Numerical methods for complementarity problems, and optimal control problems under complementarity constraints

El Hassene Osmani

THÈSE DE DOCTORAT DE

L'INSA RENNES

L'UNIVERSITÉ FERHAT ABBAS SÉTIF 1

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Mathématiques et leurs Interactions*

Par

El Hassene OSMANI

Numerical methods for complementarity problems, and optimal control problems under complementarity constraints

Thèse présentée et soutenue à l'UFA, Sétif 1 le 24/10/2022
Unité de recherche : IRMAR-UMR CNRS 6625-LMFN
Thèse N° : 22ISAR 23 / D22 - 23

Rapporteurs avant soutenance :

Didier AUSSEL	PR.	Université de Perpignan
Le Thi Hoai An	PR.	Université de Lorraine

Composition du Jury :

Président :	Bachir MERIKHI	PR.	Université Ferhat Abbas Sétif 1
Examinateurs :	Zakia KEBBICHE	PR.	Université Ferhat Abbas Sétif 1
	Carine LUCAS	MCF.	Université d'Orléans
	Aude RONDEPIERRE	MCF.	INSA de Toulouse
	Didier AUSSEL	PR.	Université de Perpignan
	Le Thi Hoai An	PR.	Université de Lorraine
Dir. de thèse :	Mounir HADDOU	PR.	INSA de Rennes
Co-dir. de thèse :	Naceurdine BENSALEM	PR.	Université Ferhat Abbas Sétif 1

Intitulé de la thèse :

Numerical methods for complementarity problems, and optimal control problems under complementarity constraints.

El Hassene OSMANI

En partenariat avec :

Remerciements

Cette thèse devient une réalité avec le soutien et l'aide de nombreuses personnes. Ces lignes me donnent l'occasion de les remercier.

Je tiens en premier lieu à remercier messieurs Mounir HADDOU et Naceurdine BENSALEM, mes directeurs de thèse. Il est difficile de résumer ces trois années en quelques mots, mais je vous remercie sincèrement pour votre écoute, votre patience, votre disponibilité, votre soutien et la confiance que vous m'avez accordée durant ces trois années. Merci de m'avoir encouragé, laissé expérimenter, et guidé. J'ai énormément progressé et appris grâce à vous et j'espère être un jour aussi expérimenté et clairvoyant que vous. Ce fut un réel plaisir de travailler avec vous et j'espère que cela pourra continuer.

Je souhaite remercier tout particulièrement Mr. Didier AUSSEL professeur à l'université de Perpignan et Mme. Le Thi Hoai An professeur à l'université de Lorraine eu la gentillesse de rapporter cette thèse. Je remercie Mr. B. MERIKHI professeur à l'université Ferhat Abbas Sétif 1, de m'avoir fait l'honneur de faire partie de ce jury et d'en être le président. Je désire remercier également : Mme. Z. KEBBICHE professeur à l'université Ferhat Abbas Sétif 1 et Mme. C. LUCAS maître de conférences HDR à l'université d'Orléans et Mme. A. RONDEPIERRE maître de conférences HDR à l'INSA de Toulouse d'avoir accepté de juger ce travail et d'en être les examinateurs.

Je remercie également mon comité de suivi de thèse : messieurs Marquis LUDOVIC et Abdeslam KADRANI d'avoir eu le temps de suivre ma thèse et de me donner des conseils précieux.

Je remercie la composante IRMAR-Insa qui m'a accueilli et m'a permis de réaliser les travaux présents dans ce manuscrit dans de bonnes conditions. Mes remerciements vont à tous les membres de la composante ainsi que son personnel administratif. Je voudrais adresser mes remerciements aussi à tous les membres du laboratoire de mathématiques fondamentales et numériques, et à toute l'équipe administrative du département de mathématiques à l'université Ferhat Abbas Sétif 1.

Une partie des résultats obtenus au cours de la recherche a été motivée par des discussions avec d'autres chercheurs autres mes encadrants dont Mme. Lina ABDALLAH que je remercie pour son soutien et les discussions qu'on a eues ensemble. De même je tiens à remercier Mr. Djamel BENTERKI professeur à l'université Ferhat Abbas Sétif 1, pour sa gentillesse et ses précieux conseils pendant toutes mes années de recherche.

Merci à toutes les personnes que j'ai eues la chance de rencontrer dans le cadre de ma thèse pour les discussions et les conseils reçus.

Je remercie les doctorants que j'ai rencontrés et avec lesquels nous avons partagé des moments utiles et agréables. Un grand merci notamment à Larbi, Ahmed, Bilel, Essoham, Huan, Mériadec et Trinh pour toutes nos discussions qui, je l'espère, continueront.

Pour finir, je souhaite remercier de tout cœur ma grande famille et en particulier mes parents pour leur soutien et leur confiance depuis toujours. Merci à mon frère El Housseine et mes soeurs Khawla et Safia. Je me sens extrêmement chanceux d'être entouré d'une famille aussi merveilleuse; je vous aime infiniment.

À mes parents,

ceux qui m'ont fait naître dans ce monde,
et ceux qui m'ont aidé à y grandir.

Abstract

Complementarity problems occur in many scientific fields: economics, physics, transport, game theory, and mathematics.

In this thesis, we offer several theoretical, algorithmic, and numerical contributions to solve the complementarity problems and optimal control problems under complementarity constraints. We are particularly interested in the regularization methods for the numerical resolution of these types of problems, we have proposed new regularization techniques.

Indeed, In the first part, we focused on optimal control problems under complementarity constraints. We studied optimal control problems governed by semilinear elliptic variational inequalities involving constraints on the state. We presented a new regularisation schema for the complementarity constraints. We proved that Lagrange multipliers exist.

Then, in the second part, we have studied linear complementarity problems (LCPs) and nonlinear complementarity problems (NCPs) by proposing new methods of regularisation to solve these kind of problems. The idea of these methods takes inspiration from interior point methods.

Throughout this manuscript, we have focused on the theoretical properties of algorithms and their digital applications.

Key words: Interior points methods, Linear complementarity problem, Nonlinear complementarity problem, Optimal control, Regularization methods, θ-function, Newton's method, Semismooth analysis, Global convergence, Local convergence.

Résumé

Les problèmes de complémentarité interviennent dans de nombreux domaines scientifiques : économie, physique, transport, théorie des jeux et mathématiques.

Dans cette thèse, on apporte plusieurs contributions théoriques, algorithmiques et numériques pour résoudre des problèmes de complémentarité et de contrôle optimal sous contraintes de complémentarité. On s'intéresse plus particulièrement aux méthodes de régularisation pour la résolution numérique de ces deux types de problèmes, où nous avons proposé de nouvelles techniques de régularisation.

En effet, dans la première partie, nous nous sommes intéressés aux problèmes de contrôle optimal sous contraintes de complémentarité. Nous avons étudié les problèmes de contrôle optimal régis par les inégalités variationnelles elliptiques semi-linéaires impliquant des contraintes sur la variable d'état. Nous avons présenté un nouveau schéma de régularisation pour la contrainte de complémentarité. Nous avons prouvé l'existence de multiplicateurs de Lagrange.

Ensuite, dans la deuxième partie, nous avons étudié les problèmes de complémentarité linéaire et non linéaire en proposant de nouvelles méthodes de régularisation pour résoudre ce genre de problèmes. L'idée de ces méthodes prend inspiration de la méthode des points intérieurs.

Dans ce travail nous nous sommes concentrés sur les propriétés théoriques des algorithmes et leurs applications numériques.

Mots clés : Méthode de point intérieur, Problème de complémentarité linéaire, Problème de complémentarité non linéaire, Contrôle optimal, Méthodes de régularisation, θ-fonction, Méthode de Newton, Analyse semi-lisse, Convergence globale, Convergence locale.

Notations

We consider here classical notations:

$\mathbb{R}^n$	:	The n-dimensional real Euclidian vector space
$\mathbb{R}^n_+$	:	The nonnegative orthant of $\mathbb{R}^n$
$\mathbb{R}^n_{++}$	:	The positive orthant of $\mathbb{R}^n$
$\mathbb{R}^{n\times n}$	:	The set of all $n \times n$ squared real matrices
A	=	(a_{ij}), a matrix with entries a_{ij}
$\det A = \|A\|$	:	The determinant of a matrix A
$\mathrm{tr}A$	=	$\sum_{i=1}^n a_{ii}$, the trace of a matrix A
A^T	:	The transpose of a matrix A
A^{-1}	:	The inverse of a matrix A
$A_{.\alpha}$	:	The columns of A indexed by α
$A_{\alpha.}$	:	The rows of A indexed by α
$A_{\alpha\beta}$	:	Submatrix of A with rows and columns indexed by α and β, respectively
I_k	:	Identity matrix of order k
$\mathrm{diag}(a)$	:	The diagonal matrix with diagonal elements equal to the components of the vector a
x_i	:	The i-th component of x
x^T	:	The transpose of vector x
x^+	:	The nonnegative part of a vector, $x^+ = \max(x, 0)$
x^-	:	The nonpositive part of a vector, $x^- = \max(-x, 0)$
x^{-1}	=	$\left(\frac{1}{x_1}, ..., \frac{1}{x_n}\right)^T \quad x_i \neq 0$ for all $i = 1, ..., n$
$\mathbf{e}$	:	The n-dimensionzl vector of ones, $\mathbf{e} = (1,\ ...,\ 1)^T$
$x.y$	:	The Hadamart product of x and y, $x.y = (x_1y_1,\ ...,\ x_ny_n)^T$
$\frac{x}{y}$	=	$\left(\frac{x_1}{y_1}, ..., \frac{x_n}{y_n}\right)^T,\quad y_i \neq 0$ for all $i = 1, ..., n$
$\log(x)$	=	$(\log(x_1), ..., \log(x_n))^T,\quad x_i > 0$ for all $i = 1, ..., n$
e^x	=	$(e^{x_1}, ..., e^{x_n})^T$
$< x,\ y >$	:	The standard inner product of vector in $\mathbb{R}^n$, $< x,\ y >= x^T y$
$x \perp y$	:	x and y are perpendicular
$\|x\|_p$	:	The l_p-norm of a vector $x \in \mathbb{R}^n$, $\|x\|_p = \left(\sum_{i=1}^n \|x_i\|^p\right)^{1/p}$

$\|x\|$	:	The l_2-norm of a vector $x \in \mathbb{R}^n$, unless otherwise specified
$\|x\|$	=	$(\|x_1\|, ... \|x_n\|)^T$, the componentwise absolute value of a vector $x \in \mathbb{R}^n$
$\|x\|_\infty$	:	The l_∞-norm of a vector $x \in \mathbb{R}^n$, $\|x\|_\infty = \max_{1 \leqslant i \leqslant n} \|x_i\|$
$x \geqslant y$	:	The (usual) partial ordering, $x_i \geqslant y_i, \quad i = 1, ..., n$
$x > y$	:	The strict ordering, $x_i > y_i, \quad i = 1, ..., n$
$sign(x)$	:	Denotes a vector with the components equal to $-1,\ 0$ or 1
$\min(x, y)$	:	The vector whose i-th component is $\min(x_i, y_i)$
$\max(x, y)$	:	The vector whose i-th component is $\max(x_i, y_i)$
$\lambda_i(A)$	:	The eigenvalues of $A \in \mathbb{R}^{n \times n},\ i = 1, ..., n$
$\lambda_{\max}(A)$	:	The largest eigenvalues of $A \in \mathbb{R}^{n \times n},\ i = 1, ..., n$
$\lambda_{\min}(A)$	:	The smallest eigenvalues of $A \in \mathbb{R}^{n \times n},\ i = 1, ..., n$
$\sigma_{\min}(A)$	:	The largest singular value of $A \in \mathbb{R}^{n \times n},\ i = 1, ..., n$
$\sigma_{\min}(A)$	:	The smallest singular value of $A \in \mathbb{R}^{n \times n},\ i = 1, ..., n$
$\rho(A)$	=	$\max(\lambda_i(A))$, the spectral radius of A
$\|A\|_2$	=	$\sqrt{\rho(A^T A)}$, the spectral norm of A
∇u	=	$\left(\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n}\right)$, the gradient of a scalar function u
$C^\infty(\Omega)$	=	$\left(u : \mathbb{R} \to \mathbb{R} \mid u^{(n)} \text{ exists for all } n\right)$, is the space of all functions that are smooth (infinitely continuously differentiabl) on Ω
$C_0^\infty(\Omega)$	:	The space of real valued smooth functions with compact support in Ω
$L^p(\Omega),\ 1 \leqslant p \leqslant \infty$	:	The space of Lebesgue-mesurable functions with finite norm defined as follows $L^p(\Omega) = \left(u : \Omega \to \mathbb{R},\ u \text{ is measurable and } \int_\Omega \|u\|^p dx < \infty\right)$ with the norm $\|u\|_{L^p} = \left(\int_\Omega \|u\|^p dx\right)^{\frac{1}{p}}$
$W^{m,p}(\Omega)$	:	The Sobolev space of $u \in L^p(\Omega)$ functions where $D^k u$ is also in $L^p(\Omega)$ for $k \leqslant m$, with the norm $\|u\|_{W^{m,p}} = \sum_{0 \leqslant k \leqslant m} \|D^k u\|_{L^p}$
$W_0^{m,p}(\Omega)$	:	The completion of $C_0^\infty(\Omega)$ in the norm $\|.\|_{W^{p,m}}$
$H_0^1(\Omega)$	:	The Sobolev space $W_0^{m,p}(\Omega)$, by taking $m = 1$ and $p = 2$, and denote the corresponding norm as follows $\|u\| = \left(\int_\Omega \|\nabla u\|^2 dx\right)^{1/2}$

Acronyms

NLP	Nonlinear program
MPCC	Mathematical program with complementarity constraints
VI	Varitional inequalities
CP	Complementarity problems
LCP	Linear pomplementarity problems
NCP	Nonlinear complementarity problems
MiCP	The mixed complementarity problems
AVE	Absolute value equation
KKT	Karush-Kuhn-Tucker
SQO	Sequential quadratic programming
C-function	Complementarity function
IPOPT	The Interior Point Optimizer
KNITRO	Nonlinear Interior point Trust Region Optimization
SNOPT	Sequential Quadratic Optimization Technique

List of Figures

List of Tables

Contents

1 Introduction

L'optimisation est une branche des mathématiques et de l'informatique en tant que disciplines. Elle intervient pratiquement dans tous les processus de modélisation actuels et elle joue un rôle très important dans beaucoup de domaines. Qu'il s'agisse de problèmes de la recherche opérationnelle, de mathématiques appliquées, d'analyse, d'analyse numérique, de statistiques, de théorie des jeux, de programmation linéaire ou encore en théorie du contrôle.

Le problème d'optimisation consiste à déterminer une solution qui maximise ou minimise l'objectif quantitatif tout en respectant éventuellement certaines contraintes. Les problèmes d'optimisation sont très divers par leurs natures et leurs structures, alors chaque type de ces problèmes sera résolu d'une manière différente [50, 103, 108]. Les problèmes d'optimisation sont classés selon leurs fonctions objectifs et leurs contraintes : optimisation linéaire, optimisation non-linéaire, optimisation linéaire quadratique et optimisation convexe...etc.

En optimisation, le problème de complémentarité en dimension finie consiste à résoudre un système fini d'inéquations tout en respectant une équation particulière qui exprime la complémentarité entre les composantes. C'est cette caractéristique importante qui distingue le problème de complémentarité du système d'inéquations traditionnel. Dans ce cas, le problème de complémentarité consiste à trouver $x \in \mathbb{R}^n$ qui satisfait la condition suivante :

$$0 \leqslant G(x) \perp H(x) \geqslant 0, \tag{CP}$$

où la notation $\perp$ signifie perpendiculaire, $G, H : K \to \mathbb{R}^n$ deux fonction et K un cône, c'est-à-dire que si $x \in K$ alors $\tau x \in K$ pour tout $\tau \geqslant 0$. La condition de complémentarité peut ainsi être réécrite comme $G_i(x)H_i(x) = 0$ pour tout $i = 1, ..., n$.

Bien que le problème de complémentarité ne soit pas un problème d'optimisation mais simplement un problème de réalisabilité et est d'un grand intérêt pour l'optimisation. En effet, les conditions nécessaires d'optimalité de nombreux problèmes d'optimisation peuvent être représentées sous la forme (CP).

L'intérêt d'étudier ce type de problème a commencé en 1964 lorsqu'il a été introduit par Richard W. Cottle dans sa thèse de doctorat puisque les applications sont nombreuses et dans plusieurs domaines différents.

Tout d'abord, les problèmes de complémentarité sont apparus dans les conditions d'optimalité de Karush-Kuhn-Tucker [64, 70] mais peuvent aussi servir à modéliser certains phénomènes décrits par des systèmes d'équations qui sont en quelque sorte en compétition.
Quelques exemples d'applications sont les problèmes d'équilibre économique [42], les jeux bimatriciels [71], le problème d'équilibre du trafic de Wardrop [46], les problèmes d'écoulement diphasiques [21, 24, 25, 49] et les simulations de contacts et de mouvements de fluides [37]. Une raison importante pour laquelle les problèmes de complémentarité sont si répandus dans l'ingénierie et l'économie est que la notion de complémentarité est synonyme d'équilibre du système étudié. L'équilibre de l'offre et la demande est au centre de tous les systèmes économiques. La complémentarité est également au cœur des problèmes d'optimisation avec contraintes.

Au fil des années, le sujet est devenu une discipline proprement dite des mathématiques. La littérature des problèmes de complémentarité a bénéficié des contributions apportées par les mathématiciens (pure, appliquée et informatique), les informaticiens et les différents ingénieurs (génie civil, électrique, mécanique et systèmes). Plusieurs livres et plus d'un millier d'articles concernant ce sujet ont été publiés [34, 39, 42, 83]. Beaucoup de résultats théoriques de base pour les problèmes de complémentarité sont connus depuis longtemps; une excellente étude dans ce domaine peut-être trouvée dans [56]. Autres références et des travaux plus récents peuvent également être trouvés dans [42, 89].

