



UNIVERSITÉ FERHAT ABBAS SÉTIF-1
FACULTÉ DES SCIENCES
DÉPARTEMENT DE MATHÉMATIQUES



THÈSE

Présentée par :

BACHIR CHERIF Larbi

Pour obtenir le titre de Doctorat LMD

OPTION

Optimisation et contrôle

Thème

*Technique des fonctions approximantes et leurs applications en
problème d'optimisation*

Sous la direction de Pr. MERIKHI Bachir

Membres du jury :

Président :	Mr. BENSALÉM Naceurdine	Professeur	UFA. Sétif-1
Rapporteur :	Mr. MERIKHI Bachir	Professeur	UFA. Sétif-1
Co-Rapporteur :	Mr. BENTERKI Djamel	Professeur	UFA. Sétif-1
Examineur :	Mr. ADDOUNE Smail	M.C.A	Univ. de B.B. Arréridj
Examineur :	Mr. BOUAFIA Mousaab	M.C.A	Université de Guelma

2021 /2022

Remerciements

J'adresse mes remerciements aux personnes qui m'ont aidé dans la réalisation de cette thèse.

Je tiens à exprimer ma reconnaissance et mes remerciements les plus profonds à mon **Pr B. MERIKHI**, professeur à l'université de Sétif-1, pour m'avoir proposé ce passionnant sujet, pour sa disponibilité, sa patience, son suivi, ses conseils et son soutien constant tout au long de ce travail.

Je remercie Monsieur **N. BENSALEM**, Professeur à l'université de Sétif-1, de m'avoir fait l'honneur de faire partie de ce jury et d'en être le président.
je remercie Monsieur **D. BENTERKI**, Professeur à l'université de Sétif-1, pour son aide et ses conseils.

Je remercie aussi Monsieur **S. ADDOUNE**, M.C.A à l'université de BBA pour ses conseils, son aide précieuse et pour tous les encouragements et les discussions fructueuses, de même je remercie Monsieur **M. BOUAFIA**, M.C.A à l'université de Guelma pour avoir accepté de participer au jury de cette thèse.
Je remercie également, Monsieur **M. ACHACHE**, Professeur à l'université de Sétif-1, pour ses encouragements.

Enfin, je ne saurais oublier d'adresser un gentil mot à mes chers parents.
Je remercie très spécialement ma femme **K. BECHANE**, qui a toujours été là pour moi.
Je tiens à remercier tous mes amis, pour leurs amitiés et leurs encouragements, spécialement Docteur **L. ZIAD**.

Table des matières

1	Notions de base	10
1.1	Définitions générales	10
1.1.1	Ensembles convexes	10
1.1.2	Fonctions convexes	14
1.1.3	Position du problème	16
1.1.4	Transformations simples	17
1.2	Qualification des contraintes	20
1.3	Résultats d'existence et d'unicité	21
1.4	Conditions d'optimalité	22
2	Application des fonctions approximantes	24
2.1	Application en programmation linéaire (PL)	24
2.1.1	Description de l'algorithme	27
2.1.2	Simulations numériques	33
2.2	Application en SDP	36
2.2.1	Problème dual	37
2.2.2	Méthode de points intérieurs pour résoudre (SDP)	37
2.2.3	Méthode barrière logarithmique via les fonctions approximantes	38

2.2.4	Calcul du pas de déplacement par des fonctions minorantes . . .	40
2.2.5	Calcul du pas de déplacement via des fonctions majorantes . . .	47
3	Une variante de la méthode des pénalités logarithmiques pour la programmation convexe	53
3.1	Introduction	53
3.2	La pénalisation	54
3.2.1	Le problème perturbé	54
3.2.2	Convergence	55
3.2.3	Le principe de la méthode	57
3.3	Résolution du problème perturbé	57
3.3.1	La direction de descente	58
3.3.2	Une fonction majorante	59
3.3.3	Minimisation d'une fonction auxiliaire	60
3.3.4	Cas linéaire.	61
3.3.5	Cas convexe	61
3.4	L'algorithme	62
3.5	Simulations numériques	63
3.5.1	Fonction objectif convexe non linéaire	63
3.5.2	Problème à tailles variables	64
4	Application de la nouvelle variante dans le cas quadratique	68
4.1	La programmation quadratique	68
4.1.1	La programmation quadratique sans contraintes	68
4.1.2	La programmation quadratique avec contraintes	69
4.2	Méthode du gradient projeté via les fonctions majorantes	70

Liste des tableaux

2.1	Tests numériques des exemples de tailles fixes	35
2.2	Tests numériques de l'exemple 2.5	36
2.3	Tests numériques pour les exemples 2.6, 2.7, 2.8 et 2.9 par les fonctions minorantes	46
2.4	Tests numériques (via fonctions majorantes) pour l'exemple 2.10 . . .	52
3.1	Tests numériques pour les exemples 3.1 et 3.2.	63
3.2	Tests numériques pour l'exemple 3.3	64
3.3	Tests numériques pour l'exemple 3.4 dans le cas $(a_i = 1, b_j = 6)$. . .	65
3.4	Tests numériques pour l'exemple 3.4 dans le cas $(a_i = 2, b_j = 4)$. . .	66
3.5	Tests numériques pour l'exemple 3.5	67
4.1	Tests numériques pour les exemples 3.1 et 3.2	74
4.2	Tests numériques de l'exemple 3.5	74

Liste des Algorithmes

1	Algorithme de Karmarkar via la fonction majorante	32
2	Résolution d'un problème <i>SDP</i> par les fonctions minorantes	44
3	Résolution d'un problème <i>SDP</i> par les fonctions majorantes	50
4	Algorithme du gradient projeté via la fonction majorante	73

Introduction générale

L'optimisation mathématique, regroupe un ensemble de sujets dans l'étude de problèmes d'optimisation. Une fonction objectif, est donnée, et le problème consiste à trouver un point minimisant (ou maximisant) cette fonction. Parfois, tous les points de l'espace sont candidats, parfois, des contraintes limitent le domaine de recherche. Les fonctions utilisées sont continues, même différentiables. On utilise des résultats d'analyse mathématique pour caractériser les points candidats, un premier pas consiste donc à obtenir des conditions satisfaites par les minima ou maxima recherchés. Et finalement un algorithme itératif réduisant (pour la minimisation) progressivement la fonction objectif. Dans notre étude, on s'intéresse à la méthode de pénalisation qui est un concept simple permettant de transformer un problème d'optimisation avec contraintes en un problème où en une suite de problèmes d'optimisation sans contraintes ou avec des contraintes simples. C'est un concept qui a une utilité à la fois théorique et numérique. Ce que le qualificatif « simple » signifie dépendra du contexte, des questions que l'on se pose, de la disponibilité d'algorithmes de résolution. En analyse, l'approche par pénalisation est parfois utilisée pour étudier un problème d'optimisation dont certaines contraintes sont difficiles à prendre en compte, alors que le problème pénalisant ces contraintes difficiles a des propriétés (l'existence de solution par exemple) mieux comprises ou plus simples à mettre en évidence. Si l'on a de la chance ou si la pénalisation est bien choisie, des passages à la limite parfois délicats permettent d'obtenir des propriétés du problème original. Par exemple, on peut obtenir des conditions d'optimalité d'un problème avec contraintes, à partir des conditions d'optimalité des problèmes pénalisés. D'autre part, comme nous allons le souligner ci-dessous, la pénalisation est un outil permettant d'étudier les problèmes d'optimisation avec et sans contraintes dans un même formalisme. D'un point de vue numérique, cette transformation permet d'utiliser des algorithmes d'optimisation sans contraintes pour obtenir la solution de problèmes dont l'ensemble admissible peut avoir une structure complexe. Cette approche est de ce fait très souvent utilisée.

L'algorithme itératif réduit à chaque itération l'objectif perturbé (non linéaire ou quadratique), et ce en produisant une suite d'itérés convergeant vers la solution optimale, suivant une direction de descente, calculée par la méthode de Newton ou gradient projeté, et un pas de déplacement calculé par une nouvelle technique basée sur les fonctions majorantes.

La thèse est divisée en quatre chapitres à savoir :

Le premier chapitre, est destiné à introduire les concepts de base de l'optimisation continue sans et avec contraintes.

Le deuxième chapitre, donne un rappel sur l'application de la technique des fonctions approximantes pour le calcul du pas de déplacement, et ce, en optimisation linéaire et semi-définie (*SDP*).

Une variante de pénalité logarithmique, fait le sujet du troisième chapitre, dans lequel on fait une analyse complète de cette dernière, allant de l'étude théorique à l'implémentation numérique.

On termine par le dernier chapitre, où on résout un problème quadratique sous des contraintes linéaires : La direction de descente est calculée par la méthode du gradient projeté, pour le calcul du pas de déplacement on utilise la technique de fonctions majorantes.

Pour tester la robustesse de nos nouvelles variantes, des simulations numériques importantes sont effectuées à la fin des deux derniers chapitres.

Notations

\mathbb{N}	Ensemble des entiers naturels.
\mathbb{R}	Ensemble des nombres réels.
\mathbb{C}	Ensemble des nombres complexes.
\mathbb{R}^n	Ensemble des nombres réels d'ordre n .
V_x	Un voisinage de $x \in \mathbb{R}^n$.
$\text{Im}(f)$	Ensemble des images de f , $\text{Im}(f) = \{f(x), x \in E\}$.
$\frac{\partial f}{\partial x_i}$	Dérivée de f par rapport à x_i .
∇f	Opérateur gradient. Soit f une fonction définie sur \mathbb{R}^n à valeurs dans \mathbb{R} , le gradient $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est l'opérateur défini par

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

$\nabla^2 f$	Hessien de la fonction f définie sur \mathbb{R}^n à valeurs dans \mathbb{R} , c'est l'opérateur $\nabla^2 f : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ défini par
--------------	--

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}.$$

$\ \cdot\ $	désigne une norme euclidienne sur \mathbb{R}^n .
$\langle \cdot, \cdot \rangle$	désigne le produit scalaire canonique sur \mathbb{R}^n , tel que $\langle x, y \rangle = y^T x$, pour $x, y \in \mathbb{R}^n$, où y^T désigne le transposé de y .

- $M_n(\mathbb{R})$ Ensemble des matrices carrés réelles de taille n .
- S_n Ensemble des matrices symétriques.
- S_n^+ Ensemble des matrices symétriques semi-définies positives.
- S_n^{++} Ensemble des matrices symétriques définies positives.

Chapitre 1

Notions de base

1.1 Définitions générales

1.1.1 Ensembles convexes

Un sous-ensemble S de \mathbb{R}^n est dit convexe si est seulement si

$$\lambda x + (1 - \lambda)y \in S, \quad \text{pour tout } x, y \in S, \text{ et pour tout } \lambda \in [0, 1].$$

Géométriquement, $z(\lambda) = \lambda x + (1 - \lambda)y$ est le barycentre des points x et y affectés (des poids) respectifs λ et $(1 - \lambda)$. Lorsque λ décrit l'intervalle $[0, 1]$, $z(\lambda)$ décrit le segment fermé d'extrémités x et y que l'on notera $[x, y]$. Dire que S est convexe, c'est donc dire que S contient le segment dès qu'il contient ses extrémités (voir Figure 1.1).

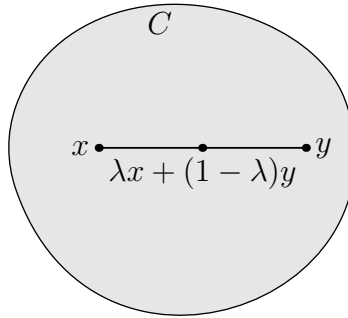


FIGURE 1.1: Ensemble convexe

Dans la proposition suivante on regroupe quelques propriétés algébriques et topologiques concernant les ensembles convexes.

Proposition 1.1.

1. Si $(S_i)_{i \in I}$ est une famille quelconque d'ensembles convexes alors leur intersection $\bigcap_{i \in I} S_i$ est convexe.
2. Le produit cartésien de deux ensembles convexes est convexe.
3. L'image directe (resp. réciproque) d'un convexe par une application affine est convexe.
4. La somme (de Minkowski) de deux convexes est convexe.

Définition 1.1 (Combinaison convexe). Pour toute famille x_1, \dots, x_m de points de \mathbb{R}^n et tout système $\lambda_1, \dots, \lambda_m$ de nombres réels positifs ou nuls avec $\sum_{i=1}^m \lambda_i = 1$, le point $x = \sum_{i=1}^m \lambda_i x_i$ s'appelle combinaison convexe des points x_1, \dots, x_m .

Définition 1.2 (L'enveloppe convexe). Soit A une partie de \mathbb{R}^n . L'enveloppe convexe de A noté $\text{conv}(A)$ est l'ensemble de toutes les combinaisons convexes d'éléments de A .

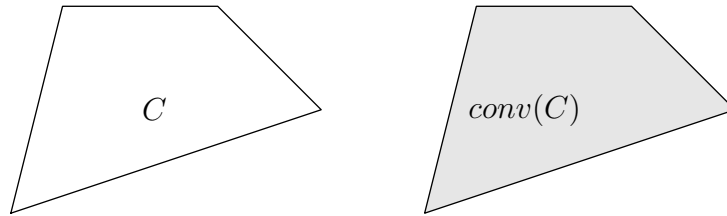


FIGURE 1.2: Un ensemble et son enveloppe convexe.

Propriétés 1.1. Soit A une partie de \mathbb{R}^n .

1. $\text{conv}(A)$ est convexe,
2. A est inclus dans $\text{conv}(A)$,
3. si C est un sous-ensemble convexe de \mathbb{R}^n contenant A , alors $\text{conv}(A)$ est inclus dans C ,
4. $\text{conv}(\emptyset) = \emptyset$.
5. $\text{conv}(A)$ est l'intersection de toutes les parties convexes de \mathbb{R}^n contenant A .
6. A est convexe si et seulement si $\text{conv}(A) = A$.

Le théorème suivant nous dit qu'en dimension finie le nombre de points qui contribuent dans la représentation d'un élément quelconque de l'enveloppe convexe d'un ensemble donné ne peut pas dépasser la dimension de l'espace plus un.

Théorème 1.1 (Carathéodory). Dans un espace affine de dimension n , l'enveloppe convexe d'un sous-ensemble A est l'ensemble des combinaisons convexes de familles de $n + 1$ points de A .

On obtient comme résultat du théorème précédent le corollaire suivant :

Corollaire 1.1.1. Dans un espace affine de dimension finie, l'enveloppe convexe d'un compact est compacte.

On note que l'enveloppe convexe d'un fermé (non borné) n'est pas en général fermé.

Un point $x \in S$ est dit extrémal si on ne peut pas l'exprimer comme combinaison convexe de deux autres points distincts de S , c'est à dire

$$\text{si } x = (1 - \lambda)y + \lambda z \text{ avec } y, z \in S \text{ et } \lambda \in]0, 1[\text{ alors } x = y = z.$$

Définition 1.3 (Ensembles affines). *Un sous ensemble S de \mathbb{R}^n est dit affine si*

$$\lambda x + (1 - \lambda)y \in S, \quad \forall x, y \in S, \forall \lambda \in \mathbb{R}.$$

Définition 1.4. *Une application $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est dite affine si :*

$$T[(1 - \lambda)x + \lambda y] = (1 - \lambda)T[x] + \lambda T[y] \quad \forall x, y \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}.$$

Définition 1.5 (Hyperplan). *On appelle hyperplan de \mathbb{R}^n toute partie affine de dimension $(n - 1)$. Ce sont des espaces orthogonaux à des espaces de dimension un. Tout hyperplan affine peut être représenté par un vecteur non nul $p \in \mathbb{R}^n$ et un scalaire $\alpha \in \mathbb{R}$. Plus précisément, H est un hyperplan de \mathbb{R}^n si et seulement s'il existe un vecteur $p \neq 0$ et un nombre réel α tels que*

$$H = \{x \in \mathbb{R}^n : p^\top x = \alpha\}.$$

Définition 1.6 (Polyèdre). *Un sous ensemble S de \mathbb{R}^n est dit polyèdre si on peut l'écrire sous la forme*

$$S = \{x \in \mathbb{R}^n : Ax \leq b\}$$

où A est une matrice réelle $m \times n$ et $b \in \mathbb{R}^m$.

Les polyèdres bornés sont appelés *polytopes*, on peut les représenter comme étant l'enveloppe convexe de leurs points extrêmes.

Un simplexe est un polytope qui est engendré par un nombre fini de points affinement indépendants.

