\alpha \mu)^2 = 2(\alpha \mu)^2 \end{aligned}$$

□

Lemme 2 :  $\left\| (XS)^{-\frac{1}{2}} \right\|^2 \leq [(1 - \alpha)\mu]^{-1}$

Preuve :

$$\text{on a : } (XS)^{-\frac{1}{2}} = \begin{pmatrix} \left(\frac{1}{x_1 s_1}\right)^{\frac{1}{2}} & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \left(\frac{1}{x_n s_n}\right)^{\frac{1}{2}} \end{pmatrix}$$

Par conséquent, puisque  $(XS)^{-\frac{1}{2}}$  est une matrice diagonale,

$$\left\| (XS)^{-\frac{1}{2}} \right\|^2 = \max_{i=1, \dots, n} \left( \frac{1}{x_i s_i} \right)$$

De plus, puisque  $(x, s)$  est solution courante de l'algorithme,

$(x, s)$  vérifie :  $\|Xs - \mu e\| \leq \alpha \mu$ .

Donc,  $\forall i = 1, \dots, n$   $-\alpha \mu \leq x_i s_i - \mu (\leq \alpha \mu)$

c'est-à-dire  $x_i s_i \geq (1 - \alpha)\mu$

ou encore  $\frac{1}{x_i s_i} \leq [(1 - \alpha)\mu]^{-1}$ .

Finalement, on obtient  $\left\| (XS)^{-\frac{1}{2}} \right\|^2 \leq [(1 - \alpha)\mu]^{-1}$ .

□

Lemme 3 : Soit  $D = X^{\frac{1}{2}}S^{-\frac{1}{2}}$

$$\text{On a : } \|Dd_s\|^2 + \|D^{-1}d_x\|^2 \leq 4 \cdot \mu \left[ \frac{\alpha^2}{1-\alpha} \right]$$

Preuve :

$$\begin{aligned} d_x^t d_s &= d_x^t [\nabla f(x + d_x) - \nabla f(x) - A^t d_y] \\ &= d_x^t [\nabla f(x + d_x) - \nabla f(x)] \quad \text{car } d_x^t A^t = (A d_x)^t = 0 \\ &= d_x^t \int_0^1 \nabla^2 f[t(x + d_x) - (1-t)x] d_x dt \quad \text{par le théorème de la moyenne} \\ &= \int_0^1 d_x^t \nabla^2 f[(x + d_x) - (1-t)x] d_x dt \\ &\geq 0 \quad \text{car } \nabla^2 f \text{ est définie positive . (*)} \end{aligned}$$

$$\begin{aligned} \text{Par le lemme 1, } \sqrt{2} \alpha \mu &\geq \|\sigma \mu e - Xs\| \\ &\geq \|Xd_s + Sd_x\| - \|r(d_x, d_y)\| \quad \text{par définition de } r(d_x, d_y) \\ &\geq \|Xd_s + Sd_x\| - \frac{1}{2} \alpha \mu. \end{aligned}$$

$$\begin{aligned} \text{On obtient donc : } \|Xd_s + Sd_x\| &\leq \left( \frac{1}{2} + \sqrt{2} \right) \alpha \mu \\ &\leq 2 \alpha \mu. \end{aligned}$$

$$\begin{aligned} \text{Par conséquent, } 4\alpha^2 \mu^2 &\geq \|Xd_s + Sd_x\|^2 \\ &= \left\| (XS)^{\frac{1}{2}} (Dd_s + D^{-1}d_x) \right\|^2 \quad \text{par définition de } D \\ &\geq (1-\alpha) \mu \|Dd_s + D^{-1}d_x\|^2 \quad \text{par le lemme 2} \\ &\geq (1-\alpha) \mu \left[ \|Dd_s\|^2 + \|D^{-1}d_x\|^2 \right] \end{aligned}$$

$$\text{car } \|Dd_s + D^{-1}d_x\|^2 = \|Dd_s\|^2 + \|D^{-1}d_x\|^2 + 2d_s^t \cdot d_x \quad \text{et } d_s^t \cdot d_x \geq 0 \quad \text{par (*).}$$

$$\text{Finalement, } \|Dd_s\|^2 + \|D^{-1}d_x\|^2 \leq 4 \cdot \mu \cdot \left[ \frac{\alpha^2}{1-\alpha} \right].$$

□

Lemme 4 :  $\|S^{-1}d_s\|^2 + \|X^{-1}d_x\|^2 \leq \left(\frac{2\alpha}{1-\alpha}\right)^2 \triangleq T_1(\alpha)$

Preuve :

$$\begin{aligned} S^{-1}d_s &= (XS)^{-\frac{1}{2}}X^{\frac{1}{2}}S^{-\frac{1}{2}}d_s \\ &= (XS)^{-\frac{1}{2}}Dd_s \quad \text{par définition de } D. \end{aligned}$$

$$\begin{aligned} \text{Par conséquent,} \quad \|S^{-1}d_s\|^2 &= \left\| (XS)^{-\frac{1}{2}}Dd_s \right\|^2 \\ &\leq \left\| (XS)^{-\frac{1}{2}} \right\|^2 \|Dd_s\|^2. \end{aligned}$$

$$\begin{aligned} \text{De même,} \quad X^{-1}d_x &= (XS)^{-\frac{1}{2}}X^{-\frac{1}{2}}S^{\frac{1}{2}}d_x \\ &= (XS)^{-\frac{1}{2}}D^{-1}d_x \end{aligned}$$

$$\text{et donc,} \quad \|X^{-1}d_x\|^2 \leq \left\| (XS)^{-\frac{1}{2}} \right\|^2 \|D^{-1}d_x\|^2$$

En regroupant ces deux résultats, on obtient :

$$\|S^{-1}d_s\|^2 + \|X^{-1}d_x\|^2 \leq \left\| (XS)^{-\frac{1}{2}} \right\|^2 \left[ \|Dd_s\|^2 + \|D^{-1}d_x\|^2 \right].$$

La thèse découle alors directement des lemmes 2 et 3 .

□

Lemme 5 :  $d'_x d_s \leq \frac{2 \cdot \alpha^2 \mu}{(1-\alpha)^2}$  et  $\|D_x d_s\| \leq \frac{2 \cdot \alpha^2 \mu}{(1-\alpha)^2}$

Preuve :

$$\max(d'_x d_s, \|D_x d_s\|) = \max(d'_x D^{-1} D d_s, \|D_x D^{-1} D d_s\|)$$

Par l'inégalité de CAUCHY-SCHWARZ, on a d'une part :

$$\begin{aligned} d'_x D^{-1} D d_s &\leq \|D^{-1}d_x\| \|Dd_s\| \\ &\leq \frac{1}{2} \left( \|D^{-1}d_x\|^2 + \|Dd_s\|^2 \right) \end{aligned}$$

$$\text{et d'autre part,} \quad \|D_x D^{-1} D d_s\| \leq \|D^{-1}d_x\| \|Dd_s\|$$



$$\leq \frac{1}{2} \left( \|D^{-1}d_x\|^2 + \|Dd_s\|^2 \right).$$

$$\text{Donc } \max(d'_x d_s, \|D_x d_s\|) \leq \frac{1}{2} \left( \|D^{-1}d_x\|^2 + \|Dd_s\|^2 \right)$$

Par le lemme 3, on obtient la thèse.

□

$$\text{Lemme 6 : } \left( 1 - \frac{3\alpha}{2\sqrt{n}} \right) \mu \leq \bar{\mu} \leq \left[ 1 - \frac{\alpha}{2\sqrt{n}} + \frac{2\alpha^2}{(1-\alpha)^2 n} \right] \mu \triangleq T_3(\alpha, n)$$

Preuve :

$$\begin{aligned} \bar{X}\bar{S} &= (X + D_x)(s + d_s) \\ &= Xs + Sd_x + Xd_s + D_x d_s \\ &= r(d_x, d_y) + \sigma\mu e + D_x d_s \quad \text{par définition de } r(d_x, d_y). \end{aligned}$$

$$\text{Or, } \bar{\mu} = \frac{\bar{X}'\bar{S}}{n} = e' \frac{\bar{X}\bar{S}}{n}.$$

$$\begin{aligned} \text{Donc, } \bar{\mu} &= e' [r(d_x, d_y) + \sigma\mu e + D_x d_s] \\ &= \sigma\mu \frac{e'e}{n} + \frac{1}{n} e' r(d_x, d_y) + \frac{1}{n} e' D_x d_s \\ &= \sigma\mu + \frac{1}{n} e' r(d_x, d_y) + \frac{1}{n} e' D_x d_s. \end{aligned}$$

Puisque  $d'_x d_s \geq 0$ , on a :  $e' D_x d_s \geq 0$

$$\text{et donc, } \bar{\mu} \geq \sigma\mu + \frac{1}{n} e' r(d_x, d_y)$$

$$\geq \sigma\mu - \frac{1}{n} \|e\| \|r(d_x, d_y)\| \quad \text{par l'inégalité de CAUCHY-SCHWARZ.}$$

On sait, par le critère de l'algorithme, que  $\|r(d_x, d_y)\| \leq \frac{1}{2}\alpha\mu$ .

$$\text{De plus, } \|e\| = \sqrt{n} \text{ et } \sigma = 1 - \frac{\alpha}{\sqrt{n}}.$$

$$\begin{aligned} \text{Donc, } \bar{\mu} &\geq \left(1 - \frac{\alpha}{\sqrt{n}}\right) \mu - \frac{1}{n} \cdot \sqrt{n} \cdot \frac{1}{2} \alpha \mu \\ &= \left(1 - \frac{\alpha}{\sqrt{n}} - \frac{\alpha}{2\sqrt{n}}\right) \cdot \mu = \left(1 - \frac{3\alpha}{2\sqrt{n}}\right) \cdot \mu \end{aligned}$$

et la première inégalité du lemme est ainsi établie .

Pour démontrer la deuxième inégalité, reprenons le résultat :

$$\bar{\mu} = \sigma \mu + \frac{1}{n} e^t r(d_x, d_y) + \frac{1}{n} e^t D_x d_s.$$

$$\text{On a : } \bar{\mu} = \mu - \frac{\alpha \mu}{\sqrt{n}} + \frac{1}{n} e^t r(d_x, d_y) + \frac{1}{n} d_x^t d_s \text{ par définition de } \sigma$$

$$\leq \mu - \frac{\alpha \mu}{\sqrt{n}} + \frac{1}{n} \|e\| \|r(d_x, d_y)\| + \frac{1}{n} d_x^t d_s$$

par l'inégalité de CAUCHY-SCHWARZ.

$$\leq \mu - \frac{\alpha \mu}{\sqrt{n}} + \frac{1}{\sqrt{n}} \cdot \frac{\alpha \cdot \mu}{2} + \frac{1}{n} d_x^t d_s$$

$$\text{car } \|e\| = \sqrt{n} \text{ et } \|r(d_x, d_y)\| \leq \frac{1}{2} \alpha \mu$$

$$\leq \mu - \frac{\alpha \mu}{\sqrt{n}} + \frac{\alpha \cdot \mu}{2\sqrt{n}} + \frac{1}{n} \frac{2\alpha^2 \mu}{(1-\alpha)^2} \text{ par le lemme 5.}$$

$$= \left[1 - \frac{\alpha}{2\sqrt{n}} + \frac{2\alpha^2}{n(1-\alpha)^2}\right] \mu$$

D'où la thèse.

□

$$\text{Lemme 7: } \|\bar{X}\bar{s} - \bar{\mu}e\| \leq \left[\frac{1}{2}\alpha + \frac{2\alpha^2}{(1-\alpha)^2}\right] \left[1 - \frac{3\alpha}{2\sqrt{n}}\right]^{-1} \bar{\mu} \triangleq T_2(\alpha, n)$$

Preuve :

Les vecteurs  $\bar{X}\bar{s} - \bar{\mu}e$  et  $e$  sont orthogonaux.

$$\text{En effet : } \langle \bar{X}\bar{s} - \bar{\mu}e, e \rangle = e^t \bar{X}\bar{s} - \bar{\mu} e^t e$$

$$\begin{aligned}
&= \bar{x}'\bar{s} - n\bar{\mu} \\
&= \bar{x}'\bar{s} - n\frac{\bar{x}'\bar{s}}{n} \text{ par définition de } \bar{\mu} . \\
&= 0
\end{aligned}$$

Par conséquent,  $\|\bar{X}\bar{s} - \bar{\mu}e\|^2 = \langle \bar{X}\bar{s} - \bar{\mu}e, \bar{X}\bar{s} - \bar{\mu}e \rangle$

$$\begin{aligned}
&= \langle \bar{X}\bar{s} - \bar{\mu}e, \bar{X}\bar{s} \rangle \text{ car } \bar{X}\bar{s} - \bar{\mu}e \perp e \\
&= \|\bar{X}\bar{s}\|^2 - \bar{\mu}e'\bar{X}\bar{s} \\
&= \|\bar{X}\bar{s}\|^2 - \bar{\mu} \cdot n\bar{\mu} \\
&= \|\bar{X}\bar{s}\|^2 - n\bar{\mu}^2
\end{aligned}$$

On a également :  $\|\bar{X}\bar{s} - \sigma\mu e\|^2 = \|\bar{X}\bar{s}\|^2 + (\sigma\mu)^2 n - 2\sigma\mu n\bar{\mu}$ .

Donc,  $\|\bar{X}\bar{s} - \sigma\mu e\|^2 - \|\bar{X}\bar{s} - \bar{\mu}e\|^2 = n[(\sigma\mu)^2 - 2\sigma\mu\bar{\mu} + \bar{\mu}^2]$

$$= n(\sigma\mu - \bar{\mu})^2 \geq 0.$$

Ainsi,  $\|\bar{X}\bar{s} - \bar{\mu}e\| \leq \|\bar{X}\bar{s} - \sigma\mu e\|$

$$\begin{aligned}
&= \|Xs + Xd_s + Sd_x + D_x d_s - \sigma\mu e\| \\
&= \|r(d_x, d_y) + D_x d_s\| \\
&\leq \|r(d_x, d_y)\| + \|D_x d_s\|.
\end{aligned}$$

Par le lemme 5,  $\|D_x d_s\| \leq \frac{2\alpha^2\mu}{(1-\alpha)^2}$ .

De plus,  $\|r(d_x, d_y)\| \leq \frac{1}{2}\alpha\mu$ .

Par conséquent,  $\|\bar{X}\bar{s} - \bar{\mu}e\| \leq \left[ \frac{1}{2}\alpha + \frac{2\alpha^2}{(1-\alpha)^2} \right] \mu$

où, par la lemme 6,  $\mu \leq \left( 1 - \frac{3\alpha}{2\sqrt{n}} \right)^{-1} \bar{\mu}$ .

On obtient finalement :  $\|\bar{X}\bar{s} - \bar{\mu}e\| \leq \left[ \frac{1}{2}\alpha + \frac{2\alpha^2}{(1-\alpha)^2} \right] \left[ 1 - \frac{3\alpha}{2\sqrt{n}} \right]^{-1} \bar{\mu}$ .

□

Grâce à ces lemmes préliminaires, le théorème de convergence se démontre aisément.

Preuve du théorème 3 :

1) Par le lemme 4, pour montrer l'inégalité  $\|S^{-1}d_s\|^2 + \|X^{-1}d_s\|^2 < 1$ ,

il suffit de prouver que si  $0 < \alpha < 0.1$ , alors  $T_1(\alpha) = \left(\frac{2\alpha}{1-\alpha}\right)^2 < 1$ .

Or,  $T_1(\alpha)$  est une fonction croissante de  $\alpha$ . En effet,

$$T_1'(\alpha) = \frac{8\alpha}{(1-\alpha)^3} > 0 \text{ pour } 0 < \alpha < 0.1.$$

Donc, la valeur maximale qu'atteint  $T_1(\alpha)$  pour  $0 < \alpha < 0.1$  est

$$T_1(0.1) = 0.04938 < 1.$$

2) Afin d'établir que  $\|X^{k+1}S^{k+1} - \mu^{k+1}e\| \leq \alpha \cdot \mu^{k+1}$ , montrons que

$$T_2(\alpha, n) = \left[ \frac{1}{2}\alpha + \frac{2\alpha^2}{(1-\alpha)^2} \right] \left[ 1 - \frac{3\alpha}{2\sqrt{n}} \right]^{-1} \mu^{k+1} \leq \alpha \mu^{k+1} \text{ pour } 0 < \alpha < 0.1$$

On a :  $T_2(\alpha, n) \leq \alpha \mu^{k+1}$

si et seulement si  $\frac{2\alpha}{(1-\alpha)^2} + \frac{3\alpha}{2\sqrt{n}} \leq \frac{1}{2}$ .

$$\text{En effet, } \left[ \frac{1}{2}\alpha + \frac{2\alpha^2}{(1-\alpha)^2} \right] \left[ 1 - \frac{3\alpha}{2\sqrt{n}} \right]^{-1} \mu^{k+1} \leq \alpha \mu^{k+1}$$

$$\Leftrightarrow \frac{1}{2}\alpha + \frac{2\alpha^2}{(1-\alpha)^2} \leq \alpha \left( 1 - \frac{3\alpha}{2\sqrt{n}} \right) \text{ car } 1 - \frac{3\alpha}{2\sqrt{n}} \geq 0$$

$$\Leftrightarrow \frac{1}{2} + \frac{2\alpha}{(1-\alpha)^2} \leq 1 - \frac{3\alpha}{2\sqrt{n}} \text{ car } 0 < \alpha < 0.1$$

$$\Leftrightarrow \frac{2\alpha}{(1-\alpha)^2} + \frac{3\alpha}{2\sqrt{n}} \leq \frac{1}{2}.$$

Le membre de gauche de l'inégalité est croissant en  $\alpha$  et décroissant en  $n$ . En posant  $\alpha = 0.1$  et  $n = 1$ , on obtient donc son maximum :

$$T_2(0.1, 1) = 0.3969 \text{ qui est bien strictement inférieur à } \frac{1}{2}.$$

3) Il reste à démontrer que  $\mu^{k+1} \leq \left(1 - \frac{0.253 \cdot \alpha}{\sqrt{n}}\right) \mu^k$ .

