



THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Comparaison de trois méthodes de points intérieurs en programmation semi-définie

Lamotte, Véronique

Award date:
2002

[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

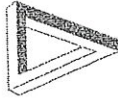
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



FUNDP
Faculté des Sciences
Département de Mathématique
Rempart de la Vierge, 8
B-5000 Namur Belgique

Comparaison de trois méthodes de points intérieurs en programmation semi-définie

Mémoire présenté pour l'obtention
du grade de
Licencié en Sciences Mathématiques
par



Lamotte Veronique

Année académique 2001-2002

Promoteur : J.-J. Strodiot

Je pense enfin à ma famille et à mon entourage qui m'ont encouragée et soutenue tout au long de mes études.

Je remercie également les personnes rencontrées au cours de ces quatre belles années aux facultés, pour leur générosité.

J'y associe tous les professeurs qui ont contribué à ma formation.

Je tiens à remercier bien sincèrement mon promoteur, Monsieur J.-J. Strodhot, pour la confiance qu'il m'a témoignée. Ses conseils, son soutien et sa disponibilité m'ont été précieux.

Semidefinite programming consists in optimizing a linear function subject to linear and matrix constraints. It is a generalization of the linear programming. This paper describes this class of optimization problems and gives several standard examples. It studies also the duality theory. Primal-dual interior-point methods for semidefinite programming are considered. Several variants are studied, based on Newton's method applied to three equations: primal feasibility, dual feasibility, and some form of centering condition. The conditions of uniqueness of search directions are described and a comparison of different choices of these is established.

Abstract

La programmation semi-définie consiste à optimiser une fonction linéaire en tenant compte de contraintes linéaires et matricielles. Elle est une généralisation de la programmation linéaire. Ce mémoire décrit cette classe de problèmes d'optimisation et donne quelques exemples standards. Il étudie aussi la théorie de la dualité. La méthode de points intérieurs primale-duale pour la programmation semi-définie est considérée. Plusieurs variantes sont étudiées, basées sur la méthode de Newton appliquée à trois équations: l'admissibilité primaire, l'admissibilité duale, et différentes formes de la condition de centralité. Les conditions d'unicité des directions de recherche sont décrites et une comparaison de différents choix de celles-ci est établie.

Résumé

Table des matières

Introduction	5
1 La programmation semi-définie.	7
1.1 Quelques notations.	7
1.2 Le problème de programmation semi-définie.	8
1.3 Exemples.	10
1.3.1 Minimisation de la plus grande valeur propre.	10
1.3.2 Minimisation de la norme-2 d'une matrice.	11
1.3.3 La programmation linéaire.	13
1.3.4 Un problème de programmation non linéaire quasi-convexe.	13
1.3.5 La programmation avec contraintes quadratiques convexes.	15
1.3.6 Le problème "max-cut".	17
2 La dualité et la trajectoire centrale.	20
2.1 La dualité.	20
2.2 La fonction barrière logarithmique.	28
2.3 La trajectoire centrale.	33
3 Méthodes de points intérieurs primales-duales.	39
3.1 Les équations Newton et les directions de recherche symétriques.	39
3.2 La notion du produit de Kronecker symétrisé.	43
3.3 L'unicité des directions de recherche symétriques.	44
4 Différentes directions de recherche. Présentation et développement.	49
4.1 La direction de Nesterov-Todd.	49
4.1.1 Définition.	49

89	Bibliographie
88	B Le produit de Kronecker.
87	A.3 Le problème aux moindres carrés.
86	A.2 Propriétés d'algèbre linéaire.
85	A.1 Trace et normes d'une matrice.
85	A Notions sur les matrices.
83	Conclusion
82	5.3 Conclusion de la comparaison.
79	5.2.2 Exemples numériques.
78	5.2.1 Algorithme prédicteur-correcteur de Mehrotra.
77	prédicteur-correcteur.
74	5.2 Comparaison des méthodes par l'application d'un algorithme
74	5.1.2 L'invariance sous changement d'échelle.
72	5.1.1 La symétrie primale-duale.
72	5.1 Deux propriétés importantes.
72	duales.
72	5 Comparaison des méthodes de points intérieurs primales-
71	4.3.2 Implémentation.
65	4.3.1 Une alternative au théorème d'unicité.
65	4.3 La direction d'Alizadeh-Haerberly-Overton.
63	4.2.2 Implémentation.
60	4.2.1 Définition.
60	4.2 Les directions HKM.
54	4.1.3 Implémentation.
52	4.1.2 La direction NT comme une direction Newton.

Introduction

La programmation semi-définie consiste à choisir une matrice symétrique qui optimise une fonction linéaire. Ce choix implique de tenir compte de contraintes linéaires et de la contrainte importante que la matrice doit être semi-définie positive. La programmation semi-définie recouvre de nombreux problèmes, dont la programmation linéaire.

Ce mémoire, basé principalement sur les articles de M. J. Todd [8] et de M. J. Todd, K.-C. Toh, et R. H. Tuntunou [9], propose d'étudier différents choix de directions de recherche pour la méthode de points intérieurs primale-duale.

Le premier chapitre débute par la définition du problème de programmation semi-définie sous ses formes primale et duale et introduit quelques notations. Afin de mieux rendre compte de l'utilité de la programmation semi-définie et de la raison de son étude, plusieurs applications sont présentées.

Le deuxième chapitre étudie la théorie de la dualité et établit les conditions suffisantes pour obtenir la dualité forte. Ce chapitre introduit la fonction barrière logarithmique pour le cône des matrices semi-définies positives et utilise celle-ci pour définir la trajectoire centrale.

Le troisième chapitre décrit plusieurs méthodes de points intérieurs primales-duales permettant d'obtenir des directions de recherche symétriques. Il donne également les conditions sous lesquelles la solution est unique.

Les deux derniers chapitres s'attardent plus particulièrement sur trois méthodes de points intérieurs. Celles-ci sont les plus connues et les plus fréquemment utilisées actuellement. Nous les connaissons sous le nom de la

méthode NT et les méthodes basées sur les directions AHO et HKM. Une comparaison de ces méthodes peut être établie par leurs propriétés et par l'application de l'algorithme prédicteur-correcteur de Mehrotra. Nous avons tenté de la réaliser dans le dernier chapitre.

Chapitre 1

La programmation semi-définie.

1.1 Quelques notations.

Avant de présenter le problème sur lequel nous allons travailler tout au

long de ce mémoire, nous devons nous accorder sur quelques notations.

Nous travaillerons principalement avec des matrices symétriques d'ordre n .

Nous noterons $\text{SR}^{n \times n}$, l'espace de ces matrices. La notation $U \bullet V$ sera utilisée pour la trace $\text{tr}(U^T V)$. Les normes associées sont la norme de Frobenius

$\|U\|^F = (U \bullet U)^{1/2}$ et la norme-2 d'une matrice, notée simplement $\|U\|$.

Nous écrirons $U \succeq 0$ ($U \succ 0$) pour dire que la matrice U est semi-définie

positive (définie positive) et nous noterons $\text{SR}_+^{n \times n}$ ($\text{SR}_{++}^{n \times n}$), l'espace des

matrices symétriques semi-définies positives (définies positives) d'ordre n . Si

$U \succeq 0$, nous écrirons $U^{1/2}$ pour la racine carrée semi-définie positive de la

matrice U .

Nous marquerons $\text{diag}(U)$ pour le vecteur des entrées diagonales de $U \in$

$\text{SR}^{n \times n}$, et $\text{Diag}(u)$ pour la matrice diagonale avec le vecteur $u \in \mathbb{R}^n$ sur sa

diagonale. Nous pouvons étendre cette notation aux matrices diagonales par

blocs: si U_1, U_2, \dots, U_k sont des matrices symétriques, alors $\text{Diag}(U_1, U_2, \dots, U_k)$

est la matrice par blocs U_i .

Nous avons rappelé, en annexe, les propriétés de la trace d'une matrice, la

norme de Frobenius et la norme-2 ainsi qu'une série de propriétés d'algèbre

linéaire régulièrement utilisées dans ce mémoire.

1.2 Le problème de programmation semi-définie.

Le problème de programmation semi-définie (problème SDP) sous la forme standard primale est

$$(P) \begin{cases} \min_X C \bullet X \\ \text{s.c.} & A_i \bullet X = b_i, i = 1, \dots, m \\ & X \succeq 0, \end{cases}$$

où tous les $A_i \in \text{SIR}^{n \times n}$, $b \in \mathbb{R}^m$, $C \in \text{SIR}^{n \times n}$ sont donnés, et $X \in \text{SIR}^{n \times n}$ est la variable.

Nous allons aussi considérer le problème SDP sous la forme standard duale

$$(D) \begin{cases} \max_{y,S} b^T y \\ \text{s.c.} & \sum_{i=1}^m y_i A_i + S = C \\ & S \succeq 0, \end{cases}$$

où $y \in \mathbb{R}^m$ et $S \in \text{SIR}^{n \times n}$ sont les variables. Il peut encore s'écrire

$$\begin{cases} \max_{y,S} b^T y \\ \text{s.c.} & \sum_{i=1}^m y_i A_i \preceq C \end{cases}$$

ou

$$\begin{cases} \max_{y,S} b^T y \\ \text{s.c.} & C - \sum_{i=1}^m y_i A_i \succeq 0, \end{cases}$$

mais nous verrons que la matrice S sera utile pour les algorithmes. C'est la raison pour laquelle nous considérons la forme (D) comme duale de (P) . Nous avons la proposition suivante:

Proposition 1.2.1 (Dualité faible)

Si X est admissible pour (P) et (y,S) pour (D) , alors

$$C \bullet X - b^T y = X \bullet S \succeq 0. \quad (1.1)$$

nière suivante:

En utilisant cette notation, nous pouvons réécrire nos problèmes de la ma-

$$\mathcal{A}^*v = \sum_{i=1}^m v_i A_i.$$

Ainsi l'adjoint de \mathcal{A} est une application de \mathbb{R}^m vers $\mathbb{R}^{n \times n}$ donnée par

$$\begin{aligned} (\mathcal{A}X)^T v &= \sum_{i=1}^m (A_i \bullet X) v_i \\ &= \left(\sum_{i=1}^m v_i A_i \right) \bullet X. \end{aligned}$$

Notons que pour tout $X \in \text{SR}^{n \times n}$ et $v \in \mathbb{R}^m$,

$$\mathcal{A}X := (A_i \bullet X)_{i=1}^m \in \mathbb{R}^m.$$

Pour plus de facilité par la suite, nous allons introduire quelques notations. Nous définissons l'opérateur linéaire $\mathcal{A} : \text{SR}^{n \times n} \rightarrow \mathbb{R}^m$ par

car $\text{tr}(PQ) = \text{tr}(QP)$ et $X^{1/2} S X^{1/2}$ est semi-définie positive puisque S l'est et que la trace d'une matrice semi-définie positive est positive ou nulle. \square

$$\begin{aligned} X \bullet S &= \text{tr}(XS) \\ &= \text{tr}(X^{1/2} X^{1/2} S) \\ &= \text{tr}(X^{1/2} S X^{1/2}) \\ &\geq 0 \end{aligned}$$

De plus, puisque la matrice X est semi-définie positive, elle possède une ra-

cine carrée $X^{1/2}$, et donc

$$\begin{aligned} C \bullet X - b^T y &= \left(\sum_{i=1}^m y_i A_i + S \right) \bullet X - b^T y \\ &= \sum_{i=1}^m (A_i \bullet X) y_i + S \bullet X - b^T y \\ &= S \bullet X \\ &= X \bullet S. \end{aligned}$$

Nous avons successivement

Preuve:

Supposons que nous avons une matrice symétrique $M(z)$ dépendant de manière linéaire (affine) d'un vecteur z . Nous voulons minimiser la plus grande valeur propre de $M(z)$ en choisissant le $z \in \mathbb{R}^m$. Notons que

1.3.1 Minimisation de la plus grande valeur propre.

Dans cette section, nous allons présenter une série d'exemples de problèmes SDP.

1.3 Exemples.

Nous appelons saut de dualité, la différence entre la valeur optimale de (P) et celle de (D) . Par la Proposition 1.2.1, le saut de dualité est toujours positif ou nul. La dualité forte affirme que le saut de dualité est nul et que les deux problèmes atteignent leur optimum lorsque les deux problèmes sont admissibles, mais cette dualité forte n'a pas lieu pour tous les problèmes SDP. Il faut imposer certaines conditions. Nous en parlerons en détail dans le Chapitre 2.

La chaîne d'équations de la dualité faible peut alors s'écrire

$$\begin{aligned}
 C \bullet X - b^T y &= (A^* y + S) \bullet X - b^T y \\
 &= (AX)^T y + S \bullet X - b^T y \\
 &= S \bullet X \\
 &= X \bullet S.
 \end{aligned}$$

$$(D) \left\{ \begin{array}{l} \max_{y,S} \\ \text{s.c.} \\ A^* y + S = C \\ S \succeq 0. \end{array} \right.$$

$$(P) \left\{ \begin{array}{l} \min \\ \text{s.c.} \\ AX = b \\ X \succeq 0, \end{array} \right.$$

et

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^T A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix} \begin{pmatrix} I & A^{-1} B \\ 0 & I \end{pmatrix}$$

Preuve:

Alors

Si $A \succ 0$.

Proposition 1.3.1 (Complément de Schur)

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succeq 0 \text{ (} \succ 0 \text{)}$$

si et seulement si

$$C - B^T A^{-1} B \succeq 0 \text{ (} \succ 0 \text{)}.$$

suivante:

Soit une matrice $P \in \mathbb{R}^{m \times n}$. Considérons tout d'abord la proposition

1.3.2 Minimisation de la norme-2 d'une matrice.

$$\left\{ \begin{array}{l} \max -\lambda \\ \text{s.c. } \lambda I - M(z) + S = 0 \\ S \succeq 0. \end{array} \right.$$

où la variable est $y := (\lambda; z)$. Ce problème peut encore s'écrire

$$(1.2) \quad \left\{ \begin{array}{l} \max -\lambda \\ \text{s.c. } \lambda I - M(z) \succeq 0, \end{array} \right.$$

Nous obtenons le problème SDF suivant, sous forme duale:

$$\begin{aligned} \lambda^{max}(M(z)) &\leq \lambda \\ \Leftrightarrow \lambda^{max}(M(z) - \lambda I) &\leq 0 \\ \Leftrightarrow \lambda^{min}(\lambda I - M(z)) &\geq 0 \\ \Leftrightarrow \lambda I - M(z) &\succeq 0. \end{aligned}$$

où la variable est $y := (\lambda; z)$.

$$(1.3) \quad \left\{ \begin{array}{l} \max -\lambda \\ \text{s.c.} \end{array} \right. \left(\begin{array}{cc} P(z)^T & P(z) \\ \lambda I & \lambda I \end{array} \right) \succeq 0,$$

Nous pouvons alors résoudre le problème de minimisation de $\|P(z)\|_2$ où $P(z)$ dépend de manière affine de z , en résolvant le problème SDF suivant:

$$\begin{aligned} & \iff \lambda \geq \|P\|_2 \\ & \iff \lambda^2 \geq \lambda_{\max}(P^T P) \\ & \iff \lambda^2 I - P^T P \succeq 0 \\ & \iff \lambda I - \frac{1}{\lambda} P^T P \succeq 0 \\ & \iff \left(\begin{array}{cc} \lambda I & P \\ P^T & \lambda I \end{array} \right) \succeq 0 \end{aligned}$$

nous

Supposons maintenant que $\lambda > 0$, alors par la Proposition 1.3.1, nous obtenons

$$\begin{aligned} & \iff \left(\begin{array}{cc} 0 & P \\ P^T & 0 \end{array} \right) \succeq 0 \\ & \iff P = 0 \\ & \iff \|P\|_2 \leq 0 \end{aligned}$$

Alors,

où $\|P\|_2 = \max \sqrt{\lambda(P^T P)}$. En effet, supposons tout d'abord que $\lambda = 0$.

$$\left(\begin{array}{cc} \lambda I & P \\ P^T & \lambda I \end{array} \right) \succeq 0,$$

si et seulement si

$$\lambda \geq \|P\|_2$$

Dès lors, nous pouvons voir que

est (semi-)définie positive si et seulement si $C - B^T A^{-1} B$ l'est également. \square

$$\begin{pmatrix} \lambda & b^T y \\ b^T y & d^T y \end{pmatrix} \succeq 0.$$

si et seulement si

$$\lambda \geq \frac{d^T y}{(b^T y)^2}$$

où nous supposons que $d^T y > 0$ pour tout y admissible. Puisque la fonction objectif peut s'écrire comme $(b^T y)(d^T y)^{-1}(b^T y)$, en comparaison avec la Proposition 1.3.1, nous avons pour y admissible

$$\left\{ \begin{array}{l} \min \frac{d^T y}{(b^T y)^2} \\ \text{s.c. } A^T y \leq c, \end{array} \right.$$

Considérons le problème

convexe.

1.3.4 Un problème de programmation non linéaire quasi-

tard. Notons que le dual de ce problème nous donne une matrice $X \in \mathbb{R}^{n \times n}$ et donc étant la $i^{\text{ème}}$ colonne de A^T .
 où, dans nos notations standards, $C = \text{Diag}(c)$ et $A_i = \text{Diag}(a_i)$, avec a_i

$$\left\{ \begin{array}{l} \max b^T y \\ \text{s.c. } \text{Diag}(c - A^T y) \succeq 0, \end{array} \right.$$

peut s'écrire comme un problème SDF sous sa forme duale:

$$(LP) \left\{ \begin{array}{l} \max b^T y \\ \text{s.c. } A^T y \leq c \end{array} \right.$$

le problème linéaire

Soient la matrice $A \in \mathbb{R}^{m \times n}$ et les vecteurs $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$,

1.3.3 La programmation linéaire.

La question est de savoir si nous pouvons aussi réduire X à un élément diagonale par blocs avec des blocs 1×1 , et donc nous retrouvons le dual de (LP)

$$\left\{ \begin{array}{l} \min_{X \in S} \mathbb{R}^{n \times n} \\ C \bullet X \\ AX = b \\ X \succeq 0. \end{array} \right.$$

Son dual est

$$\left\{ \begin{array}{l} \max_{(y,S) \in \mathbb{R}^m \times S} \\ b^T y \\ A^* y + S = C \\ S \succeq 0. \end{array} \right.$$

où $M_j \in \mathbb{R}^{n_j \times n_j}$ pour $j = 1, \dots, k$. Supposons que C et tous les A_i se trouvent dans S . Alors, tout S admissible du problème dual, avec $A^* y + S = C$, se trouve aussi dans S . Le problème (D) peut alors s'écrire

$$M = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & M_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & M_{kk} \end{pmatrix},$$

Dans les deux derniers exemples, nous avons vu le cas où les matrices C et A_i partagent la même structure diagonale par blocs. Soit S , l'espace des matrices symétriques diagonales par blocs de la forme

où C et les A_i sont des diagonales par blocs, avec m blocs 1×1 et un bloc 2×2 .

$$\text{Diag} \left(\text{Diag}(c - A^T y), \begin{pmatrix} \lambda & b^T y \\ b^T y & d^T y \end{pmatrix} \right) \succeq 0,$$

Ce problème possède deux contraintes semi-définies positives qui peuvent être combinées en une seule:

$$\left\{ \begin{array}{l} \max_{\lambda, y} -\lambda \\ \text{s.c.} \\ \text{Diag}(c - A^T y) \preceq 0 \\ \begin{pmatrix} \lambda & b^T y \\ b^T y & d^T y \end{pmatrix} \succeq 0. \end{array} \right.$$

Notre problème non linéaire peut alors s'écrire comme suit:

$$\begin{pmatrix} I & y^T G^i \\ G^i y & d_i^T y + \epsilon_i \end{pmatrix} \succeq 0$$

où $f_i(y) := y^T C_i y - d_i^T y - \epsilon_i$, $C_i \succeq 0$, $i = 1, \dots, m$. Soit $C_i = G_i^T G_i$, $i = 1, \dots, m$. Alors, par la Proposition 1.3.1, $f_i(y) \leq 0$ peut encore s'écrire

$$\left\{ \begin{array}{l} \max y^T y \\ \text{s.c. } f_i(y) \leq 0, i = 1, \dots, m, \end{array} \right.$$

problème

Pour cet exemple, nous considérons l'optimisation d'une fonction linéaire avec des contraintes quadratiques convexes. Ainsi, nous nous intéressons au

CONVEXES.