Définition 1.7 (Projection sur un convexe). *Soit X un convexe fermé, non vide de \mathbb{R}^n . La projection d'un point $x \in \mathbb{R}^n$ sur X , notée $p_X(x)$, est obtenue comme solution du problème d'optimisation suivant*

$$\text{Minimiser } \frac{1}{2}\|x - y\|_2^2 \text{ sous la contrainte } y \in X.$$

La fonction $f(y) = \frac{1}{2}\|x - y\|_2^2$ étant convexe et $\nabla f(y) = y - x$, une condition nécessaire et suffisante pour que $x^ = p_X(x)$ soit solution de problème précédent est*

$$(x - x^*)^\top (y - x^*) \leq 0, \quad \forall y \in X.$$

On remarque en particulier que si $x \in X$ alors on a nécessairement $x^* = x$.

1.1.2 Fonctions convexes

Définition 1.8 (Fonction convexe). Soit S une partie convexe non vide de \mathbb{R}^n et $f : S \rightarrow \mathbb{R}$ une fonction. La fonction f est dite convexe sur S si et seulement si pour tout $x, y \in S$ et $\lambda \in]0, 1[$ on a

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (1.1)$$

1. Si l'inégalité (1.1) est stricte pour tout $x \neq y$ alors la fonction f est dite strictement convexe.
2. La fonction f est dite concave (resp. strictement concave) si $-f$ est convexe (resp. strictement convexe).

f est dite fortement convexe sur S s'il existe $\alpha > 0$ tel que pour tout $\lambda \in]0, 1[$ et $x, y \in S$ on a

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{1}{2}\alpha\lambda(1 - \lambda)\|x - y\|^2.$$

f est dite quasi-convexe sur S si

$$f(\lambda x + (1 - \lambda)y) \leq \max(f(x), f(y)), \quad \forall \lambda \in [0, 1], \forall x, y \in S.$$

Si f est une fonction différentiable sur un ouvert contenant S alors on obtient les caractérisations suivantes :

▷ f est convexe sur S si et seulement si

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle, \quad \text{pour tout } x, y \in S.$$

▷ f est convexe sur S si et seulement si ∇f est monotone sur S , c'est à dire

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0, \quad \text{pour tout } x, y \in S.$$

▷ f est fortement convexe sur S si et seulement s'il existe $\alpha > 0$ tel que

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2, \quad \text{pour tout } x, y \in S.$$

▷ f est fortement convexe sur S si et seulement si ∇f est fortement monotone sur S , c'est à dire il existe $\alpha > 0$, tel que :

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|y - x\|^2, \quad \text{pour tout } x, y \in S.$$

Théorème 1.2 (Relation entre la matrice hessienne et la convexité). *Soit S un convexe de \mathbb{R}^n et $f : S \rightarrow \mathbb{R}$ une fonction deux fois différentiable sur un ouvert contenant S . Alors f est convexe sur S si et seulement si la matrice hessienne $\nabla^2 f(x)$ est semi-définie positive pour tout $x \in S$, c'est à dire*

$$\langle y, \nabla^2 f(x)y \rangle \geq 0 \quad \text{pour tout } y \in \mathbb{R}^n.$$

Remarque 1.1. *Si la matrice hessienne de f est définie positive sur S alors f est strictement convexe sur S . La réciproque n'est pas en général vraie (la fonction $x \mapsto x^4$ est strictement convexe sur \mathbb{R} mais sa dérivée seconde n'est pas strictement positive sur \mathbb{R}).*

Définition 1.9. *Soit A une partie non bornée de \mathbb{R}^n . On dit qu'une fonction f définie sur A à valeurs dans \mathbb{R} est coercive sur A si*

$$\lim_{\substack{\|x\| \rightarrow +\infty \\ x \in A}} f(x) = +\infty.$$

Définition 1.10 (Hypographe). *Soit f une fonction définie sur un ensemble $A \subset \mathbb{R}^n$ à valeurs dans \mathbb{R} . L'hypographe de f est l'ensemble, noté $\text{hyp} f$, défini par*

$$\text{hyp} f = \{(x, \alpha) \in A \times \mathbb{R} : f(x) \geq \alpha\}.$$

L'hypographe strict de f est l'ensemble, noté $\text{hyp}_s f$, défini par

$$\text{hyp}_s f := \{(x, \alpha) \in A \times \mathbb{R} : f(x) > \alpha\}.$$

Définition 1.11. *On dit qu'une fonction f définie au voisinage d'un point x_0 est semi-continue inférieurement (s.c.i) en ce point si pour tout $\epsilon > 0$, il existe un voisinage U de x_0 tel que*

$$f(x) \geq f(x_0) - \epsilon \quad \text{pour tout } x \in U.$$

La fonction f est dite semi-continue inférieurement sur une partie fermée $A \subset \mathbb{R}^n$ si l'une des propriétés équivalentes suivantes est vérifiée

- ▷ la fonction f est semi-continue inférieurement en tout point de A .
- ▷ pour tout réel α , l'ensemble de sous-niveau $\{x \in A \mid f(x) \leq \alpha\}$ est fermé.
- ▷ l'épigraphe $\{(x, \alpha) \in A \times \mathbb{R} \mid f(x) \leq \alpha\}$ est fermé dans $\mathbb{R}^n \times \mathbb{R}$.

Définition 1.12. On dit que f est semi-continue supérieurement (s.c.s) en x_0 si pour tout $\epsilon > 0$, il existe un voisinage U de x_0 tel que

$$f(x) \leq f(x_0) - \epsilon \quad \text{pour tout } x \in U.$$

Comme pour les fonctions s.c.i, si A est une partie fermée de \mathbb{R}^n alors on a équivalence entre les assertions suivantes

- ▷ f est semi-continue supérieurement en tout point de \mathbb{R}^n .
- ▷ pour tout réel α , l'ensemble de sur-niveau $\{x \in A \mid f(x) \geq \alpha\}$ est fermé.
- ▷ l'hypographe $\{(x, \alpha) \in A \times \mathbb{R} \mid f(x) \geq \alpha\}$ est fermé dans $\mathbb{R}^n \times \mathbb{R}$.

1.1.3 Position du problème

Définition 1.13 (Qu'est-ce-qu'un problème d'optimisation?). *Un problème d'optimisation est défini par la donnée d'un ensemble $C \subset \mathbb{R}^n$ (ensemble des actions ou de stratégies) et une fonction f (critère ou fonction objectif) définie sur ce dernier. La question est de déterminer, s'il existe, une action dans C qui rend f optimale.*

D'un point de vue mathématique, le problème se formule de la façon suivante

problème sans contraintes

$$(P_1) \quad \min_{x \in \mathbb{R}^n} f(x)$$

problème avec contraintes

$$(P_2) \quad \min_{x \in C} f(x).$$

L'ensemble C désigne les contraintes du problème qui sont souvent de type inégalité

$$C = \{x \in \mathbb{R}^n : g(x) \leq 0\}$$

ou de type égalité

$$C = \{x \in \mathbb{R}^n : h(x) = 0\}$$

ou une combinaison des deux. Les fonctions g et h sont définies sur \mathbb{R}^n à valeurs dans \mathbb{R}^m et \mathbb{R}^p respectivement.

Définition 1.14 (Extrema locaux et globaux). Soit $f : D \rightarrow \mathbb{R}$ une fonction définie sur une partie $D \subset \mathbb{R}^n$.

1. On dit que f admet un maximum (resp. minimum) global au point $a \in D$ si on a

$$f(x) \leq f(a) \text{ (resp. } f(x) \geq f(a)) \text{ pour tout } x \in D. \quad (1.2)$$

Le maximum (resp. minimum) est appelé strict si l'inégalité (1.2) est stricte pour tout $x \in D \setminus \{a\}$.

2. On dit que f admet un maximum (resp. minimum) local au point $a \in D$ si on peut trouver un nombre $r > 0$ tel que $x \in D$ et $\|x - a\| < r$ entraîne $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

Les extrema globaux sont appelés aussi extrema absolus.

1.1.4 Transformations simples

Voici quelques transformations simples qui sont souvent utilisées pour transformer un problème en un autre plus "facile" à résoudre et ayant le même ensemble de solutions.

1. Soit le problème d'optimisation

$$\min_{x \in X \subseteq \mathbb{R}^n} f(x),$$

où X est un sous-ensemble de \mathbb{R}^n . Soit une fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ strictement croissante sur $f(X) = \{f(x) : x \in X\}$, c'est-à-dire, pour tout $z_1, z_2 \in f(X)$ tel que $z_1 > z_2$ on a $g(z_1) > g(z_2)$. Alors,

$$\arg \min_{x \in X} f(x) = \arg \min_{x \in X} g(f(x)).$$

En particulier, le fait d'ajouter ou retrancher une constante à la fonction objectif d'un problème d'optimisation ne modifie pas sa solution :

$$\forall c \in \mathbb{R}, \quad \arg \min_{x \in X} f(x) = \arg \min_{x \in X} (f(x) + c).$$

De même, si la fonction f ne prend que des valeurs positives, le fait de prendre le logarithme de la fonction objectif ne modifie pas la solution

$$\arg \min_{x \in X} f(x) = \arg \min_{x \in X} \ln(f(x)).$$

On peut aussi la mettre au carré sans modifier la solution, car $g(x) = x^2$ est strictement croissante pour $x \geq 0$. Cela arrive typiquement lorsque $f(x)$ s'exprime comme une racine carrée.

2. Un problème de maximisation dont la fonction objectif est $f(x)$ est équivalent à un problème de minimisation dont la fonction objectif est $-f(x)$

$$\max_{x \in X} f(x) = -\min_{x \in X} -f(x)$$

et

$$\arg \max_{x \in X} f(x) = \arg \min_{x \in X} -f(x).$$

De la même manière, on a

$$\min_x f(x) = -\max_x -f(x)$$

3. Une contrainte définie par une inégalité de type inférieur ou égal peut être transformée en une contrainte de type supérieur ou égal comme suit

$$g(x) \leq 0 \iff -g(x) \geq 0.$$

4. Une contrainte définies par une égalité peut être remplacée par deux contraintes de type inégalité

$$g(x) = 0 \iff \begin{cases} g(x) \leq 0 \\ g(x) \geq 0 \end{cases}$$

Notons que cette transformation est principalement utilisée lorsque les contraintes sont linéaires. Lorsque $g(x)$ est non linéaire, cette transformation n'est pas en général conseillée.

5. Certains logiciels exigent que toutes les variables de décision soient positives. Or pour certains problèmes de telles restrictions n'ont pas lieu d'être. Si une variable x peut prendre toute valeur réelle elle est alors remplacée par deux variables artificielles, notées x^+ et x^- , telles que $x = x^+ - x^-$. Dans ce cas, on peut satisfaire les exigences du logiciel, et imposer $x^+ \geq 0$ et $x^- \geq 0$, sans perte de généralité.
6. En présence d'une contrainte $x \geq a$, avec $a \in \mathbb{R}^n$, un simple changement de variable

$$x = \tilde{x} + a$$

transforme la contrainte en

$$\tilde{x} \geq 0.$$

Définition 1.15 (Fonction Lagrangienne). *Soit le problème d'optimisation*

$$(P) \quad \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ h_i(x) = 0 \quad i = 1, \dots, m, \\ g_j(x) \leq 0 \quad j = 1, \dots, p. \end{cases}$$

La fonction $L : \mathbb{R}^{n+m+p} \rightarrow \mathbb{R}$ définie par

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x)$$

est appelée Lagrangien ou fonction lagrangienne du problème (P).

Définition 1.16 (Fonction duale). *Au problème (P) on associe la fonction $q : \mathbb{R}^{m+p} \rightarrow \mathbb{R}$ définie par*

$$q(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu),$$

c'est la fonction duale du problème. Les paramètres λ et μ sont appelés variables duales.

Proposition 1.2 (Borne sur la fonction duale). *Soit x un point de \mathbb{R}^n vérifiant $h(x) = 0$ et $g(x) \leq 0$ et $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$. Alors on a*

$$q(\lambda, \mu) \leq f(x).$$

En particulier, on obtient

$$\sup_{\substack{(\lambda, \mu) \\ \mu \geq 0}} q(\lambda, \mu) \leq \inf_{x \in X} f(x)$$

où $X = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$.

Les valeurs de la fonction duale fournissent des bornes inférieures de la valeur optimale du problème.

Définition 1.17 (Problème dual). Soit (P) le problème d'optimisation $\min f(x)$ sous contraintes $h(x) = 0$ et $g(x) \leq 0$ et $q(\lambda, \mu)$ sa fonction duale. Soit $X_q \subseteq \mathbb{R}^{m+p}$ le domaine de q , c'est à dire $X_q = \{(\lambda, \mu) : q(\lambda, \mu) > -\infty\}$. Le problème d'optimisation

$$(D) \quad \max_{\lambda, \mu} q(\lambda, \mu) \quad \text{s.c.} \quad \mu \geq 0 \text{ et } (\lambda, \mu) \in X_q$$

est appelé "problème dual" du problème (P) qui est dit "primal".

Définition 1.18 (Fonction barrière). Soit $C \subset \mathbb{R}^n$ un ensemble d'intérieur non vide. On appelle fonction barrière associée à C toute fonction B_r définie sur $\text{int}(C)$ telle que :

1. B_r est continue sur $\text{int}(C)$.
2. $B_r(x) \geq 0$, pour tout $x \in \text{int}(C)$.
3. $\lim_{x \rightarrow x^*} B_r(x) = +\infty$ pour tout $x^* \in \text{Fr}(C)$.

Les fonctions barrières les plus utilisées sont les fonctions logarithmiques définies par

$$B_r(x) = - \sum_{i=1}^m \ln(h_i(x))$$

où $C = \{x \in \mathbb{R}^n : h(x) \geq 0\}$ est tel que $\text{int}(C) = \{x \in \mathbb{R}^n : h(x) > 0\}$.

1.2 Qualification des contraintes

Définition 1.19 (Point admissible). Un point $x \in \mathbb{R}^n$ est dit admissible du problème (P) s'il vérifie toutes les contraintes.

Définition 1.20 (Contraintes actives). Soit $g : \mathbb{R}^n \rightarrow \mathbb{R}$. Une contrainte d'inégalité $g(x) \leq 0$ est dite active (resp. inactive) en x^* si $g(x^*) = 0$, (resp. $g(x^*) < 0$).

L'ensemble des indices des contraintes actives en x^* sera généralement noté $\mathcal{A}(x^*)$.

Indépendance linéaire des contraintes

Soit le problème de minimiser $f(x)$ sous les contraintes $h(x) = 0$ et $g(x) \leq 0$. Soit x^* un point admissible, nous dirons que la condition d'indépendance linéaire des contraintes est vérifiée en x^* si les gradients des contraintes d'égalité et les gradients des contraintes d'inégalité actives en x^* sont linéairement indépendants.

Par abus de langage, nous dirons parfois simplement que les contraintes sont linéairement indépendantes.

Directions admissibles

Soit le problème général d'optimisation et $x \in \mathbb{R}^n$ un point admissible. Un vecteur $d \in \mathbb{R}^n$ sera dit admissible en x s'il existe $\eta > 0$ tel que $x + \alpha d$ soit admissible pour tout $0 < \alpha \leq \eta$.

Proposition 1.3 (Directions admissibles : cas linéaire). Soit $X = \{Ax = b \text{ et } x \geq 0\}$ et x^* un point admissible. Un vecteur d est admissible à X en x^* si et seulement si

1. $Ad = 0$,
2. $d_i \geq 0$ si $x_i^* = 0$.

1.3 Résultats d'existence et d'unicité

Théorème 1.3 (Théorème de Weierstrass). Soit f une fonction s.c.i. sur un domaine compact S alors il existe $x^* \in S$ une solution du problème (P).

Dans le cas où l'ensemble S n'est pas borné, le corollaire suivant nous donne un résultat d'existence en imposant une hypothèse de coercivité à la fonction f .

Corollaire 1.3.1 (Existence). *Soit f une fonction définie sur une partie fermée S de \mathbb{R}^n . Si f est coercive et s.c.i sur S alors elle admet au moins un minimum, c'est à dire il existe $x^* \in S$ tel que $f(x^*) \leq f(x)$ pour tout $x \in S$.*

Toutefois, ce théorème n'assure pas l'unicité de la solution. Pour cela énonçons le théorème suivant

Théorème 1.4 (Unicité). *Soit S un ensemble convexe et soit $f : S \rightarrow \mathbb{R}$ une fonction strictement convexe. Alors le problème (P) admet au plus une solution.*

Maintenant, pour terminer. Donnons un théorème d'existence et d'unicité.