Les difficultés majeures pour résoudre le problème de complémentarité (CP) viennent de deux aspects essentiellement géométriques. D'une part, l'ensemble des solutions de ce problème n'est en général pas convexe et pas connexe. D'autre part, l'intérieur relatif de l'ensemble des solutions est vide, c'est-à-dire qu'il n'existe pas de x^* solution de (CP) tel que $G(x^*) > 0, H(x^*) > 0$.

Diverses méthodes numériques existent pour résoudre ce problème. Parmi celles-ci on peut citer les méthodes de reformulation qui transforment (CP) comme un système d'équations sans contraintes ou encore les méthodes d'activation de contraintes qui utilisent une procédure combinatoire pour déterminer les contraintes actives. Au vu des difficultés géométriques énoncées plus haut, une approche naturelle est d'utiliser des techniques de relâchement, autrement appelées techniques de régularisation. Ces techniques relâchent les contraintes du problème pour le rendre plus simple puis tentent de se rapprocher du problème initial. Ce processus mène bien souvent à des méthodes itératives. Ce sont ces méthodes qui sont au coeur de ce manuscrit. Parmi les méthodes de régularisation les plus connues, on peut citer les méthodes de point-intérieur et les méthodes de pénalisations ou de fonctions de mérites. Ces dernières transforment le problème de complémentarité (CP) comme un problème d'optimisation avec une fonction objectif qui incite à faire respecter les contraintes du problème de complémentarité. La méthode des points-intérieurs peut aussi être interprétée comme une reformulation avec une pénalité logarithmique.

Une généralisation naturelle du problème (CP) est de considérer la résolution d'un problème d'optimisation avec un problème de complémentarité inclus dans les contraintes. On appelle problème d'optimisation sous contrainte de complémentarité le problème qui consiste à minimiser une fonction $f : \mathbb{R}^n \to \mathbb{R}$ telle que

$$\begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.à } g(x) \leqslant 0,\ h(x) = 0, \\ 0 \leqslant G(x) \perp H(x) \geqslant 0, \end{cases} \tag{MPCC}$$

pour des fonctions de contraintes $g, h, G, H : \mathbb{R}^n \to \mathbb{R}^n$. De nombreuses applications utilisent le problème (MPCC) par exemple en contrôle optimal, en physique ou encore en recherche opérationnelle.

Cette thèse comporte 6 chapitres. Tout au long de ce document, nous nous intéresserons aux techniques de régularisation pour les problèmes de complémentarité et de contrôle optimal sous contraintes de complémentarité. Ces techniques de régularisation ont notamment permis de développer différentes méthodes qui seront abordées dans chacun des chapitres qui composent ce manuscrit. Nous résumons le contenu des différents chapitres ainsi que les résultats obtenus.

Dans le chapitre 2, nous introduisons les outils mathématiques qui seront nécessaires dans ce travail de thèse. D'abord, nous présentons quelques rappels essentiels sur l'optimisation et les problèmes de complémentarité, à savoir : l'analyse convexe, quelques résultats de programmation mathématique et un rappel de certains aspects de cône. Puis, nous énonçons des définitions et les propriétés de quelques classe de matrices qui interviennent de manière essentielle dans ce manuscrit. Ensuite, nous introduisons les notations habituelles et générales de l'analyse et de l'optimisation non lisse et semi-lisse. Enfin, nous présentons quelques techniques de régularisation introduite dans [4, 52] pour les problèmes de complémentarité et établissent différentes propriétés qui seront utiles pour notre thèse. Ce chapitre ne présente aucune contribution théorique, cependant nous avons donné des résultats que nous utilisons dans la suite. Les trois autres chapitres contiennent nos contributions. Nous résumons ci-dessous le contenu de chacun.

Nous terminerons enfin par une synthèse de différents apports et contributions de cette thèse, et les perspectives qui peuvent s'en dégager.

Partie I : Problème de contrôle optimal sous contraintes de complémentarité

Cette première partie se concentre sur l'étude de problème de contrôle optimal sous contraintes de complémentarité.

Chapitre 3 : Une nouvelle méthode de régularisation pour les problèmes de contrôle optimal avec des obstacles régis par des inégalités variationnelles elliptiques semi-linéaires

Dans le chapitre 3, qui est notre première contribution de cette thèse, nous nous sommes intéressés aux problèmes de contrôle optimal régis par les inégalités variationnelles elliptiques semi-linéaires impliquant des contraintes sur la variable d'état. Nous avons présenté un nouveau schéma de régularisation pour la contrainte de complémentarité. Nous allons écrire notre problème sous la forme :

$$\min \left\{ J(y,v) = \frac{1}{2}\int_\Omega (y - z_d)^2 dx + \frac{\nu}{2}\int_\Omega (v - v_d)^2 dx \right\}, \qquad (\mathcal{P})$$

$$Ay + g(y) = f + v + \xi \ \text{ dans } \Omega, \quad y = 0 \quad \text{sur} \quad \partial\Omega,$$

$$(y, v, \xi) \in \mathcal{D},$$

où

$$\mathcal{D} = \{(y, v, \xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega) \mid v \in U_{ad},\ y \geqslant 0,\ \xi \geqslant 0,\ (y, \xi)_2 = 0\}.$$

En effet, la difficulté vient du fait que l'ensemble réalisable $\mathcal{D}$ est non convexe puisqu'on a une contrainte de complémentarité et donc on ne peut pas utiliser directement les méthodes d'analyse convexe ni même les outils de programmation non linéaire non convexes.
Pour écrire des conditions d'optimalité, nous avons utilisé des méthodes adaptées et plus générales, malheureusement, l'ensemble des contraintes est d'intérieur vide vu la contrainte de complémentarité. Pour pallier ce problème, nous avons relaxé la contrainte de complémentarité.

$$(\mathcal{P}^\alpha) \quad \begin{cases} \min\ J(y,v) \\ Ay + g(y) = f + v + \xi \text{ dans } \Omega,\ y \in H_0^1(\Omega), \\ (y, v, \xi) \in \mathcal{D}_\alpha \end{cases}$$

avec

$$\mathcal{D}_\alpha = \Big\{ (y, v, \xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega) \ / \\ v \in U_{ad},\ y \geqslant 0,\ \xi \geqslant 0,\ \frac{y}{y+\alpha} + \frac{\xi}{\xi+\alpha} \leqslant 1,\ a.e.\ in\ \Omega \Big\}.$$

Nous avons prouvé l'existence de multiplicateurs de Lagrange. Leur existence est un outil important pour résoudre les problèmes relaxés. Pour l'implémentation de l'algorithme, nous avons opté pour l'environnement et le langage AMPL et sur des méthodes de points intérieurs.

Contributions du chapitre :

→ Nous avons proposé un nouveau schéma de régularisation pour la contrainte de complémentarité $y \geqslant 0,\ \xi \geqslant 0, < y, \xi >= 0$.

→ Nous avons montré que notre problème relaxé est une bonne approximation du problème initial.

→ Nous avons démontré un théorème qui consiste à prouver l'existence de multiplicateurs de Lagrange de notre problème.

→ Le problème relaxé proposé a également été étudié numériquement au moyen de simulations avec AMPL.

Partie III : Problème de complémentarité

Cette deuxième partie se concentre sur de nouvelles méthodes numériques pour résoudre les problèmes de complémentarité linéaire et non linéaire.

Chapitre 4 : Une approche lisse pour résoudre un problème de complémentarité non linéaire

La deuxième contribution de cette thèse est décrite dans le chapitre 4. Nous nous intéressons désormais à la résolution des problèmes de complémentarité non linéaire (NCP), autrement dit on cherche $x \in \mathbb{R}^n$ qui satisfait l'équation non linéaire suivante:

$$x \geqslant 0, \quad F(x) \geqslant 0, \quad x^T F(x) = 0. \tag{NCP}$$

où $F : \mathbb{R}^n \to \mathbb{R}^n$.
Nous reformulons notre problème NCP, d'une façon équivalente, en un système d'équations lisses.

Comme une première étape, nous écrivons

$$x_i^{(r)} \geqslant 0, \quad F_i(x^{(r)}) \geqslant 0 \qquad \text{et} \qquad \theta_r(x_i^{(r)}) + \theta_r(F_i(x^{(r)})) = 1, \quad i = 1, ..., n$$

où $\theta_r : \mathbb{R} \to]-\infty, 1]$ une famille de fonctions qui satisfait les propriétés suivantes :

1. θ_r deux fois continument différentiable;
2. $\theta_r(0) = 0$;
3. θ_r est une fonction croissante et concave;
4. θ_r est négative sur $\mathbb{R}_-$;
5. $\lim_{r \to 0} \theta_r(x) = 1 \quad \forall x > 0.$

Nous donnons ici quelques exemples de fonctions θ_r sur $\mathbb{R}$:

- $\theta_r^1(x) = \dfrac{x}{x + r}$;
- $\theta_r^2(x) = 1 - \exp(-x/r)$;
- $\theta_r^{\log}(x) = \dfrac{\log(1 + x)}{1 + x + r}$.

En utilisant les fonctions $\psi_r = 1 - \theta_r$ avec $\psi_r : \mathbb{R} \to]0, +\infty[$, nous allons ainsi résoudre le système lisse suivant

$$G_r(x, F(x)) = 0,$$

où

$$G_r(x, y) := (G_r(x_i, y_i))_{i=1,...,n} := \left(r\psi^{-1}\left[\psi\left(\frac{x_i}{r}\right) + \psi\left(\frac{y_i}{r}\right) \right] \right)_{i=1,...,n}, \quad \forall x, y \in \mathbb{R}^n, \text{ et } r > 0.$$

Nous avons étudié les problèmes de complémentarité non linéaires (NCP) en proposant une nouvelle méthode pour résoudre ce genre de problème. L'idée de cette méthode prend inspiration de la méthode des points intérieurs en créant de nouvelles techniques de régularisation. La différence majeure de nos méthodes est le fait qu'on a besoin d'aucun processus pour mettre à jour le paramètre de régularisation r qu'on considère comme une nouvelle variable d'où NCP est équivalent au problème suivant :

$$\mathbb{H}_\theta(\mathbb{X}) = \begin{bmatrix} G_r(x, F(x)) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|F^-(x)\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0,$$

où

$$\|x^-\|^2 = \sum_{i=1}^{n} \min^2(x_i, 0), \quad \text{et} \quad \|F^-(x)\|^2 = \sum_{i=1}^{n} \min^2(F_i(x), 0).$$

Contributions du chapitre :

→ Nous avons proposé une nouvelle méthode pour résoudre les NCP basé sur une technique de régularisation.

→ Nous avons présenté un algorithme non paramétrique pour résoudre les NCP.

→ Nous avons montré que notre problème relaxé est une bonne approximation du problème initial.

→ Nous avons prouvé que tout point limite d'une suite $\{x_k\}$ générée par notre algorithme correspond à une solution de NCP.

→ Nous avons montré que notre matrice jacobienne est inversible si et seulement si la fonction F est $\mathcal{P}_0$-fonction.

→ Nous avons prouvé la convergence locale et globale.

→ Nous avons montré que la nouvelle matrice jacobienne associée à notre approche est inversible si et seulement si la matrice jacobienne associée au problème de point intérieurs est inversible.

Chapitre 5 : Nouvelles méthodes lisse pour résoudre les problèmes de complémentarité linéaire

Dans un second travail dans la partie III, nous avons étudié le problème de complémentarité linéaire (LCP) en proposant deux nouvelles méthodes pour résoudre ce genre de problèmes. Ce problème consiste à trouver $x \in \mathbb{R}^n$ tel que

$$0 \leqslant (Mx + q) \perp x \geqslant 0, \tag{LCP}$$

pour une matrice M d'ordre n et un vecteur $q \in \mathbb{R}^n$. L'idée de ces deux méthodes prend inspiration de la méthode de points intérieurs en créant de nouvelles techniques de régularisation de la condition de complémentarité. Il est clair que ce problème n'est pas simple, car la condition de complémentarité n'est pas différentielle. Pour pallier cette difficulté, nous introduisons une famille de fonctions θ voir [53]. En utilisant cette classe de fonctions, LCP est régularisé pour $r > 0$ comme

$$y = (Mx + q) \geqslant 0, \quad x \geqslant 0, \quad \theta_r(x_i) + \theta_r(z_i) \leqslant 1, \quad i = 1, ..., n.$$

D'après les propriétés des fonctions θ_r, lorsque r tend vers 0, ce problème régularisé devrait être équivalent à LCP.
Nous introduisons aussi une notre formulation pour notre problème LCP. On remarque que

$$0 \leqslant x \perp z \geqslant 0 \iff \forall \rho > 0, \quad x - max(0, x - \rho z) = 0,$$

où la fonction max est une fonciton de $\mathbb{R}^n$ dans $\mathbb{R}^n$ composante par composante, c'est-à-dire que pour un vecteur $x \in \mathbb{R}^n$ on a $\max(x) = (max(x_i))_{i=1,...,n}$. Nous avons approximé la fonction max par une fonction différentiable d'où le problème LCP est régularisé pour $r > 0$ comme

$$y = (Mx + q) \geqslant 0, \quad x \geqslant 0, \quad x_i - r \log\left(1 + e^{\frac{x_i - \rho z_i}{r}}\right) = 0, \quad i = 1, ..., n.$$

La différence majeure de nos méthodes est le fait qu'on a besoin d'aucun processus pour mettre à jour le paramètre de régularisation r qu'on considère comme une nouvelle variable d'où LCP est équivalent aux deux problèmes suivants :

$$\mathbb{F}_\theta(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ r\left(\theta_r(x) + \theta_r(z) - \mathbf{e}\right) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0,$$

et

$$\mathbb{F}_s(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ \left(x - r \log\left(\mathbf{e} + e^{\frac{x - \rho z}{r}}\right)\right) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0,$$

où les fonctions $\theta_r, \log$ et e sont des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}^n$ composante par composante, c'est-à-dire que pour un vecteur $x \in \mathbb{R}^n$ on a

$$\theta_r(x) = (\theta_r(x_i))_{i=1,...,n}, \ \log(x) = (\log(x_i))_{i=1,...,n} \quad \text{et} \quad e^x = (e^{x_i})_{i=1,...,n}.$$

Nous nous intéressons aussi à la résolution d'équation en valeur absolue (AVE), autrement dit on cherche $x \in \mathbb{R}^n$ qui satisfait l'équation non linéaire suivante :

$$Ax - |x| = b. \tag{AVE}$$

En utilisant une décomposition de la valeur absolue, on se ramène facilement à un problème de complémentarité. Soit $x^+ = \max(x, 0)$ et $x^- = \max(-x, 0)$, il vient que $x = x^+ - x^-$ et que $|x| = x^+ + x^-$ pour tout $x \in \mathbb{R}^n$ si x^+ et x^- sont orthogonaux d'où AVE est équivalent au problème de réalisabilité suivant

$$A(x^+ - x^-) - (x^+ + x^-) = b, \quad 0 \leqslant x^+ \perp x^- \geqslant 0,$$

c'est un problème des valeurs absolues (AVE) qui a été reformulé vers le problème LCP.

Contributions du chapitre :

→ Nous avons proposé deux nouvelles méthodes pour résoudre les LCP en se basant sur deux techniques de régularisation.

→ Nous avons présenté un algorithme non paramétrique pour résoudre les LCP

→ Nous avons montré que notre problème relaxé est une bonne approximation du problème initial.

→ Nous avons prouvé que tout point limite d'une suite $\{x_k\}$ générée par notre algorithme correspond à une solution de LCP.

→ Nous avons montré que notre matrice jacobienne est inversible si et seulement si la fonction F est $\mathcal{P}_0$-fonction.

→ Nous avons prouvé la convergence locale et globale.

→ Nous avons montré que les deux nouvelles matrices jacobiennes associées à notre approche sont inversibles si et seulement si la matrice jacobienne associée au problème de point intérieurs est inversible.

Nos travaux de thèse ont donné lieu à des articles et conférences dont :

Articles

- E. H. Osmani, M. Haddou and N. Bensalem. A new relaxation method for optimal control of semilinear elliptic variational inequalities obstacle problems. *Numerical Algebra, Control & Optimization*, **2021**, (DOI:10.3934/naco.2021061).
- E. H. Osmani, M. Haddou, L. Abdallah and N. Bensalem (2021). New smoothing methods for solving the linear complementarity problem with $\mathcal{P}_0$-matrix, *article soumis.* Disponible sous archives-ouvertes.fr/hal-03516404.
- E. H. Osmani, M. Haddou, N. Bensalem and L. Abdallah. A new smoothing method for nonlinear complementarity problems involving $\mathcal{P}_0$-function. Statistics, Optimization & Information Computing, **2022**, (DOI:10.19139/soic-2310-5070-1493).

Conférences

- E. H. Osmani, M. Haddou and N. Bensalem. A smooth approach to the solution of nonlinear complementarity problems involving $\mathcal{P}_0$-function. *8th International Conference on Optimization and Applications (ICOA)*. Sestri Levante, Italy, **06-07 octobre 2022, Publisher: IEEE**, (DOI:10.1109/ICOA55659.2022.9934180).
- E. H. Osmani, M. Haddou and L. Abdallah. A new approach for solving the linear complementarity problem using smoothing functions. *7th International Conference on Optimization and Applications (ICOA)*. Wolfenbüttel, Germany, **31 mai 2021, Publisher: IEEE**, (DOI:10.1109/ICOA51614.2021.9442649).
- E. H. Osmani, M. Haddou and N. Bensalem. Solving optimal control of semilinear elliptic variational inequalities obstacle problems usingsSmoothing functions. *ICDDPOC: XV. International Conference on Deterministic Dynamic Programming and Optimal Control*. Vienna, Austria, **29-30 juillet 2021, Vol. 15, No. 7**, (2888-2415-2759-2981).
- E. H. Osmani, N. Bensalem. Les méthodes directes en contrôle optimal et transfert d'un problème du contrôle optimal en problème d'optimisation non linéaire. *Un seminaire 'Optimisation Combinatoire et Continué: Méthodes et Applications'*. Sétif, Algérie, **2-3 décembre 2018**.