1.3.5 La programmation avec contraintes quadratiques

Donc, si X est admissible pour (P) , alors $\tilde{X} \in S$ l'est aussi avec la même valeur objective. Nous pouvons, dès lors, sans perte de généralité, réduire X à un élément de S .

$$\tilde{X} = \begin{pmatrix} X_{11} & 0 & \dots & 0 \\ 0 & X_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{kk} \end{pmatrix} \succeq 0.$$

Ainsi, si $X \succeq 0$, alors $X_{jj} \succeq 0$ pour $j = 1, \dots, k$, et

$$C \bullet X = C_{11} \bullet X_{11} + \dots + C_{kk} \bullet X_{kk}.$$

$$A_i \bullet X = A_{i11} \bullet X_{11} + \dots + A_{ikk} \bullet X_{kk} \text{ pour tout } i,$$

alors,

$$X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & \dots & X_{kk} \end{pmatrix},$$

nière identique à M :

habituel. Considérons une matrice $X \in \text{SIR}^{n \times n}$ quelconque, et partageons-la de ma-

$$\begin{pmatrix} \alpha I & \alpha \\ y^T & 0 \end{pmatrix} \succeq 0$$

Ce qui n'est pas possible puisque $A \succeq 0$.
L'avantage de cette seconde formulation est que la contrainte

$$\begin{vmatrix} 0 & 2 \\ 2 & 0 \end{vmatrix} = -4.$$

dont le mineur principal suivant est négatif:

$$A = \begin{pmatrix} 0 & \begin{pmatrix} 2G^i y \\ 2 \end{pmatrix}^T \\ \begin{pmatrix} 2G^i y \\ 2 \end{pmatrix} & 0 \end{pmatrix} \succeq 0$$

et $A \succeq 0 \iff 1 + d_{TT}^i y + \epsilon_i > 0$, si $1 + d_{TT}^i y + \epsilon_i = 0$, nous obtenons la matrice

$$f_i(y) \leq 0 \iff d_{TT}^i y + \epsilon_i \geq 0 \iff 1 + d_{TT}^i y + \epsilon_i > 0$$

De plus, nous avons

$$\begin{aligned} & \iff f_i(y) \leq 0. \\ & \iff 4(d_{TT}^i y + \epsilon_i) - 4y^T C^i y \geq 0 \\ & \iff (1 + d_{TT}^i y + \epsilon_i)^2 - (1 - d_{TT}^i y - \epsilon_i)^2 - 4y^T C^i y \geq 0 \\ & \iff 1 + d_{TT}^i y + \epsilon_i - \frac{1 + d_{TT}^i y + \epsilon_i}{1 - d_{TT}^i y - \epsilon_i} [(1 - d_{TT}^i y - \epsilon_i)^2 + 4y^T C^i y] \geq 0 \\ & \iff 1 + d_{TT}^i y + \epsilon_i - \begin{pmatrix} 1 - d_{TT}^i y - \epsilon_i \\ 2G^i y \end{pmatrix}^T \begin{pmatrix} 1 + d_{TT}^i y + \epsilon_i \\ -1 \end{pmatrix} I^{-1} \begin{pmatrix} 1 - d_{TT}^i y - \epsilon_i \\ 2G^i y \end{pmatrix} \geq 0 \end{aligned}$$

$$A \succeq 0$$

En effet, si $1 + d_{TT}^i y + \epsilon_i > 0$,

$$A = \begin{pmatrix} (1 + d_{TT}^i y + \epsilon_i) I & \begin{pmatrix} 1 - d_{TT}^i y - \epsilon_i \\ 2G^i y \end{pmatrix}^T \\ \begin{pmatrix} 1 - d_{TT}^i y - \epsilon_i \\ 2G^i y \end{pmatrix} & 1 + d_{TT}^i y + \epsilon_i \end{pmatrix} \succeq 0.$$

ou encore, de façon alternative

$$(IQP) \left\{ \begin{array}{l} \max x^T C x \\ \text{s.c. } x_i \in \{-1, +1\} \quad i \in \mathbb{N} \end{array} \right.$$

Puisque chaque vecteur- (± 1) correspond à une coupure, le problème "max-cut" peut s'écrire comme un problème de programmation quadratique aux nombres entiers

$$w(\delta(K)) = \frac{1}{4} \sum_i \sum_j w_{ij} (1 - x_i x_j) = x^T C x.$$

Alors, nous avons

$$C_{ii} = \sum_j \frac{1}{4} w_{ij} \quad \forall i.$$

$$C_{ij} = -\frac{1}{4} w_{ij} \quad \forall i \neq j,$$

Définissons $C \in \mathbb{S}^n$ tel que

$$x_i x_j = \begin{cases} 1 & \text{si } ij \in \delta(K), \\ -1 & \text{sinon.} \end{cases}$$

Supposons que nous avons un graphe non orienté $G = (N, E)$ et un vecteur $w = (w_{ij}) \in \mathbb{R}_+^E$. Pour $K \subseteq N$, $\delta(K) = \{ij \in E : i \in K, j \notin K\}$ est la coupure déterminée par K , avec un poids égal à $w(\delta(K)) := \sum_{ij \in \delta(K)} w_{ij}$. Nous voulons trouver la coupure de poids maximal. Nous pouvons supposer que le graphe est complet (chaque noeud est adjacent à tous les autres) en posant $w_{ij} = 0$ pour les arêtes ij inexistantes, nous posons aussi que $w_{ii} = 0$ pour tout i . Nous utilisons $x \in \mathbb{R}^n$, tel que chaque $x_i = \pm 1$, pour représenter la coupure $\delta(K)$, où $x_i = 1$ si et seulement si $i \in K$. Il est évident que

1.3.6 Le problème "max-cut".

où K_2 est le cône de second ordre ou de Lorentz. En général, les cônes du second ordre sont préférés aux cônes semi-définis positifs.

$$\left(\begin{array}{c} \alpha \\ v \end{array} \right) \in K_2 := \left\{ \left(\begin{array}{c} \beta \\ w \end{array} \right) : \beta \geq \|w\|_2 \right\}$$

peut s'exprimer comme suit:

chaque edge est compté deux fois.

$$h(y) := \max_{x \in \mathbb{R}^n} x^T C x - \sum_{i=1}^m y_i (x_i^2 - 1) = e^T y - \min_{x \in \mathbb{R}^n} x^T (\text{Diag}(y)) x - C(x).$$

la fonction

du problème (NQCCP). Le problème dual est de minimiser, sur les $y \in \mathbb{R}^m$, La troisième manière d'obtenir un problème SDP est de considérer le dual problème devient le même que le problème (1.4). positif et qu'une telle matrice peut toujours être factorisée comme $P P^T$, le où e est un vecteur composé uniquement de 1. Puisque $P P^T$ est semi-définie $C \bullet (x x^T)$ par $C \bullet (P P^T)$, et les contraintes $x_i \in \{+1, -1\}$ par $\text{diag}(P P^T) = e$ vecteur de dimension 1. Nous pouvons alors remplacer la fonction objectif vecteurs. (P correspond au vecteur x , dont les colonnes correspondent aux dimension n à chaque noeud, et soit P , la matrice dont les colonnes sont ces de dimension 1 à chaque noeud. Associons maintenant un vecteur p_i unité de Ensuite, dans le problème (IQP), nous associons un vecteur x_i (± 1) unité

$$(1.4) \quad \left\{ \begin{array}{l} \max C \bullet X \\ \text{s.c. } X_{ii} = 1, i \in \mathbb{N} \\ X \succeq 0. \end{array} \right.$$

SDP:

Si nous supprimons la dernière contrainte, nous obtenons alors un problème

$$\left\{ \begin{array}{l} \max C \bullet X \\ \text{s.c. } X_{ii} = 1, i \in \mathbb{N} \\ X \succeq 0 \\ \text{rg}(X) = 1. \end{array} \right.$$

problème suivant:

x . Puisque $x^T C x = C \bullet (x x^T)$, le problème (NQCCP) est équivalent au ment, une telle matrice de rang 1 est de la forme $x x^T$ pour tout vecteur (± 1) entrées de X , X_{ij} , afin de ne pas avoir de confusion avec le vecteur x .) Inverse-La matrice $X \in \mathbb{R}^{n \times n}$ est telle que $X_{ii} = 1 \forall i$ et $X \succeq 0$. (Nous écrivons les les produits $x_i x_j$, lesquels sont les entrées d'une matrice $X = x x^T$ de rang 1. Tout d'abord, nous remarquons que le problème (NQCCP) est linéaire en problèmes sous forme de problèmes SDP.

Nous allons maintenant montrer trois façons différentes de mettre ces pro-

$$(NQCCP) \quad \left\{ \begin{array}{l} \max x^T C x \\ \text{s.c. } x_i^2 = 1 \quad i \in \mathbb{N}. \end{array} \right.$$

ou comme un problème quadratique avec contraintes quadratiques non convexes

Le minimum vaut 0 si $\text{Diag}(y) - C$ est semi-définie positive et $-\infty$ sinon. Le problème dual devient

$$\begin{cases} \min e^T y \\ \text{s.c. } \text{Diag}(y) - C \succeq 0. \end{cases}$$

C'est un problème SDP sous forme duale et son dual est précisément le problème SDP (1.4).

Chapitre 2

La dualité et la trajectoire centrale.

2.1 La dualité.

Nous allons, dans cette section, discuter de la relation entre les problèmes (P) et (D). Nous avons déjà montré la dualité faible et nous allons donner les conditions pour que la dualité forte ait lieu.

Nous savons déjà que (D) est le dual lagrangien de (P). Montrons que (P) est le dual lagrangien de (D). Puisque $A^*y + S = C$ est une équation entre matrices symétriques, la variable duale sera aussi une matrice symétrique que l'on note X . Le lagrangien s'écrit:

$$L(y, S, X) = b^T y - (A^*y + S - C) \bullet X.$$

Le problème dual lagrangien de (D) s'écrit

$$\left\{ \begin{array}{l} \min_X h(X) \\ \text{s.c. } X \in \text{STR}^{n \times n}, \end{array} \right.$$

où

$$h(X) = \sup_{y \in \mathbb{R}^n, S \succeq 0} L(y, S, X) = C \bullet S - \min_{y \in \mathbb{R}^n} (AX - b)^T y - \min_{S \succeq 0} S \bullet X.$$

Il est facile de voir que $h(X)$ vaut $+\infty$ sauf si $AX - b = 0$ et si $X \succeq 0$ puisque si $X \succeq 0$, $S \bullet X \succeq 0$ et $AS \succeq 0$. D'où, nous obtenons, comme dual lagrangien

avec une valeur optimale égale à 0. Pour ce problème, il n'y a pas de saut de qualité mais une des valeurs optimales n'est pas atteinte.

$$X = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

pour lequel la seule solution admissible (et donc optimale) est

$$\left\{ \begin{array}{l} \min \\ \text{s.c.} \end{array} \right. \begin{cases} 0 \bullet X \\ 1 \bullet X \\ -1 \bullet X \\ 0 \bullet X \\ 0 \bullet X \\ 0 \bullet X \\ 0 \bullet X \\ 0 \bullet X \\ 0 \bullet X \\ X \succeq 0 \end{cases}$$

problème est Nous pouvons voir que la région admissible est $\{(y_1, y_2) : y_1 > 0, y_2 > 0, y_1 y_2 \geq 1\}$. La valeur optimale est 0, mais n'est pas atteinte. Le dual de ce

$$\begin{pmatrix} y_1 & 1 \\ 1 & y_2 \end{pmatrix} \succeq 0.$$

De façon équivalente, nous pouvons écrire la contrainte

$$\left\{ \begin{array}{l} \max -y_1 \\ \text{s.c.} \end{array} \right. \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

◊ Considérons tout d'abord le problème

Nous allons maintenant présenter quelques exemples pour montrer que la dualité forte peut ne pas avoir lieu.

$$(P) \left\{ \begin{array}{l} \min_x \\ \text{s.c.} \end{array} \right. \begin{cases} C \bullet X \\ AX = b \\ X \succeq 0. \end{cases}$$

de (D), le problème

$$\begin{pmatrix} 0 & 0 & 0 \\ -y_2 & 0 & 0 \\ -y_1 & -y_2 & 0 \end{pmatrix} \succeq 0,$$

De façon équivalente, la contrainte peut s'écrire

$$\left\{ \begin{array}{l} \max 2y_2 \\ \text{s.c.} \end{array} \right. \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} y_2 \succeq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

avec une valeur optimale égale à 1. Le problème dual s'écrit

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

pour les ξ_i qui conviennent. Une solution optimale est

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & \xi_3 & \xi_4 \\ 0 & \xi_4 & 1 \end{pmatrix}$$

et puisqu'il doit être semi-défini positif,

$$\begin{pmatrix} 0 & \xi_1 & \xi_2 \\ \xi_1 & \xi_3 & \xi_4 \\ \xi_2 & \xi_4 & 1 - \xi_1 \end{pmatrix}$$

Tout X admissible pour les contraintes d'égalité est de la forme

$$\left\{ \begin{array}{l} \min \\ \text{s.c.} \\ X \succeq 0. \end{array} \right. \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \bullet X \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \bullet X = 0 \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \bullet X = 2$$

◊ Le second exemple est le suivant:

$$F(P) := \{X \in \text{SR}^{n \times n} : AX = b, X \succeq 0\},$$

$$F^0(P) := \{X \in F(P) : X \succ 0\},$$

Ayant donné des exemples où la dualité forte n'a pas lieu, nous allons maintenant nous intéresser aux conditions pour qu'elle ait lieu. C'est ce que nous appellerons la condition de Slater ou de stricte admissibilité. Définissons:

et donc, $y_2 = 0$ et $y_1 \geq 0$. Le vecteur $y = (0,0)$ est optimal avec une valeur optimale nulle. Il existe un saut de dualité infini entre les valeurs optimales.

$$\begin{pmatrix} -y_1 & -y_2 \\ 0 & -y_2 \end{pmatrix} \succeq 0$$

où la contrainte est équivalente à

$$\left\{ \begin{array}{l} \max \quad 2y_2 \\ \text{s.c.} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{array} \right.$$

qui ne peut pas être semi-défini positif. L'ensemble admissible étant vide, la valeur optimale est $+\infty$ par convention. Le problème dual s'écrit

$$\begin{pmatrix} 0 & 1 \\ 1 & \xi \end{pmatrix}$$

Tout X admissible pour les contraintes d'égalité est de la forme

$$\left\{ \begin{array}{l} \min \\ \text{s.c.} \end{array} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \bullet X \right. \\ \left. \begin{array}{l} \bullet X = 0 \\ \bullet X = 2 \\ X \succeq 0. \end{array} \right.$$

◊ Notre dernier exemple s'écrit

et donc, $y_2 = 0$ et $y_1 \geq 0$. Le vecteur $y = (0,0)$ est optimal avec une valeur optimale nulle. Les deux problèmes atteignent leurs valeurs optimales mais le saut de dualité n'est pas nul.

$\exists X \succeq 0, X \neq 0$ tel que $\forall \alpha > 0, \forall Z \in \mathcal{G}_2, Z + \alpha X \in \mathcal{G}_2$.

Comme $S \succeq 0$, l'ensemble $\{X \succeq 0 : S \bullet X \leq \beta\}$ est borné pour tout β positif ou nul. Dès lors, le problème a une région admissible compacte. La fonction objectif étant continue, ceci implique la première partie du théorème. Il reste à montrer que les valeurs optimales de (P) et de (D) sont les mêmes. Soit ζ^* la valeur optimale de (P) et ϵ positif. Nous voulons montrer l'existence d'une solution admissible de (D) avec la valeur de la fonction objectif au moins égale à $\zeta^* - \epsilon$. Considérons les deux ensembles $\mathcal{G}_1 := \text{SIR}^+_{n \times n}$ et $\mathcal{G}_2 := \{X \in \text{SIR}^{n \times n} : AX = b, C \bullet X \leq \zeta^* - \epsilon\}$. Ces deux ensembles sont convexes, disjoints, non vides et n'ont aucune direction de récession commune. En effet, s'il existe une direction de récession commune à \mathcal{G}_1 et à \mathcal{G}_2 , alors

$$(P') \left\{ \begin{array}{l} \min_X C \bullet X \\ \text{s.c.} \\ AX = b \\ \widehat{S} \bullet X \leq \widehat{S} \bullet \widehat{X} \\ X \succeq 0. \end{array} \right.$$

Preuve: Soit $\widehat{X} \in F(P)$ et $(\widehat{y}, \widehat{S}) \in F^0(D)$. Alors, nous pouvons ajouter la contrainte $C \bullet X \leq C \bullet \widehat{X}$ à (P) sans changer sa valeur optimale ou l'ensemble de ses solutions optimales. Mais, en utilisant (1.1), cette inégalité est équivalente à $\widehat{S} \bullet X = C \bullet X - b^T \widehat{y} \leq C \bullet \widehat{X} - b^T \widehat{y} = \widehat{S} \bullet \widehat{X}$. Donc, (P) a la même valeur optimale et le même ensemble de solutions optimales que

Théorème 2.1.1 (Dualité forte)
 Supposons que $F(P)$ et $F^0(D)$ sont non vides.
 Alors, (P) possède un ensemble compact non vide de solutions optimales
 et les valeurs optimales de (P) et (D) sont les mêmes.

Nous disons aussi que la condition d'indépendance linéaire est satisfaite si les matrices A_1, \dots, A_m sont linéairement indépendantes dans $\text{SIR}^{n \times n}$. Le théo-

$$F(D) := \{(y, S) \in \mathbb{R}^m \times \text{SIR}^{n \times n} : A^*y + S = C, S \succeq 0\}$$

$$F^0(D) := \{(y, S) \in F(D) : S \succ 0\}.$$

rème de dualité forte s'écrit:

Ce résultat est symétrique entre (P) et (D). En effet, chacun de ces problèmes peut se réécrire de la même manière que l'autre. Pour ce faire,

de dualité. \square
 nous avons une solution admissible pour (D) avec une valeur de la fonction objectif au moins égale à $\zeta^* - \epsilon$. Puisque ϵ est arbitraire, il n'y a pas de saut $\eta = 1$. Nous obtenons alors $C - A^*y = S \succeq 0$ et $b^T y \geq \zeta^* - \epsilon - \sigma \geq \zeta^* - \epsilon$, et strictement positif, et en divisant y , S et σ par η , nous pouvons supposer que $\widehat{X} \bullet (-A^*y) = \widehat{X} \bullet S \geq 0$, ce qui est contradictoire. Cela implique que η est Supposons que $\eta = 0$, alors, $-b^T y \leq \sigma < 0$ et aussi $-b^T y = -(AX)^T y =$

$$C\eta - A^*y = S, \quad (\zeta^* - \epsilon)\eta - b^T y \leq \sigma.$$

Si nous faisons tendre λ vers l'infini, nous obtenons $u^T S u \geq 0$. Cette inégalité étant valable pour tout $u \in \mathbb{R}^n$, nous en déduisons que $S \succeq 0$. D'autre part, $AX = b$ et $C \bullet X \leq \zeta^* - \epsilon$ impliquent que $S \bullet X \leq \sigma$. Par le théorème de la dualité en programmation linéaire (il n'existe pas de contraintes semi-définies ici), il existe $y \in \mathbb{R}^m$ et $\eta \geq 0$ avec

$$u^T S u > \frac{\lambda}{\sigma}$$

et donc

$$\sigma > S \bullet \lambda u u^T = \lambda S \bullet u u^T = \lambda u^T S u$$

En effet, nous avons

Comme $0 \in \mathcal{G}_1$, σ est négatif et comme $\lambda u u^T \in \mathcal{G}_1$, $\forall \lambda > 0$ et $u \in \mathbb{R}^n$, $S \succeq 0$.

$$S \bullet X > \sigma \quad \forall X \in \mathcal{G}_1, S \bullet X < \sigma \quad \forall X \in \mathcal{G}_2.$$

Par le théorème de séparation des convexes (voir Rockafellar [6], Corollaire 11.4.1), il existe $S \in S\mathbb{P}^{n \times n}$ et $\sigma \in \mathbb{R}$ avec compact.