Théorème 1.5. *Soit f une fonction de classe \mathcal{C}^1 de \mathbb{R}^n dans \mathbb{R} . On suppose qu'il existe $\alpha > 0$ tel que*

$$\forall x, y \in \mathbb{R}^n \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2.$$

Alors f est fortement convexe (donc strictement convexe) et coercive, en particulier le problème (P) admet une solution unique.

1.4 Conditions d'optimalité

Nous supposons dans cette section que la fonction f est suffisamment différentiable. Nous commençons par donner des conditions nécessaires de premier et second ordre dans le cas avec et sans contraintes. Pour pouvoir donner des conditions suffisantes d'optimalité dans le cas général on doit faire appel à la dérivée seconde de la fonction f .

Théorème 1.6 (Conditions nécessaires de premier ordre). *Soit S une partie de \mathbb{R}^n et $f : S \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^1 sur un voisinage de S . Si \hat{x} est un minimum local de f alors*

$$\langle \nabla f(\hat{x}), d \rangle \geq 0 \quad \text{pour tout } d \in R_S(\hat{x}) \tag{1.3}$$

où $R_S(\hat{x})$ est le cône des directions admissibles à S en \hat{x} .

Si l'ensemble S est convexe alors la relation (1.3) devient

$$\langle \nabla f(\hat{x}), x - \hat{x} \rangle \geq 0 \quad \text{pour tout } x \in S. \quad (1.4)$$

Dans le cas où l'ensemble S est un ouvert de \mathbb{R}^n ($S = \mathbb{R}^n$ par exemple) la condition (1.3) s'écrit

$$\nabla f(\hat{x}) = 0.$$

Théorème 1.7 (Conditions nécessaires de second ordre). *Soit S un ouvert de \mathbb{R}^n et $f : S \rightarrow \mathbb{R}$ une fonction de classe \mathcal{C}^2 . Si $\hat{x} \in S$ est un minimum local de f sur S alors $\nabla f(\hat{x}) = 0$ et la matrice $\nabla^2 f(\hat{x})$ est semi-définie positive, c'est à dire*

$$\langle d, \nabla^2 f(\hat{x})d \rangle \geq 0 \quad \text{pour tout } d \in \mathbb{R}^n.$$

Nous énonçons à présent des conditions suffisantes pour qu'un point \hat{x} soit un minimum local dans le cas sans contraintes pour une fonction f suffisamment régulière.

Théorème 1.8 (Conditions suffisantes). *Soit f une fonction définie sur \mathbb{R}^n de classe \mathcal{C}^1 et deux fois différentiable en \hat{x} . Si $\nabla f(\hat{x}) = 0$ et $\nabla^2 f(\hat{x})$ est définie positive alors le point \hat{x} est un minimum local strict.*

Chapitre 2

Application des fonctions approximantes

2.1 Application en programmation linéaire (PL)

La programmation linéaire est un domaine de l'optimisation où la fonction objectif est linéaire et les contraintes sont affines. Un problème de programmation linéaire peut être écrit sous différentes formes équivalentes, la plus utilisée en pratique est la forme standard

$$\begin{cases} \min & \langle c, x \rangle, \\ & Ax = b, \\ & x \geq 0. \end{cases} \quad (PL)$$

où $c, x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ et A est une matrice $m \times n$.

Le programme dual de (PL) est défini comme suit

$$\begin{cases} \max & \langle b, y \rangle, \\ & A^T y \leq c, \\ & y \in \mathbb{R}^m. \end{cases} \quad (DL)$$

Dans ce qui suit, on s'intéresse à la résolution du problème de programmation linéaire

$$\begin{cases} \min \langle c, x \rangle, \\ Ax = 0, \\ x \geq 0. \end{cases}$$

La méthode de résolution base sur le même principe que celle de Karmarkar pour le calcul de la direction de descente, le pas de déplacement est calculé par la technique des fonctions majorantes.

Karmarkar traite le problème linéaire sous la forme réduite suivante :

$$\begin{cases} \min \langle c, x \rangle = 0, \\ Ax = 0, \\ x \in S_n = \{x \in \mathbb{R}_+^n, \langle e, x \rangle = 1\}. \end{cases}$$

Remarque 2.1. On peut ramener un programme linéaire général quelconque mis sous la forme standard suivant :

$$\begin{cases} \min \langle c, x \rangle = z^*, \\ Ax = b, \\ x \geq 0, \end{cases} \quad (PL)$$

sous la forme réduite de Karmarkar.

Pour ce faire, Karmarkar utilise la transformation projective définie par

$$\begin{aligned} T_a : \mathbb{R}_+^n &\rightarrow S_{n+1} \\ x &\mapsto y = T_a(x), \end{aligned}$$

tel que

$$y_i = \begin{cases} \frac{x_i/a_i}{1 + \sum_{i=1}^n \frac{x_i}{a_i}} & \text{si } i = 1, \dots, n, \\ 1 - \sum_{i=1}^n y_i & \text{si } i = n + 1, \end{cases}$$

où $a = (a_1, a_2, \dots, a_n)^T$ est une solution strictement réalisable de (PL) .

Cette transformation nous permet d'associer à (PL) le programme linéaire suivant

$$\begin{cases} \min \langle c', y \rangle = 0, \\ A'y = 0, \\ y \in S_{n+1} = \{y \in \mathbb{R}_+^{n+1}, \langle e_{n+1}, y \rangle = 1\}, \end{cases}$$

où

$$A' = [AD_a \quad -b], \text{ et } c' = \begin{bmatrix} D_a c \\ -z^* \end{bmatrix},$$

avec $D_a = \text{diag}(a)$ est la matrice diagonale dont les éléments diagonaux sont les composantes du vecteur a .

Remarque 2.2. La transformation T_a est inversible et on a :

$$T_a^{-1}(y) = x = \frac{D_a y[n]}{y_{n+1}}, \text{ avec } y[n] = [y_1 \ y_2 \ \dots \ y(n)]^T$$

Présentation du problème :

A ce stade de la procédure, on adopte les conventions suivantes : Le vecteur $e = [1, \dots, 1]^T \in \mathbb{R}^n$, étant donné un vecteur $x \in \mathbb{R}^n$, X est la matrice diagonale dont les éléments diagonaux sont les composantes de x .

L'algorithme de Karmarkar s'applique au problème suivant

$$\text{Trouver } \bar{x} \text{ tel que } \bar{x} \geq 0, \bar{x} \neq 0, A\bar{x} = 0 \text{ et } \langle c, \bar{x} \rangle = 0, \quad (Ka)$$

où $c \in \mathbb{R}^n$ et A est une matrice $m \times n$ de plein rang.

Les hypothèses suivantes sont assumées :

1. $Ax = 0$ et $x \geq 0 \implies \langle c, x \rangle \geq 0$.
2. On connaît un point $\hat{x} > 0$ tel que $A\hat{x} = 0$ et $\langle c, \hat{x} \rangle > 0$.
3. On sait que le problème (Ka) a des solutions.

On pose

$$C = \{x : x \geq 0, x \neq 0, Ax = 0\}.$$

Pour résoudre ce type de problème, plusieurs algorithmes de points intérieurs utilisent l'une des deux fonctions auxiliaires appelée fonction potentielle multiplicative et fonction potentielle logarithmique, définie pour $x \in C$ et $x > 0$ par

$$f(x) = \frac{\langle c, x \rangle^n}{\prod_{i=1}^n x_i},$$

$$\pi(x) = \ln(f(x)) = n \ln(\langle c, x \rangle) - \sum_{i=1}^n \ln(x_i).$$

La fonction f vérifie les propriétés suivantes

1. $0 < f(x) < +\infty$ si $x > 0$ et $Ax = 0$.
2. $f(x) = +\infty$ si x appartient à la frontière relative de C sans être solution de (Ka) .
3. $f(x) = 0$ si x est solution de (Ka) ou si $x = 0$.
4. $f(kx) = f(x)$ pour tout $x \in C$ et tout $k > 0$.

Ainsi, le problème (Ka) consiste à trouver les solutions optimales de

$$0 = \min_x [f(x) : x \geq 0, Ax = 0]. \quad (Km)$$

2.1.1 Description de l'algorithme

Commençant par un point \hat{x} dans l'intérieur relatif de C connu, l'algorithme de Karmarkar est une méthode de descente qui génère une suite de points tous contenus dans l'intérieur relatif de C , en raison du caractère barrière de la fonction objectif f , d'où la dénomination de méthode de points intérieurs. Nous allons décrire le passage de l'itéré x à l'itéré suivant \tilde{x} . On suppose que l'itéré x vérifie $x > 0$ et $Ax = 0$

Normalisation

On normalise x de manière à avoir $\langle x, x \rangle = n$, par la relation

$$x = \sqrt{\frac{n}{\langle x, x \rangle}} x.$$

Direction de descente

On choisit la direction de descente de manière à ce que le nouvel itéré appartienne à l'hyperplan passant par x et orthogonal à x , il suit alors une version normalisée du problème (Km)

$$0 = \min_y [f(y) : y \geq 0, Ay = 0, \langle x, y \rangle = n]. \quad (Kx)$$

On a x est une solution réalisable de ce problème. Il est facile de voir que l'on a

$$\frac{f(y)}{f(x)} = g(z) \text{ avec } z = X^{-1}y, g(z) = \frac{\langle b, z \rangle^n}{\prod_{i=1}^n z_i} \text{ et } b = \frac{1}{\langle c, x \rangle} Xc.$$

Les conditions $Ay = 0$ et $\langle x, y \rangle = n$ se transposent en $AXz = 0$ et $\langle e, z \rangle = n$. Posons $B = AX$. On associe à (Kx) le problème auxiliaire suivant

$$0 = \min_z [g(z) : z \geq 0, Bz = 0, \langle e, z \rangle = n]. \quad (Kz)$$

e est une solution réalisable de ce problème et on a $g(e) = \langle b, e \rangle = 1$. Il est facile de voir que la matrice $[A^T \ x]$ est de rang $m + 1$, il en est de même de la matrice $[B^T \ e]$. L'obtention d'une direction de descente de Newton au point e pour le problème (Kz) s'obtient en résolvant le problème quadratique

$$\min_{\delta} \left[\frac{1}{2} \langle \nabla^2 g(e) \delta, \delta \rangle + \langle \nabla g(e), \delta \rangle : B\delta = 0, \langle e, \delta \rangle = 0 \right].$$

Pour cela, introduisons la matrice

$$P = I - [B^T \ e] \left[\begin{bmatrix} B \\ e^T \end{bmatrix} [B^T \ e] \right]^{-1} \begin{bmatrix} B \\ e^T \end{bmatrix}$$

qui correspond à la matrice de projection sur le sous espace linéaire

$$\{z : Bz = 0, \langle e, z \rangle = 0\}.$$

On a $P^2 = P = P^T$, $PB^T = 0$ et $Pe = 0$. Il est facile de voir que l'on a

$$P\nabla g(e) = Pb \text{ et } P\nabla^2 g(e)P = I + n(n-1)Pbb^T P.$$

Le problème quadratique est équivalent au problème

$$\min_{\delta} \left[\frac{1}{2} \langle \nabla^2 g(e) \delta, \delta \rangle + \langle \nabla g(e), \delta \rangle : \delta = P\delta \right],$$

dont la solution optimale est colinéaire à $d = -Pb = -P\nabla g(e)$. La direction d coïncide donc avec la direction donnée par le gradient projeté. Nous nous intéressons maintenant à quelques propriétés de d . Tout d'abord on a,

$$\langle d, e \rangle = -\langle Pb, e \rangle = -\langle b, Pe \rangle = 0.$$

On observe ensuite que l'on a d'une part

$$\left\{ z : \begin{array}{l} \langle e, z \rangle = n \\ \|z - e\|^2 \leq \frac{n}{n-1} \end{array} \right\} \subset \left\{ z : \begin{array}{l} \langle e, z \rangle = n \\ z \geq 0 \end{array} \right\} \subset \left\{ z : \begin{array}{l} \langle e, z \rangle = n \\ \|z - e\|^2 \leq n(n-1) \end{array} \right\},$$

et d'autre part

$$-1 = \min_z [\langle b, z - e \rangle : z \geq 0, Bz = 0, \langle e, z \rangle = n],$$

et, puisque $P(e - z) = e - z$,

$$\langle b, z - e \rangle = \langle Pb, z - e \rangle.$$

On obtient

$$\|Pb\| \sqrt{\frac{n}{n-1}} \leq 1 \leq \|Pb\| \sqrt{n(n-1)}.$$

Donc en résumé,

$$-\langle d, b \rangle = -\|d\|^2 = -\|Pb\|^2, \langle d, e \rangle = 0,$$

et

$$\frac{1}{n(n-1)} \leq \|d\|^2 \leq \frac{n-1}{n}.$$

Dans ce qui suit, on dénote par \bar{d} et σ la moyenne et l'écart type du n-uple (d_1, \dots, d_n) .

On a

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = 0 \text{ et } \frac{1}{n^2(n-1)} \leq \sigma^2 = \frac{\|d\|^2}{n} - \|\bar{d}\|^2 \leq \frac{n-1}{n^2}.$$

Calcul du pas de déplacement par la technique des fonctions majorantes :

Elle consiste à obtenir une valeur $t > 0$ telle que l'on ait $e + td > 0$ et qui donne une décroissance significative de $\mu_0(t) = g(e + td)$ ou de façon équivalente de $\theta_0(t) = \ln(g(e + td))$. Puisque l'équation $\theta'_0(t) = 0$ ne peut pas être résolue explicitement, il est normal de penser à des méthodes itératives de résolution, on peut aussi appliquer à θ_0 une méthode de type Armijo-Golstein-Price. Dans les deux cas, cela nécessite plusieurs évaluations de la fonction θ_0 et de sa dérivée, ce qui est coûteux. Notre démarche consiste à minimiser une majorante θ de la fonction θ_0 dont on peut obtenir le minimum de façon explicite. Rappelons que l'on a

$$\begin{aligned}\theta_0(t) &= n \ln(1 - t\|d\|^2) - \sum_{i=1}^n \ln(1 + td_i) \\ &= n \ln(1 - tn\sigma^2) - \sum_{i=1}^n \ln(1 + td_i),\end{aligned}$$

il est clair que $\theta_0(0) = 0$.

En raison du théorème [10] on a

$$\sum_{i=1}^n \ln(1 + td_i) \geq (n-1) \ln\left(1 + \frac{\sigma t}{\sqrt{n-1}}\right) + \ln(1 - \sigma t \sqrt{n-1}),$$

et donc

$$\theta_0(t) \leq \theta(t) = n \ln(1 - nt\sigma^2) - (n-1) \ln\left(1 + \frac{\sigma t}{\sqrt{n-1}}\right) - \ln(1 - \sigma t \sqrt{n-1}).$$

On a

$$\theta'(t) = \frac{n^2\sigma^2}{1 - nt\sigma^2} - \frac{(n-1)\sigma}{\sqrt{n-1} + \sigma t} + \frac{\sigma\sqrt{n-1}}{1 - \sigma t \sqrt{n-1}}.$$

On en déduit que la fonction θ atteint son minimum au point

$$\bar{t} = \frac{n\sqrt{n-1}}{\sqrt{n-1} + n(n-2)\sigma}.$$

On prend pour nouvel itéré

$$\tilde{x} = X(e + \bar{t}d) = x + \bar{t}Xd.$$

Par construction on a $\tilde{x} > 0$ et $A\tilde{x} = 0$.

La décroissance :

En remplaçant \bar{t} par sa valeur on obtient

$$1 - n\bar{t}\sigma^2 = \frac{\sqrt{n-1} + n(n-2)\sigma - n^2\sigma^2\sqrt{n-1}}{\sqrt{n-1} + n(n-2)\sigma},$$

$$1 + \frac{\sigma\bar{t}}{\sqrt{n-1}} = \frac{\sqrt{n-1} + n(n-1)\sigma}{\sqrt{n-1} + n(n-2)\sigma},$$

$$1 - \sigma\bar{t}\sqrt{n-1} = \frac{\sqrt{n-1} - n\sigma}{\sqrt{n-1} + n(n-2)\sigma}.$$

D'où

$$\theta(\bar{t}) = (n-1) \ln \left(\frac{\sqrt{n-1} + n(n-2)\sigma - n^2\sigma^2\sqrt{n-1}}{\sqrt{n-1} + n(n-1)\sigma} \right) + \ln \left(\frac{\sqrt{n-1} + n(n-2)\sigma - n^2\sigma^2\sqrt{n-1}}{\sqrt{n-1} - n\sigma} \right).$$

Finalement on a

$$\theta(\bar{t}) = (n-1) \ln \left(1 - \frac{n\sigma}{\sqrt{n-1}} \right) + \ln(1 + n\sigma\sqrt{n-1}).$$

La quantité $\theta(\bar{t}) - \theta(0) = \theta(\bar{t})$ dépend de σ .