2 Mathematical background

In this chapter, we present basic results from convex analysis, variational analysis, and nonlinear programming that are used in the following chapters. Among the major references that have been used while studying these topics, we may cite some important books such as [96] for convex analysis, [97] for variational analysis, [40] for variational inequalities and complementarity problems, and finally [5] for various subjects on optimization.

Contents

2.1 Convex analysis and cones

In this section, we introduce elementary notions about convex sets, convex functions and cones.

2.1.1 Convex sets

Definition 2.1.1. *A subset C of $\mathbb{R}^n$ is convex if $\forall x,\ y \in C,\ \forall t \in [0,1]$*

$$tx + (1-t)y \in C.$$

A geometrical interpretation of this definition is that a set is convex if any segment joining two points of this set also belongs to this set as illustrated on Figure 2.1.

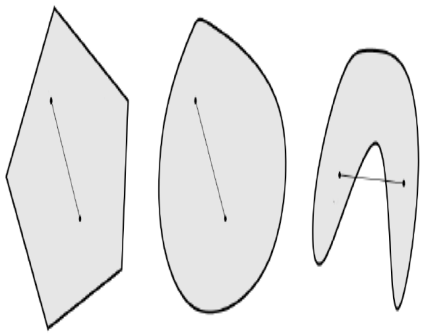

Figure 2.1: Some examples of two convex sets and a non-convex set.

Some elementary operations that preserve the convexity of a set are collected in the following proposition

Proposition 2.1.2. *For any collection $\{C_i | i \in I\}$ of convex sets $C_i \subset \mathbb{R}^{n_i}$ we have:*

1. *$C_1 \times ... \times C_m$ is convex in $\mathbb{R}^{n_1} \times ... \times \mathbb{R}^{n_m}$;*
2. *$\cap_{i \in I} C_i$ is convex, here with $n_i = n$ for all i;*
3. *The finite sum $\sum_{i=1}^{m}$ is convex, with $n_i = n$ for all i.*

For a set $C \subset \mathbb{R}^n$, the convex hull of C, denoted by *conv* C, is the intersection of all convex sets containing C.

Proposition 2.1.3. *For a set $C \subset \mathbb{R}^n$, it holds that conv C is the set of all convex combinations of elements of C, i.e.,*

$$conv\ C = \left\{ \sum_{i=1}^{m} t_i x_i \mid x_i \in C, t_i \geqslant 0,\ \sum_{i=1}^{m} = 1 \right\}.$$

2.1.2 Cones and polyhedral sets

Cones are fundamental geometric objects associated with sets. They play a key role in several aspects of mathematics.

Definition 2.1.4. *A set $K \subset \mathbb{R}^n$ is called a cone if $tx \in K$ for all $x \in K$ and for all $t > 0$.*

It can be observed that if K is a non-empty closed cone then $0 \in K$. Examples of convex cones include linear subspaces of $\mathbb{R}^n$ and the non-negative orthant $\mathbb{R}^n_+ := \{x \mid x_i \geqslant 0,\ i = 1, ..., n\}$.

Definition 2.1.5. *Given a cone $K \subset \mathbb{R}^n$, the polar of K is the cone defined by*

$$K^\circ = \left\{ y \in \mathbb{R}^n \mid y^T x \leqslant 0,\ \forall x \in K \right\}.$$

2.1.3 Convex functions

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$. An important and useful set associated with a function f is the epigraph defined by

$$epi\ f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \alpha \geqslant f(x)\}.$$

The epigraph is thus a subset of $\mathbb{R}^{n+1}$ that consists of all points of $\mathbb{R}^{n+1}$ lying on or above the graph of f. An optimisation problem can thus be expressed equivalently in terms of its epigraph as

$$\inf f = \inf \{\alpha \mid (x,\ \alpha) \in epi\ f\}.$$

For an extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, f is convex if and only if

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y), \quad \forall x, y \in \mathbb{R}^n,\ \forall t \in (0, 1).$$

The function is called strictly convex if the above inequality is strict for all $x, y \in \mathbb{R}^n$ with $x \neq y$ and $t \in (0, 1)$. A function is concave whenever $-f$ is convex.
A characterisation of the convexity of a function f can also be done using the epigraph of the function.

Definition 2.1.6. *A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$ with $f \neq \infty$ is called convex if epi f is a non-empty convex set.*

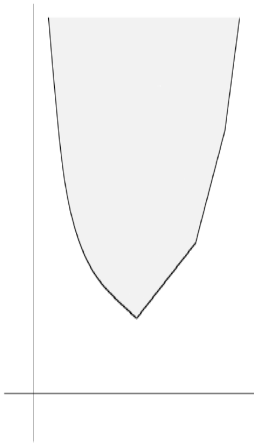

Figure 2.2: Epigraph of a convex function.

An illustration is given on Figure 2.2 and now we give some examples of convex and concave functions.

Example 2.1.7. *Examples of convex functions are:*

$$x \in \mathbb{R} \mapsto |x|, \ x \in \mathbb{R} \mapsto \|x\|_2, \ x \in \mathbb{R} \mapsto \exp(x).$$

Examples of concave functions are:

$$x \in \mathbb{R}_+ \mapsto \sqrt{x}, \ x \in \mathbb{R}_+ \mapsto \frac{x}{x+1}, \ x \in \mathbb{R} \mapsto \exp(-x).$$

Examples of functions that are neither convex nor concave:

$$x \in \mathbb{R} \mapsto \sin x, \ x \in \mathbb{R} \mapsto \cos x, \ x \in \mathbb{R} \mapsto \frac{1}{x}.$$

Definition 2.1.8. *The function $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$ is lower semi-continuous at x if*

$$f(x) = \lim_{y \to x} \inf f(y),$$

and lower semi-continuous on $\mathbb{R}^n$ if this holds for every $x \in \mathbb{R}^n$.

Example 2.1.9. *The function*

$$x \in \mathbb{R}^n \mapsto \begin{cases} 0 & \text{if } x = 0, \\ 1 & \text{otherwise}, \end{cases}$$

is not continuous but lower semi continuous.

2.2 Optimisation and variational inequalities

We discuss in this section the definition of an optimisation problem, a local minimum and some related problems called variational inequalities and complementarity problems.

2.2.1 Optimisation problems and local minima

Consider the problem of minimising a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ over a compact set $C \subset \mathbb{R}^n$ denoted by

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

Existence of minimisers is given by the the classical Weierstrass Theorem.

Theorem 2.2.1. *Let $f : K \subset \mathbb{R}^n \to \mathbb{R}$ be a continuous function defined on a compact set C. Then, there exists a global minimiser $x^* \in C$ of f on C, that is,*

$$f(x^*) \leqslant f(x), \forall x \in C.$$

It is to be mentioned here that the continuity hypothesis on f may be reduced to lower semi-continuity. Minimising a non-linear smooth function over an arbitrary compact set is already a very hard problem. An example is given in Figure 2.3. A more realistic and more accessible goal for numerical methods is to compute a local minimum. A point $x^* \in C$ is a local minimum of f over C if there exists $\varepsilon > 0$ such that for all $x \in V_\varepsilon(x) \cap C$ it holds that

$$f(x^*) \leqslant f(x),$$

where $V_\varepsilon(x)$ denotes a neighbourhood centred in x^* of radius ε.

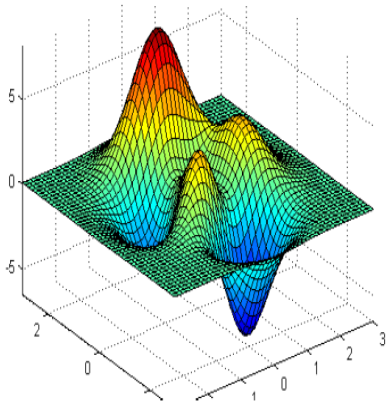

Figure 2.3: How to find the minimum of a function?

2.2.2 Variational inequalities and properties

Variational inequalities are intimately connected with optimisation problems. First, let us defined a variational inequality.

Definition 2.2.2. *Let K be a subset of $\mathbb{R}^n$ and F be a map from K into $\mathbb{R}^n$. The variational inequality, denoted (VI) is to find the vectors $x \in K$ such that*

$$(y - x)^T F(x) \geqslant 0, \quad \forall y \in K. \tag{VI}$$

The set of solutions to this probelm is denoted $SOL(K, F)$.
It is obvious that, when $K = \mathbb{R}^n$ then $x \in SOL(K, F)$ if and only if $F(x) = 0$. When K be a cone, i.e., if $x \in K$ then $tx \in K$ for all $t \geqslant 0$ then (VI) admits an equivalent form that is known as a complementarity problem.

Definition 2.2.3. *Let K be a cone and $F : K \to \mathbb{R}^n$. The complementarity problem, denoted (CP) is to find $x \in \mathbb{R}^n$ such that*

$$K \ni x \perp F(x) \in -K^\circ, \tag{CP}$$

where the notation $\perp$ means that $x^T F(x) = 0$.

Proposition 2.2.4. *[40] Let K be a cone in $\mathbb{R}^n$. Then $x \in \mathbb{R}^n$ is a solution of (VI) if and only if x is a solution of (CP).*

There are many special cases of (CP) which are very important in modeling. We now introduce the most important ones. One of this case, where K is the nonnegative orthant of $\mathbb{R}^n$, i.e., $K = \{x \in \mathbb{R}^n : x \geqslant 0\}$. We call this case the nonlinear complementarity problem or (NCP) for short.

Definition 2.2.5. *Let $F : \mathbb{R}^n \to \mathbb{R}^n$. The non-linear complementarity problem, denoted (NCP) is to find $x \in \mathbb{R}^n$ such that*

$$x \geqslant 0, \quad F(x) \geqslant 0, \quad x^T F(x) = 0 \quad \text{or} \quad 0 \leqslant x \perp F(x) \geqslant 0. \tag{NCP}$$

This problem formulation can be written equivalently as

$$x_i \geqslant 0, \ F_i(x) \geqslant 0, \quad x_i F_i(x) = 0, \quad i = 1, ..., n.$$

It is important to discriminate two types of solutions for NCPs. In the degenerate solution x^* is a component i_0 such that $x^*_{i_0} = F_{i_0}(x^*) = 0$ holds. For the nondegenerate solution x^* no such component exists, i.e. for $i = 1, ..., n$ it holds $x^*_{i_0} + F_{i_0}(x^*) > 0$. The degenerate case is numerically more difficult. Examples which lead to complementarity problems are the Nash equilibrium problem, the barrier problem and KKT conditions. We show a simple case of the last one.

Example 2.2.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex and continuously differentiable function. Then the simple optimization problem*

$$\min f(x) \quad s.t. \ x \geqslant 0,$$

is equivalent to the KKT-conditions

$$x \geqslant 0, \ \nabla f \geqslant 0, \ x^T \nabla f(x) = 0,$$

which obviously forms a complementarity problem.

In the special case of F being an affine function given by:

$$F(x) = Mx + q,$$

for some vector $q \in \mathbb{R}^n$ and matrix $M \in \mathbb{R}^{n \times n}$, we get the linear complementarity problem.

Definition 2.2.7. *Given a vector $q \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{n \times n}$, the (LCP) is to find a vector $x \in \mathbb{R}^n$ satisfying*

$$0 \leqslant x \perp (Mx + q) \geqslant 0. \tag{LCP}$$

A generalization of the (NCP) is the mixed complementarity problem abbreviated as (MiCP).

Definition 2.2.8. *Let G and H be two mappings from $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_+$ into $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_+$, respectively. The* (MiCP) *is to find $(u, v) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_+$ such that*

$$G(u, v) = 0 \quad \text{and} \quad 0 \leqslant v \perp H(u, v) \geqslant 0. \tag{MiCP}$$

We have the following relation between problem classes. The notation $P \longrightarrow Q$ means that we can derive problem Q by specializing problem P, i.e., problem Q is a special case of P.

$$\boxed{\text{VI} \longrightarrow \text{CP} \longrightarrow \text{MiCP} \longrightarrow \text{NCP} \longrightarrow \text{LCP}}$$

See [40] for more details and further results.

2.2.3 Classes of matrices

In this section, we will introduce some classes of matrices which play an important role in studying the (LCP).

Definition 2.2.9. *$M \in \mathbb{R}^{n \times n}$ is a $\mathcal{P}_0$-matrix if one of the following equivalent properties is satisfied.*

- $\forall I \subset \{1, 2, ..., n\}$, $\det(M_{II}) \geqslant 0$.
- $\forall x \in \mathbb{R}^n \backslash \{0\}$, *there exists an index i such that $x_i \neq 0$ and $x_i(Mx)_i \geqslant 0$.*
- $\forall I \subset \{1, 2, ..., n\}$, *the real eigenvalues of M_{II} are nonnegative.*

Example 2.2.10. *Here is an example of $\mathcal{P}_0$-matrix,*

$$\begin{pmatrix} 4 & 5 \\ 0 & 4 \end{pmatrix}.$$

One of the most important classes of matrices is $\mathcal{P}$-matrix that we define as following.

Definition 2.2.11. *$M \in \mathbb{R}^{n \times n}$ is a $\mathcal{P}$-matrix if one of the following equivalent properties is satisfied.*

- $\forall I \subset \{1, 2, ..., n\}$, $\det(M_{II}) > 0$.
- $\forall x \in \mathbb{R}^n$, *such that $xMx \leqslant 0$, we have $x = 0$.*
- $\forall I \subset \{1, 2, ..., n\}$, *the real eigenvalues of M_{II} are positive.*
- $\forall q$ *there exists a unique solution to the* (LCP).

Example 2.2.12. *Here is an example of $\mathcal{P}$-matrix*

$$\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}.$$

Definition 2.2.13. *$M \in \mathbb{R}^{n\times n}$ is a $\mathcal{S}$-matrix if it satisfies one of the following equivalent conditions.*

- *$\forall q \in \mathbb{R}^n$, the (LCP) is feasible.*
- *There exists $x \geqslant 0$ such that $Mx > 0$.*
- *There exists $x > 0$ such that $Mx > 0$.*

Example 2.2.14. *Here is an example of $\mathcal{S}$-matrix*

$$\begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}.$$

Definition 2.2.15. *A matrix $M \in \mathbb{R}^{n\times n}$ is a $\mathcal{Z}$-matrix if $M_{ij} \leqslant 0$ for all $i \neq j$.*

Example 2.2.16. *Here is an example of $\mathcal{Z}$-matrix*

$$\begin{pmatrix} 1 & -2 \\ 0 & 4 \end{pmatrix}.$$

Definition 2.2.17. *$M \in \mathbb{R}^{n\times n}$ is $\mathcal{M}$-matrix if M is a $\mathcal{Z}$-matrix and M satisfies one of the follow equivalent conditions.*

- *M is a $\mathcal{P}$-matrix.*
- *M is invertible and $M^{-1} \geqslant 0$, i.e., positive semi-definite, $x^T M^{-1} x \geqslant 0$ for all $x \in \mathbb{R}^n$.*
- *All eigenvalues of M have a positive real part.*
- *M is a $\mathcal{S}$-matrix.*

Example 2.2.18. *Here is an example of $\mathcal{M}$-matrix*

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

We have the following relation between classes of matrices.

$$\begin{array}{ccccc} \mathcal{M} & \rightarrow & \mathcal{P} & \rightarrow & \mathcal{P}_0 \\ \downarrow & & \downarrow & & \\ \mathcal{Z} & & \mathcal{S} & & \end{array}$$

Here, the relation $A \rightarrow B$ means that $A \subset B$, may be strictly. For more classes of matrices, see Figure 2.2.1 of [6].

2.2.4 $\mathcal{P}_0$ and $\mathcal{P}$ functions

We consider a function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and we define two properties that is used in this thesis.

Definition 2.2.19. *A map F is said to be a $\mathcal{P}_0$-function if it satisfies the following condition*

$$\forall x \neq y, \quad \exists i \in \{1,\ 2, ..., n\}, \quad (x_i - y_i)(F_i(x) - F_i(y)) \geqslant 0.$$

Here the index i may depend on $x,\ y$.

Definition 2.2.20. *A map F is said to be a $\mathcal{P}$-function if it satisfies the following condition*

$$\forall x \neq y, \quad \exists i \in \{1,\ 2, ..., n\}, \quad (x_i - y_i)(F_i(x) - F_i(y)) > 0.$$

Here the index i may depend on $x,\ y$.

Definition 2.2.21. *A map F is said to be a unifromly $\mathcal{P}$-function if there exists a constant $\mu > 0$ such that*

$$\exists i \in \{1,\ 2, ..., n\}, \quad (x_i - y_i)(F_i(x) - F_i(y)) \geqslant \mu \|x - y\|_2^2.$$

Here the index i may depend on $x,\ y$.

Proposition 2.2.22. *[40] Every uniformly $\mathcal{P}$-function must be $\mathcal{P}$-function, which in turn must be $\mathcal{P}_0$-function. If F is a $\mathcal{P}$-function then it is a $\mathcal{P}_0$-function. If F is a $\mathcal{P}_0$-function then $F + \varepsilon Id$, with $\varepsilon > 0$ is a $\mathcal{P}$-function.*

2.3 Nonlinear optimization problem

Consider the following nonlinear minimization or maximization problem (NLP):

$$\begin{aligned} &\text{optimize } f(x) \\ &\text{subject to} \\ &g_i(x) \leqslant 0, \\ &h_j(x) = 0. \end{aligned} \tag{NLP}$$

Where $x \in C$ is the optimization variable chosen from a convex subset of $\mathbb{R}^n$, f is the objective or utility function, $g_i (i = 1, ..., m)$ are the inequality constraint functions and $h_j (j = 1, ..., l)$ are the equality constraint functions. The numbers of inequalities and equalities are denoted by m and l respectively. We present in this section optimality conditions for NLP. We introduce the classical Karush Kuhn Tucker (KKT) optimality conditions in Sect. 2.3.1. These optimality conditions require some hypotheses on the set $\{x \in \mathbb{R}^n \mid g(x) \leqslant 0,\ h(x) = 0\}$ that are called constraint qualifications. A short review of some of these constraint qualifications is presented in [81].
Let the generalised Lagrangian $L(x, \mu, \lambda)$ be

$$L(x, \mu, \lambda) = f(x) + \mu^T g(x) + \lambda^T h(x),$$

where $g(x) = (g_1(x), ..., g_m(x))^T$, $h(x) = (h_1(x), ..., h_l(x))^T$ and (μ, λ) is the vector of Lagrange multipliers.