$\forall \alpha > 0$. Ce qui contredit le fait que l'ensemble des solutions optimales est D'où, $C \bullet X = 0$. Mais alors $Y + \alpha X$, avec $X \neq 0$, est une solution de (P)

$$C \bullet Y + \alpha C \bullet X \geq C \bullet Y.$$

(P), Alors $Y + \alpha X$ est admissible $\forall \alpha > 0$ et $\zeta^* - \epsilon$ et que $C \bullet Z + \alpha C \bullet X \leq \zeta^* - \epsilon$, $C \bullet X \leq 0$. Soit Y une solution de Comme $AZ = b$ et que $A(Z + \alpha X) = b$, $AX = 0$. De plus, comme $C \bullet Z \leq$

$$(P) \left\{ \begin{array}{l} \min_X C \bullet X \\ AX = b \\ X \succeq 0 \end{array} \right. \text{ s.c.}$$

D'où

$$\begin{aligned} C \bullet D &= C \bullet (X + (D - X)) \\ &= C \bullet X + C \bullet (D - X) \\ &= C \bullet X + C \bullet \sum_{j=1}^k w_j G_j \\ &= C \bullet X + \sum_{j=1}^k w_j C \bullet G_j \\ &= C \bullet X + h^T w. \end{aligned}$$

De plus,

$$D = X + \sum_{j=1}^k w_j G_j \text{ avec } w \in \mathbb{R}^k.$$

puisque $A(D - X) = AD - AX = b - b = 0$. D'où, nous pouvons écrire

$$D = X + (D - X) \text{ avec } D - X \in \ker A$$

En effet, pour tout X admissible pour (P) , nous avons

$$C \bullet D - \max_{w \in \mathbb{R}^k, X \in \text{SIR}^{n \times n}} \{h^T w : \sum_{j=1}^k w_j G_j + X = D, X \succeq 0\}.$$

admissible. Avec cette hypothèse, nous choisissons $D \in \text{SIR}^{n \times n}$ satisfaisant $AD = b$ et soit G_1, \dots, G_k une base du complément orthogonal de l'espace engendré par les matrices $A_i, i = 1, \dots, m$, dans $\text{SIR}^{n \times n}$. De plus, soit $h_j := C \bullet G_j, j = 1, \dots, k$. Alors, le problème (P) est équivalent au problème

linéaire est satisfait. En effet, si les $A_i, i = 1, \dots, m$, sont linéairement dépendantes et que $A^*y = 0$ implique $b^T y = 0$, nous pouvons choisir une base, disons $A_i, i = 1, \dots, k$, et supprimer les $m - k$ dernières contraintes primales et variables duales, pour obtenir des problèmes où l'hypothèse d'indépendance linéaire est vérifiée. Si $A^*y = 0$ et $b^T y > 0$, alors (P) est non admissible et (D) non borné s'il est

nous supposons, sans perte de généralité, que la condition d'indépendance

Nous pouvons dès lors utiliser cette construction pour passer entre les deux formes. Notons de plus que, étant donné la condition d'indépendance linéaire, pour tout S , il y a au plus un y tel que (y, S) est admissible pour (D) , ce qui nous permet d'étendre les résultats vrais pour S à la paire (y, S) . Si nous appliquons le Théorème de dualité forte 2.1.1 à la nouvelle construction, nous obtenons

$$C \bullet D - \min_{S \in \mathbb{R}^{n \times n}} \{ D \bullet S : G_j \bullet S = h_j, j = 1, \dots, k, S \succeq 0 \}.$$

est équivalent à

$$(D) \left\{ \begin{array}{l} \max_{y, S} b^T y \\ \text{s.c.} \\ A^* y + S = C \\ S \succeq 0 \end{array} \right.$$

D'où

$$\begin{aligned} C \bullet D &= S \bullet D + A^* y \bullet D \\ &= S \bullet D + (AD)^T y \\ &= S \bullet D + b^T y. \end{aligned}$$

De plus,

$$\begin{aligned} \Leftrightarrow C - S \in \text{span}\{A_i\} \\ \Leftrightarrow G_j \bullet (C - S) = 0, j = 1, \dots, k \\ \Leftrightarrow G_j \bullet S = G_j \bullet C, j = 1, \dots, k. \end{aligned}$$

En effet, nous avons d'abord

$$C \bullet D - \min_{S \in \mathbb{R}^{n \times n}} \{ D \bullet S : G_j \bullet S = h_j, j = 1, \dots, k, S \succeq 0 \}.$$

De plus, le problème (D) est équivalent au problème

$$C \bullet D - \max_{w \in \mathbb{R}^k, X \in \mathbb{R}^{n \times n}} \{ h^T w : \sum_j w_j G_j + X = D, X \succeq 0 \}.$$

est équivalent à

Pour $n = 1$, nous obtenons une fonction simple $-\ln x$, qui est définie sur l'axe positif et tend vers $+\infty$ lorsque x tend vers 0 par valeurs positives. De la

$$f(X) := \begin{cases} -\ln \det X & \text{si } X \succ 0, \\ +\infty & \text{sinon.} \end{cases}$$

Nous définissons la fonction barrière logarithmique pour le cône $\text{STR}_{n \times n}^+$ des matrices semi-définies positives par

2.2 La fonction barrière logarithmique.

La dernière équation supposant que le saut de dualité est nul, $X \bullet S = 0$, a été remplacée par la condition similaire que le produit des matrices est nul. En effet, si deux matrices U et V sont semi-définies positives, alors $U \bullet V = 0$ si et seulement si $UV = 0$.

$$(2.1) \quad \begin{aligned} A^*y + S &= C, & S \succeq 0, \\ AX &= b, & X \succeq 0, \\ XS &= 0. \end{aligned}$$

Finalement, si la dualité forte a lieu, nous avons les conditions nécessaires et suffisantes

Corollaire 2.1.2
 Supposons que la condition d'indépendance linéaire est satisfaite et que (P) et (D) possèdent des solutions strictement admissibles.
 Alors, chacun possède un ensemble compact non vide de solutions optimales et il n'existe pas de saut de dualité.

Nous trouvons aussi

Corollaire 2.1.1
 Supposons que la condition d'indépendance linéaire est satisfaite et que $F^0(P)$ et $F(D)$ sont non vides.
 Alors, (D) possède un ensemble compact non vide de solutions optimales, et il n'existe pas de saut de dualité.

Preuve: Soit la fonction ϕ telle que $\forall X \succ 0$ et $\forall H \in \text{SIR}_{n \times n}^+, \phi(\alpha) = f(X + \alpha H)$.

Théorème 2.2.1 $f(X) := -\ln \det X$ est une barrière n -normale pour $\text{SIR}_{n \times n}^+$.

Nous avons alors

$$f(\alpha X) = f(X) - \theta \ln \alpha, \forall X \succ 0, \alpha > 0.$$

De plus, f est une barrière θ -normale pour $\text{SIR}_{n \times n}^+$ (ou pour $\text{SIR}_{n \times n}^+$) si f est convexe, self-concordant, vérifie la propriété barrière, et est logarithmiquement homogène de degré θ :

$$|\phi'''(0)| \leq 2[\phi'(0)]^{3/2}.$$

et $\forall H, \phi$ satisfait: que f est self-concordant si f est convexe, trois fois différentiable et si, $\forall X$ si $\forall X \succ 0$ et $\forall H \in \text{SIR}_{n \times n}^+, \phi(\alpha) = f(X + \alpha H)$ est convexe en α . Nous disons Introduisons la notion de "self-concordance". Nous savons que f est convexe,

$$\begin{aligned} f''(X)[H] &= X^{-1} H X^{-1} \text{ et } D^2 f(X)[H, J] = X^{-1} H X^{-1} \bullet J, \\ f'''(X)[H, J] &= -X^{-1} H X^{-1} J X^{-1} - X^{-1} J X^{-1} H X^{-1}, \\ f'(X) &= -X^{-1} \text{ et } Df(X)[H] = -X^{-1} \bullet H, \end{aligned}$$

Soient $X \succ 0, H \in \text{SIR}_{n \times n}$ et $J \in \text{SIR}_{n \times n}$, nous avons positivité étant imposée par la propriété barrière de f . où il n'y a qu'une seule contrainte linéaire, la contrainte implicite de définie

$$BP(v) \begin{cases} \min_x C \bullet X + v f(X) \\ \text{s.c. } AX = b \quad (X \succ 0), \end{cases}$$

barrières paramétrisées par $v > 0$) que X soit semi-définie positive, par la séquence de problèmes (problèmes est de remplacer le problème (P) , avec la contrainte quelque peu gênante positives. Nous disons que f possède la propriété barrière pour $\text{SIR}_{n \times n}^+$. L'idée $+\infty$ lorsque X approche une matrice $\underline{X} \in \partial \text{SIR}_{n \times n}^+$ par des valeurs définies même manière, f est définie pour les matrices définies positives et tend vers

$$\begin{aligned}
 \phi''(0) &= \text{tr}(\text{Diag}(\lambda)\text{Diag}(\lambda)) = \sum \lambda_i^2 \\
 &= \sum \lambda_i^2 \\
 &= -2(X^{-1}HX^{-1}) \bullet H \\
 &= -2\text{tr}(X^{-1}HX^{-1}HX^{-1}H) \\
 &= -2\text{tr}(V^2HV^2H) \\
 &= -2\text{tr}([VHV][VHV])
 \end{aligned}$$

et

Si on note $\lambda := \lambda(VHV)$, les valeurs propres de VHV , nous obtenons successivement

$$\phi''(0) = \text{tr}([VHV][VHV]).$$

Prenons maintenant $V := X^{-1/2} > 0$. Nous avons que $f''(X)$ est un opérateur défini positif et donc non singulier. nul, alors ϕ et donc f sont strictement convexes. Nous avons montré aussi D'où, comme $\phi''(\alpha) \geq 0$, ϕ est convexe. En fait, $\|VHV\|_F^2 > 0$ si H est non

$$\begin{aligned}
 \phi''(\alpha) &= (V^2HV^2) \bullet H \\
 &= \text{tr}(V^2HV^2H) \\
 &= \text{tr}([VHV][VHV]) \\
 &= \|VHV\|_F^2 \\
 &\geq 0.
 \end{aligned}$$

où $\underline{X} := X + \alpha H$. Comme \underline{X} est définie positive, $\exists V = \underline{X}^{-1/2}$ tel que

$$\phi''(\alpha) = D^2f(\underline{X})[H, H] = (\underline{X}^{-1}H\underline{X}^{-1}) \bullet H$$

nous avons

Elle est bornée sur l'ensemble des α vérifiant $X + \alpha H \succ 0$. Sur cet ensemble,

Nous avons encore

$$\|H\|_X = \|\lambda(X)H X^{-1/2} H X^{-1/2}\|_2$$

$$= \|X^{-1/2} H X^{-1/2} H X^{-1/2}\|_F.$$

$$\|J\|_X^* := \|(f''(X))^{-1} J \bullet J\|_{1/2}.$$

et la norme $-X$ duale d'une matrice symétrique par

$$\|H\|_X := (f''(X) H \bullet H)_{1/2}$$

d'une matrice symétrique par

Comme l'opérateur $f''(X)$ est défini positif, nous pouvons définir la norme $-X$

□

$$f(\alpha X) = -\ln \det(\alpha X)$$

$$= -\ln \alpha^n \det X$$

$$= -\ln \det X - \ln \alpha^n$$

$$= f(X) - n \ln \alpha.$$

Ce qui démontre que f est self-concordant. La fonction f vérifie évidemment la propriété barrière et

$$|\phi'''(0)| = 2 \left| \sum \lambda_i^2 \right|$$

$$\leq 2 \sum |\lambda_i^2|$$

$$= 2 \|\lambda\|_3^2$$

$$\leq 2 \|\lambda\|_2^2$$

$$= 2 [\phi''(0)]^{3/2}.$$

D'où,

Preuve: $f'(\alpha X) = \alpha^{-1} f'(X)$:

Proposition 2.2.1

Pour $\alpha > 0$, X d'ordre n ,

$$f'(\alpha X) = \alpha^{-1} f'(X), \quad f''(\alpha X) = \alpha^{-2} f''(X),$$

$$f'(X) \bullet X = -n, \quad f''(X) X = -f'(X),$$

$$\|X\|_X = \sqrt{n}, \quad \|f'(X)\|_X^* = \sqrt{n}.$$

Notons que $|J \bullet H| \leq \|J\|_X^* \|H\|_X$ comme dans l'inégalité de Cauchy-Schwarz. Nous avons les propriétés suivantes:

$$\begin{aligned} \|J\|_X^* &:= [f''(X)]^{-1} J \bullet J \\ &= X J X \bullet J \\ &= \text{tr}(X J X J) \\ &= \text{tr}([X^{1/2} J X^{1/2}][X^{1/2} J X^{1/2}]) \\ &= \|X^{1/2} J X^{1/2}\|_2^2. \end{aligned}$$

En effet, nous avons successivement

$$\|J\|_X^* = \|X^{1/2} J X^{1/2}\|_2.$$

De plus,

$$\begin{aligned} \|H\|_X^* &:= f''(X) H \bullet H \\ &= X^{-1} H X^{-1} \bullet H \\ &= \text{tr}(X^{-1} H X^{-1} H) \\ &= \text{tr}([X^{-1/2} H X^{-1/2}][X^{-1/2} H X^{-1/2}]) \\ &= \|X^{-1/2} H X^{-1/2}\|_2^2 \\ &= \|\lambda(X^{-1/2} H X^{-1/2})\|_2^2. \end{aligned}$$

En effet, nous pouvons écrire successivement

Théorème 2.3.1
 Supposons que $F^0(P)$ et $F^0(D)$ sont non vides et que l'hypothèse d'indépendance linéaire est satisfaite.
 Alors, $\forall v > 0$, il existe une solution unique $(X(v), y(v), S(v))$ de $CPE(v)$.
 De plus, $X(v)$ est solution unique de $BP(v)$ et $(y(v), S(v))$ de $BD(v)$.
 Finalement, si les problèmes primal et dual ne sont pas strictement admissibles, $CPE(v)$, $BP(v)$ et $BD(v)$ n'ont pas de solution.

Ce sont les équations de la trajectoire centrale. Notons que, excepté le membre de gauche de la dernière équation, ces équations coïncident avec les conditions d'optimalité (2.1). La dernière équation est appelée condition de centralité. Si $BD(v)$ possède une solution optimale (y, S) , un raisonnement similaire montre que, pour $X > 0$, les équations (2.2) restent valables. Nous obtenons le théorème suivant:

$$(2.2) \quad CPE(v) \begin{cases} AX \\ A^*y + S = C, S > 0, \\ = b, X > 0, \\ XS = vI. \end{cases}$$

Posez $S := vX^{-1} > 0$. Alors, $(y, X) \in F^0(D)$ et nous avons une solution pour l'ensemble d'équations

$$C - vX^{-1} - A^*y = C + v f'(X) - A^*y = 0.$$

$y \in \mathbb{R}^m$, nous avons et par le théorème de Lagrange (le gradient du lagrangien est nul), pour Supposons que $BP(v)$ possède une solution optimale X . Alors, $X \in F^0(P)$ de l'autre à une constante additive près. Il n'est pas difficile de vérifier que chacun d'eux est en fait le dual lagrangien

$$BD(v) \begin{cases} \max_{y,S} b^T y - v f(S) \\ \text{s.c. } A^*y + S = C \quad (S > 0). \end{cases}$$

et

$$BP(v) \begin{cases} \min_X C \bullet X + v f(X) \\ \text{s.c. } AX = b \quad (X > 0) \end{cases}$$

(paramétrisés par $v > 0$) comme

Preuve:

- Montrons tout d'abord l'existence d'une solution.

Soit $\hat{X} \in F^0(P)$ et $(\hat{y}, \hat{S}) \in F^0(D)$ et considérons le problème $BP(v)$. Supposons $\sigma := \lambda_{min}(\hat{S}) > 0$. \hat{X} est admissible pour $BP(v)$, et par (1.1), pour X admissible, $C \bullet X$ diffère d'une constante par rapport à $\hat{S} \bullet X$. D'où, tout comme dans la preuve du Théorème de dualité forte 2.1.1, $BP(v)$ possède le même ensemble de solutions optimales que le problème suivant:

$$BP'(v) \left\{ \begin{array}{l} \min_X \hat{S} \bullet X + v f(X) \\ \text{s.c. } AX = b \\ \hat{S} \bullet X + v f(X) \leq \hat{S} \bullet \hat{X} + v f(\hat{X}) \quad (X \succ 0). \end{array} \right.$$

Le but est de montrer que le problème $BP(v)$ revient à la minimisation d'une fonction continue sur un ensemble compact, ce qui donne l'existence de solutions.

Supposons que X est admissible pour $BP'(v)$ et posons $\lambda := \lambda(X)$, les valeurs propres de X , et $e \in \mathbb{R}^n$, le vecteur composé uniquement de 1. Alors, nous avons $\lambda > 0$ et

$$\sigma e^T \lambda - v \sum \ln \lambda_j = \sigma I \bullet X + v f(X)$$

$$\begin{aligned} \hat{S} \bullet X + v f(X) &\leq \hat{S} \bullet X + v f(X) \quad \text{car } \sigma := \lambda_{min}(\hat{S}) \\ &\leq \hat{S} \bullet \hat{X} + v f(\hat{X}) \end{aligned}$$

$$=: \alpha$$

où la première égalité est vraie par

$$\begin{aligned} I \bullet X &= \text{tr}(X) \\ &= \sum \lambda_j \\ &= e^T \lambda \end{aligned}$$

et

$$\begin{aligned} f(X) &= - \ln \det X \\ &= - \ln(\prod \lambda_j) \\ &= - \sum \ln \lambda_j. \end{aligned}$$

Si le problème (P) n'est pas strictement admissible, il n'existe pas de solution admissible, aucune solution n'est possible.

- Finalement, montrons que si nous n'avons pas l'hypothèse de stricte admissibilité, nous n'avons pas l'unicité. Les équations $CPE(v)$ fournissent aussi les conditions nécessaires et suffisantes pour le problème barrière dual. $\mathcal{A}y + S = C$ ainsi que l'hypothèse d'indépendance linéaire impliquent que S est unique. L'équation $XS = vI$ montre que S est unique et l'équation pour $BP(v)$. De plus, l'objectif est strictement convexe, d'où le minimum pour $BP(v)$. Ainsi, toute solution de $CPE(v)$ est un minimum. Puisque le problème barrière est convexe, ces conditions sont aussi suffisantes pour l'optimalité. Montrons maintenant l'unicité.

- Montrons maintenant l'unicité. Nous venons de voir que la fonction objectif de $BP'(v)$ est continue sur l'ensemble admissible compact, nous avons donc l'existence d'un minimum pour $BP(v)$. Un tel minimum doit satisfaire les conditions nécessaires et donc compact.

et nous avons un ensemble admissible borné. De plus, $\lambda_j \geq \tau > 0 \forall j$ implique que $S \bullet X + v f(X)$ est continue sur cet ensemble et que celui-ci est aussi fermé

$$\|X\|_F = \|\lambda\|_2 \leq \sqrt{n\tau}$$

Nous obtenons alors

$$\sigma \lambda_j - v \ln \lambda_j \leq \alpha - \underbrace{\sum_{i \neq j} (\sigma \lambda_i - v \ln \lambda_i)}_{\geq \beta} \leq \alpha - (n-1)\beta.$$

Alors, l'inégalité (2.3) implique que $\lambda_j \in [\tau, T] \forall j$. En effet, nous avons

$$\sigma \tau - v \ln \tau > \alpha - (n-1)\beta \text{ pour } \tau \in (0, T] \text{ ou } \tau \in [\tau, +\infty).$$

minimum et supposons que

La fonction $\sigma \tau - v \ln \tau$ possède un unique minimum $\tau^* = \frac{\sigma}{v}$ et tend vers $+\infty$ quand τ tend vers 0 ou vers $+\infty$. Posons β , la valeur de la fonction au

$$\sum (\sigma \lambda_j - v \ln \lambda_j) \leq \alpha \tag{2.3}$$

D'où,

optimales ne change pas si nous remplaçons la fonction objectif de (P) par 2.3.1 en choisissant $\hat{X} \in F_0(P)$ et $(\hat{y}, \hat{S}) \in F_0(D)$. L'ensemble des solutions borné, nous procédons de la même manière que pour la preuve du Théorème Pour montrer que (P) possède un ensemble de solutions optimales non vide et 0 quand ν tend vers 0.

La dernière partie résulte de l'existence de la trajectoire centrale puisque par (2.4), le saut de dualité associé à $X(\nu)$ et $(y(\nu), S(\nu))$ est $n\nu$, et approche de

Preuve:

Théorème 2.3.3
L'existence de solutions strictement admissibles pour (P) et (D) et l'hypothèse d'indépendance linéaire impliquent que (P) et (D) possèdent des ensembles de solutions optimales non vides et bornés avec un saut de dualité nul.