Puisque l'on a

$$\frac{1}{n\sqrt{n-1}} \leq \sigma \leq \frac{\sqrt{n-1}}{n},$$

il est utile de poser

$$u = n\sigma\sqrt{n-1},$$

on a alors $1 \leq u \leq n-1$ et

$$\theta(\bar{t}) = (n-1) \ln \left(1 - \frac{u}{n-1} \right) + \ln(1+u) = \xi(u).$$

La fonction ξ est concave et donc

$$\xi(u) \leq \xi(1) + (u-1)\xi'(1) = (n-1) \ln \left(1 - \frac{1}{n-1} \right) + \ln(2) - \frac{n(u-1)}{2(n-2)}.$$

On en déduit que dans le pire des cas (celui où $u=1$) on a

$$f(\tilde{x}) \leq \frac{2}{e}f(x) = 0,735f(x).$$

La décroissance est bien supérieure en général.

La convergence de l'algorithme :

lemme 2.1. *A la $k^{\text{ème}}$ itération on a :*

$$f(x_k) < 0.735^k f(x_0).$$

Donc $f(x_k)$ converge linéairement vers 0. On en déduit que toute valeur d'adhérence de la suite $\{x_k\}$ est solution du problème (Ka).

Description de l'algorithme

Algorithme 1 : Algorithme de Karmarkar via la fonction majorante

Données : $A, c, \epsilon > 0$ est une précision donnée,

Résultat : x^*

Initialisation : On part de $x > 0$ tel que $Ax = 0$.

1 tant que $c^T x \geq \epsilon$ faire

2 Normalisation :

$$x = \sqrt{\frac{n}{\langle x, x \rangle}} x.$$

3 Direction de descente : On prend b et B comme suit

$$b = \frac{1}{\langle c, x \rangle} Xc, \quad B = AX.$$

4 On détermine d projection de b sur le sous espace linéaire

$$\{z : Bz = 0, \langle e, z \rangle = 0\}.$$

5 Finalement la direction de descente est

$$\delta = -Xd.$$

6 On calcule le pas :

$$\bar{t} = \frac{n\sqrt{n-1}}{\sqrt{n-1} + n(n-2)\sigma}, \quad \text{où } \sigma = \frac{\|d\|}{\sqrt{n}}.$$

7 Le nouvel itéré est $x = x + \bar{t}\delta$.

8 $x^* = x$.

2.1.2 Simulations numériques

On fait une études comparatives dont la quelle le pas de déplacement soit fixe ou calculé par la technique des fonctions majorantes.

Dans les tableaux suivants, on note par

TFM : calcul du pas de déplacement par la technique des fonctions majorantes.

pas fixe : le pas de déplacement est fixé.

Exemples de tailles fixes :

Exemple 2.1.

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 \\ 3 & 4 & -1 & -6 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \text{et} \quad c = \begin{bmatrix} 0 & 1 & 2 & 0 \end{bmatrix}^T.$$

La valeur optimale $z^* = 0$.

La solution optimale exacte est : $x^* = \begin{bmatrix} 1.5 & 0 & 0 & 0.25 \end{bmatrix}^T$.

La solution optimale avec le pas =0.5 est

$$x^* = \begin{bmatrix} 1.499705 & 0.000186 & 0.000093 & 0.249961 \end{bmatrix}^T.$$

La solution optimale avec TFM est

$$x^* = \begin{bmatrix} 1.495181 & 0.003057 & 0.001655 & 0.249353 \end{bmatrix}^T.$$

Exemple 2.2.

$$A = \begin{bmatrix} 2 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{et} \quad c = \begin{bmatrix} 3 & -1 & 1 & 0 & 0 & 0 \end{bmatrix}^T.$$

La valeur optimale $z^* = -0.5$.

La solution optimale exacte est $x^* = \begin{bmatrix} 0 & 0.5 & 0 & 0.5 & 0 & 0 \end{bmatrix}^T$.

La solution optimale avec le pas =0.5 est

$$x^* = \begin{bmatrix} 0.000031 & 0.499618 & 0.000090 & 0.499681 & 0.000245 & 0.000335 \end{bmatrix}^T.$$

La solution optimale avec TFM est

$$x^* = [0.000831 \quad 0.489945 \quad 0.002367 \quad 0.491607 \quad 0.006441 \quad 0.008809]^T.$$

Exemple 2.3.

$$A = \begin{bmatrix} -1 & 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 2 & -3 & 2 & 0 & 1 & 0 \\ -3 & 2 & 1 & 0 & 0 & 0 & 1 \\ 3 & 5 & 4 & 0.5 & 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 2 \end{bmatrix} \quad \text{et } c = [1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad -2]^T.$$

La valeur optimale $z^* = 0$.

La solution optimale exacte est $x^* = [0.5 \quad 0 \quad 0 \quad 1 \quad 2.5 \quad 0 \quad 1.5]^T$.

La solution optimale avec le pas =0.5 est

$$x^* = [0.5000 \quad 0.000026 \quad 0.000026 \quad 0.999322 \quad 2.499306 \quad 0.001383 \quad 1.5000]^T.$$

La solution optimale avec TFM est

$$x^* = [0.5000 \quad 0.000006 \quad 0.000006 \quad 0.999852 \quad 2.499848 \quad 0.000302 \quad 1.5000]^T.$$

Exemple 2.4.

$$A = \begin{bmatrix} 0 & 1 & 2 & -1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & -2 & 1 & 2 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & -1 & -2 & 0 & 0 & 0 & 1 \\ 1 & 3 & 4 & 2 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 1 \end{bmatrix} \quad \text{et } c = [1 \quad 0 \quad -2 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]^T.$$

La valeur optimale $z^* = -0.5$.

La solution optimale exacte est $x^* = [0 \quad 0 \quad 0.25 \quad 0 \quad 0 \quad 0.5 \quad 1.25 \quad 3.5 \quad 2]^T$.

La solution optimale avec le pas =0.5 est

$$x^* = [0.000124 \quad 0.000124 \quad 0.249799 \quad 0.000093 \quad 0.000124 \quad 0.500 \quad 1.249985 \quad 3.49938 \quad 1.999969]^T.$$

La solution optimale avec TFM est

$$x^* = [0.000029 \quad 0.000029 \quad 0.249954 \quad 0.000021 \quad 0.000029 \quad 0.500 \quad 1.249996 \quad 3.499857 \quad 1.999993]^T.$$

TABLE 2.1: Tests numériques des exemples de tailles fixes

Exemple	TFM			pas fixe=0.5		
	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)
2.1	0.0063667	10	0.03	0.000372	16	0.01
2.2	-0.48508	15	0.01	-0.499434	20	0.01
2.3	0.000151	18	0.01	0.000691	21	0.01
2.4	-0.499828	17	0.03	-0.466256	22	0.03

Exemple de tailles variables :

Exemple 2.5. *Considérons les problèmes linéaires suivants*

$$\begin{cases} \min c^T x, \\ Ax = b, \\ x \geq 0. \end{cases}$$

$$A[i, j] = \begin{cases} 1 & \text{si } j = i \text{ ou } j = i + m, \\ 0 & \text{sinon.} \end{cases}$$

$c[i] = -1, c[i + m] = 0, b[i] = 2$, pour $i = 1, \dots, m$.

La valeur optimale $z^* = -2m$.

Solution optimale $x_i^* = \begin{cases} 2, & \text{si } i=1, \dots, m. \\ 0, & \text{si } i=m+1, \dots, 2m. \end{cases}$

TABLE 2.2: Tests numériques de l'exemple 2.5

Taille	TFM			pas fixe=0.5		
	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)
5	-9.994930	3	0.05	-9.998961	28	0.02
10	-19.994561	6	0.06	-19.998186	41	0.05
25	-49.99432	13	0.14	-49.995149	65	0.21
50	-99.970680	14	0.38	-99.990118	92	1.81
75	-149.936459	15	0.76	-149.985448	113	4.89

2.2 Application en SDP

Les problèmes semi-définis, notés SDP, sont des problèmes d'optimisation, le mot SDP vient du mot anglais Semi-Definite Problem. Où les variables de décision sont des matrices semi-définies positives. Le critère à minimiser est linéaire et les contraintes sont affines, ainsi on peut présenter un problème SDP comme suit :

$$\begin{cases} \min [\langle C, X \rangle = \text{trace}(CX)], \\ \langle A_i, X \rangle = b_i, i = 1, \dots, m, \\ X \in S_n^+(\mathbb{R}), \end{cases} \quad (\text{SDP})$$

où $C, A_i \in S_n(\mathbb{R}), b = (b_1, \dots, b_m)^T \in \mathbb{R}^m$.

Remarque 2.3. Si C est non symétrique, on peut ramener tout problème d'optimisation à un problème SDP en posant C par $\frac{1}{2}(C + C^T)$ de même pour les A_i . Tout programme linéaire peut se ramener facilement à un problème SDP.

Définition 2.1. Une matrice $X \in S_n$ est dite réalisable (resp. strictement réalisable) pour (SDP) si $\langle A_i, X \rangle = b_i, i = 1, \dots, m, X \in S_n^+$ (resp. $\langle A_i, X \rangle = b_i, i = 1, \dots, m$ et $X \in S_n^{++}$).

Définition 2.2. La valeur optimale primale de (SDP) est définie par

$$p^* = \inf\{\langle C, X \rangle, X \in S_n^+ \text{ et } \langle A_i, X \rangle = b_i, i = 1, \dots, m\}.$$

X est une solution optimale primale de (SDP) si

$$X \text{ est réalisable et } \langle C, X \rangle = p^*.$$

2.2.1 Problème dual

La méthode la plus simple pour obtenir le problème dual de (SDP) , est d'utiliser la fonction Lagrangienne $q : \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$

$$q(y) = \inf_{X \in S_n^+} \left[\langle C, X \rangle + \sum_{i=1}^m (b_i - \langle A_i, X \rangle) y_i \right],$$

donc

$$q(y) = \begin{cases} \sum_{i=1}^m b_i y_i & \text{si } C - \sum_{i=1}^m y_i A_i \in S_n^+, \\ -\infty & \text{sinon.} \end{cases}$$

D'où par convention le dual du problème (SDP) est un problème SDP défini par :

$$\begin{cases} \max b^T y, \\ C - \sum_{i=1}^m y_i A_i \in S_n^+, \\ y \in \mathbb{R}^m. \end{cases}$$

2.2.2 Méthode de points intérieurs pour résoudre (SDP)

Dans la littérature, on retrouve bien évidemment des méthodes qui sont les plus utilisées et les plus efficaces pour la résolution du problème (SDP) parmi eux on cite les méthodes de points intérieurs, ces dernières sont classifiées en catégories comme suit

1. Les méthodes affines.
2. Les méthodes de réduction de potentiel.
3. Les méthodes de trajectoire centrale.
4. Méthode barrière logarithmique.

L'importance de ces méthodes vient du fait qu'elles génèrent des itérés successifs qui se situent à l'intérieur relatif du domaine réalisable convergeant vers une solution optimale.

2.2.3 Méthode barrière logarithmique via les fonctions approximantes

Le but de ce paragraphe est d'utiliser l'approche de barrière logarithmique pour la résolution des problèmes (*SDP*). Nous utilisons la méthode de Newton pour traiter les équations perturbées associées pour obtenir une direction de descente. Pour le calcul du pas de déplacement, on propose une nouvelle variante basée sur la technique des fonctions approximantes, nous considérons le problème de programmation semi-définie suivant

$$\begin{cases} \min b^T x, \\ \sum_{i=1}^m x_i A_i - C \in S_n^+, \\ x \in \mathbb{R}^m. \end{cases} \quad (2.1)$$

Le problème (2.1) est le dual du problème suivant

$$\begin{cases} \max \langle C, Y \rangle, \\ \langle A_i, Y \rangle = b_i, \forall i = 1, \dots, m, \\ Y \in S_n^+. \end{cases} \quad (2.2)$$

Le problème (2.1) est approximé par le problème (*SDP*) $_{\eta}$ suivant

$$\begin{cases} \min f_{\eta}(x), \\ x \in \mathbb{R}^m. \end{cases} \quad (SDP)_{\eta}$$

avec $\eta > 0$ un paramètre de pénalisation, et $f : \mathbb{R}^m \rightarrow]-\infty, +\infty]$ la fonction barrière définie par

$$f_{\eta}(x) = \begin{cases} b^T x + n\eta \ln \eta - \eta \ln \left[\det \left(\sum_{i=1}^m x_i A_i - C \right) \right] & \text{si } x \in \hat{X}, \\ +\infty & \text{sinon.} \end{cases}$$

avec

$$X = \left\{ x \in \mathbb{R}^m : \sum_{i=1}^m x_i A_i - C \in S_n^+ \right\}, \text{ l'ensemble des solutions réalisables de (2.1).}$$

$$\hat{X} = \left\{ x \in \mathbb{R}^m : \sum_{i=1}^m x_i A_i - C \in S_n^{++} \right\}, \text{ l'ensemble des solutions strictement réalisables de (2.1).}$$

La fonction f_η est deux fois continûment différentiables sur \hat{X} . Plus précisément, pour tout $x \in \hat{X}$ on a

1. $\nabla f_\eta(x) = b - \eta b(x)$,
2. $\nabla^2 f_\eta(x) = \eta \Delta(x)$,
3. La matrice $\Delta(x)$ est définie positive,

où $b(x)$ est un vecteur de \mathbb{R}^m et $\Delta(x)$ est une matrice symétrique de taille m , défini comme suit

$$\begin{aligned} b_i(x) &= \text{trace}(\hat{A}_i(x)) = \text{trace}(A_i B(x)^{-1}), \\ \Delta_{ij}(x) &= \text{trace}(B(x)^{-1} A_i B(x)^{-1} A_j) = \text{trace}(\hat{A}_i(x) \hat{A}_j(x)), \\ B(x) &= \sum_{i=1}^m x_i A_i - C = L(x) L^T(x), \\ \hat{A}_i(x) &= [L(x)]^{-1} A_i [L^T(x)]^{-1}. \end{aligned}$$

comme f_η est strictement convexe, $(SDP)_\eta$ a au plus une solution optimale.

Intéressons nous maintenant au calcul de la direction de Newton d en $x \in \hat{X}$, pour cela il faut trouver la solution du système linéaire donnée par

$$\nabla^2 f_\eta(x) d = -\nabla f_\eta(x),$$

ce système est équivalent au système suivant

$$\Delta(x) d = b(x) - \frac{1}{\eta} b. \quad (*)$$

Soit d la solution de l'équation (*), passons maintenant au calcul du pas de déplacement \bar{t} (\bar{t} est strictement positif) qui donne une décroissance significative à la fonction G tel que

$$G(t) = \frac{1}{\eta} [f_\eta(x + td) - f_\eta(x)], \quad x + td \in \hat{X},$$

qu'on peut écrire

$$G(t) = t \frac{1}{\eta} b^T d - \ln \det(B(x + td)) + \ln \det(B(x)).$$

lemme 2.2. Soit $\hat{t} = \sup\{t : 1 + \lambda_i t > 0, i = 1, \dots, n\}$
pour tout $t \in [0, \hat{t}]$, la fonction $G(t)$ est bien définie

$$G(t) = \sum_{i=1}^n [(\lambda_i - \lambda_i^2) t - \ln(1 + \lambda_i t)],$$

avec λ_i les valeurs propres de $E = L^{-1}H(L^{-1})^T$ sachant que

$$H = \sum_{i=1}^m d_i A_i \text{ et } LL^T = B(x) = \sum_{i=1}^m x_i A_i - C.$$

2.2.4 Calcul du pas de déplacement par des fonctions minorantes

Avant de donner un aperçu général sur le formalisme des fonctions minorantes, nous rappellons d'abord que le calcul du pas de déplacement dans les recherches linéaires impose une difficulté au niveau de la définition de f entraîne un coût très élevé dans les procédures classiques de la recherche linéaire. Dans notre approche, au lieu de minimiser f le long de la direction de descente d au point courant x , on va minimiser une nouvelle fonction minorante notée \hat{G} telle que

$$\hat{G}(t) \leq G(t), \text{ pour tout } t > 0,$$

avec

$$G(0) = \hat{G}(0) = 0, \quad G'(0) = \hat{G}'(0) < 0.$$

On choisit \hat{G} suffisamment proche de G de telle manière que le calcul de \bar{t} soit facile, pour donner à chaque itération une décroissance significative à f . On commence par proposer des fonctions \hat{G} pour lesquelles la solution optimale est obtenue explicitement.