En se basant sur le lemme 6, il suffit de voir que

$$T_3(\alpha, n) = \left[1 - \frac{\alpha}{2\sqrt{n}} + \frac{2\alpha^2}{(1-\alpha)^2 n}\right] \mu^k \leq \left(1 - \frac{0.253 \cdot \alpha}{\sqrt{n}}\right) \mu^k$$

pour  $0 < \alpha < 0.1$ . Réécrivons  $T_3(\alpha, n)$  sous la forme :

$$\begin{aligned} T_3(\alpha, n) &= \left[1 - \frac{\alpha}{2\sqrt{n}} + \frac{2\alpha^2}{(1-\alpha)^2 n}\right] \mu^k \\ &= \left[1 - \frac{\alpha}{2\sqrt{n}} \left(1 - \frac{4\alpha}{(1-\alpha)^2 \sqrt{n}}\right)\right] \mu^k. \end{aligned}$$

On sait que  $\frac{4\alpha}{(1-\alpha)^2 \sqrt{n}}$  est croissant en  $\alpha$  et décroissant en  $n$ .

Par conséquent, cette fonction atteint sa valeur maximale pour  $\alpha = 0.1$  et  $n = 1$ . Cette valeur est 0.4938.

On a alors :  $\frac{4\alpha}{(1-\alpha)^2 \sqrt{n}} \leq 0.4938$  pour tout  $n$  et pour  $0 < \alpha < 0.1$ .

$$\begin{aligned} \text{Donc, } T_3(\alpha, n) &\leq \left[1 - \frac{\alpha}{2\sqrt{n}} (1 - 0.4938)\right] \mu^k \\ &= \left[1 - \frac{0.253 \cdot \alpha}{\sqrt{n}}\right] \mu^k. \end{aligned}$$

□

Ce théorème montre que l'algorithme de KORTANEK-POTRA-YE résout un problème d'optimisation convexe sous contraintes d'égalités linéaires à un taux de convergence globale linéaire. A chaque itération, le saut de dualité est au moins réduit d'un facteur de  $\left(1 - \frac{\beta}{\sqrt{n}}\right)$ .

Mais en pratique, l'application de cet algorithme est rendue difficile par le caractère non-linéaire du système d'équations à résoudre à chaque itération.

ZHU a considérablement amélioré l'algorithme en linéarisant le système d'équations. Pour ce faire, il a été amené à introduire la condition SLC (Scaled Lipschitz Condition). Sa méthode fera l'objet du paragraphe suivant.

## 6. Méthode de ZHU.

La méthode de ZHU consiste à effectuer un pas de la méthode de Newton sur le système d'équations de KUHN-TUCKER. Nous devons donc résoudre le système :

$$(*) \quad \begin{cases} Ad_x = 0 \\ [\nabla^2 f(x) + X^{-1}S]d_x - A^t d_y = \mu X^{-1}e - s \end{cases}$$

Définition 3 : On dit que  $f(x)$  satisfait la condition SLC (Scaled Lipschitz Condition) si

$\forall \beta \in ]0,1[$ , il existe  $M > 0$  tel que, pour tout  $x < 0$ ,

pour tout  $d_x$  satisfaisant  $\|X^{-1}d_x\| \leq \beta$ ,

nous avons :  $\|X[\nabla f(x + d_x) - \nabla f(x) - \nabla^2 f(x) \cdot d_x]\| \leq M \cdot d_x^t \nabla^2 f(x) d_x$ .

Remarque : Il est évident que les fonctions linéaires et convexes quadratiques remplissent la condition SLC avec  $M=0$ .

Plus généralement, afin de satisfaire la condition SLC,  $f(x)$  doit au minimum être convexe et deux fois différentiable.

Nous allons montrer que si  $f(x)$  remplit la condition SLC et si  $(d_x, d_y)$  est solution du système linéaire (\*) ci-dessus, alors  $(d_x, d_y)$  est solution du système non-linéaire :

$$(**) \quad \begin{cases} Ad_x \\ \|r(d_x, d_y)\| \leq \frac{1}{2} \cdot \alpha \cdot \mu \end{cases}$$

Il suffira, dès lors, de remplacer, dans l'algorithme de KORTANEK-POTRA-YE, le système (\*\*) par le système (\*).

### A. Algorithme de ZHU.

Nous supposons désormais que la fonction objectif  $f(x)$  remplit la condition SLC. Le lemme suivant nous permettra de démontrer le résultat annoncé.

Lemme 8 : si  $(d_x, d_y)$  est la solution du système (\*),

$$\text{alors } \|X^{-1}d_x\| \leq \sqrt{2} \left( \frac{\alpha}{1-\alpha} \right).$$

Preuve :

$$\text{soit } \xi = \nabla^2 f(x) d_x - A' d_y$$

La deuxième équation du système (\*) peut se réécrire :

$$X \cdot \nabla^2 f(x) d_x + S d_x - X A' d_y = \sigma \mu e - X s$$

ou encore, puisque  $\xi = \nabla^2 f(x) d_x - A' d_y$ ,

$$X \xi + S d_x = \sigma \mu e - X s .$$

Soit définie comme précédemment  $D = X^{\frac{1}{2}} S^{-\frac{1}{2}}$ .

Par le lemme 1, nous savons que  $\|\sigma \mu e - X s\|^2 \leq 2 \cdot (\alpha \mu)^2$

$$\begin{aligned} \text{Par conséquent,} \quad 2(\alpha \mu)^2 &\geq \|\sigma \mu e - X s\|^2 \\ &= \|X \xi + S d_x\|^2 \\ &= \left\| (XS)^{\frac{1}{2}} (D \xi + D^{-1} d_x) \right\|^2 \text{ par définition de } D. \end{aligned}$$

$$\begin{aligned} \text{Or,} \quad \xi' d_x &= d_x' \xi = d_x' \nabla^2 f(x) d_x - d_x' A' d_y \\ &= d_x' \nabla^2 f(x) d_x \quad \text{car } A d_x = 0 \\ &\geq 0 \quad \text{car } \nabla^2 f \text{ est définie positive.} \end{aligned}$$

$$\begin{aligned} \text{Donc,} \quad \|D \xi + D^{-1} d_x\|^2 &= \|D \xi\|^2 + \|D^{-1} d_x\|^2 + 2 \langle D \xi, D^{-1} d_x \rangle \\ &\geq \|D \xi\|^2 + \|D^{-1} d_x\|^2. \end{aligned}$$

De plus, par le lemme 2,  $\left\| (XS)^{\frac{1}{2}} \right\|^2 \geq (1 - \alpha) \mu$ .

$$\begin{aligned} \text{Nous obtenons donc :} \quad 2(\alpha \mu)^2 &\geq (1 - \alpha) \mu \left( \|D \xi\|^2 + \|D^{-1} d_x\|^2 \right) \\ &\geq (1 - \alpha) \mu \|D^{-1} d_x\|^2 \\ &= (1 - \alpha) \mu \left\| (XS)^{\frac{1}{2}} X^{-1} d_x \right\|^2 \text{ par définition de } D \\ &\geq (1 - \alpha)^2 \mu^2 \|X^{-1} d_x\|^2 \text{ par le lemme 2.} \end{aligned}$$

$$\text{Finalement,} \quad \|X^{-1} d_x\|^2 \leq \frac{2\alpha^2}{(1 - \alpha)^2}$$

□

Remarque : Nous avons déjà obtenu le résultat :  $\|X^{-1}d_x\| \leq \sqrt{2} \frac{\alpha}{1-\alpha}$ .

Or,  $\frac{\alpha}{1-\alpha}$  est une fonction croissante de  $\alpha$  (car sa dérivée est  $\frac{1}{(1-\alpha)^2} \geq 0$ )

et  $\sqrt{2} \frac{\alpha}{1-\alpha} < 1$  si  $\alpha \leq \bar{\alpha} = 0.41$ .

Donc,  $\forall \alpha \leq \bar{\alpha}$ , on a :  $\|X^{-1}d_x\| \leq \sqrt{2} \frac{\alpha}{1-\alpha} \leq \sqrt{2} \frac{\bar{\alpha}}{1-\bar{\alpha}} < 1$ , ce qui

assure que  $\bar{x} = x^{k+1} > 0$ .

Posons  $\bar{\beta} = \sqrt{2} \frac{\bar{\alpha}}{1-\bar{\alpha}} = 0.98 < 1$ .

Du caractère SLC de  $f(x)$ , on déduit :

$\exists M > 0$  tel que  $\forall x > 0$ ,

$\forall d_x$  tel que  $\|X^{-1}d_x\| \leq \bar{\beta}$ ,

$\|X[\nabla f(x+d_x) - \nabla f(x) - \nabla^2 f(x)d_x]\| \leq Md_x \cdot \nabla^2 f(x)d_x$ .

Nous pouvons désormais démontrer le résultat majeur de ce paragraphe, à savoir la possibilité de remplacer le système (\*\*) par le système (\*) .

**Théorème 4 :** soit  $\alpha \leq \min\left(\frac{1}{4M+1}, 0.41\right)$ .

Si  $(d_x, d_y)$  est une solution du système (\*\*),

alors  $\|r(d_x, d_y)\| \leq \frac{1}{2} \alpha \cdot \mu$ .

Preuve :

Par définition,

$$\begin{aligned} r(d_x, d_y) &= X[\nabla f(x+d_x) - \nabla f(x) - A'y] + Sd_x + Xs - \sigma\mu e \\ &= X[\nabla f(x+d_x) - \nabla f(x)] - XA'y + Sd_x + Xs - \sigma\mu e + X\nabla^2 f(x)d_x - X\nabla^2 f(x)d_x \\ &= X[\nabla f(x+d_x) - \nabla f(x) - \nabla^2 f(x)d_x] + X\nabla^2 f(x)d_x + Sd_x - XA'y - (\sigma\mu e - Xs). \end{aligned}$$

Or, par le système (\*),  $X\nabla^2 f(x)d_x + Sd_x - XA'y = \sigma\mu e - Xs$ .



Donc,  $r(d_x, d_y) = X[\nabla f(x + d_x) - \nabla f(x) - \nabla^2 f(x)d_x]$  .

Puisque  $f(x)$  satisfait la condition SLC , nous avons :

$$\|X[\nabla f(x + d_x) - \nabla f(x) - \nabla^2 f(x)d_x]\| \leq Md'_x \nabla^2 f(x)d_x \text{ .}$$

Ceci implique que :

$$\|r(d_x, d_y)\| \leq Md'_x \nabla^2 f(x)d_x \text{ .} \quad (1)$$

Reprenant la deuxième équation du système (\*), nous obtenons en prenant le produit scalaire par  $d_x$  des deux côtés de l'égalité :

$$d'_x \nabla^2 f(x)d_x + d'_x X^{-1} S d_x - d'_x A^t d_y = d'_x X^{-1} (\sigma \mu e - Xs)$$

où  $d'_x A^t d_y = 0$  puisque  $Ad_x = 0$ .

Comme  $X^{-1}S$  est une matrice définie positive (car  $x$  et  $s$  sont positifs), et donc que  $d'_x X^{-1} S d_x \geq 0$  , nous pouvons déduire que :

$$\begin{aligned} d'_x \nabla^2 f(x)d_x &\leq d'_x X^{-1} (\sigma \mu e - Xs) \\ &\leq \|X^{-1} d_x\| \|\sigma \mu e - Xs\| \text{ par l'inégalité de CAUCHY-SCHWARZ.} \end{aligned}$$

Par le lemme 8,  $\|X^{-1} d_x\| \leq \sqrt{2} \frac{\alpha}{1-\alpha}$

et par le lemme 1,  $\|\sigma \mu e - Xs\| \leq \sqrt{2} \alpha \cdot \mu$  .

Donc,  $d'_x \nabla^2 f(x)d_x \leq \frac{2\alpha^2}{1-\alpha} \cdot \mu$ . (2)

Combinant (1) et (2) , nous obtenons :  $\|r(d_x, d_y)\| \leq M \frac{2\alpha^2}{1-\alpha} \mu$  .

Il suffit désormais de trouver la valeur de  $\alpha$  pour laquelle

$$M \frac{2\alpha^2}{1-\alpha} \mu \leq \frac{1}{2} \alpha \mu$$

c'est-à-dire  $4M \frac{\alpha}{1-\alpha} \leq 1$

c'est-à-dire  $(4M + 1)\alpha \leq 1$

ou encore  $\alpha \leq \frac{1}{4M + 1}$

ce qui termine la démonstration .

□

Le système (\*) étant linéaire, nous pouvons, en fait, exprimer explicitement la solution  $(d_x, d_y)$  du système .

En effet,

$$\text{le système (*) est } \begin{cases} (\nabla^2 f(x) + X^{-1}S)d_x - A'd_y = \sigma\mu X^{-1}e - s \\ Ad_x = 0 \end{cases}$$

$$\text{Posons } H = (\nabla^2 f(x) + X^{-1}S)^{-1} .$$

$$\text{Nous devons alors résoudre } \begin{cases} H^{-1}d_x - A'd_y = \sigma\mu X^{-1}e - s & (3) \\ Ad_x = 0 & (4) \end{cases}$$

$$\text{Nous extrayons } d_x \text{ de (3) : } H^{-1}d_x = (\sigma\mu X^{-1}e - s) + A'd_y$$

$$\text{c'est-à-dire } d_x = H(\sigma\mu X^{-1}e - s) + HA'd_y .$$

$$\text{Par (4) , nous obtenons : } Ad_x = AH(\sigma\mu X^{-1}e - s) + AHA'd_y = 0$$

$$\text{Donc, } AHA'd_y = -AH(\sigma\mu X^{-1}e - s)$$

$$\text{c'est-à-dire } d_y = -(AHA')^{-1}AH(\sigma\mu X^{-1}e - s)$$

$$\text{Et donc, } d_x = [H - HA'(AHA')^{-1}AH](\sigma\mu X^{-1}e - s) .$$

Nous pouvons, maintenant, sur base des résultats précédents, présenter l'algorithme de ZHU.

### *L' algorithme de ZHU .*

On suppose que la fonction objectif  $f(x)$  satisfait la condition SLC.

Initialisation: soit  $\varepsilon > 0$  le paramètre de précision de l'algorithme

$$\text{soit } 0 < \alpha < \min\left(0.1, \frac{1}{4M+1}\right)$$

$$\text{soit } (x^0, s^0) \text{ un point de départ tel que } (x^0, s^0) \in \mathcal{O}(\alpha)$$

$$\text{Posons } k = 0$$

$$\text{Pas 1: } \text{ si } \mu^k = \frac{(x^k)^t \cdot s^k}{n} \leq \varepsilon \text{ alors stop}$$

sinon effectuer le pas 2 .

Pas 2:      poser  $\sigma = 1 - \frac{\alpha}{\sqrt{n}}$

Calculer       $H = \left( \nabla^2 f(x^k) + (X^{-1})^k S^k \right)^{-1}$  ;       $\xi = (\sigma \mu^k (X^k)^{-1} e - s^k)$

$$d_x = \left[ H - HA'(AHA')^{-1}AH \right] \xi$$

$$d_y = - \left[ (AHA')^{-1}AH \right] \xi$$

Mettre à jour       $x^{k+1} = x^k + d_x$

$$y^{k+1} = y^k + d_y$$

$$s^{k+1} = \nabla f(x^{k+1}) - A'y^{k+1}$$

$$k = k + 1$$

Revenir au pas 1 .

Nous avons déjà remarqué que les fonctions linéaires et convexes quadratiques remplissaient la condition SLC. Dans le paragraphe suivant, nous mettrons en évidence d'autres classes de fonctions satisfaisant la condition SLC.

### B. Classes de fonctions satisfaisant la condition SLC.

Soit  $\mathcal{F} = \{f(x) \text{ telle que } f(x) \text{ satisfait la condition SLC}\}$ .

**Théorème 5** : soit  $f(x)$  une fonction d'entropie ayant la forme :

$$f(x) = \sum_{i=1}^n x_i \ln \frac{x_i}{u_i} \text{ où } u_i > 0 \text{ pour tout } i$$

Alors  $f(x)$  satisfait la condition SLC.

Preuve :

Observons en premier lieu que :

$$\frac{\partial f(x)}{\partial x_i} = \ln x_i - \ln u_i + 1 \quad \text{pour } i=1, \dots, n$$

et

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} = \begin{cases} \frac{1}{x_i} & \text{si } i=j \\ 0 & \text{sinon} \end{cases} \quad \text{pour } i = 1, \dots, n .$$

Posons  $\gamma = X[\nabla f(x+d_x) - \nabla f(x) - \nabla^2 f(x)d_x]$  où  $\|X^{-1}d_x\| \leq \beta < 1$ .

Nous obtenons :

$$\begin{aligned} |\gamma_i| &= \left| x_i \left[ \ln(x_i + d_{x_i}) - \ln u_i + 1 - \ln x_i + \ln u_i - 1 - \frac{d_{x_i}}{x_i} \right] \right| \\ &= \left| x_i \left[ \ln(x_i + d_{x_i}) - \ln x_i - \frac{d_{x_i}}{x_i} \right] \right| \end{aligned}$$

Développant  $\ln(x_i + d_{x_i}) - \ln x_i - \frac{d_{x_i}}{x_i}$  en série, nous déduisons qu'il existe  $t \in [0, 1]$  tel que :

$$|\gamma_i| = \left| x_i \cdot \frac{1}{2} d_{x_i}^2 (x_i + t \cdot d_{x_i})^{-2} \right|.$$

Donc, 
$$|\gamma_i| = \left| \frac{1}{2} \frac{1}{x_i} (d_{x_i})^2 \left( 1 + t \frac{d_{x_i}}{x_i} \right)^{-2} \right|.$$

On a :  $\|X^{-1}d_x\| \leq \beta$  et donc aussi  $\|tX^{-1}d_x\| \leq \beta$  puisque  $0 \leq t \leq 1$ .