Nous pouvons dès lors utiliser la trajectoire centrale pour prouver la dualité forte. En effet, nous avons

$$C \bullet X(\nu) - b^T y(\nu) = X(\nu) \bullet S(\nu) = n\nu. \quad (2.4)$$

Théorème 2.3.2
Supposons que (P) et (D) possèdent tous deux des solutions strictement admissibles et que l'hypothèse d'indépendance linéaire est satisfaite. Si $X(\nu, y(\nu), S(\nu))$ résolvent les équations (2.2) pour une valeur particulière positive ν , alors, $X(\nu)$ est une solution strictement admissible pour (P) et $(y(\nu), S(\nu))$ est une solution strictement admissible pour (D) , avec un saut de dualité de valeur:

Nous venons d'établir l'existence d'une solution unique de $CPE(\nu)$ pour chaque ν positif. En prenant la trace de la dernière équation de (2.2), nous obtenons le théorème suivant:

Nous pouvons appliquer un argument similaire au dual. □

strictement admissible.

Le minimum de la fonction continue $\hat{S} \bullet X$ sur cet ensemble compact (contenant \hat{X}) est atteint, et l'ensemble de solutions optimales est borné et non vide. La preuve que l'ensemble de solutions duales optimales est borné et non vide, est similaire: nous partons avec le fait que la fonction objectif de maximisation $b^T y$ peut être remplacée par la minimisation de $\hat{X} \bullet S$ en utilisant (1.1). \square

$$\|X\|_F = \|\lambda\|_2 \leq \sqrt{n} \frac{\sigma}{(\hat{S} \bullet \hat{X})}.$$

$\hat{S} \bullet X$ et ajoutons la contrainte $\hat{S} \bullet X \leq \hat{S} \bullet \hat{X}$. Cette contrainte implique que toutes les valeurs propres de X sont bornées par $(\hat{S} \bullet \hat{X})/\sigma$ où $\sigma > 0$ est la plus petite valeur propre de \hat{S} . Cela montre que toutes les solutions optimales de (P) se trouvent dans un ensemble compact de solutions admissibles. En effet,

Chapitre 3

Méthodes de points intérieurs primales-duales.

Dans ce chapitre, nous nous intéressons aux méthodes de points intérieurs primales-duales, générant une séquence d'itérés approximant la trajectoire centrale et convergeant vers les solutions primale et duale. Nous verrons sous quelles conditions nous avons l'unicité des directions de recherche. Pour ce faire, nous supposons l'hypothèse que la condition d'indépendance linéaire est satisfaite, c'est-à-dire que les matrices A_1, \dots, A_m sont linéairement indépendantes dans $\text{SIR}^{n \times n}$.

3.1 Les équations Newton et les directions de recherche symétriques.

Comme nous l'avons vu dans le chapitre précédent, en supposant que (P) et (D) possèdent des solutions strictement admissibles, la trajectoire centrale est définie comme l'ensemble des solutions des équations

$$\begin{aligned} A_i \bullet X &= b_i, \quad i = 1, \dots, m, \\ \sum_{i=1}^m y_i A_i + S &= C, \\ XS &= \nu I, \end{aligned} \tag{3.1}$$

pour tout $\nu > 0$ (ce qui demande que X et S soient symétriques et définies positives).

$$H^p(M) := \frac{1}{2} [PMP^{-1} + P^{-T}M^TPT] \quad (3.6)$$

Une des symétrisations la plus importante est celle de Y. Zhang [10]. Définitions la transformation linéaire suivante:

gées. la solution est approchée. C'est pourquoi d'autres linéarisations sont envisagées. pas aussi précise que l'application $(X, S) \rightarrow XS$, particulièrement lorsque le cône symétrique semi-défini positif). Cependant, l'application inverse n'est $X + \nu f'(S) = 0$ où $f(X) := -\ln \det X$ est la fonction barrière n-normale pour $S - \nu X^{-1} = 0$ ou $X - \nu S^{-1} = 0$ (ce qui correspond à $S + \nu f'(X) = 0$ ou le résultat se trouve dans $S\mathbb{R}^{n \times n}$. Par exemple, elle peut se réécrire comme Un remède naturel est de réécrire la dernière équation de (3.1) de manière que symétrique même si X et S le sont.

général inexact pour ΔX . En effet, le produit XS n'est pas nécessairement ΔS , qui satisfait le système ci-dessus, est symétrique. Cependant, c'est en $(X + \Delta X, y + \Delta y, S + \Delta S) \in S\mathbb{R}^{n \times n} \times \mathbb{R}^m \times S\mathbb{R}^{n \times n}$. Par l'équation (3.3), tout à-dire de partir d'un itéré $(X, y, S) \in S\mathbb{R}^{n \times n} \times \mathbb{R}^m \times S\mathbb{R}^{n \times n}$ vers un point Un point essentiel des méthodes de points intérieurs pour la programmation semi-définie est de générer des matrices ΔX et ΔS symétriques c'est-à-dire de partir d'un itéré $(X, y, S) \in S\mathbb{R}^{n \times n} \times \mathbb{R}^m \times S\mathbb{R}^{n \times n}$ vers un point

$$\Delta XS + X\Delta S = \sigma \mu I - XS \quad (3.5)$$

La linéarisation de (3.4) donne

$$\Delta XS + X\Delta S = \sigma \mu I - (XS + \Delta X\Delta S). \quad (3.4)$$

$$\sum_{i=1}^m \Delta y_i A_i + \Delta S = C - S - \sum_{i=1}^m y_i A_i \quad (3.3)$$

$$A_i \bullet \Delta X = b_i - A_i \bullet X, \quad i = 1, \dots, m, \quad (3.2)$$

est $(X + \Delta X, y + \Delta y, S + \Delta S)$, nous avons Supposons que nous essayons d'approximer le point sur la trajectoire centrale correspondant à une valeur $\nu = \sigma \mu$ où $\mu = X \bullet S/n$ et $\sigma \in [0, 1]$. Si la solution troisième équation de ce système.

intérieurs pour les problèmes SDF, se sont portées sur la linéarisation de la système équivalent. Plusieurs recherches, à propos des méthodes de points un simple pas Newton au système non linéaire d'équations (3.1) ou à un coup d'algorithme de points intérieurs général itéré suivant en appliquant Etant donné un itéré (X, y, S) pour les problèmes SDF primal et dual, beau-

$$M = \nu I.$$

Puisque le spectre de M est réel, nous devons avoir que $G = 0$. Sinon M devrait avoir une valeur propre $\nu + \phi_k(G)$ pour un nombre non nul imaginaire pure $\phi_k(G)$ (en effet, les valeurs propres d'une matrice anti-symétrique sont imaginaires pures). Ce qui contredit l'hypothèse réel du spectre de M .
 Dès lors nous obtenons

$$\begin{aligned} &[-(PMP^{-1})^T + \nu I]^T = [-P^{-T}M^T P^T + \nu I]^T \\ &= -PMP^{-1} + \nu I \\ &= (PMP^{-1})^T - \nu I \\ &= -[-(PMP^{-1})^T + \nu I]. \end{aligned}$$

anti-symétrique:
 il suffit de poser $G = -(PMP^{-1})^T + \nu I$. Vérifions que $-(PMP^{-1})^T + \nu I$ est pour une certaine matrice G anti-symétrique (c'est-à-dire $G = -G^T$), en effet

$$PMP^{-1} = \nu I + G$$

Nous pouvons écrire

$$PMP^{-1} + (PMP^{-1})^T = 2\nu I.$$

Supposons que l'égalité de gauche est vérifiée. Nous avons

Preuve:

Théorème 3.1.1
 Pour $M \in \mathbb{R}^{n \times n}$ avec un spectre réel, $P \in \mathbb{R}^{n \times n}$ non singulière, et un scalaire ν ,
 $H_P(M) = \nu I \Leftrightarrow M = \nu I$.

le théorème suivant:
 En effet, puisque les valeurs propres d'une matrice symétrique sont réelles et que deux matrices semblables ont le même spectre, nous pouvons appliquer

$$H_P(M) = \nu I \Leftrightarrow M = \nu I.$$

M est semblable à une matrice symétrique définie positive, alors pour une matrice P inversible donnée. Observons que si P est inversible et

En particulier, $P = W^{-1/2}$.

$$W = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2} \\ = S^{-1/2}(S^{1/2}XS^{1/2})^{1/2}S^{-1/2}.$$

définie par

P satisfaisant à $P^T P = W^{-1}$ où W est la matrice de mise à échelle
- Dans [9], Todd présente la direction NT, elle correspond à une matrice
Nous parlerons des directions HKM.

- Dans [4] et [10], Monteiro et Zhang analysent le cas où $P = X^{-1/2}$ et
Nous parlons de direction AHO.

$$(\Delta XS + X \Delta S) + (S \Delta X + \Delta SX) = 2\sigma \mu I - XS - SX.$$

- Dans [1], Alizadeh, Haebberly et Overton utilisent la linéarisation pure,
c'est-à-dire qu'ils posent $P = I$ dans (3.10), soit
Suivant la matrice P , nous avons des symétrisations différentes et donc des
méthodes différentes.

$$(3.10) \quad P(\Delta XS + X \Delta S)P^{-1} + P^{-T}(S \Delta X + \Delta SX)P^T \\ = 2\sigma \mu I - PXS P^{-1} - P^{-T}SXP^T.$$

ΔS symétriques. Notons que l'équation (3.9) peut s'écrire en détail comme
les conditions pour produire une solution unique $(\Delta X, \Delta y, \Delta S)$ avec ΔX et
transformation de $S \mathbb{R}^{n \times n} \times \mathbb{R}^m \times S \mathbb{R}^{n \times n}$ vers lui-même, et donc, nous avons
Le système résultant de (3.2)-(3.3)-(3.9) peut être vu comme défini par une

$$(3.9) \quad H^p(\Delta XS + X \Delta S) = \sigma \mu I - H^p(XS).$$

et

$$(3.8) \quad H^p(\Delta XS + X \Delta S) = \sigma \mu I - H^p(XS + \Delta X \Delta S)$$

et (3.5) sont remplacés par

métrique définie positive $S^{1/2}XS^{1/2}$. En utilisant cette caractérisation, (3.4)
sans affecter la trajectoire centrale, puisque XS est semblable à la matrice sy-

$$(3.7) \quad H^p(XS) = \nu I$$

Nous pouvons remplacer la dernière équation de (3.1) par

L'implication inverse est évidente. \square

3.2 La notion du produit de Kronecker symétrisé.

Pour calculer le pas Newton $(\Delta X, \Delta y, \Delta S)$, il est plus facile d'exprimer le système linéaire d'équations (3.2)-(3.3)-(3.9) sous une forme matricielle standard en utilisant le produit de Kronecker symétrisé. Tout d'abord, nous définissons un opérateur transformant des matrices symétriques en vecteurs. Si U est une matrice symétrique $n \times n$, $svec(U)$ est défini par

$$svec(U) := (u_{11}, \sqrt{2}u_{21}, \dots, \sqrt{2}u_{n1}, u_{22}, \sqrt{2}u_{32}, \dots, \sqrt{2}u_{n2}, \dots, u_{nn})^T. \quad (3.11)$$

Le facteur $\sqrt{2}$ est introduit pour que $svec$ soit une isométrie entre $SIR^{n \times n}$ et $\mathbb{R}^{n(n+1)/2}$ avec leur produit standard respectif. Nous noterons par $smat$ l'application inverse de $svec$.

Le produit de Kronecker symétrisé de deux matrices $n \times n$, G et K (pas nécessairement symétriques) est une matrice carrée d'ordre $n(n+1)/2$. Son action sur un vecteur $u := svec(U)$, où $U \in SIR^{n \times n}$ est donné par

$$(G \otimes_s K) svec(U) := \frac{1}{2} svec(KUG^T + GUK^T). \quad (3.12)$$

Les principales propriétés du produit de Kronecker symétrisé se trouvent en annexe ainsi que celles du produit de Kronecker standard. En utilisant (3.11) et (3.12), nous pouvons maintenant réécrire le système (3.2)-(3.3)-(3.9) en une équation par blocs 3×3 :

$$\begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ E & 0 & F \end{pmatrix} \begin{pmatrix} svec(\Delta X) \\ \Delta y \\ svec(\Delta S) \end{pmatrix} = \begin{pmatrix} svec(R_d) \\ r_p \\ svec(R_c) \end{pmatrix}, \quad (3.13)$$

où \mathcal{I} est la matrice identité d'ordre $n(n+1)/2$,

$$E := P \otimes_s P^{-T} S, F := P X \otimes_s P^{-T}, \quad (3.14)$$

$$r_p := b - Asvec(X), R_d := C - S - \sum_{i=1}^m y_i A_i, R_c := \sigma \mu I - H^p(XS). \quad (3.15)$$

De plus \mathcal{A} est une matrice de dimension $m \times n(n+1)/2$ définie par

$$\mathcal{A} := \begin{pmatrix} svec(A_1) \\ \vdots \\ svec(A_m) \end{pmatrix}. \quad (3.16)$$

L'hypothèse d'indépendance linéaire des A_i , $i = 1, \dots, m$ implique que \mathcal{A} est de rang plein.

3.3 L'unicité des directions de recherche symétriques.

Pour continuer notre analyse, il est essentiel de connaître les conditions sous lesquelles le système (3.13) possède une solution unique. Nous observons tout d'abord que les matrices E et F sont inversibles lorsque X et S sont définies positives. En effet, par les propriétés du produit de Kronecker symétrisé, nous avons

$$(3.17) \quad E = P \otimes_s P^{-T} S = (I \otimes_s P^{-T} S P^{-1})(P \otimes_s P)$$

et

$$(3.18) \quad F = P X \otimes_s P^{-T} = (P X P^T \otimes_s I)(P^{-T} \otimes_s P^{-T}),$$

donc, E et F sont inversibles puisque chacun des facteurs est inversible. Nous avons dès lors le théorème suivant qui assure l'unicité des solutions pour le système (3.13):

Théorème 3.3.1
 Supposons que X et S sont symétriques définies positives. Alors, le système (3.13) possède une solution unique $(\Delta X, \Delta y, \Delta S) \in S\mathbb{R}^{n \times n} \times \mathbb{R}^m \times S\mathbb{R}^{n \times n}$ si $E^{-1}F$ est définie positive (pas nécessairement symétrique).
 En particulier, le résultat est vérifié lorsque X , S et $H_P(XS)$ sont semi-définies positives.

Preuve: Pour montrer l'unicité, il nous suffit de démontrer que le système de blocs 3×3 suivant détient la seule solution nulle:

$$(3.19) \quad \begin{pmatrix} 0 & \mathcal{A}^T & \mathcal{I} \\ \mathcal{A} & 0 & 0 \\ E & 0 & F \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \Delta y \\ \text{vec}(\Delta S) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Puisque, par l'équation (3.17), E est inversible, nous pouvons utiliser l'élimination de Gauss par blocs pour réduire le système (3.19) à l'équation suivante:

$$(3.20) \quad (\mathcal{A}E^{-1}F\mathcal{A}^T)\Delta y = 0.$$

Cela montre que la solution de (3.13) existe et est unique. Pour démontrer que $H^p(XS)$ est semi-définie positive implique que $E^{-1}F$ soit définie positive, nous considérons un vecteur arbitraire non nul g de dimension $n(n+1)/2$. Nous allons montrer que $g^T E^{-1} F g > 0$. Soit $k = E^{-T} g$ et $K := \text{mat}(k)$, tel que k est non nul. Nous avons successivement

$$\Delta X = -E^{-1} F \Delta S = 0.$$

et le troisième donne

$$\Delta S = -\text{smat}(\mathcal{A}^T \Delta y) = 0$$

système (3.19) donne

0 est la seule solution de l'équation (3.20). Le premier bloc d'équation du rang plein, cela implique que $\mathcal{A} E^{-1} F \mathcal{A}^T$ est définie positive et donc $\Delta y = 0$. Comme la matrice $E^{-1} F$ est définie positive par hypothèse et que \mathcal{A} est de

$$\begin{pmatrix} \mathcal{I} & -(E^{-1}E)^{-1} \mathcal{A}^T \\ 0 & \mathcal{A}(E^{-1}E)^{-1} \mathcal{A}^T \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \Delta y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Nous effectuons une seconde élimination pour obtenir l'équation (3.20).

$$\begin{pmatrix} -E^{-1}E & \mathcal{A}^T \\ 0 & \mathcal{A} \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \Delta y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Nous obtenons dès lors le système suivant

$$\begin{pmatrix} -E^{-1}E & \mathcal{A}^T & 0 \\ 0 & 0 & \mathcal{I} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \Delta y \\ \text{vec}(\Delta S) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

En effet, une première élimination donne

1/2

Le théorème suivant décrit une classe de matrices P qui satisfont la condition suffisante du Théorème 3.3.1.

Dès lors, l'unicité n'est pas toujours satisfaite.
 Si $P = I$, $H^p(XS) = \widehat{XS} + SX$ n'est pas nécessairement définie positive. choix de P de Monteiro [4] et Zhang [10].
 Nous avons donc une solution unique avec ΔX et ΔS symétriques pour le avons $H^p(XS) = S^{1/2}XS^{1/2}$, qui est définie positive lorsque X et S le sont. De façon similaire, lorsque $P = S^{1/2}$ ou $P = (S^{1/2}XS^{1/2})^\alpha S^{1/2}$, nous pour tout α , alors $H^p(XS) = X^{1/2}SX^{1/2}$ est définie positive lorsque X et S Notons que, si $P = X^{-1/2}$ ou plus généralement $P = (X^{1/2}SX^{1/2})^\alpha X^{-1/2}$

lité découle de l'hypothèse que $H^p(XS)$ est semi-définie positive. \square
 raisonnablement analogue peut être utilisé pour $P^{-T}XP^{-1}$. La dernière inégalité, si $X^{1/2}P^T x = 0$, alors $P^T x = 0$ et $x = 0$ puisque P est inversible. Un

$$x^T P X P^T x = \|X^{1/2} P^T x\| > 0 \forall x \neq 0,$$

effet, La stricte inégalité ci-dessus est possible, car $P X P^T \otimes_s P^{-T} S P^{-1}$ est définie positive, puisque $P X P^T$ et $P^{-T} S P^{-1}$ sont symétriques définies positives. En

$$\begin{aligned} & \geq 0. \\ & = \frac{1}{2} \text{tr}(K H^p(XS) K) \\ & = \frac{1}{4} \text{tr}(K(P X S P^{-1} + P^{-T} S X P^T) K) \\ & = \frac{1}{4} (K \bullet P X S P^{-1} K + K P^{-T} S X P^T \bullet K) \\ & = \frac{1}{4} (K \bullet P X S P^{-1} K + K \bullet K P^{-T} S X P^T) \\ & = \frac{1}{4} (\text{svec}(K))^T \text{svec}(P X S P^{-1} K + K P^{-T} S X P^T) \\ & > \frac{1}{2} k^T (P X S P^{-1} \otimes_s I) k \\ & = \frac{1}{2} k^T (P X P^T \otimes_s P^{-T} S P^{-1}) k + \frac{1}{2} k^T (P X S P^{-1} \otimes_s P^{-T} P^T) k \\ & = k^T (P X \otimes_s P^{-T}) (P^T \otimes_s S P^{-1}) k \\ & g^T E^{-1} E g = k^T F E^T k \end{aligned}$$

Théorème 3.3.2

Supposons que X et S sont des matrices symétriques définies positives et que P est une matrice inversible. Soit H et F définies par (3.14). Alors, les propriétés suivantes sont équivalentes:

1. PXP^T et $P^{-T}SP^{-1}$ commutent,
2. $PXSP^{-1}$ est symétrique,
3. FET est symétrique,
4. $E^{-1}F$ est symétrique.

La condition 1. implique que $H_P(XS)$ est définie positive. Ce qui a pour conséquence, que chacune de ces propriétés implique que $E^{-1}F$ et FET sont toutes deux définies positives.

Preuve:

L'équivalence entre 1. et 2. est immédiate. En effet, PXP^T et $P^{-T}SP^{-1}$ commutent si et seulement si

$$\begin{aligned} PXSP^{-1} &= (PXP^T)(P^{-T}SP^{-1}) \\ &= (P^{-T}SP^{-1})(PXP^T) \\ &= P^{-T}SXP^T \\ &= (PXSP^{-1})^T \end{aligned}$$

c'est-à-dire que $PXSP^{-1}$ est symétrique.