Première fonction minorante

L'approximation doit être simple et suffisamment proche de G , dans notre cas on exige

$$\hat{G}(0) = 0, \quad -\hat{G}'(0) = \hat{G}''(0) = \|\lambda\|^2,$$

pour tout $x_i = 1 + t\lambda_i$ $i = 1, \dots, m$, on a $\bar{x} = 1 + t\bar{\lambda}$ et $\sigma_x = t\sigma_\lambda$, en appliquant les inégalités [10], on obtient

$$\sum_{i=1}^n \ln(1 + \lambda_i t) \leq (n-1) \ln(1 + \beta_1 t) + \ln(1 + \gamma_1 t),$$

avec

$$\beta_1 = \bar{\lambda} - \frac{\sigma_\lambda}{\sqrt{n-1}} \text{ et } \gamma_1 = \bar{\lambda} + \sigma_\lambda \sqrt{n-1}.$$

Alors

$$n\bar{\lambda}t - \|\lambda\|^2 t - \sum_{i=1}^n \ln(1 + \lambda_i t) \geq n\bar{\lambda}t - \|\lambda\|^2 t - (n-1) \ln(1 + \beta_1 t) - \ln(1 + \gamma_1 t),$$

pour tout $t \in [0, \hat{t}_1[$ avec

$$\hat{t}_1 = \begin{cases} -\frac{1}{\beta_1} & \text{si } \beta_1 < 0, \\ +\infty & \text{sinon,} \end{cases}$$

on peut définir la fonction minorante suivante

$$\hat{G}_1(t) = \delta_1 t - (n-1) \ln(1 + \beta_1 t) - \ln(1 + \gamma_1 t),$$

avec

$$\delta_1 = n\bar{\lambda} - \|\lambda\|^2.$$

\hat{G}_1 vérifie les propriétés suivantes

- ▷ $\hat{G}_1(0) = 0$,
- ▷ $\hat{G}_1''(0) = -\hat{G}_1'(0) = \text{trace}(E^2)$,
- ▷ $\hat{G}_1(t) < 0, \forall t \in [0, \hat{t}_1[$,
- ▷ \hat{G}_1 est strictement convexe.

Puisque \hat{G}_1 est strictement convexe alors elle admet un minimum unique sur $[0, \hat{t}_1[$, que l'on trouve en résolvant l'équation $\hat{G}_1'(t) = 0$.

Pour cela, chercher les racines t , revient à résoudre une équation de deuxième degré de la forme

$$t^2 + bt + c = 0,$$

avec

$$b = \left(\frac{1}{\beta_1} + \frac{1}{\gamma_1} - \frac{n}{\delta_1} \right) \text{ et } c = \frac{-\|\lambda\|^2}{\beta_1 \gamma_1 \delta_1}.$$

On obtient les deux racines $\bar{t}_{11}, \bar{t}_{12}$, et on prend la racine positive.

Deuxième fonction minorante

On peut également penser à d'autres fonctions plus simples que \hat{G}_1 , qu'impliquent un seul logarithme. Pour cela, nous considérons les fonctions du type suivant

$$\hat{G}(t) = \hat{\delta}t - \hat{\gamma} \ln(1 + \hat{\beta}t), t \in [0, \hat{t}[.$$

Cette fonction est bien définie sur $t \in [0, \hat{t}[$, avec

$$\hat{t}_1 = \sup(t : 1 + t\hat{\beta} > 0).$$

Alors, on a la fonction de minoration suivante

$$\hat{G}_2(t) \leq G(t) \text{ pour tout } t \in [0, \hat{t}_1[.$$

tel que :

$$\hat{G}_2(t) = \delta_2 t - \gamma_2 \ln(1 + \beta_2 t),$$

avec

$$\beta_2 = \bar{\lambda} - \frac{\sigma_\lambda}{\sqrt{n-1}}, \gamma_2 = \frac{\|\lambda\|^2}{\beta_2^2}, \text{ et } \delta_2 = \gamma_2 \beta_2 - \|\lambda\|^2,$$

\hat{G}_2 vérifie les propriétés suivantes

- ▷ $\hat{G}_2(0) = 0$
- ▷ $\hat{G}_2''(0) = -\hat{G}_2'(0) = \text{trace}(E^2)$
- ▷ $\hat{G}_2(t) < 0, \forall t \in [0, \hat{t}_1[$
- ▷ \hat{G}_2 est strictement convexe.

On a \hat{G}_2 est strictement convexe alors elle admet un minimum unique sur $[0, \hat{t}_1[$, que l'on trouve en résolvant l'équation $\hat{G}_2'(t) = 0$ puis on obtient

$$\bar{t}_2 = \frac{\gamma_2}{\delta_2} - \frac{1}{\beta_2}.$$

Troisième fonction minorante

On cherche à obtenir une autre fonction minorante plus simple que les précédentes, déduite de l'inégalité

$$\left(\|\lambda\| - \sum_{i=1}^n \lambda_i \right) t - \ln(1 + t\|\lambda\|) + \sum_{i=1}^n \ln(1 + t\lambda_i) \leq 0.$$

On pose

$$\hat{G}_3(t) = \delta_3 t - \ln(1 + \beta_3 t),$$

avec

$$\delta_3 = -\|\lambda\|(\|\lambda\| - 1) \text{ et } \beta_3 = \|\lambda\|,$$

\hat{G}_3 vérifie les propriétés suivantes

- ▷ $\hat{G}_3(0) = 0$ $\hat{G}_3''(0) = -\hat{G}_3'(0) = \text{trace}(E^2)$
- ▷ $\hat{G}_3(t) < 0$
- ▷ \hat{G}_3 est convexe

On a \hat{G}_3 est convexe alors elle admet un minimum unique, que l'on trouve en résolvant l'équation $\hat{G}_3'(t) = 0$ puis on obtient

$$\bar{t}_3 = -(\|\lambda\| - 1)^{-1}$$

Description de l'algorithme

Algorithme 2 : Résolution d'un problème *SDP* par les fonctions minorantes

Initialisation : $\epsilon > 0$ est une précision donnée, et $\rho > 0, \sigma \in]0, 1[$ sont des paramètres, $\eta > 0$ paramètre de pénalisation, un point $x^0 \in \hat{X}$ et $k = 0$

étape 1 : à l'itération k , on pose

$$B(x^k) = \sum_{i=1}^m x_i^k A_i - C \text{ et } L_k \text{ tel que } L_k L_k^T = B(x^k)$$

et calculer

$$\begin{aligned}\hat{A}_i^k &= [L_k]^{-1} A_i [L_k^T]^{-1} \\ b_i^k &= \text{trace}(\hat{A}_i^k) \\ \Delta_{ij}^k &= \text{trace}(\hat{A}_i^k \hat{A}_j^k) \\ H &= \eta \Delta^k\end{aligned}$$

étape 2 : Résoudre le système linéaire $Hd = \eta b^k - b$ et prendre la nouvelle itération $x^{k+1} = x^k + \bar{t}d$, telle que \bar{t} est obtenue par l'utilisation de la stratégie de pas de déplacement de $\hat{G}_i, i = 1, 2, 3$.

- 1 **si** $n\eta > \epsilon$ ou $|b^T x^{k+1} - b^T x^k| > n\rho\eta$ **alors**
 - 2 | prendre $k = k + 1, \eta = \sigma\eta$ et aller en (1).
 - 3 **sinon**
 - 4 | Stop x^k est une solution approchée de notre problème.
-

à travers des simulations numériques sur le problème *SDP* on peut tester l'efficacité de notre approche (fonction minorante) relativement aux méthodes de la recherche linéaire classique, pour cela on donne quelques exemples d'application. A noter que les matrices C, A_i et le vecteur b correspondent au problème (*SDP*) tel que x_0 est le point de départ.

Simulations numériques

Exemple 2.6.

$$C = \text{diag}(-5, -8, -8, -5)$$
$$A_k[i, j] = \begin{cases} 1 & \text{si } i = j = k & \text{ou } i = j = k + 1 \\ -1 & \text{si } i = k, j = k + 1 & \text{ou } i = k + 1, j = k \\ 0 & \text{sinon} \end{cases}$$
$$A_4 = \text{diag}(1, 1, 1, 1)$$

$$b = (1, 1, 1, 2)^T$$
$$x^0 = (1.5, 1.5, 1.5, 1.5)^T$$

Exemple 2.7.

$$C = \text{diag}(-4, -2, -2, 0, 0, 0),$$
$$A_1 = \text{diag}(1, -1, 1, 1, 0, 0),$$
$$A_2 = \text{diag}(1, 1, 1, 0, 1, 0),$$
$$A_3 = \text{diag}(2, 2, 1, 0, 0, 1),$$
$$b = (6, 2, 4)^T,$$
$$x^0 = (0.5, 1, 1)^T.$$

Exemple 2.8.

$$C = \text{diag}(-4, -5, 0, 0, 0),$$
$$A_1 = \text{diag}(2, 1, 1, 0, 0),$$
$$A_2 = \text{diag}(1, 2, 0, 1, 0),$$
$$A_3 = \text{diag}(0, 1, 0, 0, 1),$$
$$b = (8, 7, 3)^T$$
$$x^0 = (-2, -1, -2)^T.$$

Exemple 2.9.

$$C = \text{diag}(-4, -5, -1, -3, 5, -8, 0, 0, 0, 0, 0, 0),$$

$$A_1 = \text{diag}(1, 0, -4, 3, 1, 1, 1, 0, 0, 0, 0, 0),$$

$$A_2 = \text{diag}(5, 3, 1, 0, -1, 3, 0, 1, 0, 0, 0, 0),$$

$$A_3 = \text{diag}(4, 5, 3, 3, 4, 1, 0, 0, 1, 0, 0, 0),$$

$$A_4 = \text{diag}(0, -1, 0, 2, 1, -5, 0, 0, 0, 1, 0, 0),$$

$$A_5 = \text{diag}(-2, 1, 1, 1, 2, 2, 0, 0, 0, 0, 1, 0),$$

$$A_6 = \text{diag}(2, -3, 2, -1, 4, 5, 0, 0, 0, 0, 0, 1),$$

$$b = (1, 4, 4, 5, 7, 5)^T$$

$$x^0 = \left(\frac{1}{16}, \frac{1}{16}, 1, \frac{1}{16}, \frac{1}{8}, \frac{1}{16}\right)^T.$$

Dans le tableau suivant, on note par

Méthode 1 : Calcul du pas de déplacement par la technique des fonctions minorantes.

m1 : Première fonction minorante.

m2 : Deuxième fonction minorante.

m3 : Troisième fonction minorante.

Méthode 2 : Calcul du pas de déplacement par la recherche linéaire classique (Wolfe)

TABLE 2.3: Tests numériques pour les exemples 2.6, 2.7, 2.8 et 2.9 par les fonctions minorantes

exemple	Méthode 1						Méthode 2	
	m1		m2		m3		itér	temp(s)
itér	temp(s)	itér	temp(s)	itér	temp(s)	itér		
2.6	3	0.012	3	0.014	4	0.19	5	0.25
2.7	1	0.0016	1	0.0022	2	0.0065	7	0.36
2.8	1	0.0027	1	0.0056	3	0.081	10	0.87
2.9	1	0.0037	1	0.0078	1	0.0081	10	0.87

2.2.5 Calcul du pas de déplacement via des fonctions majorantes

Cette technique est basée sur la majoration de la fonction de recherche linéaire G , tout en conservant les mêmes propriétés de cette dernière. L'avantage dans ce cas est la détermination du pas de déplacement soit facile et donnée explicitement, en utilisant la fonction majorante \tilde{G} de G , notons que l'utilisation de l'approximation majorante est mieux adaptée que l'approximation minorante et elle est intéressante car le domaine de définition de \tilde{G} est inclus de la domaine de définition de G et la décroissance de \tilde{G} nous assure la décroissance de G .

Dans notre cas on exige

$$\tilde{G}(0) = 0, -\tilde{G}'(0) = \tilde{G}''(0) = \|\lambda\|^2.$$

Première fonction majorante

Dans ce paragraphe, on définit la première fonction majorante qui semble apparaître simple et suffisamment proche de G , à l'aide de l'application du théorème [10]. Partons de ce théorème on obtient

$$-\sum_{i=1}^n \ln(1 + \lambda_i t) \leq -(n-1) \ln(1 + \beta_1 t) - \ln(1 + \gamma_1 t),$$

avec

$$\beta_1 = \bar{\lambda} + \frac{\sigma_\lambda}{\sqrt{n-1}} \text{ et } \gamma_1 = \bar{\lambda} - \sigma_\lambda \sqrt{n-1},$$

alors

$$n\bar{\lambda}t - \|\lambda\|^2 t - \sum_{i=1}^n \ln(1 + \lambda_i t) \leq n\bar{\lambda}t - \|\lambda\|^2 t - (n-1) \ln(1 + \beta_1 t) - \ln(1 + \gamma_1 t),$$

pour tout $t \in [0, \hat{t}_1[$ avec

$$\hat{t}_1 = \begin{cases} -\frac{1}{\gamma_1} & \text{si } \gamma_1 < 0, \\ +\infty & \text{sinon.} \end{cases}$$

Par conséquent, on aura la majoration suivante

$$\tilde{G}_1(t) = \delta_1 t - (n-1) \ln(1 + \beta_1 t) - \ln(1 + \gamma_1 t),$$

tel que

$$\delta_1 = n\bar{\lambda} - \|\lambda\|^2.$$

On à \tilde{G}_1 est strictement convexe, alors elle admet un minimum unique sur $[0, \hat{t}_1[$, que l'on trouve en résolvant l'équation $\tilde{G}'_1(t) = 0$.

Pour cela on cherche les racines de l'équation de deuxième degré

$$\beta_1\gamma_1\delta_1 t^2 + [(\beta_1 + \gamma_1)\delta_1 - n\beta_1\gamma_1]t + [\delta_1 - (n-1)\beta_1 - \gamma_1] = 0.$$

On obtient

$$\tilde{t}_{1i} = \frac{-1}{2\beta_1\gamma_1\delta_1} \left[(\beta_1 + \gamma_1)\delta_1 - n\beta_1\gamma_1 \pm \sqrt{[(\beta_1 + \gamma_1)\delta_1 - n\beta_1\gamma_1]^2 + 4\beta_1\gamma_1\delta_1(\beta_1 - \gamma_1)} \right], i \in \{1, 2\},$$

On prend une seule des deux racines qui appartient à $[0, \hat{t}_1[$.

Deuxième fonction majorante

On considère la fonction suivante :

$$\tilde{G}_2(t) = \gamma_2 t - \delta_2 \ln(1 + \beta_2 t), \quad t \in [0, \hat{t}[, \quad \hat{t}_2 = \sup(t : 1 + \beta_2 t > 0),$$

tels que

$$\beta_2 = \bar{\lambda} - \sigma_\lambda \sqrt{n-1},$$

et

$$\delta_2 = \frac{\|\lambda\|^2}{\beta_2^2}, \quad \gamma_2 = \delta_2 \beta_2 - \|\lambda\|^2.$$

Puisque \tilde{G}_2 est strictement convexe, alors elle admet un minimum unique sur $[0, \hat{t}_2[$, que l'on trouve en résolvant l'équation $\tilde{G}'_2(t) = 0$, puis on obtient

$$\tilde{t}_2 = \frac{\delta_2}{\gamma_2} - \frac{1}{\beta_2}.$$

Troisième fonction majorante

Il est ainsi possible de définir une nouvelle majoration plus simple par rapport à la première, et qui contient un seul logarithme.

L'idée consiste à utiliser l'inégalité

$$-\sum_{i=1}^n \ln(1 + \lambda_i t) \leq (-\|\lambda\| - \sum_{i=1}^n \lambda_i)t - \ln(1 - \|\lambda\|t).$$

Alors

$$G(t) \leq -\|\lambda\|(1 + \|\lambda\|)t - \ln(1 - \|\lambda\|t),$$

d'où

$$\tilde{G}_3(t) = -\|\lambda\|(1 + \|\lambda\|)t - \ln(1 - \|\lambda\|t),$$

définie sur $[0, \hat{t}_3[$ avec $\hat{t}_3 = \frac{1}{\|\lambda\|}$.