2.3.1 Karush-Kuhn-Tucker optimality conditions

In mathematical optimization, the Karush–Kuhn–Tucker (KKT) conditions, also known as the Kuhn–Tucker conditions, are first derivative tests (sometimes called first-order necessary conditions) for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied. The KKT conditions were originally named after Harold W. Kuhn and Albert W. Tucker, who first published the conditions in 1951 [70]. Later scholars discovered that the necessary conditions for this problem had been stated by William Karush in his master's thesis in 1939 [64]. The Karush–Kuhn–Tucker theorem then states the following.

Theorem 2.3.1. *[105] If x^*, μ^* is a saddle point of $L(x, \mu)$ in $x \in C$, $\mu \geqslant 0$, then x^* is an optimal vector for the above optimization problem. Suppose that $f(x)$ and $g(x)$, are convex in x and that there exists $x_0 \in C$ such that $g(x_0) < 0$. Then with an optimal vector x^* for the above optimization problem there is associated a non-negative vector μ^* such that $L(x^*, \mu^*)$ is a saddle point of $L(x, \mu)$.*

Since the idea of this approach is to find a supporting hyperplane on the feasible $\{x \in X : g_i(x) \leqslant 0, i = 1, ..., m\}$, the proof of the Karush–Kuhn–Tucker theorem makes use of the hyperplane separation theorem [65].
The system of equations and inequalities corresponding to the KKT conditions is usually not solved directly, except in the few special cases where a closed-form solution can be derived analytically. In general, many optimization algorithms can be interpreted as methods for numerically solving the KKT system of equations and inequalities [22].

2.3.2 Necessary conditions

Suppose that the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$ and $h_j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable at a point $x^* \in \mathbb{R}^n$. If x^* is a local optimum and the optimization problem satisfies some regularity condition (see [81]), then there exist constants $\mu_i (i = 1, ..., m)$ and $\lambda_i (j = 1, ..., l)$, called KKT multipliers, such that the following four groups of conditions hold:

Stationarity

For minimizing $f(x)$:

$$\nabla f(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0.$$

For maximizing $f(x)$:

$$-\nabla f(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0.$$

Primal feasibility

$$h_j(x^*) = 0, \quad \text{for} \quad j = 1, ..., l,$$
$$g_i(x^*) \leqslant 0, \quad \text{for} \quad i = 1, ..., m.$$

Dual feasibility

$$\mu_i \geqslant 0, \quad \text{for} \quad i = 1, ..., m.$$

Complementary slackness

$$\sum_{i=1}^{m} \mu_i g_i(x^*) = 0.$$

The last condition is sometimes written in the equivalent form: $\mu_i g_i(x^*)$, for $i = 1, ..., m$. In the particular case $m = 0$, i.e., when there are no inequality constraints, the KKT conditions turn into the Lagrange conditions, and the KKT multipliers are called Lagrange multipliers. If some of the functions are non-differentiable, subdifferential versions of Karush–Kuhn–Tucker (KKT) conditions are available, see [99].

2.4 Nonsmooth Analysis

In this section, we introduce subdifferentials and semi-smooth functions in the sense of Clarke [31, 32]. A very important class of functions for reformulation is the NCP-function. We use these functions to reformulate our problems as a non-linear unconstrained equation to which we can apply well-known numerical methods for computing solutions.

2.4.1 Lipschitz functions

In this subsection, we introduce one of the most important classes of functions which is the class of Lipschitz continuous function. Because most of NCP-functions be Lipschitz continuous functions.

Definition 2.4.1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a Lipschitz function on a subset R of $\mathbb{R}^n$ if for all $x,\ y \in R$, there exists a constant $L > 0$ such that*

$$|f(x) - f(y)| \leqslant L|x - y|.$$

The constant L is called Lipschitz constant.

Definition 2.4.2. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a locally Lipschitz function in a point $x \in \mathbb{R}^n$ if there exists a positive real number ε such that f is Lipschitz on $B(x, \varepsilon)$.*

Example 2.4.3. *Let $f_{\min},\ f_{FB} : \mathbb{R}^2 \to \mathbb{R}$,*

- $f_{\min} = \min\{a,\ b\}$;
- $f_{FB}(a,\ b) = \sqrt{a^2 + b^2} - (a + b)$.

Then these functions are Lipschitz continuous functions.

2.4.2 Subdifferentials

We introduce three subdifferentials which are interconnected. Let $U \subset \mathbb{R}^n$ be an open set and the function $G : U \to \mathbb{R}^m$ be a locally Lipschitz continuous function. We denote D_G the set where G is differentiable, for more details, see [6, 31, 32, 40].

Definition 2.4.4. *(**Bouligand-subdifferential**) The B-subdifferential of G at a point $x \in U$ is defined as*

$$\partial_B G(x) = \{H \in \mathbb{R}^{m\times n} \mid \exists\{x_n\} \subset D_G : \{x_n\} \to x \text{ and } G^{'}(x_n) \to H\}.$$

Definition 2.4.5. *(**Generalized Jacobian**) The generalized Jacobian of Clarke [32] is defined as*

$$\partial G(x) = co(\partial_B G(x)),$$

where "co" stands for the convex hull. When $m = 1$, we also call $\partial G(x)$ the generalized gradient of G, which is a row vector.

Example 2.4.6. *Let $f(x) = |x|$. This function is not differentiable at 0. Then $\partial_B f(x) = \{-1,\ 1\}$ and $\partial f(x) = [-1,\ 1]$.*

Definition 2.4.7. *(**Clarke-subdifferential**) The C-subdifferential of G at a point $x \in U$ is defined as*

$$\partial_C G(x) = [\partial G_1(x) \times \partial G_2(x) \times ... \times \partial G_m(x)]^T,$$

or

$$\partial_C G(x) = \{M^T \in (\mathbb{R}^{m\times n})^m : M = (M_1, M_2, ..., M_m),\ M_i \in \partial G_i(x),\ 1 \leqslant i \leqslant m\}.$$

Example 2.4.8. *Consider the Euclidean function*

$$\begin{aligned} G &: \mathbb{R}^n \to \mathbb{R} \\ x &\to G(x) = \|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}. \end{aligned}$$

Then

$$\begin{aligned} \partial_B G(0) &= \{x \in \mathbb{R}^n : \|x\|_2 = 1\} = \partial B(0,1) = S_{n-1}, \\ \partial G(0) &= \{x \in \mathbb{R}^n : \|x\|_2 \leqslant 1\} = \overline{B(0,1)}, \\ \partial_B G(x) &= \partial G(x) = \{\frac{1}{\|x\|_2} x^T\} \quad \forall x \neq 0. \end{aligned}$$

Where $B(0,1)$ is a unit ball and S_{n-1} is the unit sphere in $\mathbb{R}^n$.

Proposition 2.4.9. *[6, 32, 40] Let $G : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz continuous. Then*

1. *$\partial_B G(x) \subseteq \partial G(x) \subseteq \partial_C G(x)$ for all $x \in \mathbb{R}^n$.*
2. *$\partial G(x)$ is a nonempty set, convex and compact subset of $\mathbb{R}^{m\times n}$. This implies that $\partial_B G(x)$ is a nonempty set too.*

3. *G is continuously differentiable on an open set $D \subset \mathbb{R}^n$ if and only if $\partial G(x) = \{G'(x)\}$ is a singleton.*

2.4.3 Semi-smooth functions

For the problems we study in this thesis we need the notion of semi-smooth functions. The set of these functions is a subset of the set of locally Lipschitz continuous functions and a superset of the set of continuously differentiable functions. First, we recall the concept of directional derivative.

Definition 2.4.10. *A function $G : \mathbb{R}^n \to \mathbb{R}^m$ is called directionally differentiable at point $x \in \mathbb{R}^n$ if the limit*

$$G'(x,d) = \lim_{t \to 0} \frac{G(x+td) - G(x)}{t} \in \mathbb{R}^m,$$

exists for all direction $d \in \mathbb{R}^n$.

Definition 2.4.11. *Let $U \subset \mathbb{R}^n$ be open and $G : U \to \mathbb{R}^m$ be a locally Lipschitz continuous which is directionally differentiable. Then G is called semi-smooth at $x \in U$ if*

$$\lim_{d \to 0,\ H \in \partial G(x+d)} \frac{Hd - G'(x,d)}{\|d\|} = 0.$$

Definition 2.4.12. *Let G is a semi-smooth function. G is called strongly semi-smooth in $x \in U$ if*

$$\lim_{d \to 0,\ H \in \partial G(x+d)} \frac{Hd - G'(x,d)}{\|d\|^2} < \infty.$$

Proposition 2.4.13. *[23] Suppose that $G : \mathbb{R}^n \to \mathbb{R}^n$ is a locally Lipschitzian function. If each component of G is (strongly) semi-smooth at x, then G is (strongly) semi-smooth at x.*

Lemma 2.4.14. *[23] Let $U \subset \mathbb{R}^n$ be open, $x \in U$ and $G : U \to \mathbb{R}^m$ be a function. Then*

- *If G is continuously differentiable around x, then G is semi-smooth in x.*
- *If G is differentiable around x and G' is Lipschitz continuous around x, then G is strongly semi-smooth in x.*

2.4.4 NCP-functions

In this subsection we introduce the NCP-functions, which plays an important part in this thesis. Then we will show how to reformulate the complimentarity problems via NCP-functions. We start with the definition of NCP-function.

Definition 2.4.15. *A function $\phi : \mathbb{R}^2 \to \mathbb{R}$ with the property*

$$\phi(a, b) = 0 \Longleftrightarrow a \geqslant 0,\ b \geqslant 0,\ ab = 0,$$

is called NCP-function.

Here are some examples of NCP-functions.

- Min-function
$$\phi_{\min}(a, b) = \min(a, b).$$
- Fischer-Burmeister function
$$\phi_{FB} = \sqrt{a^2 + b^2} - (a + b).$$
- $\phi_1(a, b) = \xi(|a - b|) - (\xi(a) + \xi(b))$ where $\xi : \mathbb{R} \to \mathbb{R}$ be strictly increasing and $\xi(0) = 0$.
- $\phi_2 = \frac{1}{2} \min^2\{0, a + b\} - ab$.

It is not too difficult to check these functions are NCP-functions. This is obvious for the minimum function $\phi_{\min}$. We verify this for ϕ_{FB}. From squaring $\phi_{FB}(a, b) = 0$ it follows that

$$a^2 + b^2 = (a + b)^2,$$

and from this we conclude that

$$ab = 0,$$

and $\phi_{FB}(a, b) = 0$ is equivalent with

$$a + b = \sqrt{a^2 + b^2} \geqslant 0,$$

together with $ab = 0$ this means that either $a = 0,\ b \geqslant 0$ or $a \geqslant 0,\ b = 0$ holds. This is the first implication. The other implication can be directly verified with the same case distinction $a = 0,\ b \geqslant 0$ and $a \geqslant 0,\ b = 0$.

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a function and $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a NCP-function. Then we define the vector-valued function $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ as

$$\Phi(x) := \begin{bmatrix} \phi(x_1, F_1(x)) \\ \phi(x_2, F_2(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{bmatrix}.$$

The following result describes the connection of Φ to the complementarity problem.

Theorem 2.4.16. *A vector $x^* \in \mathbb{R}^n$ is a solution of the complementarity problem (NCP) (resp. (LCP)) if and only if x^* is a solution of the nonlinear equation system $\Phi(x) = 0$.*

Proof. From the definition of Φ and the property of NCP-functions it follows immediately

$$\begin{aligned}\Phi(x) = 0 &\iff \phi(x_i, F_i(x)) = 0 \quad \forall i = 1, ..., n, \\ &\iff x^i \geqslant 0,\ F_i(x) \geqslant 0,\ x_i F_i(x) = 0, \quad \forall i = 1, ..., n.\end{aligned}$$

which is the assertion of this theorem. □

With this theorem we have reduced the complementarity problem to the well known problem of solving a nonlinear system of equations. If F and the NCP-function ϕ are continuously differentiable then Φ is also continuously differentiable and we can solve the equation system $\Phi = 0$ e.g. with Newton's method. A further requirement for Newton's method is that the Jacobian $\Phi'(x^*)$ in the solution x^* has to be nonsingular. The next result shows that this might not be fulfilled in the given context.

Theorem 2.4.17. *[23] Let $F : \mathbb{R}^n \to \mathbb{R}^n$ and $\phi : \mathbb{R}^2 \to \mathbb{R}$ be differentiable and x^* be a degenerate solution of (NCP) (resp. (LCP)). Then the Jacobian $\Phi'(x^*)$ contains a zero row i, where $x_i^* = F_i(x^*) = 0$ holds, and is therefor singular.*

Proof. Since ϕ and F are differentiable the composed function $\phi(x_i, F_i(x))$ is differentiable and we calculate the Jacobian of $x \to \phi(x_i, F_i(x))$ with the chain rule. This clearly gives

$$\frac{\partial \phi(x_i, F_i(x))}{\partial x} = \frac{\partial \phi(x_i, F_i(x))}{\partial a} \frac{\partial F_i(x)}{\partial x} + \frac{\partial \phi(x_i, F_i(x))}{\partial b} e_i^T,$$

where e_i is the i-th unit vector. Since x^* is degenerate, there is an index i with $F_i(x^*) = x_i^* = 0$. Now let i be such an index. Since $\partial\phi(x_i, F_i(x))$ is the i-th component function of Φ the row vector $\frac{\partial \phi(x_i, F_i(x))}{\partial x}$ is the i-th row of the Jacobian $\Phi'(x)$. The proof is complete if we show that $\nabla\phi(0, 0) = (0, 0)^T$ holds. For the first partial derivative we have

$$\frac{\partial \phi(0,0)}{\partial a} = \lim_{t \to 0} \frac{\phi(t, 0) - \phi(0, 0)}{t} = \lim_{t \to 0} \frac{0 - 0}{t} = 0,$$

which follows from the NCP-function definition. In the same way, we have

$$\frac{\partial \phi(0,0)}{\partial b} = 0.$$

□

2.4.5 Newton method for semi-smooth functions

In this subsection, we will introduce the semi-smooth Newton method, which is a version of Newton's method, for solving nonlinear equation systems under weaker requirements. The theory of the semi-smooth Newton method was developed by Qi [93] see also [40, 91, 94] for related material. Let $G : \mathbb{R}^n \to \mathbb{R}$ be a given function and consider the problem of finding a solution $x^* \in \mathbb{R}^n$ for

$$G(x) = 0.$$

If G is differentiable we can try to solve this with the classical Newton method. It produces a sequence $(x^i) \subset \mathbb{R}^n$ according to the rule

$$x^{i+1} = x^i - G^{'}(x^i)^{-1}G(x^i), \quad i = 0, 1, 2, ...$$

for a starting vector $x^0 \in \mathbb{R}^n$. If G is not differentiable then the Jacobian $G^{'}(x^i)$ might not exist and the next iterate x^i is not defined. With the theory of subdifferentials from subsection 2.4.2, it presents itself to generalize Newton's method for locally Lipschitz continuous functions G as follows

$$x^{i+1} = x^i - H_i^{-1}G(x^i), \quad i = 0, 1, 2, ...$$

where $H_i \in \partial G(x^i)$. In the following algorithm, we restrain ourselves to the B-subdifferential but the generalized Jacobian would be equally possible.

Algorithm 2.1 (Semismooth Newton method)

---

1. (Initialization) Choose $x^0 \in \mathbb{R}^n$, $\varepsilon \geqslant 0$ ans set $k = 0$.
2. (Termination Criterion) If $\|G(x^k)\| \leqslant \varepsilon$, stop.
3. (Newton Direction Calculation) Choose a matrix $H_k \in \partial_B G(x^k)$
   and find a solution d^k of the linear system
   $$H_k d^k = -G(x^k).$$
4. (Update) Set $x^{k+1} = x^k + d^k$, $k = k + 1$, and go to 2.

In the termination criterion one can use any norm. But it is sometimes useful to choose a certain norm. For differentiable functions G this algorithm reduces to the classical one since $\partial_B G(x^i) = \{G^{'}(x^i)\}$ holds for such functions.

In the rest of this subsection, we will show that this algorithm has the same local convergence properties as the classical Newton method if certain requirements are met. Note that the use of the generalized Jacobian would not alter the convergence properties but a little the requirements thereof.

Definition 2.4.18. *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous in $x \in \mathbb{R}^n$. Then x is called BD-regular if all matrices $H \in \partial_B G(x)$ are nonsingular.*

Example 2.4.19. *We consider the scalar function $G(x) = |x|$. In the solution $x^* = 0$ we have $\partial_B G(0) = \{-1, 1\}$. Therefore $x = 0$ is BD-regular for this function. On the other hand does $\partial G(0) = [-1, 1]$ contain zero. As a result, the generalized Jacobian fails to satisfy a comparable reguarity condition, and hence fails to meet a crucial convergence requirement. This is a significant benefit of using the B-subdifferential to formulate the semismooth Newton technique.*

The next step toward a convergence result is the following Lemma.

Lemma 2.4.20. *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz continuous and let $x^* \in \mathbb{R}^n$ be a BD-regular point of G. Then there are numbers $\varepsilon > 0$ and $c > 0$ so that all matrices $H \in \partial_B G(x)$ for all points $x \in B_\varepsilon(x^*)$ are nonsingular and*

$$\|H^{-1}\| \leqslant c \quad \forall H \in \partial_B G(x) \quad \forall x \in B_\varepsilon(x^*).$$

Proof. see ([94] Proposition 3.1) □

For differentiable functions, this result simplifies to a well-known Lemma once more. Finally, the main theorem can be stated.