De plus,

$$\begin{aligned} FET &= (PX \otimes P^{-T})(P^T \otimes SP^{-1}) \\ &= \frac{1}{2}(PXP^T \otimes P^{-T}SP^{-1} + PXSP^{-1} \otimes I). \end{aligned}$$

Le premier terme de la seconde égalité est symétrique puisque PXP^T et $P^{-T}SP^{-1}$ le sont. Dès lors, FET est symétrique si et seulement si le second terme l'est, c'est-à-dire si et seulement si $PXSP^{-1}$ est symétrique. Donc, 2. et 3. sont équivalents.
De plus, $E^{-1}F = E^{-1}(FET)E^{-T}$ est symétrique si et seulement si FET l'est. Nous avons la dernière équivalence entre 3. et 4. .
Supposons maintenant que 1. est satisfaite (et donc 2.). Alors,

Algorithme 3.3.1

1. Choisir $0 \leq \sigma \leq 1$ et définir

$$\nu = \sigma \mu = \sigma \frac{X \bullet S}{n}.$$

2. Déterminer $(\Delta X, \Delta y, \Delta S)$ par le système (3.16) ou de manière équivalente par

$$AE^{-1}FA^T\Delta y = r_p + AE^{-1}(F \text{vec}(R_d) - \text{vec}(R_c)),$$

$$\text{vec}(\Delta S) = \text{vec}(R_d) - A^T\Delta y,$$

$$\text{vec}(\Delta X) = E^{-1}(\text{vec}(R_c) - F \text{vec}(\Delta S)).$$

3. Choisir des longueurs de pas α, β et mettre à jour les itérés par

$$X \leftarrow X + \alpha \Delta X,$$

$$y \leftarrow y + \beta \Delta y,$$

$$S \leftarrow S + \beta \Delta S.$$

Nous obtenons finalement l'algorithme de base de points intérieurs:

La première des conditions équivalentes du Théorème 3.3.2 est satisfaite pour $P = X^{-1/2}$ et $P = S^{1/2}$. En effet, une des deux matrices est l'identité. Elle est également satisfaite pour $P = W^{-1/2}$. En effet, nous verrons par la suite que ces deux matrices sont égales.

est le produit de deux matrices définies positives qui commutent. Cela implique que $H_p(XS)$ est définie positive et par conséquent, par le Théorème 3.3.1, $E^{-1}F$ et F^TE sont définies positives. \square

$$H_p(XS) = PXS P^{-1} \\ = (PXP^T)(P^{-T}SP^{-1})$$

Chapitre 4

Différentes directions de recherche. Présentation et développement.

Dans ce chapitre, nous allons étudier trois directions de recherche que nous pouvons obtenir suivant le choix de la matrice P . Elles portent le nom de directions de Nesterov-Todd, HKM et d'Alizadeh-Haerberly-Overton.

4.1 La direction de Nesterov-Todd.

4.1.1 Définition.

La direction NT est la solution pour un certain $\sigma \in [0,1]$ du système d'équations

$$\begin{aligned}
 A_i \bullet \Delta X &= 0, \quad i = 1, \dots, m, \\
 \sum_{i=1}^m \Delta y_i A_i + \Delta S &= 0, \\
 W^{-1} \Delta X W^{-1} + \Delta S &= \sigma \mu X^{-1} - S,
 \end{aligned}
 \tag{4.1}$$

où W est la matrice de mise à échelle définie plus loin et $\mu := \frac{X \bullet S}{n}$. Nous allons tout d'abord montrer comment nous obtenons le système (4.1). En fait, c'est un cas spécial pour la programmation semi-définie d'un système d'équations défini pour les problèmes en programmation convexe sous forme conique où le cône, disons K , possède une fonction barrière "self-scaled" f . Dans notre cas, le cône K est l'ensemble des matrices semi-définies d'ordre

$$\begin{aligned} X^{-1/2}XSX^{-1/2} &\iff X^{-1/2}XSX^{-1/2} \\ X^{-1/2}XSX^{-1/2} &\iff X^{-1/2}XSX^{-1/2} \\ X^{-1/2}XSX^{-1/2} &\iff X^{-1/2}XSX^{-1/2} \end{aligned}$$

avons successivement multipliant l'équation $WSW = X$ respectivement par $X^{-1/2}$ et $SX^{1/2}$. Nous

$$(4.4) \quad W = X^{1/2}(X^{1/2}SX^{-1/2})^{-1/2}X^{1/2} = S^{-1/2}(S^{1/2}XS^{1/2})^{1/2}S^{-1/2}.$$

Nous obtenons alors

$$(4.3) \quad W^{-1}XW^{-1} = S \iff WSW = X.$$

L'équation (4.2) nous donne

$$\begin{aligned} f''(W)[X] &= W^{-1}XW^{-1}, f''(T)[S] = T^{-1}ST^{-1}, \\ f'(X) &= -X^{-1}, f'(S) = -S^{-1}, \end{aligned}$$

Dans le cas de la programmation semi-définie, $f^*(S) = -\ln \det S - n$. Dès lors, comme dans le Chapitre 2, nous obtenons

échelle. De plus, $f''(W)f'(S) = f'(X)$.
 La théorie implique aussi que, avec $T := f'(W)$, la fonction f^* , qui est une fonction barrière self-scaled pour le cône dual K^* , satisfait $f''(T) = [f''(W)]^{-1}$ et $f''(T)[S] = X$. Nous avons dès lors un point dual de mise à

$$(4.2) \quad f''(W)[X] = S.$$

Une conséquence des fonctions barrières self-scaled (que nous ne développons pas ici) est que pour tout X à l'intérieur du cône primal K et pour tout S à l'intérieur du cône dual correspondant K^* , il existe un unique point de mise à échelle W à l'intérieur de K tel que

$$f(X) := -\ln \det X.$$

n et la fonction barrière est celle définie dans le Chapitre 2:

$$(4.6) \quad \Delta X + f''(L)[\Delta S] = -X - \sigma \mu f'_*(S).$$

Cette équation est encore équivalente à

$$(4.5) \quad f''(W)[\Delta X] + \Delta S = -S - \sigma \mu f'_*(X).$$

avec l'équation

(4.1) revient aux équations d'admissibilité pour les directions primale et duale ce qui nous démontre la première propriété du Théorème 3.3.2. Le système

$$W^{-1/2} X W^{-1/2} = W^{-1/2} S W^{-1/2}$$

Nous avons aussi

$$\begin{aligned} \Leftrightarrow W &= S^{-1/2} (S^{1/2} X S^{1/2})^{-1/2} S^{-1/2} \\ \Leftrightarrow S^{-1} S W &= S^{-1/2} (S^{1/2} X S^{1/2})^{-1/2} S^{-1/2} \\ \Leftrightarrow S^{-1} W^{-1} X &= S^{-1/2} (S^{1/2} X S^{1/2})^{-1/2} S^{-1/2} \\ S^{-1/2} S^{-1/2} W^{-1} X S^{1/2} S^{-1/2} &= S^{-1/2} (S^{1/2} X S^{1/2})^{-1/2} S^{-1/2} \end{aligned}$$

nous obtenons

et en prémultipliant et postmultipliant par $S^{-1/2}$ cette dernière équation,

$$\begin{aligned} \Leftrightarrow S^{-1/2} W^{-1} X S^{1/2} &= (S^{1/2} X S^{1/2})^{-1/2} \\ \Leftrightarrow S^{-1/2} W^{-1} X S^{1/2} S^{-1/2} W^{-1} X S^{1/2} &= S^{1/2} X S^{1/2} \\ S^{-1/2} W^{-1} X S^{1/2} W^{-1} X S^{1/2} &= S^{-1/2} S X S^{1/2} \end{aligned}$$

successivement

Montrons maintenant la seconde égalité en prémultipliant et postmultipliant l'équation $W^{-1} X W^{-1} = S$ respectivement par $S^{-1/2}$ et $X S^{1/2}$. Nous avons

$$\begin{aligned} \Leftrightarrow W &= X^{1/2} (X^{1/2} S X^{1/2})^{-1/2} X^{1/2} \\ \Leftrightarrow S^{-1} S W &= X^{1/2} (X^{1/2} S X^{1/2})^{-1/2} X^{1/2} \\ \Leftrightarrow S^{-1} W^{-1} X &= X^{1/2} (X^{1/2} S X^{1/2})^{-1/2} X^{1/2} \\ X^{1/2} X^{-1/2} S^{-1/2} W^{-1} X^{1/2} X^{1/2} &= X^{1/2} (X^{1/2} S X^{1/2})^{-1/2} X^{1/2} \end{aligned}$$

obtenons

et en prémultipliant et postmultipliant par $X^{1/2}$ cette dernière équation, nous

Cela prouve que la solution de (4.1') satisfait aussi (3.13). De plus, nous savons, par les remarques faites ci-dessus et le Théorème 3.3.1, que (3.13) possède une solution unique. Des lors, les deux systèmes sont équivalents. \square

$$W^{-1}(\Delta XS + X\Delta S)W + S\Delta X + \Delta SX = 2\sigma\mu I - 2SX.$$

en utilisant le fait que $W^{-1}SW = X$, nous obtenons l'équation (4.7):

$$W^{-1}\Delta XW^{-1}X + \Delta SX + S\Delta X + \Delta SX = \sigma\mu(I + SWX^{-1}W^{-1}) - SX - SW^{-1}SW,$$

Si nous additionnons ces deux équations,

$$S\overbrace{W^{-1}W}^I\overbrace{W^{-1}W}^I\Delta XW^{-1}W + S\overbrace{W^{-1}W}^I\Delta SXW^{-1}W = \sigma\mu SWX^{-1}W - SW^{-1}SW.$$

et postmultiplions par W :
 En effet, l'équation (4.8) est obtenue en postmultipliant la dernière équation du système (4.1) par X . Pour l'équation (4.9), nous prémultiplions par SW

$$(4.9) \quad S\Delta X + S\overbrace{W^{-1}W}^I\Delta SXW^{-1}W = \sigma\mu SWX^{-1}W - SW^{-1}SW.$$

et

$$(4.8) \quad W^{-1}\Delta XW^{-1}X + \Delta SX = \sigma\mu I - SX$$

est équivalente à chacune des deux équations suivantes:

$$W^{-1}\Delta XW^{-1} + \Delta S = \sigma\mu X^{-1} - S$$

D'un autre côté, la dernière équation de (4.1)

$$(4.7) \quad W^{-1}(\Delta XS + X\Delta S)W + S\Delta X + \Delta SX = 2\sigma\mu I - 2SX.$$

$$P^T P (\Delta XS + X\Delta S) P^{-1} P^{-T} + P^T P^{-T} (S\Delta X + \Delta SX) P^T P^{-T} = 2\sigma\mu P^T P^{-T} - P^T P^{-T} S X P^T P^{-T}$$

nous obtenons

En prémultipliant et postmultipliant cette équation respectivement par P^T et P^{-T} , et en utilisant les équations $P^T P = W$ et $P X S P^{-1} = P^{-T} S X P^T$,

$$P(\Delta XS + X\Delta S)P^{-1} + P^{-T}(S\Delta X + \Delta SX)P^T = 2\sigma\mu I - P X S P^{-1} - P^{-T} S X P^T.$$

*

4.1.3 Implémentation.

Cette section décrit la manière d'implémenter les calculs nécessaires à chaque itération de façon la plus performante possible.

Calcul de W .

Pour le calcul de W , nous utilisons deux factorisations de Cholesky et une décomposition en valeurs singulières (SVD).

Soit les factorisations de Cholesky des matrices symétriques définies positives X et S :

$$X = LL^T, \quad S = RR^T, \quad (4.10)$$

et soit $UDV^T = R^T L$, la décomposition en valeurs singulières de $R^T L$, où U et V sont orthogonales et D est diagonale. Définissons $\tilde{Q} := L^{-1} X^{1/2}$. Par le lemme suivant, \tilde{Q} est une matrice orthogonale:

Lemme 4.1.1
 Supposons que $B = CC^T$ est définie positive.
 Alors $\tilde{Q} := C^{-1} B^{1/2}$ est orthogonale.

Preuve:

Nous avons successivement

$$\begin{aligned} \tilde{Q}\tilde{Q}^T &= C^{-1} B C^{-T} \\ &= C^{-1} C C^T C^{-T} \\ &= I. \end{aligned}$$

□

où

$$(4.13) \quad G := LV D^{-1/2}.$$

$$(4.12) \quad W = G G^T$$

De cette équation (4.11), W peut être facilement calculé par

$$\begin{aligned} & \hat{Q}^T V D^{-1} V^T \hat{Q} = \hat{Q}^T V D^{-1} V^T \hat{Q} \\ & = \hat{Q}^T V D^{-2} V^T \hat{Q}. \end{aligned}$$

Nous devons montrer que $(\hat{Q}^T V) D^{-1} V^T \hat{Q}$ est la racine carrée de $(\hat{Q}^T V) D^{-2} V^T \hat{Q}$, ce qui est tout de suite vérifié puisque $(V^T \hat{Q})(\hat{Q}^T V) = I$:

$$\begin{aligned} X_{1/2} S X_{1/2} &= (\hat{Q}^T V) D_2 V^T \hat{Q} \\ \Leftrightarrow (X_{1/2} S X_{1/2})^{-1} &= (V^T \hat{Q})^{-1} D^{-2} (\hat{Q}^T V)^{-1} \\ &= (\hat{Q}^T V) D_2 V^T \hat{Q}. \end{aligned}$$

En effet, nous savons

$$(4.11) \quad (X_{1/2} S X_{1/2})^{-1/2} = (\hat{Q}^T V) D^{-1} V^T \hat{Q}.$$

Notons que $V D_2 V^T$ est la décomposition en valeurs propres de $L^T R^T L = L^T S L$, nous pouvons donc l'utiliser pour remplacer la SVD. Mais le nombre de conditionnement de cette matrice est le carré de celui de $R^T L$, ce qui peut conduire à un algorithme moins stable. Comme $\hat{Q}^T V$ est orthogonale (puisque \hat{Q} et V le sont), nous avons

$$\begin{aligned} X_{1/2} S X_{1/2} &= \hat{Q}^T L^T R^T L \hat{Q} \\ &= \hat{Q}^T (R^T L)^T (R^T L) \hat{Q} \\ &= \hat{Q}^T (U D V^T)^T (U D V^T) \hat{Q} \\ &= \hat{Q}^T V D U^T U D V^T \hat{Q} \\ &= (\hat{Q}^T V) D_2 V^T \hat{Q}. \end{aligned}$$

Puisque $L \hat{Q} = X_{1/2} = (X_{1/2})^T = \hat{Q}^T L^T$, alors

Par l'équation (4.12), nous remarquons que $G^{-T}G^{-1} = W^{-1}$, et donc que $P := G^{-1}$ satisfait $P^T P = W^{-1}$. C'est encore un autre choix de P qui fournit la direction NT. Dans ce cas, $P^{-T} S P^{-1}$ et $P X P^T$ commentent puisqu'elles ont la même matrice diagonale.

Notons encore que, si nous utilisons la factorisation "renversée" de Cholesky de S , c'est-à-dire que L est une matrice triangulaire inférieure mais que R est une matrice triangulaire supérieure dans l'équation (4.10), nous éparignons du travail dans le calcul de la matrice triangulaire inférieure $R^T L$. De plus, nous n'avons pas besoin de la matrice orthogonale U de la décomposition en valeurs singulières, V et D suffisent.

$$\begin{aligned} G^{-1} X G^{-T} &= (L V D^{-1/2})^{-1} X (D^{-1/2} V^T L^T)^{-1} \\ &= D^{1/2} V^{-1} L^{-1} L^T L^{-1} V^{-T} D^{1/2} \\ &= D. \end{aligned}$$

Ensuite,

$$\begin{aligned} G^T S G &= (L V D^{-1/2})^T S (L V D^{-1/2}) \\ &= D^{-1/2} V^T L^T R^T L V D^{-1/2} \\ &= D^{-1/2} V^T V D U^T U D V^T V D^{-1/2} \\ &= D. \end{aligned}$$

En effet, nous avons tout d'abord c'est-à-dire que G met à échelle X et S par la même matrice diagonale.

$$(4.14) \quad G^T S G = G^{-1} X G^{-T} = D$$

Notons que nous avons aussi Nous utiliserons la matrice G dans le calcul de la direction $(\Delta X, \Delta y, \Delta S)$.

$$\begin{aligned} W &= X^{1/2} (X^{1/2} S X^{1/2})^{-1/2} X^{1/2} \\ &= (L \hat{Q})(\hat{Q}^T V) D^{-1} (V^T \hat{Q})(\hat{Q}^T L^T) \\ &= L V D^{-1} V^T L^T \\ &= G G^T. \end{aligned}$$

En effet,

$$\begin{pmatrix} -F^{-1}E & \mathcal{A} & 0 \\ \mathcal{A}^T & -F^{-1}E & 0 \\ \text{vec}(\Delta X) & \text{vec}(\Delta S) & \Delta y \end{pmatrix} = \begin{pmatrix} \text{vec}(R_d) - F^{-1}\text{vec}(R_c) \\ r_p \\ \text{vec}(R_c) \end{pmatrix}.$$

Nous obtenons dès lors le système suivant

$$\begin{pmatrix} -F^{-1}E & \mathcal{A} & 0 \\ \mathcal{A}^T & -F^{-1}E & 0 \\ \text{vec}(\Delta X) & \text{vec}(\Delta S) & \Delta y \end{pmatrix} = \begin{pmatrix} \text{vec}(R_d) - F^{-1}\text{vec}(R_c) \\ r_p \\ \text{vec}(R_c) \end{pmatrix}.$$

Une première élimination de Gauss donne

En effet, nous procédons de la même manière que dans le Théorème 3.3.1.

$$(4.16) \quad (\mathcal{A}E^{-1}F\mathcal{A}^T)\Delta y = r_p + \mathcal{A}E^{-1}F\text{vec}(R_d) - \mathcal{A}E^{-1}\text{vec}(R_c).$$

comprenant que Δy comme inconnue donnée par

Gauss par blocs pour le réduire à une équation de complètement de Schur, ne méthode standard pour résoudre ce système est d'utiliser l'élimination de Le système (4.15) est un système à $m + n(n+1)$ équations linéaires. Une et \mathcal{A} est la matrice définie par (3.16).

$$\begin{aligned} &= D^2 \\ &= \frac{1}{2}(G^{-1}XG^{-T}G^TSG + G^TSGG^{-1}XG^{-T}) \\ &= \frac{1}{2}(G^{-1}XSG + G^T SXG^{-T}) \end{aligned}$$

où, par l'équation (4.14), nous utilisons

$$r_p := b - \mathcal{A}\text{vec}(X), R_d := C - S - \sum_{i=1}^m y_i A_i, R_c := \sigma \mu I - D^2,$$

$$E = G^{-1} \otimes_s G^T S, F = G^{-1} X \otimes_s G^T,$$

où

$$(4.15) \quad \begin{pmatrix} 0 & \mathcal{A}^T & I \\ \mathcal{A} & 0 & 0 \\ E & 0 & F \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \text{vec}(\Delta S) \\ \Delta y \end{pmatrix} = \begin{pmatrix} \text{vec}(R_d) \\ r_p \\ \text{vec}(R_c) \end{pmatrix},$$

la direction NT comme la solution de

En utilisant le Théorème 4.1.1 et le choix $P = G^{-1}$, nous pouvons obtenir

Calcul de la direction NT.

La matrice $E^{-1}F$ est symétrique et définie positive (puisque X et S le sont et donc W aussi), elle peut être factorisée comme $E^{-1}F = KK^T$ où $K := G \otimes_s G$, obtenue de l'égalité $W \otimes_s W = (G \otimes_s G)(G \otimes_s G)^T$. Nous appelons une telle factorisation de $E^{-1}F$ une factorisation semblable à Cholesky puisque elle est le produit d'une matrice et de sa transposée, mais ces

$$\begin{aligned} E \operatorname{vec}(\sigma \mu S^{-1} - X) &= (G^{-1} \otimes_s G^T S) \operatorname{vec}(\sigma \mu S^{-1} - X) \\ &= \frac{1}{2} \operatorname{vec}(G^T S (\sigma \mu S^{-1} - X) G^{-T} + G^{-T} (\sigma \mu S^{-1} - X) S G) \\ &= \frac{1}{2} \operatorname{vec}(\sigma \mu I - - G^T S X G^{-T} + \sigma \mu I - - G^{-1} X S G) \\ &= \operatorname{vec}(\sigma \mu I - - D_2^2) \\ &= \operatorname{vec}(R_c). \end{aligned}$$

De plus, nous obtenons successivement

$$\begin{aligned} E(W \otimes_s W) &= (G^{-1} \otimes_s G^T S)(W \otimes_s W) \\ &= \frac{1}{2} (G^{-1} W \otimes_s G^T S W + G^{-1} W \otimes_s G^T S W) \\ &= G^{-1} W \otimes_s G^T S W \\ &= G^{-1} G^T \otimes_s G^T S G G^T \\ &= G^T \otimes_s G^{-1} X G^{-T} G^T \\ &= G^T \otimes_s G^{-1} X \\ &= G^{-1} X \otimes_s G^T \\ &= F. \end{aligned}$$

En effet, nous avons tout d'abord

$$E^{-1}F = W \otimes W, E^{-1} \operatorname{vec}(R_c) = \operatorname{vec}(\sigma \mu S^{-1} - X). \quad (4.17)$$

La matrice E^{-1} ne peut pas être obtenue explicitement, mais nous avons seulement besoin de $E^{-1}F$ et de $E^{-1} \operatorname{vec}(R_c)$. En prémultipliant par E , nous pouvons vérifier les deux égalités suivantes:

$$\begin{pmatrix} \mathbf{I} & -E^{-1}E^{-1}A^T \\ 0 & A(E^{-1}E)^{-1}A^T \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\Delta X) \\ \Delta y \end{pmatrix} = \begin{pmatrix} -E^{-1}(F \operatorname{vec}(R_d) - \operatorname{vec}(R_c)) \\ r_p + AE^{-1}(F \operatorname{vec}(R_d) - \operatorname{vec}(R_c)) \end{pmatrix}.$$

Nous effectuons une seconde élimination pour obtenir l'équation (4.16).