\tilde{G}_3 vérifie les propriétés suivantes

- ▷ $\tilde{G}_3(0) = G(0) = 0$ et $\tilde{G}'_3(0) = G'(0) = -\|\lambda\|^3 < 0$,
- ▷ $\tilde{G}''_3(0) = G''(0) = \|\lambda\|^3 > 0$,
- ▷ $G(t) \leq \tilde{G}_3(t)$,
- ▷ \hat{G}_3 est strictement convexe.

Puisque \tilde{G}_3 est strictement convexe, alors elle admet un minimum unique, que l'on trouve en résolvant l'équation $\hat{G}'_3(t) = 0$ puis on obtient

$$\tilde{t}_3 = \frac{-1}{\|\lambda\| - 1}.$$

Description de l'algorithme

Algorithme 3 : Résolution d'un problème *SDP* par les fonctions majorantes

Initialisation : $\epsilon > 0$ est une précision donnée, et $\rho > 0, \sigma \in]0, 1[$ sont des paramètres, $\eta > 0$ paramètre de pénalisation, un point $x^0 \in \hat{X}$ et $k = 0$,

étape 1 : à l'itération k on prend

$$B(x^k) = \sum_{i=1}^m x_i^k A_i - C \text{ et } L_k \text{ tel que } L_k L_k^T = B(x^k),$$

et calculer

$$\begin{aligned} \hat{A}_i^k &= L_k^{-1} A_i [L_k^T]^{-1}, \\ b_i^k &= \text{trace}(\hat{A}_i^k), \\ \Delta_{ij}^k &= \text{trace}(\hat{A}_i^k \hat{A}_j^k), \\ H &= \eta \Delta^k. \end{aligned}$$

étape 2 : Résoudre le système linéaire $Hd = \eta b^k - b$, et prendre la nouvelle itération $x^{k+1} = x^k + \bar{t}d$, telle que \bar{t} est obtenu par l'utilisation de la stratégie de pas de déplacement de $\hat{G}_i, i = 1, 2, 3$.

- 1 **si** $n\eta > \epsilon$ ou $|b^T x^{k+1} - b^T x^k| > n\rho\eta$ **alors**
 - 2 | prendre $k = k + 1, \eta = \sigma\eta$ et aller en (1).
 - 3 **sinon**
 - 4 | Stop x^k est une solution approchée de notre problème.
-

Exemple 2.10 (Cube). dans ce cas $n = 2m$, C est une matrice diagonale de taille n avec les éléments diagonaux sont égaux, $b = (2, \dots, 2)^T \in \mathbb{R}^m$ et les matrices $A_k, k = 1, \dots, m$, sont données par

$$A_k[i, j] = \begin{cases} 1 & \text{si } i = j = k & \text{ou } i = j = k + m, \\ a^2 & \text{si } i = j = k + 1 & \text{ou } i = j = k + m + 1, \\ -a & \text{si } i = k, j = k + 1 & \text{ou } i = k + m, j = k + m + 1, \\ -a & \text{si } i = k + 1, j = k & \text{ou } i = k + m + 1, j = k + m, \\ 0 & \text{ailleurs.} \end{cases}$$

Avec a est un réel donné.

Dans tous les cas on prend

$$\rho = 1, \sigma = 0.125, \eta = 0.3, \epsilon = 0.1.$$

Premier cas :

$$m = 50, a = 1, C = I_n, x_0 = (3, \dots, 3, \frac{3}{2})^T.$$

Deuxième cas :

$$m = 50, a = 2, C = -I_n, x_0 = (\frac{3}{2}, \dots, \frac{3}{2})^T.$$

Troisième cas :

$$m = 50, a = 5, C = -2I_n, x_0 = (0, \dots, 0)^T.$$

Quatrième cas :

$$m = 50, a = 2, C = -2I_n, x_0 = (0, \dots, 0)^T.$$

Cinquième cas :

$$m = 50, a = 0, C = I_n, x_0 = (\frac{3}{2}, \dots, \frac{3}{2})^T.$$

Dans le tableau suivant, on note par

Méthode 1 : Calcul du pas de déplacement par la technique des fonctions majorantes.

M1 : Première fonction majorante.

M2 : Deuxième fonction majorante.

M3 : Troisième fonction majorante.

Méthode 2 : Calcul du pas de déplacement par la recherche linéaire classique (Wolfe)

TABLE 2.4: Tests numériques (via fonctions majorantes) pour l'exemple 2.10

exemple	Méthode 1						Méthode 2	
	M1		M2		M3		itér	temp(s)
	itér	temp(s)	itér	temp(s)	itér	temp(s)	itér	temp(s)
1	24	376	24	380	32	820	div	div
2	19	289	19	292	36	572	div	div
3	9	120	9	130	14	288	div	div
4	15	220	15	240	19	326	div	div
5	3	38	4	38	6	60	div	div

Chapitre 3

Une variante de la méthode des pénalités logarithmiques pour la programmation convexe

Ce chapitre présente une variante des méthodes de pénalités logarithmiques pour la programmation convexe. Si le calcul de la direction est obtenu de façon classique par une méthode newtonienne, la détermination du pas de déplacement s'appuie sur une technique dite de fonctions majorantes. Les tests numériques montrent l'efficacité de notre approche par rapport à la recherche linéaire classique.

3.1 Introduction

On s'intéresse au problème d'optimisation non linéaire

$$\alpha = \min [f(x) : x \in D] \text{ avec } D = \{x \in \mathbb{R}^n : x \geq 0, Ax = b\}, \quad (P)$$

où A est une matrice $p \times n$ et le vecteur b appartient à \mathbb{R}^p . La matrice A et le vecteur b sont donnés.

Il existe de nombreuses méthodes de résolution pour ce problème basées sur des extensions de l'algorithme de Kamarkar. Dans ce travail on propose une méthode de pénalité logarithmique utilisant une nouvelle technique pour le calcul du pas de déplacement

Dans tout ce que suite, on fait les hypothèses suivantes

1. f est convexe et deux fois différentiable sur D .
2. On connaît $x_0 > 0$ tel que $Ax_0 = b$.
3. La matrice A est de plein rang.
4. L'ensemble des solutions optimales de (P) est non vide et borné.

La dernière hypothèse est équivalente à dire que

$$\{d \in \mathbb{R}^n : f_\infty(d) \leq 0, d \geq 0, Ad = 0\} = \{0\}$$

où f_∞ désigne la fonction asymptotique de f .

Il résulte des conditions d'optimalité que \bar{x} est solution optimale de (P) si et seulement s'il existe $\bar{u} \in \mathbb{R}^p$ et $\bar{v} \in \mathbb{R}^n$ tels que

$$\nabla f(\bar{x}) + A^T \bar{u} = \bar{v} \geq 0, \quad A\bar{x} = b, \quad \langle \bar{v}, \bar{x} \rangle = 0. \quad (3.1)$$

3.2 La pénalisation

3.2.1 Le problème perturbé

La fonction $\gamma : \mathbb{R}^2 \rightarrow (-\infty, +\infty]$ définie par

$$\gamma(r, t) = \begin{cases} r \ln(r) - r \ln(t) & \text{si } t > 0 \text{ et } r > 0, \\ 0 & \text{si } r = 0 \text{ et } t \geq 0, \\ +\infty & \text{ailleurs.} \end{cases}$$

est convexe, sci et propre. On considère ensuite la fonction définie sur $\mathbb{R} \times \mathbb{R}^n$ par

$$\varphi(r, x) = \begin{cases} f(x) + \sum_{i=1}^n \gamma(r, x_i) & \text{si } x \geq 0, Ax = b, \\ +\infty & \text{sinon.} \end{cases}$$

Cette fonction est elle aussi convexe, sci et propre. Enfin, on introduit la fonction h définie par

$$h(r) = \inf_x [\varphi_r(x) = \varphi(r, x) : x \in \mathbb{R}^n]. \quad (P_r)$$

La fonction h est convexe puisque φ est convexe. Par construction, (P_0) n'est rien d'autre que le problème (P) , et donc

$$\alpha = h(0).$$

La fonction φ_r est convexe, sci et propre, sa fonction asymptotique est

$$[\varphi_r]_\infty(d) = \lim_{t \rightarrow +\infty} \frac{\varphi_r(x_0 + td) - \varphi_r(x_0)}{t}.$$

On obtient

$$[\varphi_r]_\infty(d) = \begin{cases} f_\infty(d) & \text{si } d \geq 0, Ad = 0, \\ +\infty & \text{sinon.} \end{cases}$$

Donc,

$$\{d \in \mathbb{R}^n : [\varphi_r]_\infty(d) \leq 0\} = \{d \in \mathbb{R}^n : f_\infty(d) \leq 0, d \geq 0, Ad = 0\}.$$

On sait que cet ensemble est réduit à $\{0\}$. Le problème (P_r) admet donc des solutions optimales. Lorsque $r > 0$, φ_r est fortement convexe, (P_r) admet donc une solution unique que l'on notera x_r ou $x(r)$ si on veut étudier son comportement en fonction de r .

3.2.2 Convergence

En raison des conditions nécessaires et suffisantes d'optimalité, il existe $u_r = u(r) \in \mathbb{R}^p$ tel que

$$\nabla f(x_r) - rX_r^{-1}e + A^T u_r = 0 \quad (3.2)$$

$$Ax_r - b = 0. \quad (3.3)$$

Notons que u_r est défini de façon unique en raison de l'hypothèse 3. En fait, le couple (x_r, u_r) est solution de l'équation $H(x, u) = 0$ où

$$H(x, u) = \begin{pmatrix} \nabla f(x) - rX^{-1}e + A^T u \\ Ax - b \end{pmatrix}$$

En utilisant le théorème des fonctions implicites, on voit que les fonctions $r \mapsto x(r) = x_r$ et $r \mapsto u(r) = u_r$ sont différentiables sur $(0, \infty)$ et on a

$$\begin{pmatrix} \nabla^2 f(x(r)) + rX^{-2}(r) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x'(r) \\ u'(r) \end{pmatrix} = \begin{pmatrix} X^{-1}(r)e \\ 0 \end{pmatrix}. \quad (3.4)$$

Il s'ensuit que la fonction h est différentiable sur $(0, \infty)$. Rappelons que

$$h(r) = nr \ln(r) + f(x(r)) - r \sum_{i=1}^n \ln(x_i(r)),$$

et donc

$$h'(r) = n + n \ln(r) - \sum_{i=1}^n \ln(x_i(r)) + \langle \nabla f(x(r)) - rX^{-1}(r)e, x'(r) \rangle.$$

Compte tenu de (3.2) et (3.4), on obtient

$$\begin{aligned} h'(r) &= n + n \ln(r) - \sum_{i=1}^n \ln(x_i(r)) - \langle A^T u(r), x'(r) \rangle, \\ &= n + n \ln(r) - \sum_{i=1}^n \ln(x_i(r)) - \langle u(r), Ax'(r) \rangle, \\ &= n + n \ln(r) - \sum_{i=1}^n \ln(x_i(r)). \end{aligned}$$

Puisque $x(r) \in D$ et h est convexe, on a

$$f(x(r)) \geq \alpha = h(0) \geq h(r) + (0 - r)h'(r) = f(x(r)) - nr.$$

Donc

$$\alpha \leq f(x(r)) \leq \alpha + nr. \quad (**)$$

Intéressons nous maintenant à la trajectoire $\{x(r)\}$ lorsque $r \rightarrow 0$.

i) Cas où f est fortement convexe de coefficient $\tau > 0$. Alors (P) a une solution optimale unique \bar{x} . On a

$$nr \geq f(x(r)) - f(\bar{x}) \geq \langle \nabla f(\bar{x}), x(r) - \bar{x} \rangle + \frac{\tau}{2} \|x(r) - \bar{x}\|^2.$$

Compte tenu de (3.1) on a

$$nr \geq \langle \bar{v}, x(r) \rangle + \frac{\tau}{2} \|x(r) - \bar{x}\|^2 \geq \frac{\tau}{2} \|x(r) - \bar{x}\|^2.$$

D'où

$$\|x(r) - \bar{x}\| \leq \sqrt{\frac{2nr}{\tau}}.$$

La convergence de $x(r)$ vers la solution optimale unique est d'ordre 1/2.

2) L'étude, dans le cas où f est seulement convexe, est plus complexe. Notons tout d'abord que, pour tout $r \leq 1$,

$$x(r) \in \{x : x \geq 0, Ax = b, f(x) \leq n + \alpha\}.$$

Cet ensemble est convexe, fermé et non vide. Son cône asymptotique est

$$\{d \in \mathbb{R}^n : f_\infty(d) \leq 0, d \geq 0, Ad = 0\} = \{0\}.$$

Et il est borné en raison de l'hypothèse 4. Il s'ensuit que toute valeur d'adhérence de $\{x(r)\}$ lorsque $r \rightarrow 0$ est solution optimale de (P) .

3.2.3 Le principe de la méthode

Posons $\tilde{D} = \{x \in \mathbb{R}^n : x > 0, Ax = b\}$. Le principe général de la méthode est le suivant : Soit $(x_k, r_k) \in \tilde{D} \times (0, \infty)$ l'itéré courant.

1. On fait une minimisation approchée du problème perturbé (P_{r_k}) qui donne un point x_{k+1} tel que $\varphi(r_k, x_{k+1}) < \varphi(r_k, x_k)$.
2. On prend $r_{k+1} < r_k$.

On réitère jusqu'à ce que l'on obtienne une solution optimale approchée du problème original satisfaisante.

3.3 Résolution du problème perturbé

Le problème perturbé est du type

$$\min_x [\varphi(r, x)] = \min_x [f(x) + \sum_{i=1}^n \gamma(r, x_i) : x \geq 0, Ax = b]. \quad (P_r)$$

3.3.1 La direction de descente

En $x \in \tilde{D}$, la direction de descente de Newton d est donnée en résolvant le problème d'optimisation convexe quadratique suivant

$$\min_d [\langle \nabla \varphi_r(x), d \rangle + \frac{1}{2} \langle \nabla^2 \varphi_r(x) d, d \rangle : Ad = 0].$$

Il suffit de résoudre le système linéaire à $n + p$ équations

$$\begin{pmatrix} \nabla^2 f(x) + rX^{-2} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ v \end{pmatrix} = \begin{pmatrix} rX^{-1}e - \nabla f(x) \\ 0 \end{pmatrix}. \quad (SL)$$

Il est facile de voir que le système est inversible. On a

$$(d^T, 0) \begin{pmatrix} \nabla^2 f(x) + rX^{-2} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ v \end{pmatrix} = (d^T, 0) \begin{pmatrix} rX^{-1}e - \nabla f(x) \\ 0 \end{pmatrix}.$$

On en déduit

$$\langle \nabla^2 f(x) d, d \rangle + \langle \nabla f(x), d \rangle = r[\langle X^{-1}d, e \rangle - \|X^{-1}d\|^2]. \quad (3.5)$$

Le système peut encore s'écrire

$$\begin{pmatrix} X\nabla^2 f(x)X + rI & XA^T \\ AX & 0 \end{pmatrix} \begin{pmatrix} X^{-1}d \\ v \end{pmatrix} = \begin{pmatrix} re - X\nabla f(x) \\ 0 \end{pmatrix}. \quad (3.6)$$

La direction de descente étant ainsi obtenue, il s'agit maintenant de minimiser la fonction d'une variable réelle

$$\theta(t) = \frac{1}{r}[\varphi_r(x + td) - \varphi_r(x)] = \frac{1}{r}[f(x + td) - f(x)] - \sum_{i=1}^n \ln(1 + ty_i),$$

où $y = X^{-1}d$. Cette fonction θ est convexe. On a

$$\begin{aligned} \theta'(t) &= \frac{1}{r} \langle \nabla f(x + td), d \rangle - \sum_{i=1}^n \frac{y_i}{1 + ty_i}, \\ \theta''(t) &= \frac{1}{r} \langle \nabla^2 f(x + td) d, d \rangle + \sum_{i=1}^n \frac{y_i^2}{(1 + ty_i)^2}. \end{aligned}$$

On déduit de (3.5) que l'on a $\theta'(0) + \theta''(0) = 0$, ce qui est attendu puisque d est la direction de descente de Newton.

3.3.2 Une fonction majorante

Il suffirait de résoudre l'équation $\theta'(t) = 0$ pour trouver le minimum de $\theta(t)$. Dans le cas où f est une fonction linéaire et donc $\nabla f(x + td)$ est constant, cela reviendrait à résoudre une équation polynomiale de degré $n + 1$. On pourrait évidemment utiliser une méthode itérative classique, les calculs deviennent lourds lorsque n est grand. La méthode que nous proposons est basée sur l'utilisation d'une majorante θ_1 de la fonction θ . On utilise le résultat suivant.