Theorem 2.4.21. *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be semismooth and let x^* be a BD-regular solution for $G(x) = 0$. Then there is a $\varepsilon > 0$ so that for every starting vector $x^0 \in B_\varepsilon(x^*)$ the following holds:*

1. *Algorithm 2.1 is well defined and produces a sequence (x^i) which converges to x^*.*
2. *The convergence rate is superlinear, i.e. $\|x^{i+1} - x^*\| = o\left(\|x^i - x^*\|\right)$*
3. *If G is strongly semismooth then the convergence rate is quadratic, i.e. $\|x^{i+1} - x^*\| = O\left(\|x^i - x^*\|^2\right)$*

Proof. see ([94] Theorem 3.2). □

We can solve nonlinear systems with nondifferentiable functions if they are semismooth. In particular we can solve nonlinear systems that stem from complementarity problems. These techniques may present difficulties to converge. An efficient approach is to approximate $\min(x_i, F_i(x)) = 0,\ i = 1, ..., n$, by a smooth one. The following section introduces some smoothing functions and establishes different properties that will be useful for our thesis.

2.5 θ functions

In this thesis we propose some smoothing techniques to regularize the complementarity problem or the complementarity constraints for example ($x \geqslant 0,\ z \geqslant 0$ and $x_i z_i{=}0$) or $\min(x_i, z_i) = 0$ and construct relaxed problems that are suitable for non linear programming (NLP) algorithms. Many regularization and relaxation techniques have already been proposed, here is an incomplete list of such methods

$$
\begin{aligned}
&x^T z = 0 \quad \text{is relaxed to} \quad x_i z_i \leqslant r \quad \forall i,\ [95, 101]\\
&x^T z = 0 \quad \text{is relaxed to} \quad x_i z_i = r \quad \forall i,\ [95, 101]\\
&x^T z = 0 \quad \text{is relaxed to} \quad x^T z \leqslant r \quad \forall i,\ [95, 101]\\
&x^T z = 0 \quad \text{is relaxed to} \quad \sqrt{(x_i - z_i)^2 + 4r^2} - (x_i + z_i) = 0 \quad \forall i, [38]\\
&x^T z = 0 \quad \text{is relaxed to} \quad r \ln\left\{e^{\frac{-x_i}{r}} + e^{\frac{-z_i}{r}}\right\} = 0 \quad \forall i,\ [19, 41].
\end{aligned}
$$

In almost all these techniques, the complementarity problem or the complementarity constraints ($x \geqslant 0,\ z \geqslant 0$ and $x_i z_i{=}0$) or $\min(x_i, z_i) = 0$ are replaced by some smooth approximations and maintain the positivity constraints. In our approach, we maintain the positivity constraints and interpret the complementarity constraint componnent-wise as:

$$\forall i,\ \text{At most one of } x_i \text{ or } z_i \text{ is nonzero.}$$

So, we construct some parameterized real functions that satisfy:

$$(\theta_r(x) \simeq 1 \text{ if } x \neq 0) \quad \text{and} \quad (\theta_r(x) \simeq 0 \text{ if } x = 0),$$

to count nonzeros and then replace the constraint

$$x_i z_i = 0,$$

by

$$\theta_r(x_i) + \theta_r(z_i) \leqslant 1.$$

In this section, we elaborate on how such a regularized function can be actually built up from a function that is not differentiable everywhere. Our smoothing technique is based on the continuous approximation of a more elementary object, namely the step function.

The step function is understood here to be the function $\Im : \mathbb{R}_+ \to \{0,\ 1\}$ defined as

$$\Im(t) = \begin{cases} 0 & \text{if} \quad t = 0, \\ 1 & \text{if} \quad t > 0. \end{cases} \tag{2.5.1}$$

As an indicator of positive arguments $t > 0$ over $\mathbb{R}_+$, the step function $\Im$ "descriminates" the argument $t = 0$ by assigning a zero value to it. The price to be paid for this sharp detection is the discontinuity of $\Im$ at $t = 0$. We wish to have a regularization of $\Im$, that is, a family of functions

$$\{\tilde{\Im}(., r) : \mathbb{R}_+ \to [0, 1), r > 0\}, \tag{2.5.2}$$

such that

- $\tilde{\Im}(., r)$ is a smooth function of $t \geqslant 0$, for all $r > 0$;
- $\tilde{\Im}$ is continuous with respect to r, in some functional sense;
- $\lim_{r \downarrow 0} \tilde{\Im}(., r) = \Im(.)$, in some functional sense.

To obtain such a family, we follow the methodology developed by Haddou and his coauthors [4, 52], the key ingredient of which is a smoothing function. This notion turned out to be a versatile tool in a wide variety of pure and applied mathematical problems [14, 53, 54, 81]. We begin with a "father" function, from which all other regularized functions will be generated.

Definition 2.5.1. *(θ-smoothing function). A function $\theta : \mathbb{R} \to [0, 1)$ is said to be a θ-smoothing function if it is continuous, nondecreasing, concave, and*

$$\begin{aligned} \theta(0) &= 0, \\ \lim_{t \to +\infty} \theta(t) &= 1. \end{aligned} \tag{2.5.3}$$

The two most common examples of smoothing functions are:

1. *the rational function $\theta^1 : \mathbb{R} \to (-\infty, 1)$ defined by*

$$\theta^1(t) = \frac{t}{t+1} \quad \text{for} \quad t \geqslant 0 \quad \text{and} \quad \theta^1(t) = t \quad \text{for} \quad t \leqslant 0. \tag{2.5.4}$$

2. *the exponential function $\theta^2 : \mathbb{R} \to (-\infty, 1)$ defined by*

$$\theta^2(t) = 1 - \exp(-t). \tag{2.5.5}$$

A more general "recipe" to build such function is to consider nonincreasing probability density functions
$f : \mathbb{R}_+ \to \mathbb{R}_+$ and then take the corresponding cumulative distribution function on $\mathbb{R}_+$ i.e.,

$$\theta(t) = \int_0^t f(y)dy, \quad t \geqslant 0, \tag{2.5.6}$$

we complete the definition of θ on $\mathbb{R}_-$ by $\theta(t) = t$ to get a continuous, nondecreasing function. The nonincreasing assumption on f gives the concavity of θ. Once a favorite θ-smoothing has been selected, the next step is to dilate or compress it in order to produce a family of regularized functions for the step function $\Im$.

Definition 2.5.2. *(θ-smoothing family). Let θ be a θ-smoothing function. The family of functions*

$$\left\{\theta_r(t) := \theta(\frac{t}{r}), \quad r > 0\right\}, \tag{2.5.7}$$

is said to be the θ-smoothing family associated with θ.

Obviously, θ_r is a smooth function of $t \geqslant 0$ for all $r > 0$. It is also continuous with respect to r at each fixed $r \geqslant 0$. From the defining properties (2.5.3), it can be readily shown that

$$\lim_{r \to 0} \theta_r(t) = \Im(t), \quad \forall t \geqslant 0. \tag{2.5.8}$$

In other words, $\Im$ is the limit of θ_r in the sense of pointwise convergence. Thus, $\{\Im(., r) = \theta_r,\ r > 0\}$ is a good family of regularized functions in the sense of (2.5.2). Associated with the two examples (2.5.4)-(2.5.5) are:

1. the rational family $\theta_r^1 : \mathbb{R} \to (-\infty, 1)$ defined by

$$\theta_r^1(t) = \frac{t}{t+r} \quad \text{for} \quad t \geqslant 0 \quad \text{and} \quad \frac{t}{r} \quad \text{for} \quad t \leqslant 0. \tag{2.5.9}$$

2. the exponential family $\theta_r^2 : \mathbb{R} \to (-\infty, 1)$ defined by

$$\theta_r^2(t) = 1 - \exp(-t/r). \tag{2.5.10}$$

Figure 2.4 display the two families (2.5.9)-(2.5.10) for a few values of the parameter r. We can see that the smaller r is, the steeper is the slope at $t = 0$ and the closer to $\Im$ the function is.

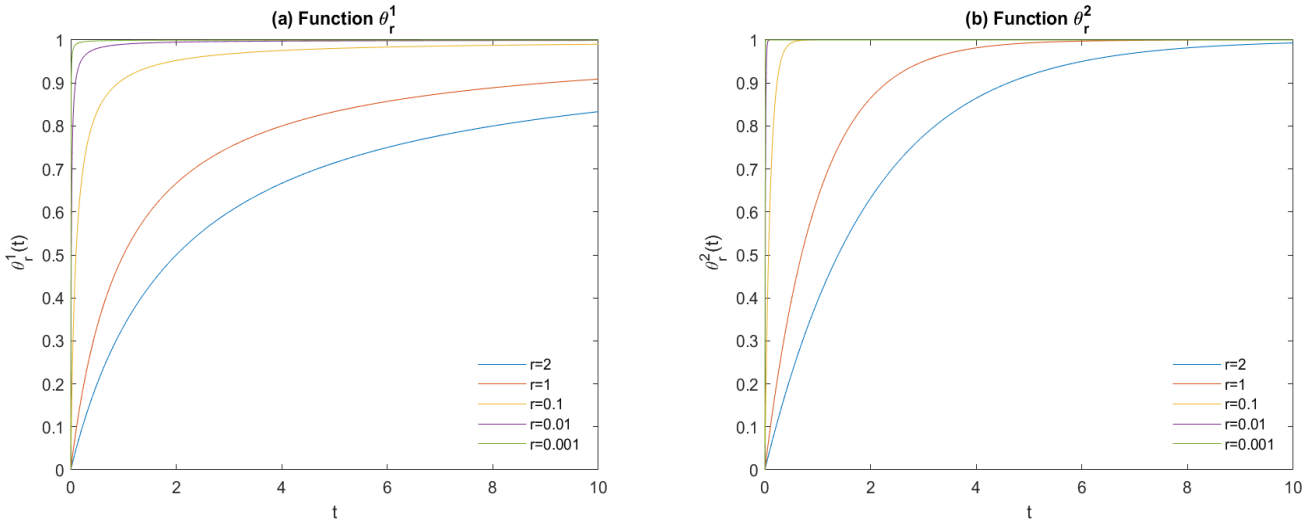

Figure 2.4: Function θ_r for a few values of r.

2.5.0.1 θ-smoothing of a complementarity condition

A θ-smoothing function paves the way for a smooth approximation of a complementarity condition. Let $(x, z) \in \mathbb{R}^2$ be two scalars such that

$$0 \leqslant x \perp z \geqslant 0, \tag{2.5.11}$$

that is,

$$x \geqslant 0, \quad z \geqslant 0, \quad xz = 0.$$

In the (x, z)-plane, the set of points obeying (2.5.11) is the union of the two semi-axes $\{x \geqslant 0,\ z = 0\}$ and
$\{x = 0,\ z \geqslant 0\}$. Visually, the nonsmoothness of (2.5.11) is manifested by the "kink" at the corner $(x, z) = (0, 0)$. It is also clear that the corresponding set is non-convex. We consider two possible smooth approximations of (2.5.11), depending how it is rewritten in terms of the step function $\Im$. Through this manuscript, we use the functions θ_r to regularise the complementarity problem. The following lemma, provides an intuition of the motivation behind such a technique and shows the link between this family of functions and the complementarity.

Lemma 2.5.3. *Assuming $x \geqslant 0$ and $z \geqslant 0$, we have the equivalence*

$$xz = 0 \iff \Im(x) + \Im(z) \leqslant 1. \tag{2.5.12}$$

The equivalence (2.5.12) suggests us to impose

$$x \geqslant 0, \quad z \geqslant 0, \quad \theta_r(x) + \theta_r(z) = 1, \tag{2.5.13}$$

for $r > 0$, as a smooth approximation of (2.5.11). Replacing $\Im$ by θ_r in (2.5.12) is logical. Replacing "$\leqslant$" by "$=$" in (2.5.12) and the (2.5.13) seems to be a bold move, but this is motivated by the fact that we want an equality to be mounted into the system of equations. Some times an additional assumption (strict complementarity $x + z > 0$) is made to get such equations.

Proof. Prove by contradiction that

$$\lim_{r\to 0}(\theta_r(x) + \theta_r(z)) \leqslant 1 \Longrightarrow x \perp z.$$

Suppose $x, z > 0$, then

$$\lim_{r\to 0}(\theta_r(x) + \theta_r(z)) = \lim_{r\to 0}\theta_r(x) + \lim_{r\to 0}\theta_r(z) = 2.$$

This leads to a contradiction and therefore $x \perp z$. Conversely it is clear that $x \perp z$ implies $x = 0$ or $z = 0$ and the result follows. □

In the case of the function θ_r^1, it holds that

$$\theta_r^1(x) + \theta_r^1(z) = 1 \iff xz = r^2.$$

A classical property shared by all concave functions that vanish in zero is the subadditivity.

Lemma 2.5.4. *θ_r is sub-additive for non-negative values, i.e. given $x \geqslant 0$ and $z \geqslant 0$, it hold that*

$$\theta_r(x) + \theta_r(z) \geqslant \theta_r(x + z). \tag{2.5.14}$$

Proof. Since θ is concave, we obtain

$$\forall x \neq z \in \mathbb{R},\ \forall t \in (0,1), \quad t\theta_r(x) + (1-t)\theta_r(z) \leqslant \theta_r(tx + (1-t)z),$$

with equality if $t = 0$ or $t = 1$ and if $x = z$. Considering $z = 0$ and $\theta_r(0) = 0$ yields

$$\theta_r(tx) = \theta_r(tx + (1-t)z) \geqslant t\theta_r(x) \quad \forall t \in (0,1),$$

with equality if $t = 0$ or 1 and if $x = 0$.

Let $i \in \{1, ..., n\}$ and suppose that $x_i + z_i \neq 0$ (the case $x_i = z_i = 0$ stay true)

$$\begin{aligned}\theta_r(x_i) + \theta_r(z_i) &= \theta_r((x_i + z_i)\frac{x_i}{x_i + z_i}) + \theta_r((x_i + z_i)\frac{z_i}{x_i + z_i}) \\ &\geqslant \frac{x_i}{x_i + z_i}\theta_r(x_i + z_i) + \frac{z_i}{x_i + z_i}\theta_r(x_i + z_i) \\ &= \theta_r(x_i + z_i),\end{aligned}$$

with equality if and only if $x_i = 0$ or $z_i = 0$,

$$x_i z_i = 0 \iff \theta_r(x_i) + \theta_r(z_i) = \theta_r(x_i + z_i).$$

This concludes the proof. □

3 A new relaxation method for optimal control of semilinear elliptic variational inequalities obstacle problems

This chapter is a paper accepted in Numerical Algebra, Control & Optimization [86].
In this chapter, we investigate optimal control problems governed by semilinear elliptic variational inequalities involving constraints on the state, and more precisely the obstacle problem. Since we adopt a numerical point of view, we first relax the feasible domain of the problem, then using both mathematical programming methods and penalization methods we get optimality conditions with smooth Lagrange multipliers. Some numerical experiments using the Interior Point Optimizer (IPOPT), Nonlinear Interior point Trust Region Optimization (KNITRO) and Sequential Quadratic Optimization Technique (SNOPT) are presented to verify the efficiency of our approach.

Contents

3.1 Introduction

In this chapter, we investigate optimal control problems where the state is described by semilinear variational inequalities. These problems involve state constraints as well. We may consider these problems from many points of view. In [10, 11], the authors provide first-order necessary optimality conditions. Indeed, they use some relaxation of the original problem governed by variational inequalities and involving constraints on both the control and the state. Their reformulation of the problem involves nonconvex coupling constraints on both the state and the control variables.

In our work, we use the methodology of [10, 11] and generalize their results to the semilinear case. It is known that Lagrange multipliers may not exist for such problems [15]. Nevertheless, providing qualification conditions, one can exhibit multipliers for relaxed problems. These multipliers usually allow getting optimality conditions of Karush-Kuhn-Tucker type. Our purpose is to get optimality conditions that are useful in practice: when the penalization parameter ε goes to 0, all the variables and multipliers exist, and remain bounded. Indeed, we have to ensure the existence of Lagrange multipliers to prove the convergence of lagrangian methods and justify their use. These kinds of problems have been extensively studied, see for instance [3, 7, 13, 45, 58, 79]. Especially, the variational iteration method solves quadratic optimal control problems of systems governed by linear partial differential equations [3, 58]. The idea consists of deriving the necessary optimality conditions by applying the minimum principle of Pontryagin, which leads to the well known Hamilton–Pontryagin equations.

In this work, we interpret the variational inequality as a state equation, introducing another control function as in [10, 11]. Then, we consider the problem as a "standard" control problem governed by a semilinear partial differential equation, involving pure and mixed control-state constraints which are not necessarily convex. In order to derive some optimality conditions, we have to "relax" the domain; so we do not solve the original problem but this point of view will be justified and commented on. Then, by the use of mathematical programming in Banach spaces methods [102, 111] and penalization techniques, we provide first-order necessary optimality conditions.

The first part of this chapter is devoted to the presentation of the problem: we recall some classical results on variational inequalities there. In section 3.3, we propose relaxations of the original problem. In section 3.4, we briefly present some mathematical programming results in Banach spaces. Next, we use a penalization technique and apply the tools of the previous section to the penalized problem. For the penalized problems, we obtain optimality conditions and assuming some qualification conditions we may pass to the limit to get optimality conditions for the original problem. In the last section, we present some numerical results and propose a conclusion.