Un rappel sur les problèmes aux moindres carrés se trouve en annexe. Deux façons de calculer Δy sont possibles. Une première façon est d'utiliser

$$\min_{\Delta y} \|B^T \Delta y - K^{-1} \text{vec}(X_r) + h\|.$$

qui sont les équations normales du problème aux moindres carrés

$$B^T \Delta y = B(K^{-1} \text{vec}(X_r) + h)$$

(4.18) comme

Le système (4.18) représente les équations normales d'un problème linéaire aux moindres carrés. Pour le voir, choisissons une matrice $X_r \in \text{SR}^{n \times n}$ telle que $\text{Avec}(X_r) = r_p$ (une telle matrice existe puisque A est de rang plein). En utilisant le fait que $BK^{-1} \text{vec}(X_r) = r_p$, nous pouvons réécrire le système

$$\begin{aligned} h &= K^T \text{vec}(R_d) - K^{-1} E^{-1} \text{vec}(R_c) \\ &= (G^T \otimes_s G^T) \text{vec}(R_d) - K^{-1} \text{vec}(\sigma \mu S^{-1} - X) \\ &= \text{vec}(G^T R_d G) - (G^{-1} \otimes_s G^{-1}) \text{vec}(\sigma \mu S^{-1} - X) \\ &= \text{vec}(G^T R_d G) - \text{vec}(\sigma \mu G^{-1} S^{-1} G^{-1} - G^{-1} X G^{-1}) \\ &= \text{vec}(G^T R_d G) - \text{vec}(\sigma \mu D^{-1} - D). \end{aligned}$$

$G \otimes_s G$, nous obtenons

En effet, en utilisant les équations (4.17), (4.14) et la définition de $K :=$

$$h = \text{vec}(G^T R_d G - \sigma \mu D^{-1} + D). \quad (4.20)$$

et

$$B^T := K^T A^T = [\text{vec}(G^T A_1 G) \dots \text{vec}(G^T A_m G)] \quad (4.19)$$

où

$$B^T \Delta y = r_p + B h \quad (4.18)$$

et l'équation (4.16) peut se réécrire comme

$$A E^{-1} F A^T = A K K^T A^T = B^T$$

matrices peuvent ne pas être triangulaires. En utilisant cette factorisation semblable à Cholesky de $E^{-1} F$, nous avons

où $P = S^{1/2}$ et $P = X^{-1/2}$,

$$P(\Delta XS + X\Delta S)P^{-1} + P^{-T}(S\Delta X + \Delta SX)P^T = 2\sigma\mu I - PXS P^{-1} - P^{-T}SXP^T,$$

Les directions HKM sont données par la linéarisation

4.2.1 Définition.

4.2 Les directions HKM.

$$\Delta X = smat[E^{-1}svect(R_c) - E^{-1}Fsvect(\Delta S)]. \quad (4.22)$$

et ΔX par le troisième bloc d'équation du système (4.15)

$$\Delta S = R_d - smat(A^T \Delta y), \quad (4.21)$$

tion du système (4.15)

lesky, nous pouvons alors calculer ΔS simplement par le premier bloc d'équation. Supposons que le pas Δy a été calculé via une factorisation QR ou de Cholesky, nous pouvons alors calculer ΔS simplement par le premier bloc d'équation. La matrice de complément de Schur.

L'approche de la factorisation QR est donc meilleure que l'approche utilisant le nombre compris entre $\sqrt{\kappa}$ et κ .

en résolvant l'équation (4.18) par une factorisation QR de B^T , dépend d'un du Δy calculé dépendante de κ . Par contre, l'erreur relative du Δy obtenu par une factorisation de Cholesky de $AE^{-1}FA^T$ produit une erreur relative nombre de conditionnement de $AE^{-1}FA^T$. Alors, résoudre l'équation (4.16) rigueur de précision finie, elles peuvent avoir une stabilité différente. Soit κ , le En arithmétique exacte, ces deux méthodes sont les mêmes. Mais en arithmétique

Golub et C. F. Van Loan [2].

par l'équation (4.18) en appliquant une factorisation QR de B^T . Voir G. H. de Cholesky à la matrice $AE^{-1}FA^T$. Mais nous pouvons aussi calculer Δy l'équation de complément de Schur (4.16) en appliquant une factorisation

$$(4.28) \quad \begin{pmatrix} 0 & A^T & I \\ E & 0 & F \\ \mathcal{A} & 0 & 0 \\ F & 0 & F \end{pmatrix} \begin{pmatrix} \text{svec}(\Delta X) \\ \text{svec}(\Delta S) \\ \Delta y \end{pmatrix} = \begin{pmatrix} \text{svec}(R_d) \\ \text{svec}(R_c) \\ r_p \end{pmatrix},$$

où $\mu := \frac{n}{X \bullet S}$, ou de manière équivalente du système

$$(4.27) \quad \begin{aligned} A_i \bullet \Delta X &= b_i - A_i \bullet X, \quad i = 1, \dots, m, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S &= C - S - \sum_{i=1}^m y_i A_i, \\ 2S\Delta XS + SX\Delta S + \Delta SXS &= 2\sigma\mu S - 2SXS, \end{aligned}$$

Dès lors, la direction de recherche HKM est la solution $(\Delta X, \Delta y, \Delta S) \in \text{SIR}^{n \times n} \times \mathbb{R}^m \times \text{SIR}^{n \times n}$ avec $\sigma \in [0, 1]$ du système linéaire

$$(4.26) \quad E \text{svec}(\Delta X) + F \text{svec}(\Delta S) = \text{svec}(R_c).$$

L'équation (4.23) devient alors

$$(4.25) \quad F = SX \otimes_s I.$$

$$(4.24) \quad E = S \otimes_s S,$$

Définissons

$$(4.23) \quad (S \otimes_s S) \text{svec}(\Delta X) + (SX \otimes_s I) \text{svec}(\Delta S) = \text{svec}(R_c).$$

En utilisant les notations du produit de Kronecker symétrique, nous obtenons

$$\text{où } R_c = \sigma\mu S - SX S = S^{1/2}(\sigma\mu I - S^{1/2}XS^{1/2})S^{1/2}.$$

$$\Leftrightarrow S\Delta XS + \frac{1}{2}SX\Delta S + \frac{1}{2}\Delta SXS = R_c,$$

$$\Leftrightarrow 2S\Delta XS + SX\Delta S + \Delta SXS = 2\sigma\mu S - 2SXS$$

$$\Leftrightarrow = 2\sigma\mu I - 2S^{1/2}XS^{1/2}$$

$$\Leftrightarrow 2S^{1/2}\Delta XS^{1/2} + S^{1/2}XS^{1/2}\Delta S + S^{-1/2}\Delta SXS^{1/2}$$

$$\Leftrightarrow = 2\sigma\mu I - S^{1/2}XS^{1/2} - S^{1/2}XS^{1/2}$$

$$\Leftrightarrow S^{1/2}\Delta XS^{1/2} + S^{1/2}XS^{1/2}\Delta S + S^{1/2}\Delta SXS^{1/2} + S^{-1/2}\Delta SXS^{1/2}$$

$$\Leftrightarrow = 2\sigma\mu I - S^{1/2}XS^{1/2} - S^{-1/2}XS^{1/2}$$

$$\Leftrightarrow S^{1/2}(\Delta XS + X\Delta S)S^{-1/2} + S^{-1/2}(S\Delta X + \Delta SX)S^{1/2}$$

Pour la première direction HKM, cette linéarisation devient

$$(4.33) \quad \begin{aligned} A_i \bullet \Delta X = b_i - A_i \bullet X, \quad i = 1, \dots, m, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S = C - S - \sum_{i=1}^m y_i A_i, \\ 2X \Delta S X + X S \Delta X + \Delta X S X = 2\sigma \mu X - 2X S X, \end{aligned}$$

Dès lors, la direction de recherche HKM duale est la solution $(\Delta X, \Delta y, \Delta S) \in \text{SIR}^{n \times n} \times \mathbb{R}^m \times \text{SIR}^{n \times n}$ avec $\sigma \in [0, 1]$ du système linéaire

$$(4.32) \quad E \text{vec}(\Delta X) + F \text{vec}(\Delta S) = \text{vec}(R_c).$$

L'équation (4.29) devient alors

$$(4.31) \quad F = X \otimes_s X.$$

$$(4.30) \quad E = X S \otimes_s I,$$

Définissons

$$(4.29) \quad (X S \otimes_s I) \text{vec}(\Delta X) + (X \otimes_s X) \text{vec}(\Delta S) = \text{vec}(R_c).$$

En utilisant les notations du produit de Kronecker symétrique, nous obtenons

$$\begin{aligned} \text{où } R_c = \sigma \mu X - X S X = X^{1/2} (\sigma \mu I - X^{1/2} S X^{1/2}) X^{1/2}. \\ \Leftrightarrow X \Delta S X + \frac{1}{2} \Delta X S X + \frac{1}{2} X S \Delta X = R_c, \\ \Leftrightarrow 2X \Delta S X + \Delta X S X + X S \Delta X = 2\sigma \mu X - 2X S X \\ \Leftrightarrow = 2\sigma \mu I - 2X^{1/2} S X^{1/2} \\ \Leftrightarrow 2X^{1/2} \Delta S X^{1/2} + X^{-1/2} \Delta X S X^{1/2} + X^{1/2} S \Delta X X^{-1/2} \\ \Leftrightarrow = 2\sigma \mu I - X^{1/2} S X^{1/2} - X^{-1/2} S X^{1/2} \\ \Leftrightarrow X^{-1/2} \Delta X S X^{1/2} + X^{1/2} \Delta S X X^{-1/2} + X^{1/2} S \Delta X X^{-1/2} + X^{-1/2} \Delta S X X^{1/2} \\ \Leftrightarrow = 2\sigma \mu I - X^{-1/2} S X^{1/2} - X^{1/2} S X^{-1/2} \\ X^{-1/2} (\Delta X S + X \Delta S) X^{1/2} + X^{1/2} (S \Delta X + \Delta S X) X^{-1/2} \end{aligned}$$

vient

Pour la seconde direction HKM, appelée direction duale, la linéarisation de-

$$\begin{aligned} \text{où} \\ E = S \otimes_s S, F = S X \otimes_s I, \\ r_p := b - A \text{vec}(X), R_d := C - S - \sum_{i=1}^m y_i A_i, R_c := \sigma \mu S - S X S. \end{aligned}$$

ce qui est immédiatement vérifié par les propriétés du produit de Kronecker symétrisé. Nous allons maintenant voir que Δy peut être calculé par une factorisation QR, tout comme pour la direction NT. Pour cela, il nous faut obtenir une

$$\begin{aligned} F = E(X \otimes_s S^{-1}) \\ \Leftrightarrow SX \otimes_s I = (S \otimes_s S)(X \otimes_s S^{-1}), \end{aligned}$$

Si nous considérons la direction HKM correspondant à $P = S^{1/2}$, nous pouvons remarquer que $E^{-1}F = X \otimes_s S^{-1}$. En effet, il nous suffit de prémultiplier cette équation par E . Nous devons obtenir

$$\begin{aligned} (AE^{-1}FA^T)\Delta y = r_p + AE^{-1}(F \text{vec}(R_d) - \text{vec}(R_c)), \\ \Delta S = R_d - \text{mat}(A^T \Delta y), \\ \text{vec}(\Delta X) = E^{-1}(\text{vec}(R_c) - F \text{vec}(\Delta S)). \end{aligned}$$

Tout comme pour la direction NT, en utilisant l'élimination de Gauss par blocs, $(\Delta X, \Delta y, \Delta S)$ est obtenu par

4.2.2 Implémentation.

Cette seconde direction de recherche HKM est appelée direction duale, car si nous commençons le rôle de X et de S , nous obtenons la première direction de recherche HKM.

$$E = XS \otimes_s I, F = X \otimes_s X, \\ r_p := b - A \text{vec}(X), R_d := C - S - \sum_{i=1}^m y_i A_i, R_c := \sigma \mu X - XSX.$$

où

$$(4.34) \quad \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ E & 0 & F \end{pmatrix} \begin{pmatrix} \text{vec}(\Delta X) \\ \Delta y \\ \text{vec}(\Delta S) \end{pmatrix} = \begin{pmatrix} \text{vec}(R_d) \\ r_p \\ \text{vec}(R_c) \end{pmatrix},$$

où $\frac{X \bullet S}{n}$, ou de manière équivalente du système

Alors que si $P = X^{-1/2}$, nous avons $E = XS \otimes_s I$, et calculer l'inverse nécessite la solution des équations de Lyapunov, c'est-à-dire des équations du type $AX + XA = B$ pour une matrice X .

$$E^{-1} = (S \otimes_s S)^{-1} = S^{-1} \otimes_s S^{-1}.$$

Nous pouvons remarquer que si nous voulons obtenir la matrice E^{-1} explicitement, en utilisant $P = S^{1/2}$, nous pouvons la calculer relativement facilement puisque par une propriété du produit de Kronecker symétrisé, nous avons Cholesky à la matrice $AE^{-1}FA^T$. Une autre manière d'implémenter Δy est d'appliquer une factorisation de pour la direction HKM duale où $P = X^{-1/2}$. Des lors, nous pouvons également utiliser l'approche de la factorisation QR symétrique.

3.3.2, les deux matrices commutent si la matrice $E^{-1}F$ est nécessairement PXP^T et $P^{-T}SP^{-1}$. Une telle base existe puisque par le Théorème nale dont les colonnes forment une base orthogonale qui diagonalise simultanément \tilde{D} est une matrice diagonale définie positive et Q est une matrice orthogonale

$$E^{-1}F = (P^{-1}Q \otimes_s P^{-1}Q) \tilde{D} (P^{-1}Q \otimes_s P^{-1}Q)^T,$$

pouvons alors exprimer cette matrice sous la forme direction solution de (3.13), lorsque la matrice $E^{-1}F$ est symétrique. Nous En fait, l'approche de la factorisation QR peut être appliquée facilement à la

$$E^{-1}F = KK^T \text{ où } K := (G \otimes_s G)M.$$

semblable à Cholesky facilement. Notons cette matrice M . Nous obtenons alors une factorisation de la matrice D . Ainsi, la racine carrée de $D \otimes_s D^{-1}$ existe et peut être calculée positives et de la forme $(d_i d_j^{-1} + d_j d_i^{-1})/2$, où les d_i sont les entrées diagonales Le second terme $D \otimes_s D^{-1}$ est une matrice diagonale dont les entrées sont

$$(4.35) \quad \begin{aligned} &= GDG^T \otimes_s GD^{-1}G^T \\ &= (G \otimes_s G)(D \otimes_s D^{-1})(G^T \otimes_s G^T). \end{aligned}$$

factorisation semblable à Cholesky de $E^{-1}F$. En utilisant l'équation (4.14) nous obtenons

Preuve: Pour montrer que le système (4.36) a une solution unique, il suffit de prouver que l'unique solution $(\Delta X, \Delta y, \Delta S) \in \text{SIR}^{n \times n} \times \mathbb{R}^m \times \text{SIR}^{n \times n}$ du système

Théorème 4.3.1

$S_i(X, y, S) \in \text{SIR}^{n \times n} \times \mathbb{R}^m \times \text{SIR}^{n \times n}$ (où X et S sont définies positives) est tel que

$$\|S^{1/2} X S^{1/2} - \nu I\| \leq \frac{\nu}{2} \forall \nu > 0. \quad (4.37)$$

Alors, le système (4.36) possède une solution unique.

Nous avons vu, par le Théorème 3.3.1 du Chapitre 3, que le système (4.36) possède une solution unique, et donc une direction de recherche AHO bien définie, si $X S + S X$ est définie positive. Or, nous savons que cette propriété n'est pas toujours vérifiée. Le théorème suivant est une condition suffisante, alternative pour que le système (4.36) possède une solution unique.

4.3.1 Une alternative au théorème d'unicité.

$$\text{où } \mu = \frac{\nu}{X \bullet S}.$$

$$\begin{aligned} A_i \bullet \Delta X &= b_i - A_i \bullet X, \quad i = 1, \dots, m, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S &= C - S - \sum_{i=1}^m y_i A_i, \\ X \Delta S + \Delta S X + S \Delta X + \Delta X S &= 2\mu I - X S - S X, \end{aligned} \quad (4.36)$$

Le système linéaire est par définition la solution $(\Delta X, \Delta y, \Delta S) \in \text{SIR}^{n \times n} \times \mathbb{R}^m \times \text{SIR}^{n \times n}$ du La direction de recherche AHO au point $(X, y, S) \in \text{SIR}^{n \times n} \times \mathbb{R}^m \times \text{SIR}^{n \times n}$ (où X et S sont définies positives), avec un paramètre de centralité $\sigma \in [0, 1]$,

4.3 La direction d'Alizadeh-Haerberly-Overton.

et donc $\widetilde{\Delta X} = 0$. Puisque J^T, X^{-1} et S sont des matrices non singulières,

$$\text{tr}(\widetilde{\Delta X})^T \widetilde{\Delta X} = \|\widetilde{\Delta X}\|_F^2 = 0$$

obtenons

J^T . Donc, $\text{tr}(\Delta X S J^T \Delta X S) = 0$, et en posant $\widetilde{\Delta X} = J^T \Delta X S$, nous

Comme $S^{-1} X^{-1}$ est définie positive, J non singulière telle que $S^{-1} X^{-1} =$

$$\implies \text{tr}(\Delta X S S^{-1} X^{-1} \Delta X S) = 0.$$

$$\implies \text{tr}(\Delta X X^{-1} \Delta X S) = 0$$

$$\text{tr}(\Delta X \Delta S) = 0$$

D'où, nous obtenons

$$\Delta S = -X^{-1} \Delta X S.$$

nous pouvons écrire

Si $V := X \Delta S + \Delta X S = 0$, alors en utilisant l'équation (4.39), nous obtenons $\Delta X = \Delta S = 0$. En effet, puisque X est définie positive et donc inversible,

$$\implies \Delta S \bullet \Delta X = 0.$$

$$\implies -\Delta S \bullet \Delta X = 0$$

$$\implies \sum_{i=1}^m \Delta y_i A_i \bullet \Delta X = 0$$

$$\implies \Delta y_i A_i \bullet \Delta X = 0 \forall i = 1, \dots, m,$$

$$A_i \bullet \Delta X = 0 \forall i = 1, \dots, m$$

En effet, nous avons successivement

$$\Delta X \bullet \Delta S = \text{tr}(\Delta X \Delta S) = 0. \quad (4.39)$$

En utilisant les deux premières équations de (4.38), nous pouvons voir que

$$X \Delta S + \Delta S X + S \Delta X + \Delta X S = 0.$$

$$\sum_{i=1}^m \Delta y_i A_i + \Delta S = 0, \quad (4.38)$$

$$A_i \bullet \Delta X = 0, \quad i = 1, \dots, m,$$

homogène suivant est $(\Delta X, \Delta y, \Delta S) = (0, 0, 0)$:

En utilisant ces propriétés, nous obtenons

$$A \bullet B \leq \|A\|_F \|B\|_F.$$

3. L'inégalité de Cauchy-Schwarz pour la norme de Frobenius:

$$2. \|B^T A B\|_F \geq \lambda_{\min}[B^T B] \|A\|_F \quad \forall A \in \mathbb{R}^{n \times n} \text{ et } B \in \mathbb{R}^{n \times n}.$$

$$1. \|A B\|_F \leq \|A\|_F \|B\|_F.$$

Notons que nous avons les propriétés suivantes sur les matrices de Frobenius (voir exercice 20 de la section 5.6 et le Corollaire 4.5.11 de [3]):

$$\begin{aligned} \text{tr}(\Delta X \Delta S) &= \text{tr}(X^{-1/2} \Delta X X^{-1/2} \Delta S X^{1/2}) \\ &= \text{tr}(X^{-1/2} \Delta X \Delta S X^{1/2}) \\ &= \text{tr}(\Delta X \Delta S X^{1/2} X^{-1/2}) \\ &= \text{tr}(\Delta X \Delta S) \\ &= 0. \end{aligned}$$