Théorème 3.1. [10] Soient w_1, w_2, \dots, w_n strictement positifs. Alors,

$$\sum_{i=1}^n \ln(w_i) \geq (n-1) \ln\left(\bar{w} + \frac{\sigma_w}{\sqrt{n-1}}\right) + \ln(\bar{w} - \sigma_w \sqrt{n-1})$$

où $\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i$ et $\sigma_w = \sqrt{\frac{1}{n} \|w\|^2 - \bar{w}^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (w_i - \bar{w})^2}$.

On pose

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}, \quad w_i = 1 + ty_i.$$

Alors, $\bar{w} = 1 + t\bar{y}$, $\sigma_w = t\sigma_y$ et $\theta(t) \leq \theta_1(t)$ où

$$\theta_1(t) = \frac{f(x + td) - f(x)}{r} - (n-1) \ln\left(1 + t\bar{y} + \frac{t\sigma_y}{\sqrt{n-1}}\right) - \ln\left(1 + t\bar{y} - t\sigma_y \sqrt{n-1}\right),$$

$$\theta'_1(t) = \frac{\langle \nabla f(x + td), d \rangle}{r} - \frac{(n-1)\bar{y} + \sigma_y \sqrt{n-1}}{1 + t\bar{y} + \frac{t\sigma_y}{\sqrt{n-1}}} - \frac{\bar{y} - \sigma_y \sqrt{n-1}}{1 + t\bar{y} - t\sigma_y \sqrt{n-1}}.$$

On remarque que l'on a

$$\theta_1(0) = \theta(0) = 0 \quad 0 < \theta''(0) = \theta''_1(0) = -\theta'_1(0) = -\theta'(0).$$

La fonction θ_1 est donc une bonne approximation de la fonction θ . Elle est strictement convexe et donc atteint son minimum en un point unique. Il est clair que l'on a

$$\min \theta(t) \leq \min \theta_1(t).$$

En choisissant pour pas de déplacement t tel que $\theta'_1(t) = 0$, on obtiendra une décroissance significative de la fonction θ .

En posant

$$\alpha = \bar{y} + \frac{\sigma_y}{\sqrt{n-1}}, \quad \beta = \bar{y} - \sigma_y \sqrt{n-1}. \quad (3.7)$$

Le domaine de définition de θ_1 est $(0, \tilde{t})$ avec

$$\tilde{t} = \min\left\{t_{max}, \frac{-1}{\alpha} : \alpha < 0, \frac{-1}{\beta} : \beta < 0\right\}, t_{max} = \min\left\{-\frac{x_i}{d_i}, d_i < 0\right\}$$

qui est contenu dans celui de la fonction θ .

3.3.3 Minimisation d'une fonction auxiliaire

On s'intéresse maintenant à la minimisation de la fonction

$$\xi(t) = n\delta t - (n-1)\ln(1+\alpha t) - \ln(1+\beta t),$$

On a alors,

$$\xi'(t) = n\delta - \frac{(n-1)\alpha}{1+\alpha t} - \frac{\beta}{1+\beta t},$$

$$\xi''(t) = \frac{(n-1)\alpha^2}{(1+\alpha t)^2} + \frac{\beta^2}{(1+\beta t)^2},$$

$$\xi(0) = 0, \quad \xi'(0) = n(\delta - \bar{y}), \quad \xi''(0) = n(\bar{y}^2 + \sigma_y^2) = \|y\|^2.$$

On impose les conditions $\xi'(0) < 0$ et $\xi''(0) > 0$. La fonction ξ est strictement convexe sur son domaine. Elle atteint son minimum en un point unique $\bar{t} > 0$. \bar{t} est l'unique racine de l'équation $\xi'(t) = 0$ qui appartient au domaine de définition de ξ qui coïncide avec celui de θ .

Les calculs montrent que \bar{t} est racine de l'équation

$$\alpha\beta\delta t^2 + t(\delta\alpha + \delta\beta - \alpha\beta) + \delta - \bar{y} = 0. \quad (3.8)$$

- Si $\alpha = 0$, on a $\bar{t} = \frac{\bar{y}-\delta}{\delta\beta}$.
- Si $\beta = 0$, on a $\bar{t} = \frac{\bar{y}-\delta}{\delta\alpha}$.

- Si $\delta = 0$, on a $\bar{t} = \frac{-\bar{y}}{\alpha\beta}$.
- Si $\alpha\beta\delta \neq 0$, \bar{t} est l'unique racine de l'équation du deuxième degré qui appartient au domaine de définition de ξ . Les deux racines sont

$$t_- = \frac{1}{2} \left(\frac{1}{\delta} - \frac{1}{\alpha} - \frac{1}{\beta} - \sqrt{\Delta} \right), \quad t_+ = \frac{1}{2} \left(\frac{1}{\delta} - \frac{1}{\alpha} - \frac{1}{\beta} + \sqrt{\Delta} \right),$$

où

$$\Delta = \frac{1}{\alpha^2} + \frac{1}{\beta^2} + \frac{1}{\delta^2} - \frac{2}{\alpha\beta} + \left(\frac{2n-4}{n} \right) \left[\frac{1}{\beta\delta} - \frac{1}{\delta\alpha} \right].$$

3.3.4 Cas linéaire.

Il existe $c \in \mathbb{R}^n$ tel que $f(x) = \langle c, x \rangle$ pour tout x . La fonction θ_1 coïncide avec la fonction ξ étudiée ci-dessus où $\delta = (nr)^{-1} \langle c, d \rangle$.

Le minimum de θ_1 est atteint en \bar{t} racine de l'équation (3.8). Remarquons que la condition $\theta_1'(0) + \theta_1''(0) = 0$ implique

$$0 < \theta_1''(0) = n(\bar{y}^2 + \sigma_y^2) = \|y\|^2 = -\theta_1'(0) = n(\bar{y} - \delta).$$

On a alors

$$\theta(\bar{t}) \leq \theta_1(\bar{t}) < \theta_1(0) = \theta(0).$$

On a obtenu une décroissance significative de la fonction θ le long de la direction d .

3.3.5 Cas convexe

$\nabla f(x+td)$ n'est plus constant et l'équation $\theta_1'(t) = 0$ ne se réduit plus à une équation du second degré. On va chercher une autre majorante θ_2 de θ .

Étant donné $\hat{t} > 0$, la convexité de f nous permet d'écrire

$$\frac{f(x+td) - f(x)}{r} \leq \frac{f(x+\hat{t}d) - f(x)}{r\hat{t}} t, \quad \text{pour tout } t \in]0, \hat{t}] \quad (3.9)$$

$$\theta(t) \leq \theta_1(t) \leq \frac{f(x+\hat{t}d) - f(x)}{r\hat{t}} t - (n-1) \ln(1+\alpha t) - \ln(1+\beta t).$$

On obtient ainsi une nouvelle fonction majorante θ_2

$$\theta_2(t) = n\delta t - (n-1)\ln(1+\alpha t) - \ln(1+\beta t) \text{ où } \delta = \frac{f(x+\hat{t}d) - f(x)}{nr\hat{t}}.$$

On a alors $\theta(t) \leq \theta_1(t) \leq \theta_2(t)$ pour tout $t \in]0, \hat{t}]$. Les domaines de définition de θ et θ_1 coïncident. La valeur \hat{t} doit être choisie dans ce domaine. On se donne une valeur $\hat{t} > 0$ appartenant au domaine de définition de θ , donné par $]0, t_{max}[$, où $t_{max} = \min\{-\frac{x_i}{d_i} : d_i < 0\}$.

1. Si 1 est dans $]0, t_{max}[$, on prend $\hat{t} = 1$:

sinon $\hat{t} = \Lambda_- t_{max}$ où $0 < \Lambda_- < 1$.

2. Calculer \bar{t} .

si $\bar{t} < \hat{t}$ on a obtenu une décroissance de θ_2 et donc de θ .

sinon on augmente \hat{t} :

$\hat{t} = \Lambda_+ \hat{t}$ où $1 < \Lambda_+ < \frac{t_{max}}{\hat{t}}$ puis retourner en 2.

3.4 L'algorithme

On se donne deux seuils ε et \hat{r} strictement positifs.

1. On a un point $x > 0$ tel que $Ax = b$ et $r > 0$.
2. On calcule la direction de descente d et $y = X^{-1}d$ au moyen de la relation (3.5) ou (3.6).
3. Si $\|y\| < \varepsilon$ on a une bonne approximation de $h(r)$.
 - (a) Si $r < \hat{r}$, STOP. on a une bonne approximation de la solution optimale.
 - (b) Si $r \geq \hat{r}$. On diminue r ($r = \omega r$) avec $0 < \omega < 1$.
4. On calcule \bar{y}, σ, α et β .
5. Selon le cas linéaire ou non linéaire, on utilise la majoration θ_1 ou θ_2 pour obtenir \bar{t} tel que $\theta(\bar{t}) < \theta(0)$. On fait $x = x + \bar{t}d = X(e + \bar{t}y)$ et on retourne au point 2.

3.5 Simulations numériques

Dans les tableaux, la colonne itér indique le nombre d'itérations, la colonne temps indique le temps d'exécution en secondes, la méthode 1 représente notre variante, la méthode 2 est celle utilisant les recherches linéaires classiques (Armijo-Goldstein).

3.5.1 Fonction objectif convexe non linéaire

Considérons les deux problèmes non linéaires convexes suivants :

Exemple 3.1. $f(x) = x_1^2 + x_2^2 - 3x_1 - 5x_2$,

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 7 \end{bmatrix}.$$

La solution optimale exacte est : $x^* = (1, \frac{3}{2})$, $z^* = -\frac{29}{4}$.

Exemple 3.2. $f(x) = x_1^2 + x_2^2 - 2x_1 - 4x_2$

$$A = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 2 & 3 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 5 \\ 6 \end{bmatrix}.$$

La solution optimale exacte est : $x^* = (\frac{13}{17}, \frac{18}{17})$, $z^* = -\frac{1104}{272}$.

TABLE 3.1: Tests numériques pour les exemples 3.1 et 3.2.

Exemple	Méthode 1			Méthode 2		
	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)
1	-7.17	4	0.6×10^{-3}	-7.18	4	0.01
2	-4.05	39	0.4×10^{-2}	-4.05	39	0.01

3.5.2 Problème à tailles variables

Cas linéaire

Exemple 3.3. *Considérons les problèmes linéaires suivants :*
$$\begin{cases} \min c^T x, \\ Ax = b, \\ x \geq 0. \end{cases}$$

$$A[i, j] = \begin{cases} 1 & \text{si } i = j \text{ ou } j = i + m, \\ 0 & \text{sinon} \end{cases}$$

$c[i] = -1, c[i + m] = 0, b[i] = 2$, pour $i = 1, \dots, m$, avec $n = 2m$.

-La valeur optimale : $z^* = -2m$.

-Solution optimale : $x_i^* = \begin{cases} 2, & \text{si } i=1, \dots, m. \\ 0, & \text{si } i=m+1, \dots, 2m. \end{cases}$

TABLE 3.2: Tests numériques pour l'exemple 3.3

Taille	Méthode 1			Méthode 2		
	Objectif	itér.	Temps(s)	Objectif	itér.	Temps(s)
5	-9.96	9	0.01×10^{-1}	-9.79	9	0.18×10^{-1}
10	-19.92	9	0.04×10^{-1}	-19.87	11	0.19×10^{-1}
25	-49.80	9	0.03	-49.67	11	0.49×10^{-1}
50	-99.60	9	0.17	-99.35	11	0.22
100	-199.21	9	1.21	-199.60	13	1.79
150	-298.82	9	3.80	-299.41	13	5.96
200	-398.43	9	11.98	-399.21	13	18.47
250	-498.04	9	17.99	-499.02	13	26.57

Cas non linéaire

Exemple 3.4 (problèmes d'Erikson [20]). *On considère le problème convexe suivant avec $n = 2m$*

$$\begin{cases} \min f(x) = \sum_{i=1}^n x_i \ln\left(\frac{x_i}{a_i}\right) \\ Ax = b \\ x \geq 0 \end{cases}$$

où $a_i \in \mathbb{R}^+$, $b_i \in \mathbb{R}$ sont fixes et,

$$A[i, j] = \begin{cases} 1 & \text{si } i = j \text{ ou } j = i + m, \\ 0 & \text{sinon.} \end{cases}$$

On va résoudre ce problème, en utilisant les deux méthodes (Méthode 1 et 2) pour différentes valeurs de n , a_i et b_j .

1) ($a_i = 1$, $b_j = 6$)

TABLE 3.3: Tests numériques pour l'exemple 3.4 dans le cas ($a_i = 1$, $b_j = 6$)

Taille	Méthode 1			Méthode 2		
	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)
10	32.95	3	0.004	32.95	4	0.018
50	164.79	4	0.025	164.79	6	0.082
100	329.58	5	0.076	329.58	12	0.230
1000	3295.83	13	44.130	3295.83	25	83.979

2) ($a_i = 2, b_j = 4$)

TABLE 3.4: Tests numériques pour l'exemple 3.4 dans le cas ($a_i = 2, b_j = 4$)

Taille	Méthode 1			Méthode 2		
	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)
10	0.66×10^{-7}	2	0.002	9.09×10^{-6}	3	0.01
50	0.11×10^{-8}	3	0.30×10^{-2}	1.86×10^{-4}	4	0.07
100	0.76×10^{-7}	3	0.04	2.05×10^{-4}	5	0.22
1000	2.44×10^{-7}	6	20.26	1.03×10^{-4}	11	38.83

Exemple 3.5. cas quadratique. On considère le problème quadratique suivant avec $n = m + 2$

$$\begin{cases} \min f(x) = \frac{1}{2} \langle x, Qx \rangle + \langle c, x \rangle \\ Ax = b \\ x \geq 0 \end{cases}$$

où ;

$$Q[i, j] = \begin{cases} 2 & \text{si } i = j = 1 \text{ ou } j = i = n, \\ 4 & \text{si } i = j \text{ et } i \neq 1, n, \\ 2 & \text{si } j = i + 1 \text{ ou } j = i - 1, \\ 0 & \text{ailleurs.} \end{cases}$$

,

$$A[i, j] = \begin{cases} 1 & \text{si } j = i, \\ 2 & \text{si } j = i + 1, \\ 3 & \text{si } j = i + 2, \\ 0 & \text{ailleurs.} \end{cases}$$

et $b_i = 1, c_i = 0$.

On va applique l'algorithme précédent à cet exemple pour différentes valeurs de n ,

TABLE 3.5: Tests numériques pour l'exemple 3.5

Taille	Méthode 1			Méthode 2		
	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)
4	0.285	9	0.007	0.285	14	0.019
50	5.372	7	0.021	5.374	14	0.053
100	10.927	7	0.080	10.930	14	0.188
500	55.372	7	13.800	55.374	14	29.815

Conclusion :

Les tests numériques effectués montrent que la technique des fonctions majorantes que nous avons proposée conduit à une diminution très significative du coût de calcul et une amélioration du résultat.

Chapitre 4

Application de la nouvelle variante dans le cas quadratique

4.1 La programmation quadratique

La programmation quadratique est une branche d'optimisation non linéaire où la fonction objective à minimiser est quadratique et les contraintes sont linéaires et/ou quadratiques.

4.1.1 La programmation quadratique sans contraintes

Un problème de programmation quadratique sans contraintes peut s'écrire sous la forme suivante :

$$\begin{cases} \min f(x) = \frac{1}{2}x^T Qx + c^T x, \\ x \in \mathbb{R}^n \end{cases}$$

où Q est une matrice carrée symétrique d'ordre n , c est un vecteur de \mathbb{R}^n et $x \in \mathbb{R}^n$ est le vecteur inconnu à déterminer.

Et si la matrice hessienne $H(x^*)$ est définie positive alors la condition dite suffisante.

Il existe plusieurs méthodes pour résoudre un programme quadratique sans contraintes, nous citons, par exemple :

1. Les méthodes du gradient.
2. Les méthodes de directions conjuguées.
3. La méthode de Newton et ses variantes (sécante, quasi-Newton), ... etc.

4.1.2 La programmation quadratique avec contraintes

Un problème de programmation quadratique avec contraintes linéaires s'écrit sous la forme suivante

$$(PQ) \quad \begin{cases} \min & f(x) = \frac{1}{2}x^T Qx + c^T x, \\ & Ax = b, \\ & x \geq 0, \end{cases}$$

où Q est une matrice carrée symétrique d'ordre n , A est une matrice (m, n) , de plein rang, c est un vecteur de \mathbb{R}^n et $x \in \mathbb{R}^n$ est le vecteur inconnu à déterminer.