3.2 Problem setting

Let Ω be an open, bounded subset of $\mathbb{R}^n$ with a smooth boundray $\partial\Omega$. For convenience, in the sequel we denote $||\ .\ ||_V$, the norm in Banach space V, and more precisely $||\ .\ ||_2$ the $L^2(\Omega)$-norm. In the same way, $\langle .,. \rangle$ denotes the duality product between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$ and $(\ .\ )_2$ the $L^2(\Omega)$-inner product. Let us set

$$K = \{y \mid y \in H_0^1(\Omega),\ y \geqslant \psi \ a.e.\ \text{in}\ \Omega\}, \tag{3.2.1}$$

where ψ is a $H^2(\Omega) \cap H_0^1(\Omega)$ function. In the sequel g is a non decreasing, C^1 real-valued function such that $g^{'}$ is bounded, locally Lipschitz continuous and

$$\exists \gamma \in \mathbb{R},\ \exists \beta \geqslant 0 \ \text{ such that } \ \forall y \in \mathbb{R} \ \ |g(y)| \leqslant \gamma + \beta|y|, \tag{3.2.2}$$

and f belongs to $L^2(\Omega)$. Moreover, U_{ad} is a non empty, closed and convex subset of $L^2(\Omega)$. For each v in U_{ad} we consider the following variational inequality problem: find $y \in K$ such that

$$\tilde{a}(y,z) + G(y) - G(z) \geqslant (v + f, y - z)_2 \quad \forall z \in K, \tag{3.2.3}$$

where G is a primitive function of g, and $\tilde{a}$ is a bilinear form defined on $H_0^1(\Omega) \times H_0^1(\Omega)$ by

$$\tilde{a}(y,z) = a(y,z) + \sum_{i=1}^{n} \int_\Omega b_i \frac{\partial y}{\partial x_i} z\ dx + \int_\Omega cyz\ dx, \tag{3.2.4}$$

where $a(y,z) = \sum_{i,j=1}^n \int_\Omega a_{ij} \frac{\partial y}{\partial x_i} \frac{\partial z}{\partial x_j}\ dx$ and a_{ij}, b_i, c belong to $L^\infty(\Omega)$. Moreover, we assume that a_{ij} belongs to $\mathcal{C}^{0,1}(\bar{\Omega})$ (the space of Lipschitz continuous functions in Ω) and that c is nonnegative.
The bilinear form $\tilde{a}(.,.)$ is continuous on $H_0^1(\Omega) \cap H_0^1(\Omega)$ [7, 45], i.e.

$$\exists M > 0,\ \forall (y,z) \in H_0^1(\Omega) \cap H_0^1(\Omega), \qquad \tilde{a}(y,z) \leqslant M\ ||\ y\ ||_{H_0^1(\Omega)}\ ||\ z\ ||_{H_0^1(\Omega)} \tag{3.2.5}$$

and coercive [7, 45], i.e.

$$\exists \delta > 0,\ \forall y \in H_0^1(\Omega), \qquad \tilde{a}(y,y) \geqslant \delta\ ||\ y\ ||^2_{H_0^1(\Omega)}\ . \tag{3.2.6}$$

We define the elliptic differential operator A from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$ by

$$\forall (z,v) \in H_0^1(\Omega) \times H_0^1(\Omega) \quad \langle Ay, z \rangle = a(y,z).$$

By the coercivity of the problem (3.2.3) in y and v, for any $v \in L^2(\Omega)$, (3.2.3) has a unique solution $y = y[v] \in H_0^1(\Omega)$ (see [8] for example).

As the obstacle function belongs to $H^2(\Omega)$, we have an additional regularity result: $y \in H^2(\Omega) \times H_0^1(\Omega)$ (see [8, 12]). Moreover (3.2.3) is equivalent to (see [79])

$$Ay + g(y) = f + v + \xi,\ y \geqslant \psi,\ \xi \geqslant 0,\ \langle \xi, y - \psi \rangle = 0, \tag{3.2.7}$$

where "$\xi \geqslant 0$" stands for "$\xi(x) \geqslant 0$ almost everywhere on Ω". The system (3.2.7) can be viewed as the optimality system for problem (3.2.3), ξ is the multiplier associated to the constraint $y \geqslant \psi$. It is a priori an element of $H^{-1}(\Omega)$ but the regularity result for y shows that $\xi \in L^2(\Omega)$, so that $\langle \xi, y - \psi \rangle = (\xi, y - \psi)_2$.

Remark. *Applying the simple transformation $y^* = y - \psi$, we may assume that $\psi = 0$. Of course, functions g and f are modified as well, but this shift preserves their generic properties (local Lipschitz-continuity, monotonicity). The second part of equation (3.2.4), is integrated into the function g.*

We denote similarly the real valued function g and the Nemitsky operator such that $g(y)(x) = g(y(x))$ for every $x \in \Omega$. Therefore we keep the same notations. Now, let us consider the optimal control problem defined as follows:

$$\min \left\{ J(y,v) \overset{def}{=} \frac{1}{2}\int_\Omega (y - z_d)^2 dx + \frac{\nu}{2}\int_\Omega (v - v_d)^2 dx \ /\ y = y[v],\ v \in U_{ad},\ y \in K \right\},$$

where $z_d, v_d \in L^2(\Omega)$ and $\nu > 0$ are given quantities.
This problem is equivalent to the problem governed by a state equation (instead of inequality) with mixed state and control constraints:

$$\min \left\{ J(y,v) = \frac{1}{2}\int_\Omega (y - z_d)^2 dx + \frac{\nu}{2}\int_\Omega (v - v_d)^2 dx \right\}, \qquad (\mathcal{P})$$

$$Ay + g(y) = f + v + \xi \ \text{ in } \Omega, \quad y = 0 \quad \text{on} \quad \partial\Omega, \tag{3.2.8}$$

$$(y, v, \xi) \in \mathcal{D}, \tag{3.2.9}$$

where

$$\mathcal{D} = \{(y, v, \xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega) \mid v \in U_{ad},\ y \geqslant 0,\ \xi \geqslant 0,\ (y, \xi)_2 = 0\}. \tag{3.2.10}$$

We assume that the feasible set $\tilde{\mathcal{D}} = \{(y, v, \xi) \in \mathcal{D} \ /\ $ relation (3.2.8) is satisfied$\}$ is non empty. We know, then that problem $(\mathcal{P})$ has at least an optimal solution (not necessarily unique) that we denote $(\bar{y}, \bar{v}, \bar{\xi})$, (by the coercivity of the problem (3.2.3) in y, and v see for instance [8, 12]).
Similar problems have been studied also in [16] but in the convex context ($\mathcal{D}$ is convex).

Here, the main difficulty comes from the fact that the feasible domain $\mathcal{D}$ is non-convex and has an empty relative interior because of the bilinear constraint "$(y,\xi)_2 = 0$". So, we cannot use generic convex analysis methods that have been used for instance in [16]. To derive optimality conditions in this case, we are going to use methods adapted to quite general mathematical programming [102, 111]. Unfortunately, the domain $\mathcal{D}$ (i.e. the constraints set) does not satisfy the usual (quite weak) assumption of mathematical programming theory. This comes essentially from the fact that L^∞-interior of $\mathcal{D}$ is empty. So, we cannot ensure the existence of Lagrange multipliers. This problem does not satisfy classical constraint qualifications (in the usual KKT sense). One can find several counter-examples in finite and infinite dimensions in [15].

3.3 A relaxed problem

In order to "relax" the complementarity constraint "$\langle y,\xi\rangle = 0$" we introduce a family of $\mathcal{C}^1$ functions $\theta_\alpha : \mathbb{R}^+ \to [0,1[$, $(\alpha > 0)$ with the following properties (see [52] for more precision on these smoothing functions):

(i) $\forall \alpha > 0$, θ_α is nondecreasing, concave and $\theta_\alpha(1) < 1$;

(ii) $\forall \alpha > 0 \qquad \theta_\alpha(0) = 0$;

(iii) $\forall x > 0 \quad \lim_{\alpha\to 0} \theta_\alpha(x) = 1 \qquad$ and $\quad \lim_{\alpha\to 0} \theta'_\alpha(0) > 0.$

Example 3.3.1. *The functions below satisfy assumption $(i-iii)$ (see [52]):*

$$\theta^1_\alpha(x) = \frac{x}{x+\alpha},$$

$$\theta^{log}_\alpha(x) = \frac{log(1+x)}{log(1+x+\alpha)}.$$

Functions θ_α are built to approximate the complementarity constraint in the following sense:

$$\forall (x,y) \in \mathbb{R}\times\mathbb{R} \quad xy = 0 \tilde{\Longleftrightarrow} \theta_\alpha(x) + \theta_\alpha(y) \leqslant 1,$$

for α small enough. More precisely, we have the following proposition:

Proposition 3.3.2. *Let $(y,v,\xi) \in \mathcal{D}$ and θ^1_α satisfying $(i-iii)$. Then*

$$(y,\xi)_2 = 0 \Longrightarrow \theta^1_\alpha(y) + \theta^1_\alpha(\xi) \leqslant 1 \qquad a.e.\ in\ \Omega.$$

The proof of the proposition it is based on the followings lemmas:

Lemma 3.3.3. *For any $\varepsilon > 0$, and $x, y \geqslant 0$, there exists $\alpha_0 > 0$ such that*

$$\forall \alpha \leqslant \alpha_0, \quad (min(x,y) = 0) \Longrightarrow (\theta_\alpha(x) + \theta_\alpha(y) \leqslant 1) \Longrightarrow (min(x,y) \leqslant \varepsilon).$$

Proof. The first property is obvious since $\theta_\alpha(0) = 0$ and $\theta_\alpha \leqslant 1$. Using assumption (iii) for $x = \varepsilon$, we have

$$\forall r > 0, \quad \exists \alpha_0 > 0 \mid \forall \alpha \leqslant \alpha_0 \quad 1 - \theta_\alpha(\varepsilon) < r,$$

so that, if we suppose that $\min(x,y) > \varepsilon$, assumption (i) gives

$$\forall r > 0, \quad \theta_\alpha(x) + \theta_\alpha(y) > 2\theta_\alpha(\varepsilon) > 2(1-r).$$

Then if we choose $r < \frac{1}{2}$, we obtain that $\theta_\alpha(x) + \theta_\alpha(y) > 1$. □

Lemma 3.3.4. *we have*

1. $\forall x \geqslant 0,\ \forall y \geqslant 0 \quad \theta^1_\alpha(x) + \theta^1_\alpha(y) \leqslant 1 \Longleftrightarrow x.y \leqslant \alpha^2$, *and*

2. $\forall x \geqslant 0,\ \forall y \geqslant 0 \quad x.y = 0 \Longrightarrow \theta^{\geqslant 1}_\alpha(x) + \theta^{\geqslant 1}_\alpha(y) \leqslant 1 \Longrightarrow x.y \leqslant \alpha^2$,

where $\theta^{\geqslant 1}_\alpha$ verifying $(i - iii)$ and $\theta^{\geqslant 1}_\alpha \geqslant \theta^1_\alpha$.

Proof. (1) We have

$$\theta^1_\alpha(x) + \theta^1_\alpha(y) = \frac{2xy + \alpha x + \alpha y}{xy + \alpha x + \alpha y + \alpha^2},$$

so that

$$\begin{aligned} \theta^1_\alpha(x) + \theta^1_\alpha(y) \leqslant 1 &\Longleftrightarrow 2xy + \alpha x + \alpha y \leqslant xy + \alpha x + \alpha y + \alpha^2 \\ &\Longleftrightarrow x.y \leqslant \alpha^2. \end{aligned} \tag{3.3.1}$$

The first part of (2) follows obviously form Lemma 3.3.3 and the second one is a direct consequence of (1) since

$$\theta^{\geqslant 1}_\alpha(x) + \theta^{\geqslant 1}_\alpha(y) \leqslant 1 \Longrightarrow \theta^1_\alpha(x) + \theta^1_\alpha(y) \leqslant 1.$$

□

More precisely, we consider the relaxed domain $\mathcal{D}_\alpha$ instead of $\mathcal{D}$, with $\alpha > 0$, using the function θ^1_α

we obtain:

$$\mathcal{D}_\alpha = \Big\{ (y,v,\xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega) \ / \\ v \in U_{ad},\ y \geqslant 0,\ \xi \geqslant 0,\ \frac{y}{y+\alpha} + \frac{\xi}{\xi+\alpha} \leqslant 1,\ a.e.\ in\ \Omega \Big\}, \tag{3.3.2}$$

and the relaxed domain $\mathcal{D}_\alpha^{\log}$ instead of $\mathcal{D}$, with $\alpha > 0$

$$\mathcal{D}_\alpha^{\log} = \Big\{ (y,v,\xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega) \ / \\ v \in U_{ad},\ y \geqslant 0,\ \xi \geqslant 0,\ \frac{\log(1+y)}{\log(1+y+\alpha)} + \frac{\log(1+\xi)}{\log(1+\xi+\alpha)} \leqslant 1,\ a.e.\ in\ \Omega \Big\}, \tag{3.3.3}$$

in the case of the function $\theta_\alpha^{\log}$.
We may justify and motivate this point of view numerically, since it is usually not possible to ensure "$(y,\xi)_2 = 0$" during a computation but rather "$\frac{y}{y+\alpha} + \frac{\xi}{\xi+\alpha} \leqslant 1$" where α is a prescribed tolerance: it may be chosen small as wanted, but strictly positive. So, the problem turns to be qualified if the bilinear constraint "$(y,\xi)_2 = 0$" is relaxed to "$\frac{y}{y+\alpha} + \frac{\xi}{\xi+\alpha} \leqslant 1$" a.e. in Ω.
In the sequel, we consider an optimal control problem $(\mathcal{P}^\alpha)$ where the feasible domain is $\mathcal{D}_\alpha$ instead of $\mathcal{D}$. Moreover, we must add a bound constraint on the control ξ to be able to ensure the existence of a solution of this relaxed problem. More precisely we consider:

$$(\mathcal{P}^\alpha) \quad \begin{cases} \min\ J(y,v) \\ Ay + g(y) = f + v + \xi \text{ in } \Omega,\ y \in H_0^1(\Omega), \\ (y,v,\xi) \in \mathcal{D}_{\alpha,R} \end{cases}$$

where $R > 0$ may be very large and

$$\mathcal{D}_{\alpha,R} = \{(y,v,\xi) \in \mathcal{D}_\alpha \mid \ ||\ \xi\ ||_2 \leqslant R\}.$$

From now on, we omit the index R since this constant is definitely fixed, such that

$$R \geqslant ||\ \bar{\xi}\ ||_2, \tag{3.3.4}$$

where $(\bar{y},\bar{v},\bar{\xi})$ is a solution of $(\mathcal{P})$.
We will denote $\mathcal{D}_\alpha := \mathcal{D}_{\alpha,R}$, and $V_{ad} = \{\xi \in L^2(\Omega) \mid \xi \geqslant 0, ||\ \xi\ ||_2 \leqslant R\}$. V_{ad} is obviously a closed, convex subset of $L^2(\Omega)$. As $(\bar{y},\bar{v},\bar{\xi}) \in \mathcal{D}$, we see with (3.3.4) that $\mathcal{D}_\alpha$ is non empty for any $\alpha > 0$.

3.3.1 Existence result

In order to prove an existence result for $(\mathcal{P}^\alpha)$, we state first a basic but essential lemma.

Lemma 3.3.5. *Assume that (y_n, v_n) is a bounded sequence in $H_0^1(\Omega) \times L^2(\Omega)$ such that $\xi_n := Ay_n + g(y_n) - f - v_n$ is bounded in $L^2(\Omega)$. Then, one may extract subsequences (still denoted similarly) such that*

- *v_n converges weakly to some $\tilde{v}$ in $L^2(\Omega)$;*
- *y_n converges strongly to some $\tilde{y}$ in $H_0^1(\Omega)$;*
- *$g(y_n)$ converges strongly to $g(\tilde{y})$ in $L^2(\Omega)$;*
- *$Ay_n + g(y_n) - f - v_n$ converges weakly to $A\tilde{y} + g(\tilde{y}) - f - \tilde{v}$ in $L^2(\Omega)$.*

Proof. Let (y_n, v_n) be a bounded sequence in $H_0^1(\Omega) \times L^2(\Omega)$; therefore (y_n, v_n) weakly converges to some $(\tilde{y}, \tilde{v})$ in $H_0^1(\Omega) \times L^2(\Omega)$ (up to a subsequence). Similarly, ξ_n weakly converges to some $\tilde{\xi}$ in $L_2(\Omega)$. Thanks to [68] (Theorem 17.5, p174), assumption (3.2.2) yields that

$$(y_n)_{n\geqslant 0} \text{ bounded in } L^2(\Omega) \Longrightarrow (g(y_n))_{n\geqslant 0} \text{ bounded in } L^2(\Omega).$$

As, y_n weakly converges to $\tilde{y}$ in $H_0^1(\Omega)$, it strongly converges in $L^2(\Omega)$ a.e. in Ω. As g is continuous, $g(y_n)$ converges a.e. in Ω as well (up to subsequences). We conclude then (Lebesgue theorem), that $g(y_n)$ strongly converges to $g(\tilde{y})$ in $L^2(\Omega)$. Moreover when $Ay_n = -g(y_n) + f + v_n + \xi_n$ is bounded in $L^2(\Omega)$ it will converge weakly to some $\tilde{z}$ in $L^2(\Omega)$. As y_n weakly converges to $\tilde{y}$ in $H_0^1(\Omega)$, then Ay_n converges to $A\tilde{y}$ in $H^{-1}(\Omega)$, so $\tilde{z} = A\tilde{y}$ and Ay_n weakly converges to $A\tilde{y}$ in $L^2(\Omega)$ as well. Therefore Ay_n strongly converges to $A\tilde{y}$ in $H^{-1}(\Omega)$.
Finally we get the weak convergence of $Ay_n + g(y_n) - f - v_n$ to $A\tilde{y} + g(\tilde{y}) - f - \tilde{v}$ in $L^2(\Omega)$ and the strong convergence of y_n to $\tilde{y}$ in $H_0^1(\Omega)$. □

So that, we can consider that problem $(\mathcal{P}^\alpha)$ is a "good" approximation of the original problem $(\mathcal{P})$ in the following sense:

Theorem 3.3.6. *For any $\alpha > 0$, $(\mathcal{P}^\alpha)$ has at least one optimal solution (denoted $(y_\alpha, v_\alpha, \xi_\alpha)$). Moreover, when α goes to 0, y_α strongly converges to $\tilde{y}$ in $H_0^1(\Omega)$ (up to a subsequence), v_α strongly converges to $\tilde{v}$ in $L^2(\Omega)$ (up to a subsequence), ξ_α weakly converges to $\tilde{\xi}$ in $L^2(\Omega)$ (up to a subsequence), where $(\tilde{y}, \tilde{v}, \tilde{\xi})$ is a solution of $(\mathcal{P})$.*