Soit $\Delta X = X^{-1/2} \Delta X X^{-1/2}$ et $\Delta S = X^{1/2} \Delta S X^{1/2}$. Par l'équation (4.39), $\text{tr}(\Delta X \Delta S) = 0$. En effet, nous obtenons successivement

$$\begin{aligned} 0 &= \|V + V^T\|_F^2 \\ &= \|V\|_F^2 + \|V^T\|_F^2 + 2\|V\|_F \|V^T\|_F \\ &= \|V\|_F^2 + \|V\|_F^2 + 2\|V\|_F^2 \\ &= 2\|V\|_F^2 + 2\|V\|_F^2 \\ &> 2\text{tr}V^2 \\ &= 2\text{tr}(X \Delta S + \Delta X S). \end{aligned} \tag{4.40}$$

Puisque les matrices A_1, \dots, A_m sont linéairement indépendantes, nous obtenons, par la seconde équation de (4.38) que $\Delta y = 0$. Donc, $V = 0$ implique que $(\Delta X, \Delta y, \Delta S) = (0, 0, 0)$. Pour terminer la démonstration, il nous faut montrer que $V = 0$. Supposons par l'absurde que $V \neq 0$. Par cette supposition et la dernière équation du système (4.38), nous avons

Finalement, nous obtenons

$$\begin{aligned} &\implies \lambda^{min}(S^{1/2}XS^{1/2}) \geq \frac{\nu}{2}. \\ &\implies -\frac{\nu}{2} \leq \lambda^{min}(S^{1/2}XS^{1/2}) - \nu \leq \frac{\nu}{2} \\ &\implies |\lambda^{min}(S^{1/2}XS^{1/2}) - \nu| \leq \frac{\nu}{2} \\ &\|S^{1/2}XS^{1/2} - \nu I\| \leq \frac{\nu}{2} \end{aligned}$$

Dès lors,

$$\|S^{1/2}XS^{1/2} - \nu I\| = \sup\{|\lambda - \nu| : \lambda \text{ est une valeur propre de } S^{1/2}XS^{1/2}\}.$$

Or, $\|S^{1/2}XS^{1/2} - \nu I\| \leq \frac{\nu}{2} \iff \lambda^{min}(S^{1/2}XS^{1/2}) \geq \frac{\nu}{2}$. En effet, puisque $S^{1/2}XS^{1/2} - \nu I$ est une matrice symétrique, nous avons

$$\begin{aligned} &\text{tr}(X\Delta S + \Delta X S)^2 \geq \|\Delta S\|_F^2 + (\lambda^{min}[S^{1/2}XS^{1/2}])^2 \|\Delta X\|_F^2 \\ &+ 2\text{tr}(\Delta S \Delta X)(X^{1/2}S^{1/2} - \nu I). \end{aligned}$$

En posant $B = X^{1/2}S^{1/2}$ et $A = \Delta X$ dans la propriété 2 citée ci-dessus, et comme $\mu > 0$, nous obtenons

$$\begin{aligned} &= \text{tr}(X\Delta S \Delta X S) = \text{tr}(X X^{-1/2} \Delta S X^{-1/2} X X^{1/2} \Delta X X^{1/2} S) \\ &= \text{tr}(X^{1/2} \Delta S \Delta X X^{1/2} S) \\ &= \text{tr}(\Delta S \Delta X X^{1/2} S X^{1/2}). \end{aligned}$$

En effet,

$$\begin{aligned} &\text{tr}(X\Delta S + \Delta X S)^2 = \text{tr}(X\Delta S)^2 + (\Delta X S)^2 + 2X\Delta S \Delta X S \\ &= \text{tr}(X^{1/2} \Delta S X^{1/2} \Delta S X^{1/2} + \text{tr}(S^{1/2} \Delta X S^{1/2} \Delta X S^{1/2}) + 2\text{tr}(X\Delta S \Delta X S)) \\ &= \|X^{1/2} \Delta S X^{1/2}\|_F^2 + \|S^{1/2} \Delta X S^{1/2}\|_F^2 + 2\text{tr}(X\Delta S \Delta X S) \\ &= \|\Delta S\|_F^2 + \|S^{1/2} \Delta X S^{1/2}\|_F^2 + 2\text{tr}(\Delta S \Delta X X^{1/2} S X^{1/2}). \end{aligned}$$

$$\|S^{1/2}XS^{1/2} - \nu I\| = \frac{2}{\nu} \text{ pour } \nu = 8.$$

Montrons que

indéfinie.

Or, $p = 1156 = 34^2$, les valeurs propres sont $\frac{2}{67}$ et $-\frac{2}{1}$, $XS + SX$ est donc

$$\iff \lambda^2 - 32\lambda - 33 = 0.$$

$$\iff (16 - \lambda)^2 - 17^2 = 0$$

$$\det \begin{pmatrix} 16 - \lambda & 17 \\ 17 & 16 - \lambda \end{pmatrix}$$

En calculant les valeurs propres, nous obtenons

$$XS + SX = \begin{pmatrix} 8 & 1 \\ 1 & 0.5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 16 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 16 \end{pmatrix} \begin{pmatrix} 8 & 1 \\ 1 & 0.5 \end{pmatrix} = \begin{pmatrix} 16 & 17 \\ 17 & 16 \end{pmatrix}.$$

Calculons $XS + SX$:

$$X = \begin{pmatrix} 8 & 1 \\ 1 & 0.5 \end{pmatrix}, S = \begin{pmatrix} 1 & 0 \\ 0 & 16 \end{pmatrix}.$$

◊ Tout d'abord, considérons les matrices

pas que $XS + SX \in \text{SIR}^{n \times n}$ soit définie positive et réciproquement: Finalement, montrons par deux exemples que la condition (4.37) n'implique

positive. Utilisant des techniques similaires à celles utilisées dans la preuve de ce théorème, il est possible de montrer que la condition (4.37) implique la condition suffisante du Théorème 3.3.1, c'est-à-dire que la matrice $H^{-1}F$ est définie

Ce qui contredit la stricte inégalité inverse des équations (4.40). Nous pouvons donc conclure que $V = 0$. □

$$\begin{aligned} \text{tr}(X\Delta S + \Delta X S)_2 &\geq \|\Delta S\|_2^2 + \frac{\nu^2}{2} \|\Delta X\|_2^2 - 2\|\Delta S\|_F \|\Delta X\|_F \|X^{1/2}S^{1/2} - \nu I\|_F \\ &\geq \|\Delta S\|_2^2 + \frac{\nu^2}{2} \|\Delta X\|_2^2 - \nu \|\Delta S\|_F \|\Delta X\|_F \\ &= \|\Delta S\|_2^2 - \frac{\nu}{2} \|\Delta X\|_2^2 \\ &\geq 0. \end{aligned}$$

La matrice $E^{-1}F$ n'étant pas symétrique, nous ne pouvons calculer Δy ni par l'approche de la factorisation QR, ni par une factorisation de Cholesky de la matrice $AE^{-1}FA^T$. Dès lors, nous utilisons une factorisation LU de $AE^{-1}FA^T$. De plus, la multiplication par la matrice E^{-1} dans le calcul de ΔX demande de calculer la solution des équations de Lyapunov.

$$svec(\Delta X) = E^{-1}(svec(R_c) - F svec(\Delta S)).$$

$$\Delta S = R_d - smat(A^T \Delta y),$$

$$(AE^{-1}FA^T)\Delta y = r_p + AE^{-1}(F svec(R_d) - svec(R_c)),$$

Tout comme pour la direction NT, en utilisant l'élimination de Gauss par blocs, $(\Delta X, \Delta y, \Delta S)$ est obtenu par

$$r_p := b - A svec(X), R_d := C - S - \sum_{i=1}^m y_i A_i, R_c := \sigma \mu I - \frac{1}{2}(XS + SX).$$

$$E = I \otimes S, F = X \otimes I,$$

où

$$(4.41) \quad \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ E & 0 & F \end{pmatrix} \begin{pmatrix} svec(\Delta X) \\ \Delta y \\ svec(\Delta S) \end{pmatrix} = \begin{pmatrix} svec(R_d) \\ r_p \\ svec(R_c) \end{pmatrix},$$

Pour obtenir la direction $(\Delta X, \Delta y, \Delta S)$, il nous faut résoudre le système

4.3.2 Implémentation.

Chapitre 5

Comparaison des méthodes de points intérieurs primales-duales.

Dans ce chapitre, nous comparons les méthodes présentées dans le chapitre précédent par leurs propriétés ainsi que par l'application d'un algorithme prédicteur-correcteur.

5.1 Deux propriétés importantes.

Nous allons voir, dans cette section, deux propriétés importantes des directions de recherche.

5.1.1 La symétrie primale-duale.

Pour n'importe quelle méthode pour obtenir une direction de recherche, il correspond une méthode duale en échangeant les rôles de X et de S . Partons des problèmes primal (P) et dual (D) de départ:

$$(P) \left\{ \begin{array}{l} \min_x C \bullet X \\ \text{s.c.} \quad A_i \bullet X = b_i, i = 1, \dots, m \\ X \succeq 0 \end{array} \right.$$

Pour la direction NT, la symétrie primale-duale est obtenue par définition en regardant les équations (4.5) et (4.6). Par contre, les directions HKM ne sont pas symétriques primales-duales mais en commutant le rôle de X et de S pour l'une de ces directions, nous obtenons l'autre.

$$R_c = \mu I - \frac{1}{2}(XS + SX).$$

et

$$E(X, S) = S \otimes_s I = F(S, X)$$

Nous remarquons immédiatement que la direction AHO est symétrique primale-duale. En effet, nous avons

désirée.

valentes, et parfois, une forme d'écriture peut ne pas montrer la symétrie dernière équation du système (3.13) peut s'écrire de plusieurs manières équivalentes, et parfois, une forme d'écriture peut ne pas montrer la symétrie $R_c(S, X)$. Ces conditions sont suffisantes mais pas nécessaires, puisque la méthode est symétrique primale-duale si $E(X, S) = F(S, X)$ et $R_c(X, S) = R_c(S, X)$. Des lors, par le système (3.13), une méthode coïncide avec la direction originale. Une méthode symétrique primale-duale est une méthode dont la direction $R_c(S, X)$ est égale à $R_c(X, S)$. Les matrices X et S sont interchangées.

Si une méthode est définie par le système (3.13) avec des matrices $E = F(X, S)$, $F = F(X, S)$ et $R_c = R_c(X, S)$, alors la méthode duale est aussi définie par le système (3.13) avec des matrices $E' = F(S, X)$, $F' = F(S, X)$ et $R_c' = R_c(S, X)$. Ce qui signifie que la direction duale est $(\Delta S, \Delta y, \Delta X)$. La direction duale est $(\Delta X, \Delta y, \Delta S)$ où Δy est choisi de manière à ce que la première équation du système (3.13) soit satisfaite.

Tout d'abord, le problème (D) peut se réécrire sous la forme du problème (P) en terme de variable S , disons (P) . Un itéré (X, y, S) pour (P) et son dual (P) peut être vu comme un itéré (S, y, X) pour (P) et son dual. Nous pouvons alors appliquer la méthode à (P) et son dual à l'itéré (S, y, X) pour obtenir la direction $(\Delta S, \Delta y, \Delta X)$. La direction duale est $(\Delta X, \Delta y, \Delta S)$ où Δy est choisi de manière à ce que la première équation du système (3.13) soit satisfaite.

$$(D) \left\{ \begin{array}{l} \max_{y, S} y^T y \\ \text{s.c.} \quad \sum_{i=1}^m y_i A_i + S = C \\ S \succeq 0. \end{array} \right.$$

et

5.1.2 L'invariance sous changement d'échelle.

Une seconde propriété est importante pour les méthodes de points intérieurs, c'est l'invariance sous changement d'échelle.

En programmation linéaire, si x est remplacé par $D^{-1}x$ où D est une matrice diagonale et définie positive, et de manière correspondante, A devient AD , c devient Dc et s devient Ds , alors, pour tout algorithme, la direction de recherche au point mis à échelle $(D^{-1}x, y, Ds)$ du problème mis à échelle est $(D^{-1}\Delta x, y, D\Delta s)$, où $(\Delta x, \Delta y, \Delta s)$ est la direction de recherche au point (x, y, s) du problème original.

Nous disons qu'une direction de recherche pour le problème SDF joint de l'invariance sous changement d'échelle si nous avons une propriété similaire. Soit U une matrice $n \times n$ inversible quelconque et considérons le changement d'échelle où la matrice X est remplacée par $\tilde{X} := U^{-1}XU^{-T}$ et de manière correspondante chaque A_i par $\tilde{A}_i := U^T A_i U$, C par $\tilde{C} := U^T C U$ et S par $\tilde{S} := U^T S U$.

Nous définissons un nouveau problème mis à échelle: si (X, y, S) est admissible pour le problème original et son dual, alors $(\tilde{X}, y, \tilde{S})$ est admissible pour le problème mis à échelle et son dual avec la même valeur de la fonction objectif. Si la direction de recherche au point (X, y, S) pour le problème original est $(\Delta X, \Delta y, \Delta S)$, nous exigeons que la direction de recherche au point $(\tilde{X}, y, \tilde{S})$ pour le problème mis à échelle soit $(\Delta \tilde{X}, \Delta y, \Delta \tilde{S}) := (U^{-1}\Delta X U^{-T}, \Delta y, U^T \Delta S U)$.

Nous pouvons voir que la direction AHO n'est pas invariante sous changement d'échelle. En effet, $X S$ se transforme en $U^{-1}X S U$ tandis que $S X$ se transforme en $U^T S X U^{-T}$, dès lors, la direction de recherche obtenue ne peut pas être $(U^{-1}\Delta X U^{-T}, \Delta y, U^T \Delta S U)$. Nous devons vérifier si les autres directions sont invariantes sous changement d'échelle.

Tout d'abord, notons que la solution du système (3.13) correspondant à P est la même que celle correspondant à $\tilde{Q}P$ pour toute matrice \tilde{Q} orthogonale. En effet, en prémultipliant et postmultipliant l'équation (3.10) respectivement par P^T et P^{-T} , nous remarquons que la direction dépend uniquement

Puisque $H^p(XS)$ est définie positive et comme \tilde{Q} est inversible, $H^p(\tilde{X}\tilde{S})$ est bien définie positive. Dès lors, par le Théorème 3.3.1, le système original et le système mis à échelle possèdent tous deux une solution unique. Par la remarque ci-dessus, la direction de recherche résultante pour le problème mis à

$$\begin{aligned}
 H^p(\tilde{X}\tilde{S}) &= \frac{1}{2}[\tilde{P}^T\tilde{X}\tilde{S}\tilde{P}^{-1} + \tilde{P}^{-T}\tilde{S}\tilde{X}^T\tilde{P}^T] \\
 &= \frac{1}{2}[\tilde{Q}^T P U U^{-1} X U^{-T} S U^{-T} (\tilde{Q} P U)^{-1} + (\tilde{Q} P U)^{-T} U^T S U U^{-1} X U^{-T} (\tilde{Q} P U)^T] \\
 &= \frac{1}{2}[\tilde{Q}^T P X S P^{-1} \tilde{Q}^{-1} + \tilde{Q}^{-T} P^{-T} S X P^T \tilde{Q}^T].
 \end{aligned}$$

Preuve: Prenons $B := \tilde{P}^T \tilde{P} = U^T P^T P U$ et appliquons deux fois le Lemme 4.1.1, nous trouvons que $\tilde{P} = \tilde{Q} P U$, pour toute matrice \tilde{Q} orthogonale. De plus, la matrice $H^p(\tilde{X}\tilde{S})$ est définie positive. En effet, nous obtenons successivement

Théorème 5.1.1
 Supposons que la matrice P est définie par X et S de telle manière que si X et S sont transformées en $\tilde{X} := U^{-1} X U^{-T}$ et $\tilde{S} := U^T S U$, alors P est transformée en \tilde{P} avec $\tilde{P}^T \tilde{P} = U^T P^T P U$.
 Supposons également que $H^p(XS)$ est définie positive.
 Alors, la direction qui résout le système (3.13) est invariante sous changement d'échelle.
 En particulier, ce résultat est exact pour les directions définies par $P = X^{1/2}$, $P = S^{1/2}$ et $P = W^{-1/2}$.

En utilisant le résultat du Lemme 4.1.1, nous pouvons démontrer le théorème suivant:

$$\begin{aligned}
 P(\Delta X S + X \Delta S) P^{-1} + P^{-T} (S \Delta X + \Delta S X) P^T &\iff = 2\sigma^{\mu I} I - P X S P^{-1} - P^{-T} S X P^T \\
 P^T P (\Delta X S + X \Delta S) (P^T P)^{-1} + (S \Delta X + \Delta S X) &\iff = 2\sigma^{\mu I} I - P X S P^{-1} - P^{-T} S X P^T
 \end{aligned}$$

de $P^T P$:

Directions	Symétrie primale-duale	Invariance échelle	$H^p(XS) \geq 0$
AHO ($P = I$)	oui	non	non
HKM ($P = X_{-1/2}$ ou $P = S_{1/2}$)	non	oui	oui
NT ($P = W_{-1}$)	oui	oui	oui

Directions	E	F	$E^{-1}F$	R_c
AHO ($P = I$)	$I \otimes S$	$X \otimes I$	$(I \otimes S)_{-1} (X \otimes I)$	$\sigma_{II} - (XS + SX)/2$
HKM ($P = S_{1/2}$)	$S \otimes S$	$SX \otimes I$	$X \otimes S_{-1}$	$\sigma_{IS} - SXS$
HKM ($P = X_{-1/2}$)	$XS \otimes I$	$X \otimes X$	$(XS \otimes I)_{-1} (X \otimes X)$	$\sigma_{IX} - XSX$
NT ($P = G_{-1}$)	$G_{-1} \otimes G_{1/2} S$	$G_{-1} X \otimes G_{1/2}$	$(M \otimes M)$	$\sigma_{II} - D^2$

échelle n'est pas modifiée si nous utilisons $\tilde{F} = PU$. De plus, il est facile de vérifier que la direction mise à échelle ($\Delta X, \Delta y, \Delta S$) := $(U^{-1} \Delta X U^{-T}, \Delta y, U^T \Delta S)$ satisfait le système (3.13). Il suffit de remplacer $X, S, \Delta X, \Delta S$ et P par leurs valeurs mises à échelle dans l'équation (3.10).
 Notons encore que les trois choix donnés de la matrice P satisfont la première condition du Théorème 3.3.2, et dès lors la matrice $H^p(XS)$ est définie positive.
 □
 Ce théorème implique que les deux directions HKM ainsi que la direction NT sont invariantes sous chargement d'échelle.
 Voici un tableau récapitulatif des différentes formules pour chaque direction ainsi qu'un tableau des propriétés des trois types de directions mentionnés.

5.2 Comparaison des méthodes par l'application d'un algorithme prédicteur-correcteur.

Dans cette section, nous allons comparer les différentes méthodes vues précédemment à l'aide de l'algorithme prédicteur correcteur de Mehrotra. Cette comparaison est basée sur l'article de M. J. Todd, K.-C. Toh, et R. H. Tuntuncu [9]. Nous ne ferons ici qu'une présentation sommaire des procédés d'implémentation repris dans cet article. Notre but consiste en une comparaison simple des méthodes de points intérieurs primales-duales. Nous ne souhaitons pas nous attarder sur une longue description des exemples numériques. De plus, l'implémentation des méthodes de points intérieurs pourrait faire l'objet d'un mémoire ultérieur. Pour une information plus complète, le lecteur peut consulter l'article cité ci-dessus.

5.2.1 Algorithme prédicteur-correcteur de Mehrotra.

Algorithme 5.2.1

Supposons l'itéré initial (X_0, y_0, S_0) comme point de départ, où X_0 et S_0 sont définies positives.

Choisir $\tau \in (0, 1)$ et $\text{expon} \in \{1, 2, 3\}$.

Pour $k = 1, 2, \dots$, soit l'itéré (X, y, S) et soit $\mu := X \bullet S / n$.

1. (Pas prédicteur)

Calculer le pas Newton $(\Delta X, \Delta y, \Delta S)$ à partir du système (3.13) en utilisant $\sigma := 0$.