Remarque 4.1. *L'ensemble des contraintes $D = \{x \in \mathbb{R}^n, Ax = b, x \geq 0\}$ est un polyèdre convexe et fermé, (PQ) est convexe si et seulement si f est convexe.*

Il existe plusieurs méthodes pour la résolution des problèmes quadratiques avec contraintes linéaires. Parmi ces méthodes, on cite

1. La méthode du gradient projeté.
2. Le lagrangien augmenté.
3. La méthode d'UZAWA et les variantes de la méthode de Newton.
4. Les méthodes de points intérieurs.

4.2 Méthode du gradient projeté via les fonctions majorantes

L'idée de base consiste à projeter le gradient de la fonction objectif sur le noyau des contraintes. La méthode du gradient projeté est essentiellement intéressante dans le cas des contraintes linéaires, car dans ce cas, la projection est facile à calculer.

Le principe de la Méthode du gradient projeté

Soit le problème donné sous la forme

$$\begin{cases} \min f(x) = \frac{1}{2}x^T Qx + c^T x, \\ Ax = b, \\ x \geq 0. \end{cases}$$

Soit x un point admissible, on cherche une direction de déplacement d qui permette de diminuer suffisamment la fonction f (qui rend $(\nabla f(x)^T d)$ minimal), tout en restant dans l'ensemble des solutions, c'est à dire : $Ad = 0$.

Théorème 4.1. *Si A est une matrice $m \times n$ de plein rang $m \leq n$, la solution optimale du problème :*

$$\begin{cases} \min \nabla f(x)^T y, \\ Ay = 0. \end{cases}$$

où d est la projection de $-\nabla f(x)$ sur

$$S = \{y / Ay = 0\}.$$

donnée par

$$d = -P\nabla f(x) = -(I - A^T[AA^T]^{-1}A)\nabla f(x)$$

P est appelée la matrice de projection sur S .

Si $d \neq 0$, on déterminera le pas \bar{t} , et ce en utilisant la technique des fonctions majorantes, le \bar{t} appartient à l'intervalle $[0, t_{max}[$ avec $t_{max} = \min\{\frac{-x_i}{d_i}, d_i < 0\}$. L'itéré suivant est alors sélectionné comme un point minimisant $f(x + td)$ sur le segment $[0, t_{max}[$.

L'algorithme se poursuit donc de cette façon tant que $d \neq 0$.

Si $d = 0$, on a

$$d = -P\nabla f(x) = 0,$$

dans ce cas on pose

$$\mu = ([AA^T]^{-1}A)\nabla f(x)$$

on voit que si $\mu \geq 0$, la condition $d = 0$ s'écrit $\nabla f(x) + A^T\mu = 0$, ce qui exprime les conditions de Karush-Kuhn-Tucker au point x , qui est donc un optimum local du problème (si la fonction f est convexe localement). Si μ a des composantes strictement négatives, la recherche peut se poursuivre en enlevant de A une des contraintes correspondant à $\mu_i < 0$.

Première fonction majorante

On utilise la nouvelle fonction majorante pour le calcul du pas de déplacement comme on l'a défini dans le troisième chapitre θ_2

$$\theta_2(t) = n\delta t - (n-1)\ln(1+\alpha t) - \ln(1+\beta t), \quad (4.1)$$

où

$$\delta = \frac{f(x + \hat{t}d) - f(x)}{nr\hat{t}}, \quad \alpha = \bar{y} + \frac{\sigma_y}{\sqrt{n-1}}, \quad \beta = \bar{y} - \sigma_y\sqrt{n-1}.$$

L'expression de δ peut être donnée par

$$\begin{aligned} \delta &= \frac{1}{nr\hat{t}} \left[\frac{1}{2}(x + \hat{t}d)^T Q(x + \hat{t}d) + c^T(x + \hat{t}d) - \frac{1}{2}x^T Qx - c^T x \right] \\ &= \frac{1}{nr\hat{t}} \left[\frac{1}{2}\hat{t}x^T Qd + \frac{1}{2}\hat{t}d^T Qx + \frac{1}{2}\hat{t}^2 d^T Qd + \hat{t}c^T d \right] \\ &= \frac{1}{nr} \left[x^T Qd + \frac{1}{2}\hat{t}d^T Qd + c^T d \right] \\ &= \frac{1}{nr} \left[x^T Q + \frac{1}{2}\hat{t}d^T Q + c^T \right] d. \end{aligned}$$

Remplaçons dans (4.1) on aura

$$\theta_2(t) = \frac{t}{r} \left[x^T Q + \frac{1}{2}\hat{t}d^T Q + c^T \right] d - (n-1)\ln(1+\alpha t) - \ln(1+\beta t).$$

En remplaçant dans (3.8) par la valeurs de δ , et en résolvant l'équation (4.1), on trouve la valeur optimale \bar{t} .

Deuxième fonction majorante

Il est possible de définir, une nouvelle majoration plus simple par rapport à la première, et qui contient un seul logarithme.

L'idée consiste à utiliser l'inégalité

$$-\sum_{i=1}^n \ln(1 + y_i t) \leq (-\|y\| - n\bar{y})t - \ln(1 - \|y\|t).$$

On sait que

$$\theta(t) = \frac{1}{r}[f(x + td) - f(x)] - \sum_{i=1}^n \ln(1 + ty_i),$$

la relation (3.9), nous donne

$$\frac{f(x + td) - f(x)}{r} \leq \frac{f(x + \hat{t}d) - f(x)}{r\hat{t}}t,$$

donc

$$\theta(t) \leq (n\delta - \|y\| - n\bar{y})t - \ln(1 - \|y\|t). \quad (4.2)$$

avec $\delta = \frac{f(x + \hat{t}d) - f(x)}{nr\hat{t}}$.

D'où, enfin, on arrive à construire une deuxième fonction majorante

$$\theta_3(t) = (n\delta - \|y\| - n\bar{y})t - \ln(1 - \|y\|t).$$

θ_3 est définie sur $[0, \hat{t}_3[$ avec $\hat{t}_3 = \frac{1}{\|y\|}$,

et vérifie les propriétés suivantes

- ▷ $\theta_3(0) = \theta_2(0) = 0$ et $\theta'_3(0) = \theta'_2(0) = n(\delta - \bar{y})$,
- ▷ $\theta''_3(0) = \theta''(0) = \|\lambda\|^2 > 0$,
- ▷ $\theta(t) \leq \theta_3(t)$,
- ▷ θ_3 est strictement convexe.

On à θ_3 est strictement convexe, alors elle admet un minimum unique, que l'on trouve en résolvant l'équation $\theta'_3(t) = 0$ puis on obtient

$$\bar{t}_3 = \frac{1}{\|y\|} - \frac{1}{\|y\| + n\bar{y} - n\delta}.$$

Algorithme

Algorithme 4 : Algorithme du gradient projeté via la fonction majorante

Données : Q, c, A et $b, \epsilon \geq 0$ (un seuil de tolérance), $x_0 \in \mathbb{R}_+^n$

Résultat : x^*

Initialisation : $k = 0$.

1 répéter

2 x^k est le point courant.

3 répéter

4 Calculer la matrice de projection :

— $P^k = I - (A^k)^T [A^k (A^k)^T]^{-1} A^k$.

— $d^k = -P^k \nabla f(x^k)$,

— $\mu = -[A^k (A^k)^T]^{-1} A^k \nabla f(x^k)$,

— $\mu_i = \min\{\mu_j / j \in I\}$.

si $d^k = 0$ et $\mu_i < 0$ **alors**

— En supprimant la ligne d'indice i de A , et calculer la nouvelle matrice de projection correspondante.

sinon

— Le point courant x^k satisfait les conditions nécessaires de (KKT).

5 jusqu'à x^k satisfait les conditions nécessaires de (KKT);

— calculer \bar{t} dans l'intervalle $[0, t_{max}[$, par la technique des fonctions majorantes, où $t_{max} = \min_i \{ \frac{-x_i}{d_i}, d_i < 0 \}$,

— $x^{k+1} = x^k + \bar{t} d^k$,

— $k = k + 1$.

6 jusqu'à $\|d\| \leq \epsilon$;

7 $x^* = x^{k+1}$

Simulations numériques

On utilise les mêmes exemples du chapitre trois (on conserve le même ordre de ces exemples).

Dans les tableaux suivants, on note par

Méthode 1 : Calcul du pas de déplacement par la technique des fonctions majorantes.

M1 : Première fonction majorante.

M2 : Deuxième fonction majorante.

Méthode 2 : Calcul du pas de déplacement par la recherche linéaire classique (Wolfe)

TABLE 4.1: Tests numériques pour les exemples 3.1 et 3.2

	Méthode 1						Méthode 2		
	M1			M2			Objectif	itér	Temps(s)
Taille	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)			
1	-7.244	11	0.013	-7.244	11	0.018	-7.2197	5	0.032
2	-4.033	7	0.006	-4.033	7	0.008	-4.037	5	0.010

TABLE 4.2: Tests numériques de l'exemple 3.5

	Méthode 1						Méthode 2		
	M1			M2			Objectif	itér	Temps(s)
Taille	Objectif	itér	Temps(s)	Objectif	itér	Temps(s)			
4	0.285	8	0.009	0.285	8	0.007	0.285	16	0.021
50	5.372	8	0.017	5.372	8	0.016	5.372	27	0.073
100	10.928	8	0.048	10.929	8	0.162	10.928	27	0.189
500	55.372	8	2.558	55.373	8	3.048	55.372	27	9.649

Conclusion :

Les simulations numériques effectuées montrent que notre méthode basée sur la technique des fonctions majorantes pour le calcul du pas de déplacement, est plus efficace que celle basée sur la règle classique de Wolfe, et surtout pour les exemples de grande taille.

Conclusion générale

Dans cette thèse, nous avons abordé des questions ouvertes concernant l'optimisation non linéaire et quadratique, à savoir :

- L'étude de la convergence de notre variante en cas d'un problème d'optimisation non linéaire.
- Réduction du nombre d'itérations en utilisant les fonctions approximantes.
- L'idée de la sécante pour le calcul du pas de déplacement.
- Calcul de la direction, par la méthode du gradient projeté dans le cas d'un programme quadratique.

Les réponses à ces questions, sont encourageantes et prometteuses, où en premier, on a établi la convergence de la méthode de pénalité logarithmique. Et par suite, l'idée de la sécante nous a permis de passer facilement à l'application des fonctions approximantes pour le calcul du pas de déplacement. Les simulations numériques effectuées, montrent que la technique des fonctions approximantes est une alternative très fiable, et elle réduit d'une façon efficace le coût calculatoire dans les deux problèmes non linéaires. Comme il montrent aussi que l'utilisation de la méthode du gradient projeté est plus efficace que celle de Newton dans le cas d'un programme quadratique. Cette étude, est d'une grande importance, et incite les scientifiques à pousser à fond leurs recherches dans ce domaine [15].

Bibliographie

- [1] F. Alizadeh. Interior point methods in semidefinite programming with application to combinatorial optimization. *SIAM Journal on Optimization*, 1995.
- [2] Mokhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. *Nonlinear Programming, Theory and Algorithms*. Wiley Interscience, third edition, 2006.
- [3] J. Frédéric Bonnans, Claude Lemarechal, J. Charles Gilbert, and Claudia A. Sagastizábal. *Numerical Optimization, Theoretical and Practical Aspects*. Springer, second edition, 2006.
- [4] L. Bachir Cherif and B. Merikhi. A penalty method for nonlinear programming. *RAIRO-Oper. Res.*, 53(1) :29–38, 2019.
- [5] E. Chouzenoux. *Recherche de pas par Majoration-Minoration. Application à la résolution de problèmes inverses*. PhD thesis, Ecole Centrale de Nantes (ECN), 2010.
- [6] J.-P. Crouzeix D. Benterki and B. Merikhi. A numerical feasible interior point method for linear semidefinite programs. *RAIRO Oper. Res.*, 41, 2007.
- [7] J. P. Haberly F. Alizadeh and M. L. Overton. Primal-dual interior-point methods for semidefinite programming, convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 1998.
- [8] C. Lemaréchal J.-F. Bonnans, J.-C. Gilbert and C. Sagastizábal. Numerical optimization, theoretical and practical aspects. *Mathematics and Applications*, 27, 2003.
- [9] A. Seeger J.-P. Crouzeix. New bounds for the extreme values of a finite sample of real numbers. *Journal of Mathematical Analysis and Applications*, 197(29) :411–426, 1996.
- [10] B. Merikhi J.-P. Crouzeix. A logarithm barrier method for semidefinite programming. *RAIRO - Operations Research*, 42(2) :123–139, 2008.

- [11] G. Jia and J. Cao. A new upper bound of the logarithmic mean. *Inequalities in Pure and Applied Mathematics*, 4(4) :article 80, 2003.
- [12] A. Keraghel. *Étude adaptative et comparative des principales variantes dans l'algorithme de Karmarkar*. PhD thesis, Institut National Polytechnique de Grenoble, France, 1989.
- [13] M. Al khraboutly. *Résolution d'un problème quadratique non convexe avec contraintes mixtes par les techniques de l'optimisation D.C.* PhD thesis, Université Le Havre Normandie – France., 2018.
- [14] A. Leulmi. *Etude d'une méthode barrière logarithmique via les fonctions minorantes pour la programmation semidéfinie*. PhD thesis, Université de Biskra, 2018.
- [15] A. Yassine M. Bouafia, D. Benterki. An efficient primal-dual interior point method for linear programming problems based on a new kernel function with a trigonometric barrier term. *J. Optim. Theory Appl*, 170(2) :528–545, 2016.
- [16] R. Hestenes Magnus and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6) :409–436, December 1952.
- [17] B. Merikhi. *Extension de quelques méthodes de points intérieurs pour la programmation semi-définie, thèse de doctorat*. PhD thesis, Université de Sétif, Sétif, 2006.
- [18] M. Gupta R. Sharma, A. Kaura and S. Ram. Some bounds on sample parameters with refinements samuelson and brunk inequalities. *Linear Algebra and Application*, 3, 2009.
- [19] K. Samia. *Généralisation d'une méthode de trajectoire centrale de points intérieurs pour la programmation semidéfinie*. PhD thesis, Université de Sétif, 2015.
- [20] E. Shannon. A mathematical theory of communication. *Bell System Technical*, 27, 1948.
- [21] H. Wolkowicz and G.-P.-H. Styan. Bounds for eigenvalues using traces. *Linear Algebra and Application*, 29, 1980.

ملخص

في هذه الأطروحة نقدم في الجزء الأول منها طريقة اللوغاريتم الحدي لحل مسألة التجويد غير الخطية المجرأة حيث الاتجاه يحسب بطريقة نيوتن و تحسب الخطوة بتقنية الدالة الحدية العليا، بينما الجزء الثاني من الأطرحة يهتم بطريقة النقطة الداخلية لحل مسألة تربيعة حيث يحسب الاتجاه بطريقة الاسقاطي و تحتفظ بنفس التقنية لحساب الخطوة. نثري في هذه الدراسات الأخيرة بتجارب عددية مهمة.

الكلمات المفتاحية : الخوارزم الجزائي، الدوال الحدية العليا، الخوارزم الحدي، برنامج محذب، طريقة النقاط الداغلية.

Résumé

Dans cette thèse, nous présentons, dans une première partie, une méthode barrière logarithmique pour la résolution d'un problème d'optimisation non linéaire pénalisé, où la direction est calculée par la méthode de Newton, et le pas est calculé par la technique de fonctions majorantes. Dans la deuxième partie, on expose une méthode de points intérieurs pour la résolution d'un programme quadratique, dont la direction est calculée par gradient projeté, tout en maintenant la même technique pour le calcul du pas.

Cette étude est soutenue par des simulations numériques importantes.

Mots clés : Pénalité logarithmique, fonction majorante, barrière logarithmique, programme convexe, Méthode de points intérieurs.

Abstract

In this thesis, we present, in the first part, a logarithmic barrier method for the resolution of a penalized nonlinear optimization problem, where the direction is calculated by the Newtons method, and the setup is calculated by the technique of upper functions. In the second part, we will present an interior point method for the resolution of a quadratic program, whose direction is calculated by projected gradient, while maintaining the same technique for the calculation of the setup.

This study is supported by important numerical simulations.

Key words : Logarithmic Penalty, Upper-Function, Logarithmic Barrier, Convex Programming, Interior point Method.