Proof. Let (y_n, v_n, ξ_n) be a minimizing sequence such that $J(y_n, v_n)$ converges to $d^\alpha = \inf(\mathcal{P}^\alpha)$. As $J(y_n, v_n)$ is bounded, there exists a constant C such that we have:

$$\forall n \quad || v_n ||_2 \leqslant C.$$

So, we may extract a subsequence (denoted similarly) such that v_n converges to v_α weakly in $L^2(\Omega)$ and strongly in $H^{-1}(\Omega)$. As U_{ad} is a closed convex set, it is weakly closed and $v_\alpha \in U_{ad}$. On the other hand, we have $Ay_n + g(y_n) - f - v_n = \xi_n$ a.e. in Ω. So

$$\langle Ay_n, y_n\rangle + \langle g(y_n), y_n\rangle = \langle f + v_n, y_n\rangle + \langle y_n, \xi_n\rangle.$$

In view of Lemma 3.3.4, we have:

$$\forall y_n \geqslant 0,\ \forall \xi_n \geqslant 0, \qquad \frac{y_n}{y_n + \alpha} + \frac{\xi_n}{\xi_n + \alpha} \leqslant 1 \Longleftrightarrow y_n\xi_n \leqslant \alpha^2,$$

the integral by the two ways, gives

$$\frac{y_n}{y_n + \alpha} + \frac{\xi_n}{\xi_n + \alpha} \leqslant 1 \Longleftrightarrow y_n\xi_n \leqslant \alpha^2 \Longrightarrow \langle y_n, \xi_n\rangle \leqslant \alpha^2 \text{Area}(\Omega).$$

So

$$\langle Ay_n, y_n\rangle + \langle g(y_n), y_n\rangle = \langle f + v_n, y_n\rangle + \langle y_n, \xi_n\rangle \leqslant \langle f + v_n, y_n\rangle + \alpha^2 \text{Area}(\Omega).$$

The monotonicity of g gives

$$\langle Ay_n, y_n\rangle \leqslant \langle Ay_n, y_n\rangle + \langle g(y_n) - g(0), y_n\rangle \leqslant \langle f + v_n - g(0), y_n\rangle + \alpha^2 \text{Area}(\Omega).$$

Using the coercivity of A, we obtain

$$\begin{aligned} \delta \,||\, y_n \,||^2_{H^1_0(\Omega)} &\leqslant ||\, f + v_n - g(0) \,||^2_{H^{-1}(\Omega)} ||\, y_n \,||_{H^1_0(\Omega)} + \alpha^2 \text{Area}(\Omega) \\ &\leqslant C \,||\, y_n \,||_{H^1_0(\Omega)} + \alpha^2 \text{Area}(\Omega). \end{aligned} \tag{3.3.5}$$

This yields that y_n is bounded in $H^1_0(\Omega)$, since Ω is bounded, so y_n converges to y_α weakly in $H^1_0(\Omega)$ and strongly in $L^2(\Omega)$. Moreover as $y_n \in K$, and K is a closed convex set, K is weakly closed and $y_\alpha \in K$. We have assumed that V_{ad} is $L^2(\Omega)$-bounded. So, we can apply Lemma 3.3.5 and obtain that ξ_n weakly converges to $\xi_\alpha = Ay_\alpha + g(y_\alpha) - f - v_\alpha \in V_{ad}$ in $L^2(\Omega)$.

Remark 3.3.7. *$\xi_n = Ay_n + g(y_n) - f - v_n$, weakly converges to $\xi_\alpha = Ay_\alpha + g(y_\alpha) - f - v_\alpha$ in $H^{-1}(\Omega)$. Unfortunately the weak convergence of ξ_n to ξ_α in $H^{-1}(\Omega)$ is not sufficient to conclude.*

We need this sequence to converge weakly in $L^2(\Omega)$. That is the reason why we have bounded ξ_n in $L^2(\Omega)$.

At last, $\frac{y_n}{y_n+\alpha} + \frac{\xi_n}{\xi_n+\alpha}$ converges to $\frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha}$ because of the strong convergence of y_n in $L^2(\Omega)$ and the weak convergence of ξ_n in $L^2(\Omega)$ and we obtain
$\frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha} \leqslant 1$, we just prove that $(y_\alpha, v_\alpha, \xi_\alpha) \in \mathcal{D}_\alpha$. The weak convergence and the lower semi-continuity of J give:

$$d^\alpha = \liminf_{n\to\infty} J(y_n, v_n) \geqslant J(y_\alpha, v_\alpha) \geqslant d^\alpha.$$

So $J(y_\alpha, v_\alpha) = d^\alpha$ and $(y_\alpha, v_\alpha, \xi_\alpha)$ is a solution of $(\mathcal{P}^\alpha)$.

- Now, let us prove the second part of the theorem, we need to proof when α goes to 0 we have y_α strongly converges to $\tilde{y}$ in $H_0^1(\Omega)$, v_α strongly converges to $\tilde{v}$ in $L^2(\Omega)$ and ξ_α weakly converges to $\tilde{\xi}$ in $L^2(\Omega)$, where $(\tilde{y}, \tilde{v}, \tilde{\xi})$ is a solution of $(\mathcal{P})$.

First we note that $(\bar{y}, \bar{v}, \bar{\xi})$ belongs to $\mathcal{D}^\alpha$ for any $\alpha > 0$. So:

$$\forall \alpha > 0 \quad J(y_\alpha, v_\alpha) \leqslant J(\bar{y}, \bar{v}) < +\infty, \tag{3.3.6}$$

and v_α and y_α are bounded respectively in $L^2(\Omega)$ and $H_0^1(\Omega)$. Indeed, we use the previous arguments since v_α is bounded in $L^2(\Omega)$ and

$$\begin{aligned} \delta \mid\mid y_\alpha \mid\mid^2_{H_0^1(\Omega)} &\leqslant \mid\mid f + v_\alpha - g(0) \mid\mid^2_{H^{-1}(\Omega)} \mid\mid y_\alpha \mid\mid_{H_0^1(\Omega)} + \alpha^2 \text{Area}(\Omega) \\ &\leqslant C \mid\mid y_\alpha \mid\mid_{H_0^1(\Omega)} + \alpha^2 \text{Area}(\Omega). \end{aligned} \tag{3.3.7}$$

So (extracting a subsequence) v_α weakly converges to some $\tilde{v}$ in $L^2(\Omega)$ and y_α converges to some $\tilde{y}$ weakly in $H_0^1(\Omega)$ and strongly in $L^2(\Omega)$.
As above, it is easy to see that ξ_α weakly converges to $\tilde{\xi} = A\tilde{y} + g(\tilde{y}) - f - \tilde{v}$ in $L^2(\Omega)$ (thanks Lemma 3.3.5), and that $\tilde{y} \in K,\ \tilde{v} \in U_{ad},\ \tilde{\xi} \in V_{ad}$. In the same way $\frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha}$ converges to $\frac{\tilde{y}}{\tilde{y}+\alpha} + \frac{\tilde{\xi}}{\tilde{\xi}+\alpha}$.
As $0 \leqslant \frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha} \leqslant 1$, from Lemma 3.3.4 we get:

$$0 \leqslant \quad \frac{y_\alpha}{y_\alpha + \alpha} + \frac{\xi_\alpha}{\xi_\alpha + \alpha} \quad \leqslant 1 \Longleftrightarrow 0 \leqslant y_\alpha \xi_\alpha \leqslant \alpha^2,$$

at the limit as $\alpha \searrow 0$ this implies that $\tilde{y}\tilde{\xi} = 0 \Longleftrightarrow \langle \tilde{y}, \tilde{\xi} \rangle = 0$. So $(\tilde{y},\ \tilde{v},\ \tilde{\xi}) \in \mathcal{D}$. This yields that

$$J(\bar{y}, \bar{v}) \leqslant J(\tilde{y}, \tilde{v}). \tag{3.3.8}$$

Once again, we may pass to the inf-limite in (3.3.6) to obtain:

$$J(\tilde{y},\tilde{v}) \leqslant \liminf_{\alpha\to 0} J(y_\alpha, v_\alpha) \leqslant J(\bar{y},\bar{v}).$$

This implies that

$$J(\tilde{y},\tilde{v}) = J(\bar{y},\bar{v}),$$

therefore $(\tilde{y},\tilde{v},\tilde{\xi})$ is a solution of $(\mathcal{P})$. Moreover, as $\lim_{\alpha\to 0} J(y_\alpha, v_\alpha) = J(\tilde{y},\tilde{v})$ and y_α strongly converges to $\tilde{y}$ in $L^2(\Omega)$, we get $\lim_{\alpha\to 0} || v_\alpha ||_2 = || \tilde{v} ||_2$, so that v_α strongly converges to $\tilde{v}$ in $L^2(\Omega)$. We already know that ξ_α weakly converges to $\tilde{\xi}$ in $L^2(\Omega)$. So

$$\xi_\alpha + v_\alpha - g(y_\alpha) + f = Ay_\alpha \quad \text{converges to} \quad \tilde{\xi} + \tilde{v} - g(\tilde{y}) + f = A\tilde{y},$$

weakly in $L^2(\Omega)$ and strongly in $H^{-1}(\Omega)$. As A is an isomorphism from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$ this yields that y_α strongly converges to $\tilde{y}$ in $H_0^1(\Omega)$. □

We see then, that solutions of problem $(\mathcal{P}^\alpha)$ are "good " approximations of the solution of problem $(\mathcal{P})$.
Now, we would like to derive optimality conditions for the problem $(\mathcal{P}^\alpha)$, for $\alpha > 0$. In the sequel, we study the unconstrained control case: $U_{ad} = L^2(\Omega)$. We first present some mathematical programming tools that allow proving the existence of Lagrange multipliers.

3.4 The mathematical programming point of view

The non-convexity of the feasible domain does not allow to use convex analysis to get the existence of Lagrange multipliers. So we are going to use quite general mathematical programming methods in Banach spaces and adapt them to our framework.
The following results are mainly due to Zowe and Kurcyusz [111] and Troltzsch [102] and we briefly present them in the following.
We will work in real Banach spaces $\mathcal{X}$, $\mathcal{U}$, $\mathcal{Z}_1$ and $\mathcal{Z}_2$. Our admissible set $\mathcal{U}_{ad}$ is a convex closed subset of $\mathcal{U}$.
In the definition of the forthcoming problem, f is a real function defined on $\mathcal{X}\times\mathcal{U}$. We assume that f is Fréchet-differentiable. T and G are Fréchet continuously differentiable transformations of $\mathcal{X}\times\mathcal{U}$ into $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively. We also consider that $\mathcal{Z}_2$ is partially ordered with respect to some given convex closed cone $P \subseteq \mathcal{Z}_2$, i.e. $x \geqslant y \Leftrightarrow x - y \in P$.

Now, consider the mathematical programming problem defined by:

$$\min \{f(x,u) \mid T(x,u)=0,\ G(x,u) \leqslant 0,\ u \in \mathcal{U}_{ad}\}. \tag{3.4.1}$$

We denote the partial Fréchet-derivative of f, T, and G with respect to x and u by a corresponding index x or u. We suppose that the problem (3.4.1) has an optimal solution that we call (x_0, u_0), and we introduce the sets:

$$\begin{aligned}
\mathcal{U}_{ad}(u_0) =& \{u \in \mathcal{U} \mid \exists \lambda \geqslant 0,\ \exists u^* \in \mathcal{U}_{ad},\ u = \lambda(u^* - u_0)\},\\
P(G(x_0,u_0)) =& \{z \in \mathcal{Z}_2 \mid \exists \lambda \geqslant 0,\ \exists p \in -P,\ z = p - \lambda G(x_0,u_0)\},\\
P^+ =& \{y \in \mathcal{Z}_2^* \mid \langle y, p\rangle \geqslant 0,\ \forall p \in P\}.
\end{aligned}$$

One may now announce the main result about the existence of optimality conditions.

Theorem 3.4.1. *[111] Let u_0 be an optimal control with corresponding optimal state x_0 and suppose that the following regularity condition is fulfilled:*

$$\forall (z_1,z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2 \quad \textit{the system} \quad \begin{cases} T'(x_0,u_0)(x,u) & = z_1,\\ G'(x_0,u_0)(x,u) - p & = z_2, \end{cases} \tag{3.4.2}$$

is solvable with $(x,u,p) \in \mathcal{X} \times \mathcal{U}_{ad}(u_0) \times P(G(x_0,u_0))$.
Then a "Lagrange multiplier" $(y_1,y_2) \in \mathcal{Z}_1^* \times \mathcal{Z}_2^*$ exists such that

$$f'_x(x_0,u_0) + T'_x(x_0,u_0) * y_1 + G'_x(x_0,u_0) * y_2 = 0, \tag{3.4.3}$$

$$\Big(f'_x(x_0,u_0) + T'_x(x_0,u_0) * y_1 + G'_x(x_0,u_0) * y_2, u - u_0\Big)_{\mathcal{X}\times\mathcal{U}} \geqslant 0, \quad \forall u \in \mathcal{U}_{ad}, \tag{3.4.4}$$

$$y_2 \in P^+, \quad (y_2, G(x_0,u_0))_{\mathcal{Z}_2^* \times \mathcal{Z}_2} = 0. \tag{3.4.5}$$

Mathematical programming theory in Banach spaces allows us to study problems where the feasible domain is not convex, this is precisely our case (and we cannot use the classical convex theory and the Gâteaux differentiability to derive some optimality conditions). The Zowe and Kurcyusz condition (3.4.2) is a very weak condition to ensure the existence of Lagrange multipliers. It is natural to try to see if this condition is satisfied for the original problem $(\mathcal{P})$, unfortunately, it is impossible (see [9]) and this is another justification (from a theoretical point of view) of the fact that we have to take $\mathcal{D}_\alpha$ instead of $\mathcal{D}$.
On the other hand, if we apply the previous general result "directly" to $(\mathcal{P}^\alpha)$ we obtain a complicated qualification condition (3.4.2) which seems difficult to ensure.
So, we would rather mix these "mathematical-programming methods" with a penalization method in

order to "relax" the state-equation as well and make the qualification condition weaker and simpler.

3.5 Penalization approach

3.5.1 The penalized problem

One of the difficulties comes from the fact that we have a coupled system. It would be easier if we had only one condition. In order to split the different constraints and make them "independent", we penalize the state equation to obtain an optimization problem with non-convex constraints. Then we apply the previous method to get optimality conditions for the penalized problem. Of course, we may decide to penalize the bilinear constraint instead of the state equation: this leads to the same results (see [7, 11] for example).
Moreover, we focus on the solution $(y_\alpha, v_\alpha, \xi_\alpha)$, so, following Barbu [7], we add some adapted penalization terms to the objective functional J.
From now on, $\alpha > 0$ is fixed, so we omit the index α when no confusion is possible. For any $\varepsilon > 0$ we define a penalized functional J_ε^α on $(H^2(\Omega) \cap H_0^1(\Omega)) \times L^2(\Omega) \times L^2(\Omega))$ as follows:

$$J_\varepsilon^\alpha(y,v,\xi) = \begin{cases} J(y,v) + \dfrac{1}{2\varepsilon} \mid\mid Ay + g(y) - f - v - \xi \mid\mid_2^2 \\ \qquad + \dfrac{1}{2} \mid\mid A(y - y_\alpha) \mid\mid_2^2 + \dfrac{1}{2} \mid\mid v - v_\alpha \mid\mid_2^2 \\ \qquad + \dfrac{1}{2} \mid\mid \xi - \xi_\alpha \mid\mid_2^2 \end{cases} \tag{3.5.1}$$

and we consider the penalized optimization problem

$$\min \{J_\varepsilon^\alpha(y,v,\xi) \mid (y,v,\xi) \in \mathcal{D}_\alpha,\ y \in H^2(\Omega) \cap H_0^1(\Omega)\} \tag{$\mathcal{P}_\alpha^\varepsilon$}$$

Theorem 3.5.1. *The penalized problem* $(\mathcal{P}_\alpha^\varepsilon)$ *has at least a solution* $(y_\varepsilon, v_\varepsilon, \xi_\varepsilon) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times L^2(\Omega) \times L^2(\Omega)$.

Proof. The proof is almost the same as the one of Theorem 3.3.6. The main difference is that we have no longer $Ay_n + g(y_n) - f - v_n - \xi_n = 0$, for any minimizing sequence. Anyway, y_n, v_n, ξ_n, Ay_n and $g(y_n)$ are bounded in $L^2(\Omega)$, and it is standard to see that any weak-cluster point of this minimizing sequence is feasible and is a solution to the problem,

$$Ay_n + g(y_n) - f - v_n - \xi_n \rightharpoonup 0, \quad \text{weakly in } \ L^2(\Omega).$$

□

Now, we may also give a result concerning the asymptotic behavior of the solutions of the penalized problems.