Par conséquent, pour tout  $i=1, \dots, n$   $-\beta \leq t \frac{d_{x_i}}{x_i} (\leq \beta)$

c'est-à-dire 
$$1 - \beta \leq 1 + t \frac{d_{x_i}}{x_i}$$

ou encore 
$$(1 - \beta)^{-2} \geq \left( 1 + t \frac{d_{x_i}}{x_i} \right)^{-2}.$$

Dès lors, 
$$|\gamma_i| \leq \frac{1}{2} (1 - \beta)^{-2} d_{x_i} \frac{1}{x_i} d_{x_i}.$$

et donc, 
$$\|\gamma\| = \|X[\nabla f(x+d_x) - \nabla f(x) - \nabla^2 f(x)d_x]\|$$

$$\leq \sum_{i=1}^n |\gamma_i|$$

$$\leq \frac{1}{2} (1 - \beta)^{-2} d_x^t \nabla^2 f(x) d_x.$$

Pour obtenir la thèse, il suffit dès lors de choisir  $M = \frac{1}{2}(1 - \beta)^{-2}$ .

□

**Théorème 6 :** soient  $\psi_i : R \rightarrow R$  pour  $i = 1, \dots, n$  des fonctions satisfaisant la condition:

il existe  $T \geq 0$  et  $p \geq 0$ , tels que pour tout  $x > 0$ , pour tout  $y > 0$ , on a:

$$y \left| \psi_i'''(y) \right| \leq T \max \left[ \left( \frac{x}{y} \right)^p, \left( \frac{y}{x} \right)^p \right] \psi_i''(x) \quad (*)$$

Alors  $f(x) = \sum_{i=1}^n \psi_i(x)$  satisfait la condition SLC.

Preuve :

$$\text{soit } |\gamma_i| = \left| x_i \left[ \psi_i'(x_i + d_{x_i}) - \psi_i'(x_i) - \psi_i''(x_i) d_{x_i} \right] \right|$$

On obtient, en développant en série :

$$\begin{aligned} |\gamma_i| &= \left| \frac{1}{2} x_i d_{x_i}^2 \psi_i'''(x_i + t d_{x_i}) \right| \text{ où } 0 \leq t \leq 1 \\ &= \frac{x_i}{2(x_i + t d_{x_i})} (x_i + t d_{x_i}) d_{x_i}^2 \left| \psi_i'''(x_i + t d_{x_i}) \right|. \end{aligned}$$

Puisque  $\|X^{-1} d_x\| \leq \beta$ , on a  $x + d_x > 0$

ou encore  $x + t d_x > 0$  car  $0 \leq t \leq 1$ .

On a donc, par (\*), avec  $x = x_i$  et  $y = \frac{(x_i + t d_{x_i})}{x_i}$  :

$$|\gamma_i| \leq \frac{x_i d_{x_i}^2}{2(x_i + t d_{x_i})} T \max \left[ \left( \frac{x_i + t d_{x_i}}{x_i} \right)^p, \left( \frac{x_i}{x_i + t d_{x_i}} \right)^p \right] \psi_i''(x_i) .$$

Et comme  $\frac{x_i + t d_{x_i}}{x_i} \leq 1 + \beta$  et  $\frac{x_i}{x_i + t d_{x_i}} \leq (1 - \beta)^{-1}$ ,

$$|\gamma_i| \leq \frac{T}{2} \max [(1 + \beta)^p, (1 - \beta)^{-p}] (1 - \beta)^{-1} d_{x_i} \psi_i''(x_i) d_{x_i} .$$

Soit  $M = \frac{T}{2} \max [(1 + \beta)^p, (1 - \beta)^{-p}] (1 - \beta)^{-1}$

On obtient alors pour tout  $x > 0$  tel que  $\|X^{-1} d_x\| \leq \beta$  :

$$\begin{aligned} \|X[\nabla f(x + d_x) - \nabla f(x) - \nabla^2 f(x) d_x]\| &= \|\gamma\| \\ &\leq \sum_{i=1}^n |\gamma_i| \\ &\leq M \sum_{i=1}^n d_{x_i} \psi_i''(x_i) d_{x_i} \end{aligned}$$

$$= M d_x^t \nabla^2 f(x) d_x$$

et donc  $f(x)$  satisfait la condition SLC.

□

De la définition de la condition SLC, on peut également déduire que toute combinaison positive de fonctions de  $\mathcal{F}$  sera un élément de  $\mathcal{F}$ .

## 7. Complexité des méthodes dans le cas linéaire.

Soit le programme linéaire : (PL) 
$$\begin{cases} \text{minimiser } c^t x \\ \text{s.c.} & Ax = b \\ & x \geq 0 \end{cases}$$

où  $c \in R^n$ ,  $b \in R^m$   
et  $A \in R^{m \times n}$  est de rang plein ( $m \leq n$ ).

On suppose que toutes les données d'entrée sont entières. La taille du problème est définie comme étant le nombre  $L = l + n + 1$  où  $l$  est la longueur totale, en représentation binaire, des données d'entrée.

Considérons tous les programmes linéaires dont la taille est  $L$ . On définit alors la complexité d'un algorithme pour cette taille  $L$  comme le comportement le plus mauvais de cet algorithme en fonction des données d'entrée. Pour rappel, le comportement d'un algorithme est le nombre total d'opérations élémentaires, c'est-à-dire le nombre total d'opérations arithmétiques, d'opérations de comparaison,...

Pour calculer la complexité d'un algorithme, il convient donc d'évaluer le nombre total d'itérations de l'algorithme ainsi que la complexité d'une de ces itérations.

### A. Nombre d'itérations.

A chaque itération, le saut de dualité est réduit d'un facteur  $\left(1 - \frac{\delta}{\sqrt{n}}\right)$  où  $\delta = 0.253 \cdot \alpha$

c'est-à-dire  $(x^k)^t s^k \leq \left(1 - \frac{\delta}{\sqrt{n}}\right) (x^{k-1})^t s^{k-1} \quad \forall k.$

On a donc :  $(x^k)^t s^k \leq \left(1 - \frac{\delta}{\sqrt{n}}\right)^k (x^0)^t s^0.$

Comme le test d'arrêt de l'algorithme est  $(x^k)^t s^k \leq n\epsilon$ , il suffit de calculer, pour obtenir le nombre d'itérations, le plus petit entier  $k$  pour lequel :

$$\left(1 - \frac{\delta}{\sqrt{n}}\right)^k (x^0)^t s^0 \leq n\epsilon.$$

Prenant le logarithme dans les deux membres, on obtient :

$$k \ln \left(1 - \frac{\delta}{\sqrt{n}}\right) + \ln (x^0)^t s^0 \leq \ln n\epsilon$$

c'est-à-dire  $k \geq \frac{\ln n\varepsilon - \ln (x^0)'s^0}{\ln \left(1 - \frac{\delta}{\sqrt{n}}\right)}$

Puisque  $\ln \left(1 - \frac{\delta}{\sqrt{n}}\right) \leq -\frac{\delta}{\sqrt{n}}$ , on a :  $\frac{1}{\ln \left(1 - \frac{\delta}{\sqrt{n}}\right)} \geq -\frac{\sqrt{n}}{\delta}$ .

Et comme  $n\varepsilon < (x^0)'s^0$  (on suppose que, pour  $k=0$ , le critère d'arrêt n'est pas satisfait),

On a :  $k \geq \frac{\ln n\varepsilon - \ln (x^0)'s^0}{-\frac{\delta}{\sqrt{n}}} = \frac{\sqrt{n}}{\delta} \ln \frac{(x^0)'s^0}{n\varepsilon}$

c'est-à-dire qu'au plus  $O\left(\frac{\sqrt{n}}{\delta} \ln \frac{(x^0)'s^0}{n\varepsilon}\right)$  itérations sont nécessaires.

Choisissons dès lors  $\varepsilon = \frac{2^{-L}}{n}$  et  $(x^0, s^0)$  tel que  $(x^0)'s^0 \leq 2^L$ .

On obtient :  $\frac{\sqrt{n}}{\delta} \ln \frac{(x^0)'s^0}{n\varepsilon} \leq \frac{\sqrt{n}}{\delta} \ln 2^{2L} = O(\sqrt{n} L)$

et donc au plus  $O(\sqrt{n} L)$  itérations sont nécessaires pour obtenir  $(x^k)'s^k \leq 2^{-L}$ .

### B. Complexité d'une itération.

Le calcul de la matrice  $[A(S^k)^{-1} X^k A^t]^{-1}$ , qui intervient dans les différentes directions de recherche, est celui qui coûte le plus d'opérations arithmétiques. Il demande  $O(n^3)$  opérations et peut être réduit à  $O(n^{2.5})$  en utilisant une méthode de mise à jour.

Par conséquent, pour obtenir  $(x^k)'s^k \leq 2^{-L}$ , nous avons besoin de  $O(n^3 L)$  opérations arithmétiques.



## **DEUXIEME PARTIE :**

**Généralisation aux contraintes d'inégalités  
convexes.**

Dans cette deuxième partie, nous nous intéresserons à la résolution du programme convexe

$$(PC) \begin{cases} \text{minimiser } f(x) \\ \text{s.c.} \quad g_i(x) \leq 0 \text{ pour } i = 1, \dots, m \end{cases}$$

où les fonctions  $f(x)$  et  $g_i(x)$ ,  $1 \leq i \leq m$ , sont supposées convexes.

Dans la littérature, plusieurs auteurs ont essayé d'obtenir des conditions assez générales qui assurent la convergence de leurs algorithmes pour la résolution du problème (PC). Utilisant

le plus fréquemment la fonction barrière logarithmique  $\varphi_\mu(x) = f(x) - \mu \sum_{i=1}^m \ln(-g_i(x))$  ou la fonction potentielle  $\varphi(x, \lambda) = -m \ln(\lambda - f(x)) - \sum_{i=1}^m \ln(-g_i(x))$ , ils ont dégagé deux conditions garantissant la convergence de la méthode de Newton : la condition de self-concordance d'une part et la condition de Lipschitz relative d'autre part. Ces deux conditions sont en fait liées. ( Nous renvoyons en annexe pour plus de détails ).

Cette partie est agencée comme suit : dans un premier chapitre, nous proposerons un algorithme à petits pas imposant à la fonction potentielle la condition de self-concordance ; dans un deuxième chapitre, nous présenterons deux algorithmes à grands pas, l'un minimisant la fonction barrière logarithmique et l'autre la fonction potentielle, pour lesquels la condition requise est la condition de Lipschitz relative.

## Chapitre 1 :

### Algorithme à petit pas.

#### 1. Présentation du problème.

##### A. Le problème.

Le problème est de trouver  $\lambda^* = \min f(x)$  sous contrainte  $x \in \mathcal{P}$   
où  $\mathcal{P} = \{x \in R^n \text{ tel que } g_i(x) \leq 0 \text{ pour } 1 \leq i \leq m\}$  et où les fonctions  $f$  et  $g_i$ ,  $1 \leq i \leq m$ , sont convexes et deux fois continûment différentiables sur  $\mathcal{P}$ .

##### B. La condition de self-concordance.

Soit  $\varphi : \mathcal{P}^\circ \rightarrow R$  une fonction.

Définition 1 : on dit que  $\varphi(x)$  remplit la condition de self-concordance sur  $\mathcal{P}^\circ$  si  $\varphi(x)$  est trois fois continûment différentiable sur  $\mathcal{P}^\circ$  et si  $\exists \alpha \geq 0$  tel que  $\forall x \in \mathcal{P}^\circ$ ,  $\forall h \in R^n$ , on a :

$$|\nabla^3 \varphi(x)[h, h, h]| \leq 2\sqrt{\alpha} (h^t \nabla^2 \varphi(x) h)^{\frac{3}{2}}.$$

Intuitivement, cette définition assure, dans le cas où le paramètre  $\alpha$  est petit, une bonne approximation de la dérivée troisième de  $\varphi(x)$  par une fonction quadratique. On vérifie immédiatement que les fonctions linéaires et convexes quadratiques satisfont cette condition avec  $\alpha = 0$ .

Définition 2 : on dit que  $\varphi(x)$  remplit la condition de self-concordance forte sur  $\mathcal{P}^\circ$  si  $\varphi(x)$  remplit la condition de self-concordance sur  $\mathcal{P}^\circ$  et si les ensembles  $\{x \in \mathcal{P}^\circ \text{ tel que } \varphi(x) \leq t\}$  sont fermés dans  $R^n$  pour tout  $t \in R$ .

Cette nouvelle définition implique que la fonction  $\varphi(x)$  tend vers  $\infty$  lorsque  $x$  se rapproche de la frontière de  $\mathcal{P}^\circ$ .

Nous avons presque immédiatement le résultat suivant :

**Proposition :** soient  $\varphi_1(x)$  et  $\varphi_2(x)$  deux fonctions self-concordantes sur  $\mathcal{P}_1^\circ$  et  $\mathcal{P}_2^\circ$  respectivement et soient  $r_1$  et  $r_2$  deux nombres positifs.  
Alors  $\varphi(x) = r_1\varphi_1(x) + r_2\varphi_2(x)$  est self-concordante sur  $\mathcal{P}^\circ = \mathcal{P}_1^\circ \cap \mathcal{P}_2^\circ$ .

Nous allons désormais étudier la méthode utilisée pour atteindre l'optimum du problème (PC).

## 2. La méthode utilisée .

Voyons, en premier lieu, quelles sont les hypothèses émises pour assurer la convergence de la méthode.

### A. Hypothèses.

Nous supposons que  $\mathcal{P}^\circ$  est non-vide et borné . Nous supposons également que pour  $\lambda > \lambda^*$  la fonction  $-\ln(\lambda - f(x))$  est self-concordante sur l'ensemble  $\{x \text{ tel que } f(x) < \lambda\}$ . Nous imposerons la même condition aux fonctions  $\varphi_i(x) = -\ln(-g_i(x))$  pour  $1 \leq i \leq m$ . Par conséquent, la fonction  $\varphi(x, \lambda) = -m \ln(\lambda - f(x)) + \sum_{i=1}^m \varphi_i(x)$  sera également self-concordante. En fait, par construction, la fonction  $\varphi(x, \lambda)$  vérifie la condition de self-concordance forte.

Sans perte de généralité, nous supposerons également que la fonction objectif  $f(x)$  est linéaire, c'est-à-dire  $f: R^n \rightarrow R : x \rightsquigarrow c^t x$ . En effet, si  $f(x)$  n'est pas linéaire, il suffira de remplacer le problème (PC) par le problème :

$$\begin{cases} \min & t \\ \text{s.c.} & f(x) \leq t \\ & g_i(x) \leq 0 \quad i = 1, \dots, m \end{cases}$$

Dans ce cas, nous imposerons à la fonction  $-\ln(-f(x) + t)$  d'être self-concordante sur le domaine  $\{(x, t) \text{ tel que } x \in \mathcal{P}^\circ \text{ et } f(x) \leq t\}$

### B. La méthode .

Pour  $\lambda > \lambda^*$ , notons  $\mathcal{P}(\lambda) = \mathcal{P} \cap \{x \in R^n \text{ tel que } f(x) \leq \lambda\}$  avec  $\mathcal{P}(+\infty) = \mathcal{P}$ .

Nous remarquons directement que lorsque  $\lambda \xrightarrow{\downarrow} \lambda^*$ ,  $\mathcal{P}(\lambda)$  décroît et

$$\inf_{\lambda > \lambda^*} \mathcal{P}(\lambda) = \mathcal{P} \cap \{x \text{ tel que } f(x) = \lambda^*\}.$$

Par conséquent, nous allons suivre le chemin  $\lambda : +\infty \rightarrow \lambda^*$  et nous ferons correspondre à chaque  $\bar{\lambda}$  un point intérieur  $x(\bar{\lambda})$  de  $\mathcal{P}(\bar{\lambda})$ . Ce point  $x(\bar{\lambda})$  sera choisi comme le centre analytique de  $\mathcal{P}(\bar{\lambda})$ .

Définition 3 : soit  $+\infty \geq \lambda > \lambda^*$ .

On appelle centre analytique de  $\mathcal{P}(\lambda)$  l'unique point  $x(\lambda)$  de  $\mathcal{P}^\circ(\lambda)$  minimisant la fonction potentielle

$$\varphi(x, \lambda) : \mathcal{P}(\lambda) \rightarrow R : x \rightsquigarrow \varphi(x, \lambda) = -m \ln(\lambda - f(x)) - \sum_{i=1}^m \ln(-g_i(x))$$

$$\text{où } \varphi(x, \infty) = -\sum_{i=1}^m \ln(-g_i(x)).$$

Remarque :  $\forall \lambda$ ,  $x(\lambda)$  est unique car  $\varphi(x, \lambda)$  est fortement self-concordante et donc strictement convexe. (voir annexe pour démonstration)

Comme nous allons le découvrir, l'approximation de Newton dans la recherche du centre analytique  $x(\lambda)$  va s'avérer très efficace.

### C. La méthode de Newton.

Dans ce paragraphe, nous fournirons, sans démonstration, des résultats traitant de la convergence de la méthode de Newton dans l'approximation du centre analytique  $\bar{x}$  d'un ensemble quelconque  $\mathcal{P}$ . Nous donnerons la vitesse de convergence de la méthode comme fonction de constantes dépendant uniquement de  $\alpha$ , le paramètre de self-concordance.

La norme utilisée est la H-norme, c'est-à-dire la norme définie  $\forall h \in R^n$  par  $\|h\|_H = (h^t H h)^{\frac{1}{2}}$ .

Lemme 9 : soit  $\varphi$  une fonction fortement self-concordante de paramètre  $\alpha$  définie sur un ensemble  $\mathcal{P}^\circ$ .

Pour  $x \in \mathcal{P}^\circ$ , on définit :  $H(x) = \nabla^2 \varphi(x)$  et  $g(x) = \nabla \varphi(x)$ .

Soit  $h = h(x) = -H(x) \cdot g(x)$  le pas de Newton de départ au point  $x$  pour rechercher le centre analytique  $\bar{x}$  de  $\mathcal{P}$ .

Soit également  $\tilde{h}$  le pas de Newton au point  $x+h$ .

Définissons  $\delta = \|h\|_{H(x)}$  et  $\tilde{\delta} = \|\tilde{h}\|_{H(x+h)}$ .

$$\text{Si } \delta < \frac{1}{\sqrt{\alpha}},$$

alors  $x+h \in \mathcal{P}^\circ$

$$\text{et } \tilde{\delta} \leq \frac{\sqrt{\alpha}}{(1 - \sqrt{\alpha} \delta)^2} \delta^2$$

Par conséquent, si  $\delta < \frac{(3 - \sqrt{5})}{2\sqrt{\alpha}}$ , la méthode de Newton converge (i.e.  $\tilde{\delta} < \delta$ )

et pour  $\delta \leq \frac{1}{4\sqrt{\alpha}}$ , on a  $\tilde{\delta} \leq \frac{16}{9}\sqrt{\alpha}\delta^2$  ce qui assure la convergence quadratique.