2. Déterminer le paramètre σ tel que

$$(5.1) \quad \sigma := \left[\frac{X \bullet S}{(X + \alpha \Delta X) \bullet (S + \beta \Delta S)} \right]^{\text{expon}}$$

où

$$(5.2) \quad \alpha := \min \left\{ 1, \frac{\lambda_{\min}(X^{-1} \Delta X)}{-\tau} \right\},$$

$$(5.3) \quad \beta := \min \left\{ 1, \frac{\lambda_{\min}(S^{-1} \Delta S)}{-\tau} \right\}.$$

3. (Pas correcteur)

Calculer le pas Newton $(\Delta X, \Delta y, \Delta S)$ à partir du système (3.13) en utilisant les paramètres σ et μ déterminés ci-dessus.

4. Mettre à jour les itérés par

$$X \rightarrow X + \alpha \Delta X,$$

$$y \rightarrow y + \beta \Delta y,$$

$$S \rightarrow S + \beta \Delta S,$$

où α et β sont déterminés par les équations (5.2) et (5.3).

Nous pouvons faire quelques remarques:

- Nous utiliserons deux valeurs de $\tau \in (0, 1)$. Une première est $\tau = 0.98$ qui s'est avérée être habituellement une bonne valeur de paramètre.

- L'algorithme *AHO-PC-Sch*: Cet algorithme utilise la direction AHO et l'implémente via une factorisation LU de la matrice $AE^{-1}FA^T$.
- L'algorithme *NT-PC-Sch*: Cet algorithme utilise également la direction NT mais calcule la direction Δy par l'équation de complètement de Schur (4.16) en appliquant une factorisation de Cholesky à $AE^{-1}FA^T$.
- L'algorithme *NT-PC-QR*: Cet algorithme utilise la direction NT et la calcule par l'approche d'une factorisation QR.

les quatre algorithmes suivants: directions de recherche et leurs implémentations. En effet, nous considérons SDP, les algorithmes prédicteurs-correcteurs obtenus suivant les différentes Nous allons comparer, en les appliquant sur des exemples de problèmes

5.2.2 Exemples numériques.

- Dans le calcul de α , cela revient moins cher de calculer la valeur propre minimale de la matrice symétrique $L^{-1}\Delta XL^{-T}$, où $X = LL^T$ est la factorisation de Cholesky de X . Cette matrice possède le même spectre que la matrice non symétrique $X^{-1}\Delta X$. Nous obtenons la même chose avec β .
- Le critère d'arrêt que nous utiliserons dans l'algorithme est le suivant: Nous arrêtons l'itération si $\sigma > 1$. Ce qui se produit habituellement lorsque le saut de dualité est dans la rangée de la précision atteignable par l'algorithme. L'algorithme se termine aussi lorsque les longueurs de pas α et β dans le pas correcteur sont plus petites que 10^{-6} .

$$expon = \begin{cases} 3 & \text{pour la méthode AHO,} \\ 1 & \text{pour les méthodes HKM,} \\ 1 & \text{pour la méthode NT.} \end{cases}$$

- Le paramètre *expon* est choisi de la manière suivante: correcteur. où α^k et β^k sont les longueurs de pas primale et duale au $k^{\text{ème}}$ pas

$$(5.4) \quad \tau^{(k+1)} = 0.9 + 0.09 \min\{\alpha^{(k)}, \beta^{(k)}\},$$

Une autre manière de choisir τ est de l'adapter aux longueurs de pas pris dans l'itération précédente c'est-à-dire en posant

Tous les calculs ont été réalisés en Matlab. Pour chacune des classes des problèmes SDP ci-dessus, dix exemples avec des matrices prises au hasard ont été effectués. Pour chaque ensemble des dix exemples, les algorithmes sont comparés en utilisant les deux valeurs du paramètre τ , c'est-à-dire $\tau = 0.98$ et τ adapté par l'équation (5.4). Pour ces deux valeurs de paramètre, les

$$\left\{ \begin{array}{l} \max e^T y \\ \text{s.c. } C - \text{Diag}(y) \succeq 0. \end{array} \right.$$

5. *Problème test instructif:*

$$\left\{ \begin{array}{l} \min L \bullet X \\ \text{s.c. } \text{diag}(X) = 1 \\ X \succeq 0. \end{array} \right.$$

4. *Problème "max-cut":*

où la minimisation porte sur la classe des polynômes monics de degré m et nous utilisons la norme-2. A est une matrice réelle $n \times n$. Le problème de Chebyshev est un cas particulier du problème de minimisation de la norme-2 et peut donc s'exprimer comme un problème SDP.

$$\min^d \|p(A)\|$$

3. *Problème d'approximation de Chebyshev pour une matrice:*

où les matrices A_k , $k = 0, \dots, m$, sont réelles de dimension $n \times n$.

$$\min_{y \in \mathbb{R}^m} \|A(y)\| = \min_{y \in \mathbb{R}^m} \|A_0 + \sum_{k=1}^m y_k A_k\|$$

2. *Problème de minimisation de la norme-2:*

1. *Un problème SDP pris au hasard.*

Les exemples que nous considérons sont les suivants:

- *L'algorithme HKM-PC-Sch:* Cet algorithme utilise la direction HKM primale ($P = S^{1/2}$) et l'implémentation en appliquant une factorisation de Cholesky à $AE^{-1}FA^T$. Cette direction est plus facile à implémenter que la direction duale ($P = X^{-1/2}$).

Nous pouvons faire les remarques suivantes :

- Du point de vue du nombre d'itérations, nous pouvons remarquer que l'algorithme AHO-PC-Sch est le plus efficace sur l'ensemble des problèmes considérés ici, mais les méthodes NT sont presque aussi efficaces, elles ne demandent que deux ou trois itérations supplémentaires pour obtenir la précision donnée sur le saut de dualité. La méthode HKM est presque aussi efficace que la méthode AHO, excepté pour le dernier exemple de problèmes SDP.
- Par contre, du point de vue du temps d'exécution, chaque itération de l'algorithme AHO-PC-Sch est beaucoup plus cher que chaque itération des méthodes HKM et NT. De plus, l'algorithme NT-PC-QR prend légèrement plus de temps que l'algorithme NT-PC-Sch qui lui-même prend plus de temps que l'algorithme HKM-PC-Sch.

τ adapté	AHO HKM NT			NT(QR)			
	AHO HKM NT			NT(QR)			
Exemples	AHO	HKM	NT	NT(QR)	AHO	HKM	NT
Problème 1	9.2	10.2	9.8	9.8	30.7	22.4	20.6
Problème 2	9.0	11.0	11.2	11.2	90.8	59.8	67.1
Problème 3	8.7	11.0	11.1	11.1	89.0	60.2	66.5
Problème 4	9.3	11.0	11.0	11.0	22.7	13.9	14.0
Problème 5	13.4	20.2	15.8	15.8	70.8	45.3	38.7
							44.2
Temps moyen d'itérations pour réduire le saut de dualité par 10^{10}	AHO HKM NT			NT(QR)			
	AHO HKM NT			NT(QR)			

$\tau = 0.98$	AHO HKM NT			NT(QR)			
	AHO HKM NT			NT(QR)			
Exemples	AHO	HKM	NT	NT(QR)	AHO	HKM	NT
Problème 1	9.0	10.4	10.1	10.1	27.6	19.7	20.2
Problème 2	9.3	11.0	11.5	11.4	90.4	58.2	66.1
Problème 3	9.4	13.0	11.4	11.5	90.3	68.2	66.4
Problème 4	9.4	11.1	11.1	11.1	22.3	13.2	14.0
Problème 5	13.7	25.5	17.2	17.1	65.7	57.1	41.6
							51.0
Temps moyen d'itérations pour réduire le saut de dualité par 10^{10}	AHO HKM NT			NT(QR)			
	AHO HKM NT			NT(QR)			

tables suivantes donnent le nombre moyen d'itérations et le temps moyen de chaque méthode pour réduire le saut de dualité par un facteur de 10^{10} .

5.3 Conclusion de la comparaison.

Même si la méthode AHO semble la plus simple et la plus efficace du point de vue du nombre d'itérations, elle présente des désavantages. En effet, cette méthode ne possède pas la propriété d'invariance sous changement d'échelle et l'unicité des directions de recherche n'est pas toujours garantie. De plus, le fait que l'implémentation demande une factorisation LU et une solution des équations de Lyapunov, ne l'avantage pas par rapport au temps nécessaire pour effectuer une itération.

Les directions HKM et NT semblent équitables. En effet, l'unicité de leur direction de recherche est toujours garantie. De plus, si l'une demande un nombre d'itération plus faible, elle demande un temps d'exécution plus long et réciproquement.

Mais, la direction NT est la plus complète puisqu'elle est la seule à posséder les deux propriétés de symétrie primale-duale et d'invariance sous changement d'échelle.

Conclusion

La programmation semi-définie possède deux avantages non négligeables. D'une part, elle recouvre une grande classe de problèmes standards tels que la programmation linéaire, la programmation quadratique ou convexe, ... D'autre part, la programmation semi-définie est une extension de la programmation linéaire. Ainsi, de nombreux résultats et de nombreuses méthodes de résolution, telles que les méthodes de points intérieurs s'y généralisent.

Dans ce mémoire, nous avons tenté de mieux comprendre en quoi consiste la programmation semi-définie et nous avons plus particulièrement développé les méthodes de points intérieurs primales-duales.

Tout d'abord, nous avons porté notre attention sur la théorie de la dualité. Les conditions de stricte admissibilité et d'indépendance linéaire sont essentielles pour obtenir la dualité forte. Par la définition et les propriétés de la fonction barrière logarithmique, nous avons obtenu la notion de trajectoire centrale.

Ensuite, nous avons souligné l'importance de la symétrisation de la dernière équation de cette trajectoire centrale. Nous avons remarqué que nous pouvons obtenir différentes directions de recherche pour les méthodes de points-intérieurs suivant la symétrisation. Nous avons ainsi établi un lien entre les différentes directions. Nous nous sommes également intéressés aux conditions suffisantes pour garantir l'unicité des directions de recherche.

Dans la dernière partie du mémoire, notre attention s'est portée sur trois méthodes de points intérieurs primales-duales appelées méthodes NT, HKM, et AHO. Nous avons tenté de les comparer en considérant les propriétés de symétrie primale-duale et d'invariance sous changement d'échelle. Nous avons

également appliqué un algorithme prédicteur-correcteur correspondant à ces méthodes, sur des exemples. Par cette comparaison, nous avons pu observer les avantages et les inconvénients de chaque méthode.

En conclusion, bien qu'étant une généralisation de la programmation linéaire, la programmation semi-définie est plus complexe. Elle comprend une multitude de choix de directions de recherche pour les méthodes de points intérieurs.

Annexe A

Notions sur les matrices.

A.1 Trace et normes d'une matrice.

Soit une matrice $A \in \mathbb{R}^{n \times n}$. La trace de A est définie par

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}.$$

Nous avons les propriétés suivantes:

$$\text{tr}(\alpha A + \beta B) = \alpha \text{tr}(A) + \beta \text{tr}(B),$$

$$\text{tr}(AB) = \text{tr}(BA),$$

$$\text{tr}(A) = \text{tr}(H A H^{-1}) \quad (H \text{ non singulière}),$$

$$\text{tr}(A) = \sum_{i=1}^n \lambda_i,$$

où les λ_i sont les valeurs propres de A . La norme de Frobenius d'une matrice $A \in \mathbb{R}^{n \times n}$ est définie par

$$\|A\|_F = \sqrt{\sum_{i,j=1}^n |a_{ij}|^2}.$$

Nous avons

$$\|A\|_2 = \sqrt{\text{tr}(A^T A)},$$

5. Si $U \succeq 0$, alors chaque $u_{jj} \geq 0$, et si $u_{jj} = 0$, $u_{jk} = u_{kj} = 0$ pour tout k . De manière similaire, si $U \preceq 0$, alors chaque $u_{jj} < 0$.
4. Pour $U \in \text{SR}^{n \times n}$, les propriétés suivantes sont équivalentes:
- $U \succeq 0$ ($U \preceq 0$),
 - $v^T U v \geq 0 \forall v \in \mathbb{R}^n$ ($v^T U v < 0 \forall v \in \mathbb{R}^n$),
 - $\lambda(U) \geq 0$ ($\lambda(U) < 0$), et
 - $U = P^T P$ pour toute matrice P ($U = P^T P$ pour toute matrice P inversible).

$$\|\lambda(U)\|_1 = \sum_{j=1}^n |\lambda_j(U)| = \sum_{j=1}^n \lambda_j(U) = I \bullet \lambda(U) = \text{tr}(U) = I \bullet U.$$

Si $U \succeq 0$,

$$\|\lambda(U)\|_2 = \|U\|_F, \|\lambda(U)\|_\infty = \|U\|_2.$$

3. Nous avons les égalités suivantes pour les normes sur $\text{SR}^{n \times n}$:
2. Toute matrice $U \in \text{SR}^{n \times n}$ peut s'écrire $U = Q \Lambda Q^T$, où Q est orthogonale et Λ est diagonale. Alors, $U Q = Q \Lambda$ où les colonnes de Q sont les vecteurs propres et les entrées diagonales de Λ sont les valeurs propres correspondantes de U . Nous écrivons $Q(U) := Q$, $\lambda(U) := \Lambda$ et $\lambda(U) := \text{diag}(\Lambda)$.
1. Si $U, V \in \text{SR}^{n \times n}$ et Q est orthogonale, Alors $U \bullet V = (Q^T U Q) \bullet (Q^T V Q)$. De façon plus générale, si P est inversible, $U \bullet V = (P U P^T) \bullet (P^{-T} V P^{-1})$.

A.2 Propriétés d'algèbre linéaire.

où les λ_i sont les valeurs propres de A .

$$\|A\| = \max |\lambda_i|,$$

De plus, si A est une matrice symétrique,

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n} \|A\|_2,$$

$$\|A\|_2 = \max \sqrt{\lambda(A^T A)},$$

Si A est de rang plein, alors il y a une solution unique x^* au problème aux moindres carrés et elle résout le système linéaire d'équations normales) $A^T A x^* = A^T b$. Ce système est défini positif.

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2.$$

Le problème aux moindres carrés de rang plein est posé comme suit: pour une matrice $A \in \mathbb{R}^{m \times n}$ (avec $m \geq n$) et un vecteur $b \in \mathbb{R}^m$ donnés, on cherche un vecteur $x \in \mathbb{R}^n$ qui soit solution de

A.3 Le problème aux moindres carrés.

- une décomposition en valeurs propres avec la même matrice Q).
13. Si $U, V \in \text{SIR}^{n \times n}$, alors U et V commutent si et seulement si U et V peuvent être simultanément diagonalisés (c'est-à-dire qu'ils possèdent
 12. Si $U, V \succeq 0$, alors $U \bullet V = 0$ si et seulement si $UV = 0$.
 11. Si $U \succ 0$, alors $U \bullet V > 0$ pour toute matrice non nulle $V \succeq 0$, et $\{V \succeq 0 : U \bullet V \leq \beta\}$ est borné pour tout β positif ou nul.
 10. Si $U \in \text{SIR}^{n \times n}$, alors $x^T U x = U \bullet x x^T$.
 9. $U \succeq 0$ ($U \succ 0$) si et seulement si tout mineur principal est positif ou nul (positif).
 8. Toute sous-matrice principale d'une matrice (semi-)définie positive est (semi-)définie positive.
 7. Si U est symétrique, alors PUP^T est symétrique. Si P est inversible et PUP^T est symétrique, alors U est symétrique.
 6. Si $U \succeq 0$, alors $PUP^T \succeq 0$ pour toute matrice P de dimension des colonnes appropriées. Si P est une matrice carrée et inversible, alors $U \succ 0$ si et seulement si $PUP^T \succ 0$.

Annexe B

Le produit de Kronecker.

Voici une série de propriétés de produits de Kronecker standard et symétrisés.

◊ Le produit de Kronecker standard:

Si U est une matrice symétrique $n \times n$, $vec(U)$ est défini par

$$vec(U) := (u_{11}, u_{21}, \dots, u_{n1}, u_{22}, u_{32}, \dots, u_{n2}, \dots, u_{nn})^T.$$

$$1. G \otimes K = [g_{ij}K].$$

$$2. (G \otimes K)vec(H) = vec(KHG^T).$$

$$3. (G \otimes K)^T = G^T \otimes K^T.$$

$$4. G \otimes I \text{ est symétrique si et seulement si } G \text{ l'est.}$$

$$5. (G \otimes K)^{-1} = G^{-1} \otimes K^{-1}.$$

$$6. (G \otimes K)(H \otimes L) = GH \otimes KL.$$

$$7. \text{ Si } \Lambda(G) = \{\lambda_i\} \text{ et } \Lambda(K) = \{\mu_j\}, \text{ alors } \Lambda(G \otimes K) = \{\lambda_i \mu_j\}. \text{ Si } x_i \text{ et } y_j \text{ sont les vecteurs propres correspondant aux valeurs propres } \lambda_i \text{ et } \mu_j$$

y_j sont les vecteurs propres correspondant aux valeurs propres λ_i et μ_j respectivement de G et de K , alors $vec(y_j x_i^T)$ est le vecteur propre de

$$8. vec(G)^T vec(K) = G \bullet K.$$

$$8. vec(G)^T vec(K) = G \bullet K.$$

◊ Le produit de Kronecker symétrisé:

Si U est une matrice symétrique $n \times n$, $svec(U)$ est défini par

$$svec(U) := (u_{11}, \sqrt{2}u_{21}, \dots, \sqrt{2}u_{n1}, u_{22}, \sqrt{2}u_{32}, \dots, \sqrt{2}u_{n2}, \dots, u_{nn})^T.$$

$$1. (G \otimes_s K)svec(H) = \frac{1}{2}svec(KHG^T + GHK^T).$$

2. $G \otimes_s K = K \otimes_s G$.
3. $(G \otimes_s K)_T = G^T \otimes_s K^T$.
4. $G \otimes_s I$ est symétrique si et seulement si G l'est.
5. $(G \otimes_s G)^{-1} = G^{-1} \otimes_s G^{-1}$.
6. $(G \otimes_s K)(H \otimes_s L) = \frac{1}{2}(GH \otimes_s KL + GL \otimes_s KH)$.
7. $(G \otimes_s K)(H \otimes_s H) = (GH \otimes_s KH)$ et $(H \otimes_s H)(G \otimes_s K) = (HG \otimes_s HK)$.
8. Si $\Lambda(G) = \{\lambda_i\}$, alors $\Lambda(G \otimes_s G) = \{\lambda_i \lambda_j\}$. Si x_i et x_j sont les vecteurs propres correspondant aux valeurs propres λ_i et λ_j de G , alors $\text{vec}(\frac{1}{2}(x_i x_j^T + x_j x_i^T))$ est le vecteur propre de la valeur propre $\lambda_i \lambda_j$ de $G \otimes_s G$.
9. Soit G et K deux matrices symétriques qui commutent avec une base commune de vecteurs propres $\{x_i\}$ correspondant respectivement aux valeurs propres λ_i et μ_i . Alors, $\Lambda(G \otimes_s K) = \{\frac{1}{2}(\lambda_i \mu_j + \lambda_j \mu_i)\}$. Nous avons aussi que $\text{vec}(\frac{1}{2}(x_i x_j^T + x_j x_i^T))$ est un vecteur propre correspondant à la valeur propre $\frac{1}{2}(\lambda_i \mu_j + \lambda_j \mu_i)$ de $G \otimes_s K$.
10. $\text{vec}(G)_T \text{vec}(K) = G \bullet K$.
11. Si G et K sont symétriques et définies positives, alors $G \otimes_s K$ l'est aussi.

Bibliographie

- [1] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. *Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results*. SIAM J. Optim., 8:746-768, 1998.
- [2] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Third edition, 1996.
- [3] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New-York, 1991.
- [4] R. D. C. Monteiro. *Polynomial convergence of primal-dual algorithms for semidefinite programming*. SIAM J. Optim., 7:663-678, 1997.
- [5] R. D. C. Monteiro and P. R. Zangiacomo. *A note on the existence of the Alizadeh-Haeberly-Overton direction for semidefinite programming*. Math. Programming, 78 (1997), 393-396.
- [6] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [7] M. J. Todd. *A study of search directions in interior-point methods for semidefinite programming*. Optim. Methods Softw., 11&12:1-46, 1999.
- [8] M. T. Todd. *Semidefinite Optimization*. Acta Numerica, 10:515-560, 2001.
- [9] M. J. Todd, K.-C. Toh, and R. H. Tuntunou. *On the Nesterov-Todd direction in semidefinite programming*. SIAM J. Optim., 8:769-796, 1998.
- [10] Y. Zhang. *On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming*. SIAM J. Optim., 8:365-386, 1998.