Ce lemme affirme que la méthode de Newton converge quadratiquement lorsque le pas de Newton de départ est assez petit. Inversément, le résultat suivant garantit que, si un point  $x$  est suffisamment proche du centre analytique de  $\bar{x}$ , alors la méthode converge encore.

Lemme 10 : si la longueur  $\delta = \|h\|_{H(x)}$  du pas de Newton de départ au point  $x$  pour la recherche du centre analytique  $\bar{x}$  de  $\mathcal{P}$  est telle que  $\delta \leq \frac{1}{9\sqrt{\alpha}}$ ,

$$\text{alors} \quad \|(x+h) - \bar{x}\|_{H(x)} \leq \frac{3}{2}\sqrt{\alpha}\delta^2.$$

□

Dans les paragraphes suivants, nous appliquerons ces résultats à la fonction potentielle  $\varphi(x, \lambda)$  pour analyser une méthode "à pas courts" suivant la trajectoire des centres analytiques.

### 3. L'algorithme à pas courts.

Initialisation : Soit  $x_o \in \mathcal{P}^o$  et  $\lambda_o > \lambda^*$  tels que  $\|h_o\|_{H(x_o, \lambda_o)} = \|h(x_o, \lambda_o)\|_{H(x_o, \lambda_o)} \leq \frac{1}{4\sqrt{\alpha}}$ .

Soit  $\varepsilon$  la précision désirée.

Pas 1 : Poser  $k=0$

$$\sigma = \frac{1}{8\sqrt{m\alpha}}.$$

Pas 2 : Calculer  $h_k = h(x_k, \lambda_k) = -\nabla^2 \varphi(x_k, \lambda_k) \cdot \nabla \varphi(x_k, \lambda_k)$ .

Poser  $x_{k+1} = x_k + h_k$ .

Si  $\lambda_k - c^t x_{k+1} \leq \frac{1}{2} \varepsilon$  alors STOP

sinon effectuer le pas 3.

Pas 3 : Calculer  $\lambda_{k+1} = \lambda_k - \sigma(\lambda_k - c^t x_{k+1})$ .

Poser  $k = k + 1$ .

Revenir au pas 2.

#### 4. Analyse de convergence.

Afin de démontrer la convergence de l'algorithme à petits pas, nous allons prouver :

1) après la mise à jour de  $\lambda_{k+1}$ , l'itéré  $x_{k+1}$  vérifie encore

$$\|h(x_{k+1}, \lambda_{k+1})\|_{H(x_{k+1}, \lambda_{k+1})} \leq \frac{1}{4\sqrt{\alpha}}.$$

2)  $\lambda_k - c^t x_{k+1} \geq \frac{17}{23}(c^t x_{k+1} - \lambda^*)$ .

Par le lemme 9, le point 1) assure que l'itéré  $x_{k+1}$  reste admissible et proche du centre analytique de  $\varphi(x, \lambda_{k+1})$ . Le point 2) garantit la décroissance de l'écart  $c^t x_{k+1} - \lambda^*$ .

Pour démontrer le point 1), nous avons besoin des lemmes suivants :

Lemme 11 : Soit  $g(x, \lambda)$  le gradient de  $\varphi(x, \lambda)$ .

$$\text{On a : } g(x_{k+1}, \lambda_{k+1}) = g(x_{k+1}, \lambda_k) + m \frac{\sigma c}{\lambda_{k+1} - c^t x_{k+1}}.$$

$$\triangleq g(x_{k+1}, \lambda_k) + m \sigma \tilde{c}.$$

Preuve :

$$g(x, \lambda) = m \frac{c}{\lambda - c^t x} + \sum_{i=1}^m \frac{\nabla g_i(x)}{-g_i(x)}.$$

$$\text{Donc, } g(x_{k+1}, \lambda_{k+1}) = m \frac{c}{\lambda_{k+1} - c^t x_{k+1}} + \sum_{i=1}^m \frac{\nabla g_i(x_{k+1})}{-g_i(x_{k+1})}.$$

$$= m \frac{c}{(1-\sigma)(\lambda_k - c^t x_{k+1})} + \sum_{i=1}^m \frac{\nabla g_i(x_{k+1})}{-g_i(x_{k+1})} \quad \text{par définition de } \lambda_{k+1}.$$

$$\begin{aligned} \text{Or, } m \frac{c}{\lambda_k - c^t x_{k+1}} - m \frac{c}{(1-\sigma)(\lambda_k - c^t x_{k+1})} &= \frac{(1-\sigma)mc - mc}{(1-\sigma)(\lambda_k - c^t x_{k+1})} \\ &= \frac{-\sigma mc}{(1-\sigma)(\lambda_k - c^t x_{k+1})} \\ &= \frac{-\sigma mc}{\lambda_{k+1} - c^t x_{k+1}}. \end{aligned}$$

$$\text{Donc, } m \frac{c}{(1-\sigma)(\lambda_k - c^t x_{k+1})} = m \frac{c}{\lambda_k - c^t x_{k+1}} + \frac{\sigma mc}{\lambda_{k+1} - c^t x_{k+1}}.$$

$$\begin{aligned} \text{et } g(x_{k+1}, \lambda_{k+1}) &= g(x_{k+1}, \lambda_k) + \frac{m\sigma c}{\lambda_{k+1} - c^t x_{k+1}}, \\ &= g(x_{k+1}, \lambda_k) + m\sigma \tilde{c}. \end{aligned}$$

□

Lemme 12 : Soit  $H(x, \lambda)$  le hessien de  $\varphi(x, \lambda)$ .

$$\text{On a : } H(x_{k+1}, \lambda_{k+1}) = H(x_{k+1}, \infty) + m\tilde{c}\tilde{c}^t.$$

Preuve :

Par définition de  $\varphi(x, \lambda)$ ,

$$\text{si } \lambda = +\infty, \quad \varphi(x, \lambda) = -\sum_{i=1}^m \ln(-g_i(x))$$

$$\text{et si } \lambda \neq +\infty, \quad \varphi(x, \lambda) = -m \ln(\lambda - c^t x) - \sum_{i=1}^m \ln(-g_i(x)).$$

$$\text{On obtient donc } g(x, \lambda) = m \frac{c}{\lambda - c^t x} + \sum_{i=1}^m \frac{\nabla g_i(x)}{-g_i(x)} \quad \text{si } \lambda \neq +\infty.$$

$$g(x, \lambda) = \sum_{i=1}^m \frac{\nabla g_i(x)}{-g_i(x)} \quad ; \text{ si } \lambda = +\infty.$$

$$\text{Par conséquent, pour } \lambda \neq \infty, \quad H(x, \lambda) = H(x, \infty) + \begin{pmatrix} \frac{mc_1^2}{(\lambda - c^t x)^2} & \dots & \dots & \frac{mc_1 c_n}{(\lambda - c^t x)^2} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{mc_n c_1}{(\lambda - c^t x)^2} & \dots & \dots & \frac{mc_n^2}{(\lambda - c^t x)^2} \end{pmatrix}$$



$$= H(x, \infty) + \frac{m}{(\lambda - c^t x)^2} c c^t$$

Donc,  $H(x_{k+1}, \lambda_{k+1}) = H(x_{k+1}, \infty) + m \tilde{c} \tilde{c}^t$  par définition de  $\tilde{c}$ .

□

Lemme 13 :  $\|g(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_{k+1})^{-1}} \leq \|g(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_k)^{-1}}$ .

Preuve :

Soit  $A$  une matrice  $(n, m)$  et  $a$  un vecteur de  $R^n$ .

Si  $1 + a^t A^{-1} a \neq 0$ , alors  $(A + a a^t)^{-1} = A^{-1} - \frac{A^{-1} a a^t A^{-1}}{1 + a^t A^{-1} a}$ .

On a :  $\|g(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_{k+1})^{-1}} \stackrel{\Delta}{=} \|g\|_{H(x_{k+1}, \lambda_{k+1})^{-1}}$ .

$= g^t H(x_{k+1}, \lambda_{k+1})^{-1} g$  par définition de la  $H$ -norme

De plus, puisque, par le lemme 12,  $H(x_{k+1}, \lambda_{k+1}) = H(x_{k+1}, +\infty) + m \tilde{c} \tilde{c}^t$

et  $H(x_{k+1}, \lambda_k) = H(x_{k+1}, +\infty) + \frac{m}{(\lambda_k - c^t x_{k+1})^2} c c^t$ ,

on a immédiatement par définition de  $\tilde{c}$  et de  $\lambda_{k+1}$  :

$H(x_{k+1}, \lambda_{k+1}) = H(x_{k+1}, \lambda_k) + m \sigma^2 \tilde{c} \tilde{c}^t$ .

Donc,  $\|g\|_{H(x_{k+1}, \lambda_{k+1})^{-1}} = g^t [H(x_{k+1}, \lambda_k) + m \sigma^2 \tilde{c} \tilde{c}^t]^{-1} g$

$$= g^t H(x_{k+1}, \lambda_k)^{-1} g - m \sigma^2 g^t \frac{H(x_{k+1}, \lambda_k)^{-1} \tilde{c} \tilde{c}^t H(x_{k+1}, \lambda_k)^{-1}}{1 + \tilde{c}^t H(x_{k+1}, \lambda_k)^{-1} \tilde{c}} g$$

$$= g^t H(x_{k+1}, \lambda_k)^{-1} g - m \sigma^2 \frac{\|g^t H(x_{k+1}, \lambda_k)^{-1} \tilde{c}\|^2}{1 + \tilde{c}^t H(x_{k+1}, \lambda_k)^{-1} \tilde{c}}$$

Puisque  $H(x, \infty)$  est définie positive, on sait que  $H(x, \lambda) = H(x, \infty) + \frac{m}{(\lambda - c^t x)^2} c c^t$

est définie positive et donc  $1 + \tilde{c}^t H(x_{k+1}, \lambda_k)^{-1} \tilde{c} > 0$ .

Par conséquent,  $\|g\|_{H(x_{k+1}, \lambda_{k+1})^{-1}} \leq g^t H(x_{k+1}, \lambda_k)^{-1} g$ .

$$= \|g\|_{H(x_{k+1}, \lambda_k)^{-1}}$$

□

Preuve de 1) :

Soit l'algorithme à l'itération  $k$ .

Supposons que  $x_k$  satisfait  $\delta = \|h_k\|_{H(x_k, \lambda_k)} \leq \frac{1}{4\sqrt{\alpha}}$ .

Soit  $x_{k+1} = x_k + h_k$ .

et montrons que  $\|h_{k+1}\|_{H(x_{k+1}, \lambda_{k+1})} \leq \frac{1}{4\sqrt{\alpha}}$ .

Par le lemme 9 qui assure la convergence de la méthode de Newton, nous savons que

$$\begin{aligned} \tilde{\delta} &= \|h(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_k)} \leq \frac{16}{9}\delta^2 \\ &\leq \frac{1}{9\sqrt{\alpha}} \quad \text{car} \quad \delta \leq \frac{1}{4\sqrt{\alpha}} \end{aligned} \quad (1).$$

Examinons l'effet sur  $h$  et sur  $H$  de la mise à jour de  $\lambda_k$ .

Par le lemme 11, nous savons déjà que  $g(x_{k+1}, \lambda_{k+1}) = g(x_{k+1}, \lambda_k) + m\sigma\tilde{c}$ .

On obtient :  $\|h_{k+1}\|_{H(x_{k+1}, \lambda_{k+1})} = [h_{k+1}' H(x_{k+1}, \lambda_{k+1}) h_{k+1}]^{\frac{1}{2}}$

par définition de la  $H$ -norme

$$\begin{aligned} &\stackrel{\Delta}{=} [h_{k+1}' H_{k+1} h_{k+1}]^{\frac{1}{2}} \\ &= \left[ \left( -H_{k+1}^{-1} g_{k+1} \right)' H_{k+1} \left( -H_{k+1} g_{k+1} \right) \right]^{\frac{1}{2}} \end{aligned}$$

par définition du pas de Newton

$$\begin{aligned} &= [g_{k+1}' H_{k+1}^{-1} g_{k+1}]^{\frac{1}{2}} \\ &= \|g_{k+1}\|_{H(x_{k+1}, \lambda_{k+1})^{-1}} \\ &\leq \|g(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_{k+1})^{-1}} + \|m\sigma\tilde{c}\|_{H(x_{k+1}, \lambda_{k+1})} \end{aligned}$$

par le lemme 11

$$\leq \|g(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_k)^{-1}} + \|m\sigma\tilde{c}\|_{H(x_{k+1}, \lambda_{k+1})}$$

par le lemme 13.

Par la propriété : si  $h$  est définie positive, alors  $\tilde{c}'(H + m\tilde{c}\tilde{c}')^{-1}\tilde{c} \leq \frac{1}{m}$ , nous

pouvons borner  $\|m\sigma\tilde{c}\|_{h(x_{k+1}, \lambda_{k+1})}$  par :

$$\begin{aligned} \|m\sigma\tilde{c}\|_{h(x_{k+1}, \lambda_{k+1})}^{-1} &= [m\sigma\tilde{c}'H(x_{k+1}, \lambda_{k+1})^{-1}m\sigma\tilde{c}]^{\frac{1}{2}} \\ &= \sigma m [\tilde{c}'(H(x_{k+1}, \infty) + m\tilde{c}\tilde{c}')^{-1}\tilde{c}]^{\frac{1}{2}} \quad \text{par le lemme 12} \\ &\leq \sigma\sqrt{m} . \end{aligned}$$

$$\begin{aligned} \text{Donc, } \|h_{k+1}\|_{H(x_{k+1}, \lambda_{k+1})} &\leq \|g(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_k)}^{-1} + \sigma\sqrt{m} \\ &= \|h(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_k)} + \sigma\sqrt{m} \\ &\leq \frac{1}{9\sqrt{\alpha}} + \sigma\sqrt{m} \quad \text{par (1)} \\ &= \frac{1}{9\sqrt{\alpha}} + \frac{\sqrt{m}}{8\sqrt{m}\sqrt{\alpha}} \quad \text{par définition de } \sigma \\ &\leq \frac{1}{4\sqrt{\alpha}} . \end{aligned}$$

□

Nous avons besoin , pour démontrer la partie 2) , des deux lemmes suivants :

Lemme 14 : (approximation ellipsoïdale d'un ensemble  $\mathcal{P}^\circ$ ) .

Soit  $\varphi : \mathcal{P}^\circ \rightarrow R$  une fonction vérifiant la condition de self-concordance forte de paramètre  $\alpha$  .

Soit  $x \in \mathcal{P}^\circ$  et  $h, z$  deux vecteurs quelconques de  $R^n$  .

Posons  $\delta = \sqrt{h'H(x)h} = \|h\|_{H(x)}$  où  $H(x) = \nabla^2\varphi(x)$ .

Si  $\delta < \frac{1}{\sqrt{\alpha}}$  ,

alors  $\bar{x} = x + h \in \mathcal{P}^\circ$  et  $\|z\|_{H(\bar{x})} \leq \frac{1}{(1 - \sqrt{\alpha}\delta)} \|z\|_{H(x)}$  .

Lemme 15 : Soit  $E_r = \{x \text{ tel que } (x - \bar{x})^t H(x - \bar{x}) \leq r^2\}$ .

où  $H$  est une matrice définitive positive symétrique.

$$\text{Alors, } \max_{x \in E_r} c^t x - c^t \bar{x} = r \sqrt{c^t H^{-1} c}.$$

Preuve :

Ecrivons les équations de K-T pour ce problème :

$$\begin{cases} c - 2uH(x - \bar{x}) = 0 \\ u \geq 0 \\ u[r^2 - (x - \bar{x})^t H(x - \bar{x})] = 0 \\ (x - \bar{x})^t H(x - \bar{x}) \leq r^2 \end{cases}$$

Par la première équation, nous obtenons :  $u \neq 0$  et  $x - \bar{x} = \frac{H^{-1}c}{2u}$ .

Utilisant ce résultat dans la quatrième équation et sachant, par la troisième équation, que le point maximum se trouve sur la frontière de l'ellipsoïde, nous cherchons :

$$\frac{(H^{-1}c)^t H H^{-1} \cdot c}{u^2} = r^2$$

$$\text{c'est-à-dire : } u = \frac{\sqrt{c^t H^{-1} c}}{2r}$$

$$\text{Donc, } x - \bar{x} = \frac{H^{-1}c}{2u}$$

$$= r \frac{H^{-1}c}{\sqrt{c^t H^{-1} c}}$$

$$\text{et } c^t x - c^t \bar{x} = r \sqrt{c^t H^{-1} c}.$$

□

Preuve de 2) :

Soit  $\bar{x}$  le centre analytique de  $\mathcal{P}(\lambda_k)$ .

$$\text{On a : } \|x_{k+1} - \bar{x}\|_{H(x_{k+1}, \lambda_k)} \leq \|h(x_{k+1}, \lambda_k)\|_{H(x_{k+1}, \lambda_k)}$$

$$+ \|x_{k+1} + h(x_{k+1}, \lambda_k) - \bar{x}\|_{H(x_{k+1}, \lambda_k)}$$

$$\begin{aligned} &\leq \frac{1}{9\sqrt{\alpha}} + \frac{3\sqrt{\alpha}}{2} \left( \frac{1}{9\sqrt{\alpha}} \right)^2 \text{ par le lemme 10} \\ &\leq \frac{7}{54\sqrt{\alpha}}. \end{aligned} \quad (1)$$

Par le lemme 14, nous savons que :

$$\begin{aligned} \|x_{k+1} - \bar{x}\|_{H(\bar{x}, \lambda_k)} &\leq \frac{1}{1 - \sqrt{\alpha} \delta} \|x_{k+1} - \bar{x}\|_{H(x_{k+1}, \lambda_k)} \\ &\leq \frac{1}{1 - \sqrt{\alpha}} \cdot \frac{7}{54\sqrt{\alpha}} \quad \text{par (1)} \\ &\leq \frac{1}{1 - \frac{7}{54}} \cdot \frac{7}{54\sqrt{\alpha}} \quad \text{par (1)} \\ &\leq 0.15 \frac{1}{\sqrt{\alpha}}. \end{aligned}$$

$$\begin{aligned} \text{Par conséquent, } c^t x_{k+1} - c^t \bar{x} &\leq \max_{\|x - \bar{x}\|_{H(\bar{x})} \leq \frac{0.15}{\sqrt{\alpha}}} (c^t x - c^t \bar{x}) \\ &= 0.15 \max_{\|x - \bar{x}\|_{H(\bar{x})} \leq \frac{1}{\sqrt{\alpha}}} (c^t x - c^t \bar{x}) \quad \text{par le lemme 15} \\ &\leq 0.15 \max_{x \in P(\lambda_k)} (c^t x - c^t \bar{x}) \end{aligned}$$

car, par le lemme 14,  $\left\{ x \text{ tel que } \|x - \bar{x}\|_{H(\bar{x})} < \frac{1}{\sqrt{\alpha}} \right\} \subset \mathcal{P}(\lambda_k)$ .

Par définition de  $\mathcal{P}(\lambda_k)$ , on a aussi :  $c^t x - c^t \bar{x} \leq \lambda_k - c^t \bar{x} \quad \forall x \in \mathcal{P}(\lambda_k)$ ,

ce qui entraîne que :  $c^t x_{k+1} - c^t \bar{x} \leq 0.15(\lambda_k - c^t \bar{x}) \quad (2)$ .

Puisque les fonctions  $g_i(x)$  sont convexes et puisque  $f(x)$  est linéaire, on a :

$$\lambda_k - c^t \bar{x} \geq \frac{1}{2}(\lambda_k - \lambda^*) \quad (3)$$

(2) peut se réécrire :  $-c^t x_{k+1} + 0.85c^t \bar{x} \geq -0.15\lambda_k$

et (3) :  $0.85\lambda_k - 0.85c^t \bar{x} \geq \frac{0.85}{2}(\lambda_k - \lambda^*)$ .

En additionnant ces deux dernières inégalités, nous obtenons :

$$\lambda_k - c^t x_{k+1} \geq \frac{0.85}{2}(\lambda_k - \lambda^*)$$

et donc :  $\lambda_k - c^t x_{k+1} \geq \frac{0.85}{2}(c^t x_{k+1} - \lambda^*)$  car  $\lambda_k \geq c^t x_{k+1}$ ,

ce qui démontre la thèse .

□

## Chapitre 2 :

### Algorithmes à pas longs.

En pratique, les algorithmes à pas longs sont les plus intéressants car ils permettent une implémentation meilleure . Nous allons présenter deux de ces algorithmes et obtenir pour chacun leur taux de convergence.

#### 1. Algorithme à pas longs se basant sur la fonction potentielle.

##### A. Le problème .

Soit le problème d'optimisation convexe suivant :

$$(CP) \text{ trouver } \lambda^* = \min_{x \in P} f(x)$$

où  $\mathcal{P} = \{x \in R^n \text{ tel que } g_i(x) \leq 0 \text{ pour } 1 \leq i \leq m\}$

et où les fonctions  $f$  et  $g_i$   $1 \leq i \leq m$ , sont supposées convexes et deux fois continûment différentiables sur  $\mathcal{P}$  .

Nous supposons également que  $\mathcal{P}^0 = \{x \in R^n \text{ tel que } g_i(x) < 0 \text{ pour } 1 \leq i \leq m\}$  est non vide et borné .

Le dual de Wolfe du problème (CP) s'écrit :

$$(D) \begin{cases} \max & f(x) + \sum_{i=1}^m y_i g_i(x) \\ \text{s.c.} & \nabla f(x) + \sum_{i=1}^m y_i \nabla g_i(x) = 0 \\ & y \geq 0 \end{cases}$$

Par le théorème de dualité faible, si  $\bar{x} \in \mathcal{P}$  et  $(x, y)$  est un point dual-admissible, alors :

$$f(\bar{x}) \geq f(x) + \sum_{i=1}^m y_i g_i(x) .$$

De plus, puisque  $\mathcal{P}^0$  est non-vidé, la condition de Slater est vérifiée et donc, par le théorème de dualité de Wolfe, si (CP) a une solution optimale, (D) en possède une également et les valeurs extrêmes sont égales.

Ces résultats de base étant acquis, nous procéderons désormais de façon identique à ce qui a été établi dans le cadre de l'algorithme à pas courts.

Nous associons au programme (CP) la même fonction potentielle  $\varphi(x, \lambda) = -m \ln(\lambda - f(x)) - \sum_{i=1}^m \ln(-g_i(x))$  définie sur l'ensemble  $\mathcal{P}(\lambda)^0$  où  $+\infty \geq \lambda > \lambda^*$  et où  $\mathcal{P}(\lambda) = \mathcal{P} \cap \{x \in R^n \text{ tel que } f(x) \leq \lambda\}$ .

Posons  $H(x, \lambda) \triangleq \nabla^2 \varphi(x, \lambda)$  et  $g(x, \lambda) = \nabla \varphi(x, \lambda)$ .

Comme nous le démontrons en annexe, la condition relative de Lipschitz (RLC) imposée aux fonctions  $g_i(x)$  ( $1 \leq i \leq m$ ) garantit le caractère strictement convexe de  $\varphi(x, \lambda)$  sur  $\mathcal{P}^0(\lambda)$ . Cette condition supplémentaire est formulée de la manière suivante :

$$\exists M > 0 \quad : \quad \forall v \in R^n, \forall x \text{ et } x+h \in R^n, \text{ on a :}$$

$$|v'[\nabla^2 g_i(x+h) - \nabla^2 g_i(x)]v| \leq M \|h\|_H v' \nabla^2 g_i(x) v.$$

où  $H$  représente  $\nabla^2 g_i(x)$

$$\text{et } \|h\|_H = \sqrt{h' H h}.$$

Puisque  $\varphi(x, \lambda)$  est strictement convexe sur son domaine  $\mathcal{P}(\lambda)^0$ , elle atteint un minimum unique noté  $x(\lambda)$ ; ce point a déjà été défini comme le centre analytique de  $\mathcal{P}(\lambda)$ . Il vérifie les conditions de Kuhn-Tucker du problème :

$$\begin{cases} \min \varphi(x, \lambda) \\ \text{s.c. } g_i(x) \leq 0 \quad i = 1, \dots, m \end{cases}$$

Ces conditions sont :

$$(*) \quad \begin{cases} g_i(x) \leq 0 & 1 \leq i \leq m \\ \nabla f(x) + \sum y_i \nabla g_i(x) = 0 \\ y_i \geq 0 & 1 \leq i \leq m \\ \frac{\lambda - f(x)}{m} = -y_i g_i(x) & 1 \leq i \leq m \end{cases}$$

Ces résultats préliminaires établis, nous pouvons dès à présent décrire l'algorithme à pas longs et établir sa convergence vers la solution optimale de (CP).

### B. Description de l'algorithme.

L'algorithme ci-dessus suit la trajectoire centrale, c'est-à-dire le chemin des centres analytiques  $x(\lambda)$ . Faisant varier le paramètre  $\lambda$  de  $+\infty$  à  $\lambda^*$ , nous minimiserons  $\varphi(x, \lambda)$  à chaque itération en effectuant un pas de Newton et éventuellement une recherche linéaire le



long de la direction de Newton si le pas est trop long, c'est-à-dire si le point engendré est trop éloigné de la trajectoire.

La direction de Newton  $h(x, \lambda)$  associée à la fonction  $\varphi(x, \lambda)$  est obtenue par la formule :

$$h(x, \lambda) \triangleq h = -H(x, \lambda)^{-1} g(x, \lambda) \triangleq H^{-1} g .$$

Le critère de proximité par rapport à la trajectoire centrale sera :

$\|h\|_H \leq \tau$  où  $\tau$  est un paramètre de tolérance . Remarquons que  $\|h\|_H = 0$  si et seulement si  $x = x(\lambda)$  .

La valeur  $\frac{1}{8(1+2\mu)}$  pour  $\tau$  s'avèrera appropriée dans l'analyse de convergence .

Initialisation: Soit  $x_0 \in \mathcal{P}^0$

$$\lambda_0 \in \mathcal{P}^0 \text{ tel que } \lambda_0 > f(x) , \lambda_0 - \lambda^* \leq \frac{1}{\varepsilon} , \|h_0\|_H \leq \tau$$

où  $\varepsilon$  est le seuil de précision de l'algorithme

$$\tau = \frac{1}{8(1+2\mu)}$$

Soit  $0 < \vartheta < 1$  le facteur de mise à jour de l'algorithme et  $t \in N$  un paramètre d'exactitude .

Pas 1 : poser  $x = x_0$

$$\lambda = \lambda_0$$

Pas 2 : tant que  $\lambda - f(x) > e^t$

(début de l'approche de la trajectoire centrale)

tant que  $\|h\|_H > \tau$

(début de la recherche linéaire)

$$\text{calculer } \bar{\alpha} = \min_{\alpha > 0} (\varphi(x + \alpha h, \lambda) : x + \alpha h \in \mathcal{P}(\lambda)^0)$$

$$\text{mettre à jour } x = x + \bar{\alpha} h$$

(fin de la recherche linéaire)

mettre à jour  $\lambda = \lambda - \mathfrak{g}(\lambda - f(x))$

(fin de l'approche de la trajectoire centrale).

### C. Analyse de convergence .

Avant de démontrer les théorèmes de convergence, nous allons fournir la preuve de quelques lemmes préliminaires .

Nous noterons  $q_x(y, \lambda)$  l'approximation quadratique de  $\varphi(x, \lambda)$  lorsque  $y$  est proche de  $x$  , c'est-à-dire :

$$q_x(y, \lambda) = \varphi(x, \lambda) + g'(y-x) + \frac{1}{2}(y-x)'H(y-x) .$$

Lemme 16 : Si les fonctions  $g_i(x)$  satisfont la condition RLC avec la constante de Lipschitz  $M > 0$  et si  $x \in \mathcal{P}(\lambda)^0$  et  $\|d\|_H < \min\left(\frac{1}{2}, \frac{1}{2M^3}\right)$ ,

alors,  $x+d \in \mathcal{P}(\lambda)^0$

$$\text{et } |\varphi(x+d, \lambda) - q_x(x+d, \lambda)| \leq \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M)$$

Preuve :

Nous ne démontrerons ce lemme que dans le cas où les fonctions sont linéaires ou quadratiques.

Développons  $\varphi(x+d, \lambda)$  en série de Taylor autour de  $x$  . Nous obtenons, puisque les fonctions  $f(x)$  et  $g_i(x)$ ,  $1 \leq i \leq m$ , sont linéaires ou quadratiques :

$$\varphi(x+d, \lambda) = q_x(x+d, \lambda) + \sum_{i=3}^{+\infty} t_i \quad (1)$$

où le terme  $t_i$  représente le terme d'ordre  $i$  de la série de Taylor .

Par construction,  $\varphi(x, \lambda)$  ne prend que des valeurs finies sur son domaine  $\mathcal{P}(\lambda)^0$  .

Par conséquent, pour montrer que  $x+d \in \mathcal{P}(\lambda)^0$ , il suffit de montrer que  $\varphi(x+d, \lambda)$  est finie, c'est-à-dire, par la relation (1), que  $\sum_{i=3}^{+\infty} t_i$  converge pour  $d$  tel que  $\|d\|_H < 1$  .

Nous pouvons montrer que  $|t_i| \leq \frac{\|d\|_H^i}{i}$  pour  $i \geq 3$  .

La preuve de cette inégalité est assez technique et sera donc omise ici . Nous renvoyons le lecteur à DEN HERTOOG, ROOS et TERLAKY <sup>1</sup> pour une preuve complète de cette inégalité .

Nous obtenons dès lors :

$$\begin{aligned} \sum_{i=3}^{\infty} |t_i| &\leq \sum_{i=3}^{\infty} \frac{\|d\|_H^i}{i} \leq \frac{1}{3} \sum_{i=3}^{+\infty} \|d\|_H^i \\ &= \frac{1}{3} \|d\|_H^3 \sum_{i=0}^{+\infty} \|d\|_H^i \\ &= \frac{\|d\|_H^3}{3(1-\|d\|_H)} \end{aligned}$$

et donc,  $x + d \in \mathcal{P}(\lambda)^0$  .

De plus, par (1) , nous obtenons :

$$\varphi(x + d, \lambda) - q_x(x + d, \lambda) = \sum_{i=3}^{+\infty} t_i .$$

$$\text{D'où la thèse : } |\varphi(x + d, \lambda) - q_x(x + d, \lambda)| \leq \frac{\|d\|_H^3}{3(1-\|d\|_H)} .$$

□

lemme 17 : Si  $\|h\|_H < \frac{1}{8(1+2M)}$  ,

$$\text{alors } \|x - x(\lambda)\|_H \leq \frac{5}{2} \|h\|_H .$$

Preuve :

Soit  $p$  un point arbitraire tel que  $\|p\|_H = \frac{3}{2} \|h\|_H$  .

Considérons alors l'ellipsoïde  $E = \{x + h + p : \|p\|_H = \frac{3}{2} \|h\|_H\}$  .

$$\begin{aligned} \text{Nous avons : } \|p + h\|_H &\leq \|h\|_H + \|p\|_H \\ &= \frac{5}{2} \|h\|_H \end{aligned} \tag{2}$$

<sup>1</sup> DEN HERTOOG D, ROOS C, TERLAKY T : A large-step analytic center method for a class of smooth convex programming problems. *SIAM Journal on optimization*, 2(1992), pp 55-70.

$$\leq \frac{5}{16(1+2M)} \quad \text{par hypothèse .}$$

$$< \frac{1}{3}$$

Par conséquent, nous pouvons appliquer le lemme 16 avec  $d = h + p$  .

$$\begin{aligned} \text{Nous obtenons : } \varphi(x+h+p, \lambda) &\geq q_x(x+h+p, \lambda) - \frac{\|p+h\|_H^3}{3(1-\|p+h\|_H)}(1+2M) \\ &> q_x(x+h+p, \lambda) - \frac{\|p+h\|_H^3}{3(1-\frac{1}{3})}(1+2M) . \end{aligned}$$

Par définition du pas de Newton  $h$ , nous savons que  $x+h = \min_y q_x(y, \lambda)$  .

$$\text{Par conséquent, } q_x(x+h+p, \lambda) \geq q_x(x+h, \lambda) + \frac{1}{2}\|p\|_H^2$$

$$\begin{aligned} \text{Donc, } \varphi(x+h+p, \lambda) &> q_x(x+h, \lambda) + \frac{1}{2}\|p\|_H^2 - \frac{\|p+h\|_H^3}{2}(1+2M) \\ &\geq q_x(x+h, \lambda) + \frac{9}{8}\|h\|_H^2 - \frac{125}{16}\|h\|_H(1+2M) \\ &\quad \text{par (2) et par définition de } p . \\ &\geq q_x(x+h, \lambda) + 9\|h\|_H^3(1+2M) - \frac{125}{16}\|h\|_H(1+2M) \\ &\quad \text{par hypothèse .} \\ &> q_x(x+h, \lambda) + \|h\|_H^3(1+2M) . \end{aligned}$$

Par le lemme 16 , nous obtenons aussi :

$$\varphi(x+h, \lambda) \leq q_x(x+h, \lambda) + \frac{1}{2}\|h\|_H^3(1+2M) .$$

Regroupant ces deux inégalités, nous obtenons alors :

$$\varphi(x+h+p, \lambda) > \varphi(x+h, \lambda)$$

ce qui montre qu'au centre  $x+h$  de l'ellipsoïde  $E$ , la fonction potentielle atteint une valeur plus petite que sur le bord de  $E$  .

Par conséquent, puisque  $\varphi(x, \lambda)$  est strictement convexe, le minimum de  $\varphi(x, \lambda)$  est à l'intérieur de l'ellipsoïde  $E$  et donc,

$$\begin{aligned} \|x-x(\lambda)\|_H &\leq \|h+p\|_H \\ &\leq \frac{5}{2}\|h\|_H \quad \text{par (2) .} \end{aligned}$$

□

Lemme 18 : Si  $\|h\|_H \geq \frac{1}{8(1+2M)}$ ,

alors la diminution  $\Delta\varphi$  de de la fonction potentielle après une recherche linéaire le long de la direction de Newton  $h$  satisfait :

$$\Delta\varphi \geq \frac{1}{140(1+2M)^2}$$

Preuve :

Soit  $\mu$  tel que  $\|\mu h\|_H \leq \min\left(\frac{1}{2}, \frac{1}{2M^{\frac{1}{3}}}\right)$

Par le lemme 16, on a donc :

$$\varphi(x + \mu h, \lambda) \leq q_x(x + \mu h, \lambda) + \frac{\mu^3 \|h\|_H^3}{3(1 - \mu \|h\|_H)} (1 + 2M).$$

Or, par définition,

$$q_x(x + \mu h, \lambda) = \varphi(x, \lambda) + \mu g^t h + \frac{1}{2} \mu^2 h^t H h$$

$$\text{Donc, } \varphi(x, \lambda) - \varphi(x + \mu h, \lambda) \geq -\mu g^t h - \frac{1}{2} \mu^2 \|h\|_H^2 - \frac{\mu^3 \|h\|_H^3}{3(1 - \mu \|h\|_H)} (1 + 2M).$$

$$\text{Et puisque } h = -H^{-1}g, \quad -g = Hh.$$

$$\text{On obtient } \varphi(x, \lambda) - \varphi(x + \mu h, \lambda) \geq \mu \|h\|_H^2 - \frac{1}{2} \mu^2 \|h\|_H^2 - \frac{\mu^3 \|h\|_H^3}{3(1 - \mu \|h\|_H)} (1 + 2M).$$

$$\text{Posons } \mu = \frac{1}{9(1+2M)\|h\|_H}.$$

$$\text{On a que } \mu \text{ vérifie } \|\mu h\|_H \leq \min\left(\frac{1}{2}, \frac{1}{2M^{\frac{1}{3}}}\right).$$

De plus,

$$\begin{aligned} \varphi(x, \lambda) - \varphi(x + \mu h, \lambda) &\geq \frac{\|h\|_H}{9(1+2M)} - \frac{1}{162(1+2M)^2} - \frac{1}{9^3(1+2M)} \cdot \frac{1}{3(1+2M)-1} \\ &\geq \frac{1}{72(1+2M)^2} - \frac{1}{162(1+2M)^2} - \frac{1}{9^3(1+2M)} \cdot \frac{1}{3(1+2M)-1} \\ &\geq \frac{1}{140(1+2M)^2}. \end{aligned}$$

□

Lemme 19 : si  $\|h\|_H \leq \frac{1}{8(1+2M)}$ ,

alors  $\varphi(x, \lambda) - \varphi(x(\lambda), \lambda) \leq 4\|h\|_H^2$ .

Preuve :

Soit  $d = x(\lambda) - x$ .

Par le lemme 17, on a :  $\|d\|_H = \|x(\lambda) - x\|_H \leq \frac{5}{2}\|h\|_H$ .

Utilisons le lemme 16, nous obtenons :

$$\begin{aligned} \varphi(x(\lambda), \lambda) &\geq q_x(x(\lambda), \lambda) - \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \\ &= q_x(x+d, \lambda) - \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \\ &= \varphi(x, \lambda) + g^t d + \frac{1}{2}d^t H d - \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \quad \text{par définition de } q_x(y, \lambda) \\ &= \varphi(x, d) - h^t H d + \frac{1}{2}d^t H d - \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \quad \text{par définition de } g. \end{aligned}$$

Par l'inégalité de CAUCHY-SCHWARZ, nous avons :

$$-h^t H d \geq \|h\|_H \|d\|_H.$$

Par conséquent,

$$\begin{aligned} \varphi(x, \lambda) - \varphi(x(\lambda), \lambda) &\leq \|h\|_H \|d\|_H - \frac{1}{2}\|d\|_H^2 + \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \\ &\leq \|h\|_H \|d\|_H + \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \\ &\leq \|d\|_H \|h\|_H + \frac{\|d\|_H^3}{24(1-\|d\|_H)\|h\|_H} \quad \text{car } \|h\|_H \leq \frac{1}{8(1+2M)} \\ &\leq \frac{5}{2}\|h\|_H^2 + \frac{\frac{125}{8}\|h\|_H^3}{24(1-\frac{5}{16})\|h\|_H} \end{aligned}$$

$$\text{car } \|d\|_H \leq \frac{5}{2}\|h\|_H \leq \frac{5}{2} \frac{1}{8(1+2M)} \leq \frac{5}{16}.$$

$$\lambda - f(x(\lambda)) \leq \lambda - f(x) \leq \left(1 + \frac{\beta}{\sqrt{m}}\right)(\lambda - f(x)) .$$

Supposons donc que  $f(x(\lambda)) < f(x)$  .

Notons  $\varphi(x, \lambda) = \sum_{i=1}^{2m} \ln(-g_i(x))$  avec, pour  $m+1 \leq i \leq 2m$ ,  $g_i(x) = f(x) - \lambda$  .

$$\text{Nous pouvons écrire : } H = \nabla^2 \varphi(x, \lambda) = \sum_{i=1}^{2m} \left[ \frac{\nabla g_i(x) \nabla g_i(x)^t}{g_i(x)^2} + \frac{\nabla^2 g_i(x)}{-g_i(x)} \right] .$$

Par définition, nous avons :

$$\begin{aligned} \beta^2 &\geq \|x - x(\lambda)\|_H^2 \\ &= (x - x(\lambda))^t \left[ \sum_{i=1}^{2m} \left( \frac{\nabla g_i(x) \nabla g_i(x)^t}{g_i(x)^2} + \frac{\nabla^2 g_i(x)}{-g_i(x)} \right) \right] (x - x(\lambda)) \end{aligned}$$

où chaque terme de la somme est défini positif par la convexité des fonctions  $g_i(x)$ ,  $1 \leq i \leq 2m$  .

$$\begin{aligned} \text{Par conséquent, } \beta^2 &\geq (x - x(\lambda))^t \left[ \sum_{i=m+1}^{2m} \left( \frac{\nabla g_i(x) \nabla g_i(x)^t}{g_i(x)^2} \right) \right] (x - x(\lambda)) \\ &= m(x - x(\lambda))^t \frac{\nabla f(x) \nabla f(x)^t}{(f(x) - \lambda)^2} (x - x(\lambda)) . \end{aligned}$$

Or,  $f(x)$  est convexe .

$$\text{Donc, } f(x) - f(x(\lambda)) \geq (x - x(\lambda))^t \nabla f(x)$$

$$\text{On obtient alors : } \beta^2 \geq m \frac{[f(x) - f(x(\lambda))]^2}{(f(x) - \lambda)^2}$$

c'est-à-dire, puisque  $f(x) > f(x(\lambda))$  :

$$f(x) - f(x(\lambda)) \leq \frac{\beta}{\sqrt{m}} (\lambda - f(x))$$

$$\text{ou encore } \lambda - f(x) + f(x) - f(x(\lambda)) \leq \left(1 + \frac{\beta}{\sqrt{m}}\right) (\lambda - f(x)) .$$

D'où la thèse .

□

Lemme 22 : Si  $\|x - x(\lambda)\|_H \leq \beta$  ,  
 Alors  $\lambda - \lambda^* \leq 2 \left( 1 + \frac{\beta}{\sqrt{m}} \right) (\lambda - f(x))$  .

Preuve :

Le centre analytique  $x(\lambda)$  minimise la fonction potentielle  $\varphi(x, \lambda)$  .  
 Ce point  $x(\lambda)$  est déterminé par le système d'équation de K-T (\*) . De ce système, nous savons que :

$$\nabla f(x(\lambda)) + \sum_{i=1}^m y_i(\lambda) \nabla g_i(x(\lambda)) = 0$$

c'est-à-dire que  $(x(\lambda), y(\lambda))$  est dual-admissible .

Par conséquent,  $\lambda^* \geq f(x(\lambda)) + \sum_{i=1}^m y_i(\lambda) g_i(x(\lambda))$

et donc, 
$$f(x(\lambda)) - \lambda^* \leq - \sum_{i=1}^m y_i(\lambda) g_i(x(\lambda))$$

$$= \lambda - f(x(\lambda)) \text{ par le système (*)}$$

On a alors :  $(\lambda - \lambda^*) - [\lambda - f(x(\lambda))] \leq \lambda - f(x(\lambda))$

c'est-à-dire :  $\lambda - \lambda^* \leq 2[\lambda - f(x(\lambda))]$

et par le lemme 21 :  $\lambda - \lambda^* \leq 2 \left( 1 + \frac{\beta}{\sqrt{m}} \right) [\lambda - f(x)]$  .

□

Lemme 23 : L'écart  $\lambda - f(x(\lambda))$  décroît de façon monotone si  $\lambda > \lambda^*$  décroît .

Preuve :

$x(\lambda)$  et  $y(\lambda)$  remplissent les conditions de KUHN-TUCKER (\*) .

Donc,  $\nabla f(x(\lambda)) + \sum_{i=1}^m y_i(\lambda) \nabla g_i(x(\lambda)) = 0$

et 
$$-y_i g_i(x(\lambda)) = \frac{\lambda - f(x)}{m} \quad i = 1, \dots, m$$
 .

Dérivons ces équations par rapport à  $\lambda$  :



$$\begin{cases} \nabla^2 f(x(\lambda))x'(\lambda) + \sum_{i=1}^m y_i'(\lambda)\nabla g_i(x(\lambda)) + \sum_{i=1}^m y_i(\lambda)\nabla^2 g_i(x(\lambda))x'(\lambda) = 0 & (1) \\ -y_i'(\lambda)g_i(x(\lambda)) - y_i(\lambda)\nabla g_i(x(\lambda))x'(\lambda) = \frac{1 - \nabla f(x(\lambda))'x'(\lambda)}{m} & (2) \end{cases}$$

Multiplions (2) par  $y_i(\lambda)$ . On obtient :

$$-y_i'(\lambda)y_i(\lambda)g_i(x(\lambda)) - y_i^2(\lambda)\nabla g_i(x(\lambda))x'(\lambda) = \frac{1 - \nabla f(x(\lambda))'x'(\lambda)}{m}y_i(\lambda)$$

c'est-à-dire, puisque  $-y_i(\lambda)g_i(x(\lambda)) = \frac{\lambda - f(x(\lambda))}{m}$  par (\*) :

$$\frac{\lambda - f(x(\lambda))}{m}y_i'(\lambda) - y_i^2(\lambda)\nabla g_i(x(\lambda))x'(\lambda) = \frac{1 - \nabla f(x(\lambda))'x'(\lambda)}{m}y_i(\lambda).$$

Multiplions par  $\nabla g_i(x(\lambda))$  cette dernière équation et sommons sur  $i$  :

$$\begin{aligned} \frac{\lambda - f(x(\lambda))}{m} \sum_{i=1}^m \nabla g_i(x(\lambda))y_i'(\lambda) - \sum_{i=1}^m y_i^2(\lambda)\nabla g_i(x(\lambda))'x'(\lambda)\nabla g_i(x(\lambda)) \\ = \frac{1 - \nabla f(x(\lambda))'x'(\lambda)}{m} \sum_{i=1}^m \nabla g_i(x(\lambda))y_i(\lambda) \end{aligned}$$

et par (\*) :

$$\begin{aligned} \frac{\lambda - f(x(\lambda))}{m} \sum_{i=1}^m \nabla g_i(x(\lambda))y_i'(\lambda) - \sum_{i=1}^m y_i^2(\lambda)\nabla g_i(x(\lambda))'x'(\lambda)\nabla g_i(x(\lambda)) \\ = -\frac{1 - \nabla f(x(\lambda))'x'(\lambda)}{m} \nabla f(x(\lambda)). \end{aligned}$$

Par (1), on obtient alors :

$$\begin{aligned} -\frac{\lambda - f(x(\lambda))}{m} \left[ \sum_{i=1}^m y_i(\lambda)\nabla^2 g_i(x(\lambda))x'(\lambda) + \nabla^2 f(x(\lambda))x'(\lambda) \right] \\ - \sum_{i=1}^m y_i^2(\lambda)\nabla g_i(x(\lambda))'x'(\lambda)\nabla g_i(x(\lambda)) = -\frac{1 - \nabla f(x(\lambda))'x'(\lambda)}{m} \nabla f(x(\lambda)). \end{aligned}$$

Prenons le produit scalaire avec  $x'(\lambda)$  :

$$\begin{aligned} \frac{\nabla f(x(\lambda))'x'(\lambda) - 1}{m} \nabla f(x(\lambda))'x'(\lambda) = -\sum_{i=1}^m y_i^2(\lambda) [\nabla g_i(x(\lambda))'x'(\lambda)]^2 \\ - \frac{\lambda - f(x(\lambda))}{m} x'(\lambda) \left[ \sum_{i=1}^m y_i(\lambda)\nabla^2 g_i(x(\lambda)) + \nabla^2 f(x(\lambda)) \right] x'(\lambda) \end{aligned}$$

Puisque les fonctions  $f(x)$  et  $g_i(x)$ ,  $1 \leq i \leq m$ , sont convexes et puisque

$y_i(\lambda) \geq 0$ ,  $1 \leq i \leq m$ , on a :

$$\frac{\nabla f(x(\lambda))'x'(\lambda) - 1}{m} \nabla f(x(\lambda))'x'(\lambda) \leq 0 .$$

Par conséquent,  $0 \leq \nabla f(x(\lambda))'x'(\lambda) \leq 1$

ou encore  $1 - \nabla f(x(\lambda))'x'(\lambda) \geq 0$  .

Et donc ,  $\lambda - f(x(\lambda))$  est croissante en  $\lambda$  .

□

Nous pouvons désormais évaluer la convergence de l'algorithme . Cette analyse se fera grâce à deux théorèmes : le premier fournira une borne supérieure au nombre d'itérations extérieures, le deuxième une borne supérieure au nombre d'itérations intérieures de l'algorithme.

**Théorème 7 :** Soit  $\beta \leq \frac{5}{16(1+2M)}$ .

Alors après au plus  $K = \frac{2(1 + \frac{\beta}{\sqrt{m}})}{\vartheta} \mathcal{O}(|\ln \varepsilon|)$  itérations extérieures, l'algorithme converge vers une solution optimale à  $\varepsilon$ -près.

Preuve :

Soit  $\lambda^k$  la borne supérieure à l'itération extérieure  $k$ .

$$\begin{aligned} \text{On a : } \frac{\lambda^k - \lambda^*}{\lambda^{k-1} - \lambda^*} &\leq \frac{\lambda^{k-1} - \vartheta(\lambda^{k-1} - f(x^{k-1})) - \lambda^*}{\lambda^{k-1} - \lambda^*} \quad \text{par définition de } \lambda^k \\ &= 1 - \vartheta \frac{\lambda^{k-1} - f(x^{k-1})}{\lambda^{k-1} - \lambda^*}. \end{aligned}$$

$$\text{Par hypothèse, } \beta = \frac{5}{16(1+2M)} > \tau = \frac{1}{8(1+2M)}.$$

Donc par le lemme 22, on a :

$$\frac{\lambda^k - \lambda^*}{\lambda^{k-1} - \lambda^*} \leq 1 - \frac{\vartheta}{2(1 + \frac{\beta}{\sqrt{m}})} \quad (1)$$

Après  $K$  itérations extérieures, on a donc :

$$\begin{aligned} f(x^K) - \lambda^* &\leq \lambda^K - \lambda^* \quad \text{puisque } \lambda^K \geq f(x^K) \\ &\leq \left[ 1 - \frac{\vartheta}{2(1 + \frac{\beta}{\sqrt{m}})} \right] (\lambda^{K-1} - \lambda^*) \quad \text{par (1).} \\ &\leq \left[ 1 - \frac{\vartheta}{2(1 + \frac{\beta}{\sqrt{m}})} \right]^K (\lambda^0 - \lambda^*) \quad \text{en appliquant } K \text{ fois l'inégalité (1).} \end{aligned}$$

$$\text{Donc, } f(x^K) - \lambda^* \leq \varepsilon \quad \text{si} \quad \left[ 1 - \frac{\vartheta}{2(1 + \frac{\beta}{\sqrt{m}})} \right]^K (\lambda^0 - \lambda^*) \leq \varepsilon$$

c'est-à-dire, en prenant le logarithme :

$$-K \ln \left[ 1 - \frac{\vartheta}{2(1 + \frac{\beta}{\sqrt{m}})} \right] \geq \ln(\lambda^0 - \lambda^*) - \ln \varepsilon$$

$$= \ln(\lambda^o - \lambda^*) + |\ln \varepsilon| .$$

Puisque  $-\ln(1 - \nu) \geq \nu$ , on obtient :

$$K > \frac{2(1 + \frac{\beta}{\sqrt{m}})}{\vartheta} [|\ln \varepsilon| + \ln(\lambda^o - \lambda^*)]$$

Par hypothèse,  $\lambda^o - \lambda^* \leq \frac{1}{\varepsilon}$ .

$$\text{Donc, } K > \frac{2(1 + \frac{\beta}{\sqrt{m}})}{\vartheta} \vartheta(|\ln \varepsilon|) .$$

□

**Théorème 8 :** Le nombre total  $P$  d'itérations intérieures lors d'une itération extérieure arbitraire vérifie :

$$P\delta \leq 1 + \frac{\vartheta m \beta}{\sqrt{m} + \beta} + \frac{m \vartheta^2}{1 - \vartheta}$$

où  $\beta \leq \frac{5}{16(1 + 2M)}$  et  $\delta$  est la diminution garantie de la fonction  $\varphi(x, \lambda)$  à

chaque itération intérieure, c'est-à-dire  $\delta = \frac{1}{140(1 + 2M)^2}$  par le lemme 18.

Preuve :

Notons  $\tilde{\lambda}$  la borne supérieure lors d'une itération extérieure quelconque, et  $\bar{\lambda}$  l'ancienne borne supérieure. On a donc :  $\tilde{\lambda} = \bar{\lambda} - \vartheta(\bar{\lambda} - f(x))$  où  $x$  est l'itéré au début de l'itération extérieure ( $x$  est donc "proche" de  $x(\bar{\lambda})$ ).

On a alors, en posant  $x_1, x_2, \dots$  les différents itérés intérieurs :

$$\varphi(x, \tilde{\lambda}) - \varphi(x(\tilde{\lambda}), \tilde{\lambda}) = \varphi(x, \tilde{\lambda}) - \varphi(x_1, \tilde{\lambda}) + \varphi(x_1, \tilde{\lambda}) - \varphi(x_2, \tilde{\lambda}) + \dots - \varphi(x(\tilde{\lambda}), \tilde{\lambda})$$

$$\leq P\delta \text{ par le lemme 18 .}$$

$$\text{Posons } \varphi(x, \tilde{\lambda}) - \varphi(x(\tilde{\lambda}), \tilde{\lambda}) = \Phi(x, \tilde{\lambda}) .$$

Par le théorème de la valeur intermédiaire, il existe  $\hat{\lambda} \in ]\tilde{\lambda}, \bar{\lambda}[$  tel que

$$\Phi(x, \tilde{\lambda}) = \Phi(x, \bar{\lambda}) + \left( \frac{d\Phi(x, \lambda)}{d\lambda} \right)_{\lambda=\hat{\lambda}} (\bar{\lambda} - \tilde{\lambda}) .$$

Calculons  $\frac{d\Phi(x, \lambda)}{d\lambda} = \frac{d\varphi(x, \lambda)}{d\lambda} - \frac{d\varphi(x(\lambda), \lambda)}{d\lambda}$

On a :  $\frac{d\varphi(x, \lambda)}{d\lambda} = \frac{m}{\lambda - f(x)}$

et 
$$\begin{aligned} \frac{d\varphi(x(\lambda), \lambda)}{d\lambda} &= m \frac{\nabla f(x(\lambda))'x'(\lambda) + 1}{\lambda - f(x(\lambda))} + \sum_{i=1}^m \frac{\nabla g_i(x(\lambda))'x'(\lambda)}{-g_i(x(\lambda))} \\ &= m \frac{\nabla f(x(\lambda))'x'(\lambda) + 1}{\lambda - f(x(\lambda))} + \frac{m}{\lambda - f(x(\lambda))} \sum_{i=1}^m y_i(\lambda) \nabla g_i(x(\lambda))'x'(\lambda) \end{aligned}$$

car, par (\*),  $\frac{m}{\lambda - f(x(\lambda))} = -y_i(\lambda)g_i(x(\lambda))$  pour  $1 \leq i \leq m$ .

Donc, 
$$\begin{aligned} \frac{d\varphi(x(\lambda), \lambda)}{d\lambda} &= \frac{m}{\lambda - f(x(\lambda))} \left[ \left( \nabla f(x(\lambda)) + \sum_{i=1}^m y_i(\lambda) \nabla g_i(x(\lambda)) \right)'x'(\lambda) + 1 \right] \\ &= \frac{m}{\lambda - f(x(\lambda))} \quad \text{par (*).} \end{aligned}$$

On obtient alors : 
$$\begin{aligned} \left( \frac{d\Phi(x, \lambda)}{d\lambda} \right)_{\lambda=\hat{\lambda}} &= m \left( \frac{1}{\lambda - f(x)} - \frac{1}{\lambda - f(x(\lambda))} \right)_{\lambda=\hat{\lambda}} \\ &\leq m \left( \frac{1}{\tilde{\lambda} - f(x)} - \frac{1}{\bar{\lambda} - f(x(\bar{\lambda}))} \right) \end{aligned}$$

puisque  $\tilde{\lambda} < \hat{\lambda}$  et, par le lemme 23,  $\lambda - f(x(\lambda))$  est croissante en  $\lambda$ .

Donc,  $\Phi(x, \tilde{\lambda}) \leq \Phi(x, \bar{\lambda}) + m \left( \frac{1}{\tilde{\lambda} - f(x)} - \frac{1}{\bar{\lambda} - f(x(\bar{\lambda}))} \right) (\bar{\lambda} - \tilde{\lambda})$ .

On a :  $\tilde{\lambda} = \bar{\lambda} - \vartheta(\bar{\lambda} - f(x))$

Donc,  $\tilde{\lambda} - f(x) = (1 - \vartheta)(\bar{\lambda} - f(x))$

et  $\bar{\lambda} - \tilde{\lambda} = \vartheta(\bar{\lambda} - f(x))$ .

Ainsi  $\Phi(x, \tilde{\lambda}) \leq \Phi(x, \bar{\lambda}) + m \left( \frac{\vartheta}{1 - \vartheta} - \frac{\vartheta(\bar{\lambda} - f(x))}{\bar{\lambda} - f(x(\bar{\lambda}))} \right)$ .

On sait, par le lemme 21, que  $\frac{(\bar{\lambda} - f(x))}{\bar{\lambda} - f(x(\bar{\lambda}))} \leq 1 + \frac{\beta}{\sqrt{m}}$ .

De plus,  $\Phi(x, \bar{\lambda}) = \varphi(x, \bar{\lambda}) - \varphi(x(\bar{\lambda}), \bar{\lambda}) < 1$  par le lemme 19.

$$\begin{aligned} \text{Donc, } \Phi(x, \tilde{\lambda}) &\leq 1 + m\vartheta \left[ \frac{1}{1-\vartheta} + \frac{1}{1 + \frac{\beta}{\sqrt{m}}} \right] \\ &= 1 + m\vartheta \left[ \frac{1}{1-\vartheta} + \frac{\beta}{\beta + \sqrt{m}} \right] \end{aligned}$$

□

Regroupant ces deux résultats, on obtient comme nombre total d'itérations :

$$\frac{2}{\delta} \left( 1 + \frac{\beta}{\sqrt{m}} \right) \left( \frac{1}{\vartheta} + \frac{m\beta}{\beta + \sqrt{m}} + \frac{\vartheta m}{1-\vartheta} \right) |\ln \varepsilon| .$$

Donc, en remplaçant  $\delta$  par sa valeur et en choisissant  $\vartheta$  tel que  $0 < \vartheta < 1$  et indépendant de  $m, M$  et  $\varepsilon$ , on obtient  $\mathcal{O}\left((1+M)^2 n \ln |\varepsilon|\right)$  comme borne d'itérations.

## 2. Algorithme à pas longs se basant sur la fonction barrière logarithmique.

### A. Généralités.

Soit le programme convexe (CP) 
$$\begin{cases} \min f(x) \\ \text{s.c. } g_i(x) \leq 0 & 1 \leq i \leq m \end{cases}$$

où les fonctions  $f(x)$  et  $g_i(x)$ ,  $1 \leq i \leq m$ , sont convexes et deux fois continûment différentiables et où les fonctions  $g_i(x)$  satisfont la condition RLC.

Sans perte de généralité, nous pouvons supposer que  $f(x)$  est linéaire, c'est-à-dire que  $f: R^n \rightarrow R: x \rightsquigarrow f(x) = c'x$  où  $c \in R^n$ .

(CP) se réécrit donc (CP) 
$$\begin{cases} \min c'x \\ \text{s.c. } g_i(x) \leq 0 & 1 \leq i \leq m \end{cases}$$

Le dual de Wolfe de (CP) s'énonce (D) 
$$\begin{cases} \max c'x + \sum_{i=1}^m y_i g_i(x) \\ \text{s.c. } \sum_{i=1}^m y_i \nabla g_i(x) + c = 0 \\ y_i \geq 0 & i = 1, \dots, m \end{cases}$$

et le théorème de dualité faible établit que, si  $x$  est une solution admissible de (CP) et  $(\bar{x}, \bar{y})$  une solution admissible de (D), alors :

$$c'x \geq c'\bar{x} + \sum_{i=1}^m \bar{y}_i g_i(\bar{x}).$$

Soit  $\mathcal{P} = \{x \in R^n \text{ tel que } g_i(x) \leq 0 \quad i = 1, \dots, m\}$  l'ensemble primal admissible et  $\mathcal{P}^0$  l'intérieur relatif de  $\mathcal{P}$ .

On associe à (CP) la fonction barrière logarithmique

$$\varphi_\mu(x): \mathcal{P}^0 \rightarrow R: x \rightsquigarrow \varphi_\mu(x) = c'x - \mu \sum_{i=1}^m \ln(-g_i(x))$$

pour  $\mu$  variant de  $+\infty$  à 0.

On sait, par définition de  $\varphi_\mu(x)$ , que  $\nabla \varphi_\mu(x) = c + \mu \sum_{i=1}^m \frac{\nabla g_i(x)}{-g_i(x)}$

$$\text{et } \nabla^2 \varphi_\mu(x) = \mu \sum_{i=1}^m \left[ \frac{\nabla^2 g_i(x)}{-g_i^2(x)} + \frac{\nabla g_i(x) \nabla g_i(x)'}{g_i^2(x)} \right].$$

De la même manière que cela a été fait auparavant, nous remplacerons le problème (CP) par les problèmes

$$(CP)_\mu \begin{cases} \min c'x - \mu \sum_{i=1}^m \ln(-g_i(x)) \\ \text{s.c. } g_i(x) \leq 0 \end{cases}$$

Comme  $\varphi_\mu(x)$  est strictement convexe sur  $\mathcal{P}^0$ , elle atteint, pour  $\mu$  fixé, un minimum unique noté  $(\mu)$ .

L'ensemble des minima  $x(\mu)$  où  $\mu$  varie de  $+\infty$  à 0 constitue la trajectoire centrale et les conditions de KUHN-TUCKER, nécessaires et suffisantes, s'écrivent avec

$$L(x, y) = c^t x - \mu \sum_{i=1}^m \ln(-g_i(x)) + \sum_{i=1}^m y_i g_i(x)$$

$$(KT) \begin{cases} c + \sum_{i=1}^m y_i \nabla g_i(x) = 0 \\ g_i(x) \leq 0 \quad i = 1, \dots, m \\ y_i \geq 0 \quad i = 1, \dots, m \\ -g_i(x) y_i = \mu \quad i = 1, \dots, m \end{cases}$$

Le saut de dualité vérifie alors :

$$\begin{aligned} c^t x - \left[ c^t x + \sum_{i=1}^m y_i g_i(x) \right] &= - \sum_{i=1}^m y_i g_i(x) \\ &= m\mu \end{aligned} \quad (1).$$

Le lemme suivant décrit le comportement des fonctions objectifs primale et duale en fonction du paramètre  $\mu$ .

Lemme 24 : i) La fonction objectif  $c^t x(\mu)$  du problème primal (CP) est strictement décroissante lorsque  $\mu$  décroît.

ii) La fonction objectif  $c^t x(\mu) + \sum_{i=1}^m y_i g_i(x)$  du problème dual (D) est croissante lorsque  $\mu$  décroît.

Preuve :

i) Il faut montrer que  $c^t x' > 0$  où  $x'$  désigne  $\frac{dx}{d\mu}$ .

Soient  $x(\mu)$  et  $y(\mu)$  satisfaisant le système (KT).

Dérivons par rapport à  $\mu$  les deux égalités de ce système. On obtient :

$$\sum_{i=1}^m y_i' \nabla g_i(x) + \sum_{i=1}^m y_i G_i x' = 0 \quad (2)$$

$$-y_i' g_i(x) - y_i \nabla g_i(x)^t x' = 1 \quad i = 1, \dots, m \quad (3)$$

où  $y_i'$  désigne  $\frac{dy_i}{d\mu}$  et  $G_i$  la matrice hessienne de  $g_i(x)$ .



Multiplions (3) par  $y_i \nabla g_i(x)$  ; on obtient, puisque  $\mu = -g_i(x)y_i$  :

$$\mu y_i' \nabla g_i(x) - y_i^2 \nabla g_i(x) \nabla g_i(x)' x' = y_i \nabla g_i(x) \quad i = 1, \dots, m$$

et sommons sur  $i$  :

$$\mu \sum_{i=1}^m y_i' \nabla g_i(x) - \sum_{i=1}^m y_i^2 \nabla g_i(x) \nabla g_i(x)' x' = \sum_{i=1}^m y_i \nabla g_i(x) \quad .$$

Par le système (KT), le second membre vaut  $-c$  . Utilisant (2) pour le premier terme du membre de gauche, cette dernière équation se réécrit :

$$-\mu \sum_{i=1}^m y_i G_i x' - \sum_{i=1}^m y_i^2 \nabla g_i(x) \nabla g_i(x)' x' = -c$$

$$\text{ou } \mu \sum_{i=1}^m y_i G_i x' + \sum_{i=1}^m y_i^2 \nabla g_i(x) \nabla g_i(x)' x' = c \quad .$$

Prenant le produit scalaire par  $x'$  des deux côtés de l'égalité, on obtient :

$$c' x' = \mu \sum_{i=1}^m y_i x'^t G_i x' + \sum_{i=1}^m y_i^2 (\nabla g_i(x)' x')^2 \quad (4)$$

Par définition de  $\nabla^2 \varphi_\mu(x)$  et par (KT), on a :

$$c' x' = \mu^2 x'^t \nabla^2 \varphi_\mu(x) x'$$

$> 0$  car  $\nabla^2 \varphi_\mu(x)$  est définie positive.

ii) Il suffit de prouver que  $\frac{d}{d\mu} \left[ c' x + \sum_{i=1}^m y_i g_i(x) \right] \leq 0$

ou encore  $\frac{d}{d\mu} \left[ -c' x - \sum_{i=1}^m y_i g_i(x) \right] \geq 0$  .

On a :  $\frac{d}{d\mu} \left[ -c' x - \sum_{i=1}^m y_i g_i(x) \right] = \frac{d}{d\mu} [-c' x + m\mu]$  par le système de K-T

$$= -c' x' + m$$

$$= -\mu \sum_{i=1}^m y_i x'^t G_i x' - \sum_{i=1}^m y_i^2 (\nabla g_i(x)' x')^2 + m \quad \text{par (4)}$$

Elevons (3) au carré . On obtient :

$$y_i'^2 g_i^2(x) + y_i^2 (\nabla g_i(x)' x')^2 + 2y_i' y_i g_i(x) \nabla g_i(x)' x' = 1$$

et sommons sur  $i$  :

$$\sum_{i=1}^m y_i'^2 g_i^2(x) + \sum_{i=1}^m y_i^2 (\nabla g_i(x)'x')^2 + 2 \sum_{i=1}^m y_i' y_i g_i(x) \nabla g_i(x)'x' = m .$$

$$\begin{aligned} \text{On a : } \sum_{i=1}^m y_i' y_i g_i(x) \nabla g_i(x)'x' &= -\mu \sum_{i=1}^m y_i' \nabla g_i(x)'x' \quad \text{par le système de K-T} \\ &= -\mu x'^t \sum_{i=1}^m y_i' \nabla g_i(x) \\ &= \mu x'^t \sum_{i=1}^m y_i G_i x' \quad \text{par (2)} \\ &= \mu \sum_{i=1}^m y_i x'^t G_i x' . \end{aligned}$$

$$\begin{aligned} \text{Donc,} \quad -c'x' + m &= -\mu \sum_{i=1}^m y_i x'^t G_i x' - \sum_{i=1}^m y_i^2 (\nabla g_i(x)'x')^2 + \sum_{i=1}^m y_i'^2 g_i^2(x) \\ &\quad + \sum_{i=1}^m y_i^2 (\nabla g_i(x)'x')^2 + 2\mu \sum_{i=1}^m y_i x'^t G_i x' \\ &= \sum_{i=1}^m y_i'^2 g_i^2(x) + \mu \sum_{i=1}^m y_i x'^t G_i x' \\ &\geq 0 \quad \text{car les fonctions } g_i(x) \text{ sont convexes.} \quad (5) \end{aligned}$$

□

Remarque : par (4) et (5), on a :  $m \geq c'x' > 0$  le long de la trajectoire centrale .  
En intégrant, on obtient :

$$0 < c'x(\mu) - c'x(\bar{\mu}) \leq m(\mu - \bar{\mu}) \quad \text{pour } \mu > \bar{\mu} \quad (6).$$

### B. L'algorithme.

Cet algorithme adopte le même principe que l'algorithme à grands pas de minimisation de la fonction potentielle. Si l'itéré est trop éloigné de la trajectoire centrale, nous effectuerons une recherche linéaire le long de la direction de NEWTON, donnée par :

$$p \stackrel{\Delta}{=} p(x, \mu) = -H^{-1}(x, \mu) g(x, \mu) \stackrel{\Delta}{=} -H^{-1}g$$

où  $H(x, \mu)$  désigne  $\nabla^2 \varphi_\mu(x)$  et  $g(x, \mu)$  désigne  $\nabla \varphi_\mu(x)$  .

Le critère de proximité est  $\|p\|_H \leq \tau$  où  $\tau$  est la tolérance et, à chaque itération externe, le paramètre  $\mu$  sera réduit d'un facteur  $1 - \theta$  où  $0 < \theta < 1$  .

Initialisation : Soit  $\varepsilon > 0$  le paramètre de précision de l'algorithme .

Soit  $\tau = \frac{1}{8(1+2M)^2}$  le paramètre de tolérance .

Soit  $x^0$  un point intérieur admissible et  $\mu^0 \leq \frac{1}{\varepsilon}$

tels que  $\|p(x^0, \mu^0)\|_{H(x^0, \mu^0)} \leq \tau$

Algorithme: Poser  $x = x^0$  ;  $\mu = \mu^0$  .

(Début de l'itération externe)

Tant que  $\mu > \frac{\varepsilon}{2n}$

(Début de la recherche linéaire)

Tant que  $\|p\|_H > \tau$

effectuer  $\bar{\alpha} = \arg \min_{\alpha > 0} [\varphi_\mu(x + \alpha p) : x + \alpha p \in \mathcal{P}^0]$

poser  $x + \alpha p$

(Fin de la recherche linéaire)

Poser  $\mu = (1 - \theta)\mu$

(Fin de l'itération externe) .

### C. Analyse de convergence .

Afin d'établir les théorèmes de convergence, nous dresserons une liste de lemmes techniques préliminaires . Ces lemmes présentent des résultats très semblables à ceux qui ont été démontrés dans le cadre de la minimisation de la fonction potentielle . Pour cette raison, nous nous contenterons d'énoncer ces lemmes sans en fournir de démonstration.

Lemme 25 : Si les fonctions  $f(x)$  et  $g_i(x)$  sont linéaires ou quadratiques avec des matrices hessiennes semi-définies positives, et si  $x \in \mathcal{P}^0$  et  $\|d\|_H < 1$ , alors  $x + d \in \mathcal{P}^0$

$$\text{et } |\varphi_\mu(x + d) - q_x(x + d, \mu)| < \frac{\|d\|_H^3}{3(1 - \|d\|_H)} .$$

Lemme 26 : Si les fonctions  $g_i(x)$  satisfont la condition RLC avec la constante  $M > 0$ , et si  $x \in \mathcal{D}^0$  et  $\|d\|_H < \min(\frac{1}{2}, \frac{1}{2M^3})$

alors  $x + d \in \mathcal{D}^0$

$$\text{et } |\varphi_\mu(x + d) - q_x(x + d, \mu)| < \frac{\|d\|_H^3}{3(1 - \|d\|_H)}(1 + 2M).$$

Lemme 27 : Si  $\|p\|_H \leq \frac{1}{8(1 + 2M)}$ ,

$$\text{Alors } \|x - x(\mu)\|_H \leq \frac{5}{2}\|p\|_H.$$

Lemme 28 : Si  $\|p\|_H \geq \frac{1}{8(1 + 2M)}$ ,

alors la décroissance  $\Delta\varphi_\mu$  de la fonction barrière  $\varphi_\mu$  après une recherche linéaire le long de la direction de NEWTON  $p$  satisfait :

$$\Delta\varphi_\mu \geq \frac{1}{140(1 + 2M)^2}.$$

Lemme 29 : Si  $\|p\|_H \leq \frac{1}{8(1 + 2M)}$ ,

$$\text{alors } \varphi_\mu(x) - \varphi_\mu(x(\mu)) \leq 4\|p\|_H^2.$$

Lemme 30 : Si  $\|p\|_H \leq \frac{1}{8(1 + 2M)}$ ,

$$\text{alors } |c^t x - c^t x(\mu)| \leq \frac{1}{2}\mu\sqrt{m}.$$

Le théorème suivant fournit une borne supérieure au nombre d'itérations externes de l'algorithme.

**Théorème 9 :** Après au plus  $O\left(\frac{1}{\vartheta} \ln\left(\frac{m}{\varepsilon}\right)\right)$  itérations externes, l'algorithme se termine avec une solution  $\varepsilon$ -optimale pour (CP) .

Preuve :

Soit  $z^*$  le coût optimal de (CP) .

Après  $K$  itérations, le saut de dualité peut être majoré par :

$$\begin{aligned} c^T x^K - z^* &= |c^T x^K - z^*| \leq |c^T x^K - c^T x(\mu^K)| + |c^T x(\mu^K) - z^*| \\ &\leq \frac{1}{2} \mu^K \sqrt{m} + m \mu^K \quad \text{par le lemme 30 et par (1)} \\ &= \mu^K \left( \frac{1}{2} \sqrt{m} + m \right) . \end{aligned}$$

Donc, si  $\mu^K \left( \frac{1}{2} \sqrt{m} + m \right) \leq \varepsilon$ , alors  $c^T x^K - z^* \leq \varepsilon$

c'est-à-dire si  $(1 - \vartheta)^K \mu^0 \left( \frac{1}{2} \sqrt{m} + m \right) \leq \varepsilon$

ou encore  $K \ln(1 - \vartheta) + \ln \mu^0 + \ln\left(\frac{1}{2} \sqrt{m} + m\right) \leq \ln \varepsilon$  .

Finalement,  $K \geq \frac{-\ln \varepsilon + \ln \mu^0 + \ln\left(\frac{1}{2} \sqrt{m} + m\right)}{-\ln(1 - \vartheta)}$  .

On a supposé que  $\mu^0 \leq \frac{1}{\varepsilon}$  c'est-à-dire  $\ln \mu^0 \leq -\ln \varepsilon$  .

De plus,  $\vartheta \leq -\ln(1 - \vartheta)$  .

D'où,  $\frac{1}{\vartheta} \geq \frac{1}{-\ln(1 - \vartheta)}$  .

Donc, si  $K \geq \frac{-2 \ln \varepsilon + \ln\left(m + \frac{1}{2} \sqrt{m}\right)}{\vartheta}$  ,

c'est-à-dire si  $K = O\left(\frac{1}{\vartheta} \ln \frac{m}{\varepsilon}\right)$ , l'inégalité est vérifiée .

□

**Théorème 10 :** Le nombre total  $P$  d'itérations intérieures pendant une itération extérieure arbitraire satisfait :

$$P\delta \leq 1 + \frac{9}{1-9}(9m + \sqrt{m})$$

$$\text{où } \delta = \frac{1}{140(1+2M)^2}$$

Preuve :

Soit une itération extérieure quelconque .

Notons  $x^0$  et  $x^P$  les valeurs de la variable primale respectivement au début de l'itération externe et après avoir effectué  $P$  itérations internes.

Par le lemme 28 ,on a, avec  $\bar{\mu} = (1-9)\mu$  ,

$$\begin{aligned} \varphi_{\bar{\mu}}(x^P) - \varphi_{\bar{\mu}}(x^0) &= \varphi_{\bar{\mu}}(x^P) - \varphi_{\bar{\mu}}(x^{P-1}) + \dots - \varphi_{\mu}(x^0) \\ &= -P\Delta\varphi_{\mu} \\ &\leq -P\delta . \end{aligned}$$

$$\text{Donc, } P\delta \leq \varphi_{\bar{\mu}}(x^0) - \varphi_{\bar{\mu}}(x^P) \tag{7}$$

$$\text{On a : } \varphi_{\bar{\mu}}(x) = \varphi_{\mu}(x) + 9\frac{c'x}{\bar{\mu}} .$$

$$\text{Substituant dans (7), on obtient : } P\delta \leq \varphi_{\mu}(x^0) - \varphi_{\mu}(x^P) + \frac{9}{\bar{\mu}}(c'x^0 - c'x^P) .$$

Comme  $x^0$  est presque centré par rapport à  $x(\mu)$ , on a, par le lemme 29 :

$$\varphi_{\mu}(x^0) - \varphi_{\mu}(x(\mu)) \leq 4 \frac{1}{8(1+2M)^2} \leq 1$$

$$\text{et donc, } \varphi_{\mu}(x^0) \leq 1 + \varphi_{\mu}(x(\mu)) .$$

$$\text{Par conséquent, } \varphi_{\mu}(x^0) - \varphi_{\mu}(x^P) \leq 1 + \varphi_{\mu}(x(\mu)) - \varphi_{\mu}(x^P)$$

$$\leq 1 \quad \text{car } x(\mu) \text{ est le minimum de } \varphi_{\mu}(x) .$$

$$\text{On obtient alors } P\delta \leq 1 + \frac{9}{\bar{\mu}}(c'x^0 - c'x^P) .$$

$$\text{Par le lemme 30 , on a : } |c'x^0 - c'x(\mu)| \leq \frac{1}{2}\bar{\mu}\sqrt{m}$$

$$\text{et } |c'x^P - c'x(\bar{\mu})| \leq \frac{1}{2}\bar{\mu}\sqrt{m} .$$

De plus, par (6),  $c^t x(\mu) - c^t x(\bar{\mu}) \leq m(\mu - \bar{\mu})$

Donc,  $c^t x^0 - c^t x^P = |c^t x^0 - c^t x^P|$  par le lemme 24

$$\leq |c^t x^0 - c^t x(\mu)| + |c^t x^P - c^t x(\bar{\mu})| + |c^t x(\mu) - c^t x(\bar{\mu})|$$

$$\leq \frac{1}{2}\mu\sqrt{m} + \frac{1}{2}\bar{\mu}\sqrt{m} + m(\mu - \bar{\mu})$$

$$\leq \mu(\vartheta m + \sqrt{m}) \quad \text{puisque } \bar{\mu} = (1 - \vartheta)\mu .$$

Donc,  $P\delta \leq 1 + \frac{\vartheta}{1 - \vartheta}(\vartheta m + \sqrt{m})$ .

□

Par conséquent, le nombre total d'itérations de l'algorithme est :

$$\frac{1}{\delta} \left( \frac{1}{\vartheta} + \frac{\vartheta m + \sqrt{m}}{1 - \vartheta} \right) O\left(\ln\left(\frac{m}{\varepsilon}\right)\right) .$$

Donc, en prenant  $\vartheta = O\left(\frac{1}{\sqrt{n}}\right)$ , on obtient  $O\left((1 + M)^2 \sqrt{n} |\ln \varepsilon|\right)$  comme borne d'itérations, c'est-à-dire exactement la même borne que pour la réduction de la fonction potentielle.

## CONCLUSIONS.

Dans ce mémoire, nous avons étudié la résolution de problèmes de programmation convexe.

Dans un premier temps, nous avons considéré le cas des contraintes d'égalités linéaires. Un algorithme primal-dual, suivant la trajectoire centrale, a été présenté, algorithme dont la convergence vers la solution optimale est assurée lorsque la fonction objectif remplit la condition SLC.

Pour le cas des contraintes d'inégalités convexes, nous avons été amené à émettre l'hypothèse de self-concordance pour un algorithme à petits pas, et la condition RLC dans le cas d'algorithmes à grands pas.

En fait, la condition la plus générale est la condition de self-concordance. Il serait donc intéressant, comme complément de ce travail, d'établir les preuves de convergence des algorithmes en fonction de cette condition.



## Annexe 1

### Liens entre les différentes conditions

#### A. Rappels des définitions.

Self concordance  $\varphi : \mathcal{P}^0 \rightarrow R$  est dite  $\alpha$ -self-concordante sur  $\mathcal{P}^0$   
si  $\forall x \in \mathcal{P}^0$  et  $\forall h \in R^n$ , on a :  
 $|\nabla^3 \varphi(x)[h, h, h]| \leq 2\alpha (h' \nabla^2 \varphi(x) h)^{\frac{3}{2}}$ .

Cette condition est imposée aux fonctions  $-\ln(-g_i(x))$  afin de permettre aux fonctions potentielle ou barrière logarithmique d'être fortement self-concordantes.

SLC  $f(x) : R_+^n \rightarrow R$  satisfait la condition SLC  
si  $\exists M > 0$  tel que  $\forall x > 0$   
 $\forall \gamma \in ]0, 1[$   
 $\forall d_x$  satisfaisant  $\|X^{-1} d_x\| \leq \gamma$ ,  
on a :  $\|X[\nabla f(x + d_x) - \nabla f(x) - \nabla^2 f(x) d_x]\| \leq M d_x' \nabla^2 f(x) d_x$ .

Pour les programmes (LCCP), ZHU impose cette condition à la fonction objectif  $f(x)$ .

RLC  $f(x) : \mathcal{P}^0 \rightarrow R$  satisfait la condition RLC  
si  $\exists M > 0$  tel que  $\forall v \in R^n$   
 $\forall x, x+h \in \mathcal{P}^0$ ,  
on a :  $|v'[\nabla^2 f(x+h) - \nabla^2 f(x)]v| \leq M \|h\|_{H(x)} v' \nabla^2 f(x) v$ .  
où  $H$  est la matrice hessienne de la fonction  $-\ln(-f(x))$ .

Cette condition est imposée aux  $m$  contraintes d'inégalités d'un programme convexe (PC).

#### B. Relation entre self-concordance et SLC.

On admettra, sans démonstration, le lemme suivant.

Lemme 1 : Soit  $f(x) \in C^3(\mathcal{P}^0)$  une fonction convexe.

Si  $\exists \beta$  tel que  $\forall x \in \mathcal{P}^0, \forall h \in R^n$ , on a :

$$\nabla^3 f(x)[h, h, h] \leq \beta h^t \nabla^2 f(x) h \sqrt{\sum_{i=1}^n \frac{h_i^2}{x_i^2}} \quad (1)$$

alors  $\varphi_\mu(x) = f(x) - \mu \sum_{i=1}^n \ln x_i$  est  $\left(1 + \frac{1}{3}\beta\right)$  self-concordante sur  $\mathcal{P}^0$ .

et  $\varphi(x, \lambda) = -\ln(\lambda - f(x)) - \sum_{i=1}^n \ln x_i$  est  $\left(1 + \frac{1}{3}\beta\right)$  self-concordante sur  $\mathcal{P}(\lambda)^0$ .

Lemme 2 : Soit  $f(x) \in C^3$  et remplissant la condition SLC.

Alors les fonctions  $\varphi_\mu(x)$  et  $\varphi(x, \lambda)$  sont  $\left(1 + \frac{2}{3}M\right)$  self-concordantes.

Preuve :

Il suffit de prouver (1).

Posons  $h = d_x$  dans la définition SLC.

Nous avons :  $\|X^{-1} d_x\| = \|X^{-1} h\|$

$$\begin{aligned} &= \left\| \left( \frac{h_1}{x_1}, \dots, \frac{h_n}{x_n} \right)^t \right\| \\ &= \sqrt{\sum_{i=1}^n \frac{h_i^2}{x_i^2}} \end{aligned}$$

Puisque  $f \in C^3$ , nous pouvons développer  $\nabla f$  en série :

$$\nabla f(x + d_x) = \nabla f(x) + \nabla^2 f(x) d_x + \frac{1}{2} \nabla^3 f(x)[d_x, d_x, \cdot] + o(\|d_x\|^2)$$

où  $\nabla^3 f(x)[d_x, d_x, \cdot]$  est un vecteur dont la  $i^e$ -composante vaut

$$\sum_{j,k=1}^n \frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_k} d_x^j d_x^k$$

En fait,  $\exists \lambda \in [0, 1]$  tel que

$$\nabla f(x + d_x) = \nabla f(x) + \lambda \nabla^2 f(x) d_x + \frac{1}{2} \lambda^2 \nabla^3 f(x)[d_x, d_x, \cdot]$$

Remplaçant  $d_x$  par  $\lambda d_x$  dans la définition, nous obtenons :

$$\|X \nabla^3 f(x)[d_x, d_x, \cdot]\| \leq 2M d_x^t \nabla^2 f(x) d_x \text{ en divisant les deux membres par } \lambda^2.$$

Par l'inégalité de CAUCHY-SCHWARZ, nous obtenons :

$$\begin{aligned} \|X\nabla^3 f(x)[d_x, d_x, \cdot]\| &\geq \frac{(X^{-1}d_x)^t}{\|X^{-1}d_x\|} X\nabla^3 f(x)[d_x, d_x, \cdot] \\ &= \frac{\nabla^3 f(x)[d_x, d_x, d_x]}{\|X^{-1}d_x\|} \end{aligned}$$

On obtient finalement :

$$\nabla^3 f(x)[d_x, d_x, d_x] \leq 2M\|X^{-1}d_x\|d_x^t \nabla^2 f(x)d_x$$

ce qui est exactement la relation (1) .

□

### C. Relation entre self-concordance et RLC.

Lemme 3 : Soit  $f(x) \in C^3(\mathcal{P}^0)$  une fonction convexe et  $\mathcal{P}f = \{x \in R^n \text{ tel que } f(x) \leq 0\}$

Si  $f(x)$  satisfait la condition RLC

alors  $\varphi(x) = -\ln(-f(x))$  est  $(1+M)^2$  self-concordante sur  $\mathcal{P}_f^0$

Preuve :

soit  $h \in R^n$  une direction arbitraire fixée et  $x \in \mathcal{P}^0$  .

Soit la fonction  $g : R \rightarrow R : \vartheta \rightsquigarrow g(\vartheta) = h^t \nabla^2 f(x)h - h^t \nabla^2 f(x + \vartheta h)h$  .

Puisque  $f(x)$  vérifie la condition RLC, on a:

$$|g(\vartheta)| \leq M\|\vartheta h\|_{H(x)} h^t \nabla^2 f(x)h \quad \text{pour } \vartheta h \text{ assez petit}$$

Par la définition de  $\|\cdot\|_{H(x)}$ , on obtient :

$$|g(\vartheta)| \leq M|\vartheta| \left[ \frac{h^t \nabla^2 f(x)h}{-f(x)} + \frac{(h^t \nabla f(x))^2}{f^2(x)} \right]^{\frac{1}{2}} h^t \nabla^2 f(x)h$$

$$\leq Md_2^{\frac{3}{2}} + (M+3)d_1d_2 + 2d_1^3 .$$

Notons  $a = d_1$  et  $b = \sqrt{d_2}$  .

$$\begin{aligned} \text{On a : } |\nabla^3\varphi(x)[h, h, h]|^2 &\leq (Mb^3 + Mab^2 + 3ab^2 + 2a^3)^2 \\ &= M^2b^6 + 2M^2ab^5 + 6Mab^5 + 4Ma^3b^3 + M^2a^2b^4 + 6Ma^2b^4 \\ &\quad + 4Ma^4b^2 + 9a^2b^4 + 12a^4b^2 + 4a^6 . \end{aligned}$$

Puisque  $2ab \leq a^2 + b^2$  , on obtient :

$$\begin{aligned} |\nabla^3\varphi(x)[h, h, h]|^2 &\leq 2M^2b^6 + 2M^2a^2b^4 + 11Ma^2b^4 + 3Mb^6 + 6Ma^4b^2 + 9a^2b^4 \\ &\quad + 12a^4b^2 + 4a^6 \\ &\leq (4 + 8M + 4M^2) (a^6 + 3a^4b^2 + 3a^2b^4 + b^6) \\ &= (4 + 8M + 4M^2) (a^2 + b^2)^3 \\ &= 4(1 + M)^2 (a^2 + b^2)^3 \\ &= 4(1 + M)^2 (d_1^2 + d_2)^3 \\ &= 4(1 + M)^2 [h^t \nabla^2 \varphi(x) h]^3 . \end{aligned}$$

Par conséquent,

$$|\nabla^3\varphi(x)[h, h]| \leq 2(1 + M) [h^t \nabla^2 \varphi(x) h]^{\frac{1}{3}} .$$

□

## Annexe 2

**Théorème 11 :** Si  $\varphi(x)$  est fortement self-concordante sur  $\mathcal{P}^0$ ,  
alors  $\varphi(x)$  est strictement convexe sur  $\mathcal{P}^0$ .

Preuve :

Soit  $x \in \mathcal{P}^0$

Montrons que  $H(x)$  est définie positive .

Supposons, par l'absurde, que  $H(x)$  n'est pas définie positive .

Comme  $H(x)$  est semi-définie positive,

$\exists h \neq 0$  tel que  $h^T H(x) h = 0$  ie  $\|h\|_{H(x)} = 0$  .

Par conséquent,  $\forall \zeta > 0$  ,  $\|\zeta h\|_{H(x)} = 0$

et par le lemme 9,  $x + \zeta h \in \mathcal{P}^0$ .

Or,  $\mathcal{P}^0$  est borné, ce qui met en évidence la contradiction .

□

## BIBLIOGRAPHIE.

DEN HERTOOG D. , ROOS C. , TERLAKY T. , A large-step analytic center method for a class of smooth convex programming problems. SIAM Journal on Optimization , 2 (1992) pp 55-70

DEN HERTOOG D. , ROOS C. , TERLAKY T. , On the classical logarithmic barrier function method for a class of smooth convex programming problems. JOTA , 73 (1992) pp 1-25

DEN HERTOOG D. , JARRE F. , ROOS C. , TERLAKY T. , A unifying investigation of interior-point methods for convex programming. Rapport Nr 92-89 , Delt University of technology (1992).

JARRE F. , Interior-point methods for convex programming. Applied Mathematics and Optimization, (1992) pp 287-311.

JARRE F. , On the method of analytic centers for solving smooth convex programs. in "Optimization" édité par S.DOLECKI , Lecture Notes in Mathematics 1405, Springer-Verlag (1989) pp 69-85.

KORTANEK K.O., POTRA F. , YE Y. , On some efficient interior-point methods for nonlinear convex programming. Linear Algebra and Its Applications, 152 (1991) pp169-189.

STRODIOT J.J. , A view on interior-point methods. Conférence donnée à Namur le 29 janvier 1993 dans le cadre du "troisième cycle en programmation mathématique".

ZHU J. , A path-following algorithm for a class of convex programming problems. ZOR-Methods and Models of Operations Research , 36 (1992) pp 359-377.