Theorem 3.5.2. *When ε goes to 0, $(y_\varepsilon, v_\varepsilon, \xi_\varepsilon)$ strongly converges to $(y_\alpha, v_\alpha, \xi_\alpha) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times L^2(\Omega) \times L^2(\Omega)$.*

Proof. The proof is quite similar to the one of Theorem 3.3.6. We have:

$$\forall \varepsilon > 0 \qquad J_\varepsilon^\alpha(y_\varepsilon, v_\varepsilon, \xi_\varepsilon) \leqslant J_\varepsilon^\alpha(y_\alpha, v_\alpha, \xi_\alpha) = J(y_\alpha, v_\alpha) = j_\alpha < +\infty. \tag{3.5.2}$$

So

$$\begin{aligned} 2J(y_\varepsilon, v_\varepsilon) + \frac{1}{\varepsilon} \mid\mid Ay_\varepsilon + g(y_\varepsilon) - f - v_\varepsilon - \xi_\varepsilon \mid\mid_2^2 + \mid\mid A(y_\varepsilon - y_\alpha) \mid\mid_2^2 \\ + \mid\mid v_\varepsilon - v_\alpha \mid\mid_2^2 + \mid\mid \xi_\varepsilon - \xi_\alpha \mid\mid_2^2 \\ \leqslant \ 2j_\alpha. \end{aligned} \tag{3.5.3}$$

Therefore $v_\varepsilon, Ay_\varepsilon$ and ξ_ε are $L^2(\Omega)$-bounded; this yields that $Ay_\varepsilon + g(y_\varepsilon) - f - v_\varepsilon$ is $L^2(\Omega)$-bounded and y_ε is $H^2(\Omega) \cap H_0^1(\Omega)$-bounded. So, using Lemma 3.3.5, we conclude that

(i) v_ε converges to some $\tilde{v}$ weakly in $L^2(\Omega)$;

(ii) y_ε converges to some $\tilde{y}$ strongly in $H_0^1(\Omega)$;

(iii) ξ_ε converges to some $\tilde{\xi}$ weakly in $L^2(\Omega)$;

(iv) $Ay_\varepsilon + g(y_\varepsilon) - f - v_\varepsilon - \xi_\varepsilon$ converges to $A\tilde{y} + g(\tilde{y}) - f - \tilde{v} - \tilde{\xi}$ weakly in $L^2(\Omega)$.

Moreover, the inequality

$$\mid\mid Ay_\varepsilon + g(y_\varepsilon) - f - v_\varepsilon - \xi_\varepsilon \mid\mid_2^2 \leqslant 2\varepsilon j_\alpha,$$

implies the strong convergence of $Ay_\varepsilon + g(y_\varepsilon) - f - v_\varepsilon - \xi_\varepsilon$ to 0 in $L^2(\Omega)$. Therefore $A\tilde{y} + g(\tilde{y}) = f + \tilde{v} + \tilde{\xi}$. It is easy to see that $\tilde{y} \in K$, $\tilde{v} \in U_{ad}$ and $\tilde{\xi} \in V_{ad}$. Moreover, as y_ε converges to $\tilde{y}$ strongly in $L^2(\Omega)$ and ξ_ε converges to $\tilde{\xi}$ weakly in $L^2(\Omega)$, we know that $\frac{y_\varepsilon}{y_\varepsilon + \alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon + \alpha} (\leqslant 1)$ converges to $\frac{\tilde{y}}{\tilde{y} + \alpha} + \frac{\tilde{\xi}}{\tilde{\xi} + \alpha}$. So $\frac{\tilde{y}}{\tilde{y} + \alpha} + \frac{\tilde{\xi}}{\tilde{\xi} + \alpha} \leqslant 1$ and $(\tilde{y}, \tilde{v}, \tilde{\xi})$ belongs to $\mathcal{D}_\alpha$. Relation (3.5.2) implies that

$$J(y_\varepsilon, v_\varepsilon) + \frac{1}{2} \mid\mid A(y_\varepsilon - y_\alpha) \mid\mid_2^2 + \frac{1}{2} \mid\mid v_\varepsilon - v_\alpha \mid\mid_2^2 + \frac{1}{2} \mid\mid \xi_\varepsilon - \xi_\alpha \mid\mid_2^2 \leqslant J(y_\alpha, v_\alpha). \tag{3.5.4}$$

Passing to the inf-limit and using the fact that $(\tilde{y}, \tilde{v}, \tilde{\xi})$ belongs to $\mathcal{D}_\alpha$, we obtain

$$\begin{aligned} J(\tilde{y},\tilde{v}) + \frac{1}{2} \mid\mid A(\tilde{y} - y_\alpha) \mid\mid_2^2 + \frac{1}{2} \mid\mid \tilde{v} - v_\alpha \mid\mid_2^2 + \frac{1}{2} \mid\mid \tilde{\xi} - \xi_\alpha \mid\mid_2^2 \\ \leqslant J(y_\alpha, v_\alpha) \\ \leqslant J(\tilde{y}, \tilde{v}). \end{aligned} \tag{3.5.5}$$

Therefore, $A(\tilde{y} - y_\alpha) = 0$ (which implies $\tilde{y} = y_\alpha$ since $A(\tilde{y} - y_\alpha) \in H_0^1(\Omega)$), $\tilde{v} = v_\alpha$ and $\tilde{\xi} = \xi_\alpha$. We just prove the weak convergence of $(y_\varepsilon, v_\varepsilon, \xi_\varepsilon)$ to $(y_\alpha, v_\alpha, \xi_\alpha)$ in $H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$, and that $\lim_{\varepsilon \to 0} J(y_\varepsilon, v_\varepsilon) = J(y_\alpha, v_\alpha)$. Relation (3.5.4) gives

$$\mid\mid A(y_\varepsilon - y_\alpha) \mid\mid_2^2 + \mid\mid v_\varepsilon - v_\alpha \mid\mid_2^2 + \mid\mid \xi_\varepsilon - \xi_\alpha \mid\mid_2^2 \leqslant 2[J(y_\alpha, v_\alpha) - J(y_\varepsilon, v_\varepsilon)], \tag{3.5.6}$$

therefore we get the strong convergence of Ay_ε towards Ay_α in $L^2(\Omega)$, that is the strong convergence of y_ε to y_α in $H^2(\Omega) \cap H_0^1(\Omega)$. We get also the strong convergence of $(v_\varepsilon, \xi_\varepsilon)$ towards (v_α, ξ_α) in $L^2(\Omega) \times L^2(\Omega)$. Let us remark, at last, that y_ε converges to y_α uniformly in $\bar{\Omega}$, since $H^2(\Omega) \cap H_0^1(\Omega) \subset \mathcal{C}(\bar{\Omega})$. □

Corollary 3.5.3. *If we define the adjoint state p_ε of the penalized problem as the solution of*

$$A^* p_\varepsilon + g'(y_\varepsilon) p_\varepsilon = y_\varepsilon - z_d \quad on \quad \Omega, \; p_\varepsilon \in H_0^1(\Omega), \tag{3.5.7}$$

then p_ε strongly converges to p_α in $H_0^1(\Omega)$, where p_α satisfies

$$A^* p_\alpha + g'(y_\alpha) p_\alpha = y_\alpha - z_d \quad on \quad \Omega. \tag{3.5.8}$$

Proof. We have seen that $\mid\mid y_\varepsilon - y_\alpha \mid\mid_\infty \to 0$. Therefore y_ε remains in a bounded set of $\mathbb{R}^n$ (independent of ε). As g is a $\mathcal{C}^1$ function, this means that $\mid\mid g'(y_\varepsilon) \mid\mid_\infty$ is bounded by a constant C which does not depend on ε. In particular $g'(y_\varepsilon)$ is bounded in $L^2(\Omega)$ and Lebesgue's theorem implies the strong convergence of $g'(y_\varepsilon)$ to $g'(y_\alpha)$ in $L^2(\Omega)$. Let p_ε be the solution of (3.5.7), this gives

$$\langle A^* p_\varepsilon, p_\varepsilon \rangle + \langle g'(y_\varepsilon) p_\varepsilon, p_\varepsilon \rangle = \langle y_\varepsilon - z_d, p_\varepsilon \rangle,$$

as $g' \geqslant 0$ and A^* is coercive we get

$$\delta \mid\mid p_\varepsilon \mid\mid_{H_0^1(\Omega)}^2 \leqslant \mid\mid y_\varepsilon - z_d \mid\mid_{H^{-1}(\Omega)} \mid\mid p_\varepsilon \mid\mid_{H_0^1(\Omega)} .$$

So, p_ε is bounded in $H_0^1(\Omega)$ and weakly converges to $\tilde{p}$ in $H_0^1(\Omega)$. Moreover, p_ε is the solution to

$$A^* p_\varepsilon = -g^{'}(y_\varepsilon)p_\varepsilon + y_\varepsilon - z_d \quad \text{on} \quad \Omega,$$

the left-hande side (weakly) converges to $-g^{'}(y_\alpha)\tilde{p} + y_\alpha - z_d$ in $L^2(\Omega)$; this completes the proof. □

3.5.2 Optimality conditions for the penalized problem

We apply Theorem 3.4.1 to the penalized problem $(\mathcal{P}_\alpha^\varepsilon)$.
We set

$$x = y, \quad u = (v, \xi), \quad (x_0, u_0) = (x_\varepsilon, v_\varepsilon, \xi_\varepsilon),$$
$$\mathcal{X} = H^2(\Omega) \cap H_0^1(\Omega), \quad \mathcal{Z}_2 = \mathcal{X} \text{ and } \mathcal{U} = L^2(\Omega) \times L^2(\Omega), \ \mathcal{U}_{ad} = U_{ad} \times V_{ad},$$
$$P = \{y \in H^2(\Omega) \cap H_0^1(\Omega) \mid y \geqslant 0\} \times \mathbb{R}^+, \ f(x,u) = J_\varepsilon^\alpha(y, v, \xi),$$
$$G(y,u) = [G_1(y,v,\xi); G_2(y,v,\xi)] = \left[-y; \left(1, \frac{y}{y+\alpha} + \frac{\xi}{\xi+\alpha}\right)_2 - \text{Area}(\Omega)\right],$$

there is no equality constraint and G is $\mathcal{C}^1$,

$$G^{'}(y_\varepsilon, v_\epsilon, \xi_\epsilon)(y,v,\xi) = \left[-y; \left(y, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2 + \left(\xi, \frac{\alpha}{(\xi_\varepsilon+\alpha)^2}\right)_2\right].$$

Here

$$\mathcal{U}_{ad}(v_\varepsilon, \xi_\varepsilon) = \{(\lambda v - \lambda v_\varepsilon, \ \mu\xi - \mu\xi_\varepsilon) \ / \ \lambda \geqslant 0, \ \mu \geqslant 0, \ v \in U_{ad}, \ \xi \in V_{ad}\},$$
$$P(G(y_\varepsilon, v_\varepsilon, \xi_\varepsilon)) = \Bigg\{ \left[-p + \lambda y_\varepsilon \ ; \ -\gamma - \lambda\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 + \lambda\text{Area}(\Omega)\right]$$
$$\in H^2(\Omega) \cap H_0^1(\Omega) \times \mathbb{R} \ / \ \gamma, \ \lambda \geqslant 0, \ p \geqslant 0 \Bigg\}.$$

Let us write the condition (3.4.2). For any (z, β) in $\mathcal{X} \times \mathbb{R}$ we must solve the system:

$$\begin{aligned} -y + p - \lambda y_\varepsilon &= z, \\ \left(y, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2 + \left(\mu(\xi - \xi_\varepsilon), \frac{\alpha}{(\xi_\varepsilon+\alpha)^2}\right)_2 + \gamma + \lambda\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 - \lambda\text{Area}(\Omega) &= \beta, \end{aligned}$$

with $\mu, \ \gamma, \ \lambda \geqslant 0, \ \xi \in V_{ad}, \ v \in U_{ad}$, and $y \in \mathcal{X}$. Taking y from the first equation into the second we have to solve:

$$\left(p - \lambda y_\varepsilon - z, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2 + \left(\mu(\xi - \xi_\varepsilon), \frac{\alpha}{(\xi_\varepsilon+\alpha)^2}\right)_2 + \gamma$$
$$+\lambda\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 - \lambda\text{Area}(\Omega) = \beta.$$

So

$$\left(p, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2 - \lambda\left(y_\varepsilon, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2 + \left(\mu(\xi-\xi_\varepsilon), \frac{\alpha}{(\xi_\varepsilon+\alpha)^2}\right)_2$$
$$+\gamma + \lambda\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 - \lambda\text{Area}(\Omega)$$
$$=\beta + \left(z, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2 = \rho,$$

with $\mu,\ \gamma,\ \lambda \geqslant 0,\ \xi \in V_{ad},\ v \in U_{ad}$. We see that we may take: $\mu = 1,\ \xi = \xi_\varepsilon,\ p = 0$, and if $\rho \geqslant 0$, we choose $\lambda = 0,\ \gamma = \rho$. If $\rho < 0$, we have two cases:

- If
$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 - \text{Area}(\Omega) = \zeta < 0,$$
then we set $\gamma = \lambda\left(y_\varepsilon, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2,\ \lambda = \frac{\rho}{\zeta}$.

- If
$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 - \text{Area}(\Omega) = 0,$$
then we set $\gamma = 0,\ \lambda = -\frac{\rho}{\eta}$, such that $\eta = \lambda\left(y_\varepsilon, \frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2$.
Indeed, we have
$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 - \text{Area}(\Omega) = 0,$$
in view of Lemma 3.3.4, we have
$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 - \text{Area}(\Omega) = 0 \Longleftrightarrow y_\varepsilon.\xi_\varepsilon = \alpha^2 \text{ a.e. in } \Omega.$$
Therefore y_ε and ξ_ε are strictly positive. (Since $\alpha > 0$ fixed). Hence, $\eta > 0$ and $\lambda > 0$.

So, condition (3.4.2) is always satisfied and we may apply Theorem 3.4.1, since J_ε^α is Fréchet differentiable, and

$$J_\varepsilon^{\alpha'}(y_\varepsilon, v_\varepsilon, \xi_\varepsilon)(y, v, \xi)$$
$$= \begin{pmatrix} (J_\varepsilon^\alpha)'_y(y_\varepsilon, v_\varepsilon, \xi_\varepsilon) & (J_\varepsilon^\alpha)'_v(y_\varepsilon, v_\varepsilon, \xi_\varepsilon) & (J_\varepsilon^\alpha)'_\xi(y_\varepsilon, v_\varepsilon, \xi_\varepsilon) \end{pmatrix} \cdot \begin{pmatrix} y \\ v \\ \xi \end{pmatrix},$$

we have:

$$J_\varepsilon^\alpha(y,v,\xi)=\begin{cases} J(y,v)+\dfrac{1}{2\varepsilon}\,||\,Ay+g(y)-f-v-\xi\,||_2^2 \\ \qquad +\dfrac{1}{2}\,||\,A(y-y_\alpha)\,||_2^2+\dfrac{1}{2}\,||\,v-v_\alpha\,||_2^2 \\ \qquad +\dfrac{1}{2}\,||\,\xi-\xi_\alpha\,||_2^2\,. \end{cases} \tag{3.5.9}$$

So,

$$\begin{aligned} (J_\varepsilon^\alpha)_y^{'}(y_\varepsilon,v_\varepsilon,\xi_\varepsilon) &= (1,y_\varepsilon-z_d)_2+\frac{1}{\varepsilon}\Big(A+g^{'}(y_\varepsilon),Ay_\varepsilon+g(y_\varepsilon)-f-v_\varepsilon-\xi_\varepsilon\Big)_2 \\ &\quad +(A,A(y_\varepsilon-y_\alpha))_2\,. \\ (J_\varepsilon^\alpha)_v^{'}(y_\varepsilon,v_\varepsilon,\xi_\varepsilon) &= (\nu,v_\varepsilon-v_d)_2+(1,v_\varepsilon-v_\alpha)_2 \\ &\quad -\frac{1}{\varepsilon}\,(1,Ay_\varepsilon+g(y_\varepsilon)-f-v_\varepsilon-\xi_\varepsilon)_2\,. \\ (J_\varepsilon^\alpha)_\xi^{'}(y_\varepsilon,v_\varepsilon,\xi_\varepsilon) &= (1,\xi_\varepsilon-\xi_\alpha)_2-\frac{1}{\varepsilon}\,(1,Ay_\varepsilon+g(y_\varepsilon)-f-v_\varepsilon-\xi_\varepsilon)_2\,. \end{aligned}$$

Therefore

$$\begin{aligned} J_\varepsilon^{\alpha^{'}}(y_\varepsilon,v_\varepsilon,\xi_\varepsilon)(y,v,\xi) &= (y,y_\varepsilon-z_d)_2+\nu\,(v,v_\varepsilon-v_d)_2+(v,v_\varepsilon-v_\alpha)_2 \\ &\quad +(\xi,\xi_\varepsilon-\xi_\alpha)_2+(Ay,A(y_\varepsilon-y_\alpha))_2+(q_\varepsilon,A_\varepsilon y-v-\xi)_2\,, \end{aligned}$$

where

$$q_\varepsilon=\frac{Ay_\varepsilon+g(y_\varepsilon)-f-v_\varepsilon-\xi_\varepsilon}{\varepsilon}\in L^2(\Omega)\qquad\text{and}\qquad A_\varepsilon=A+g^{'}(y_\varepsilon). \tag{3.5.10}$$

There exists $s_\varepsilon\in\mathcal{X}^*$ and $r_\varepsilon\in\mathbb{R}$ such that:

$$\forall y\in\mathcal{X}\quad (y,y_\varepsilon-z_d)_2+(q_\varepsilon,A_\varepsilon y)_2+(Ay,A(y_\varepsilon-y_\alpha))_2+r_\varepsilon\left(y,\frac{\alpha}{(y_\varepsilon+\alpha)^2}\right)_2-\langle s_\varepsilon,y\rangle_{\mathcal{X}^*,\mathcal{X}}=0, \tag{3.5.11}$$

$$\forall v\in U_{ad}\quad (\nu(v_\varepsilon-v_d)+v_\varepsilon-v_\alpha-q_\varepsilon,v-v_\varepsilon)_2\geqslant 0, \tag{3.5.12}$$

$$\forall\xi\in V_{ad}\quad \left(r_\varepsilon\frac{\alpha}{(\xi_\alpha+\alpha)^2}-q_\varepsilon+\xi_\varepsilon-\xi_\alpha,\xi-\xi_\varepsilon\right)_2\geqslant 0, \tag{3.5.13}$$

$$r_\varepsilon\geqslant 0,\quad r_\varepsilon\left[\left(1,\frac{y_\varepsilon}{y_\varepsilon+\alpha}+\frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2-\mathrm{Area}(\Omega)\right]=0, \tag{3.5.14}$$

$$\forall y\in\mathcal{X},\ y\geqslant 0,\quad \langle s_\varepsilon,y\rangle_{\mathcal{X}^*,\mathcal{X}}\geqslant 0,\ \ \langle s_\varepsilon,y_\varepsilon\rangle_{\mathcal{X}^*,\mathcal{X}}=0. \tag{3.5.15}$$

Here $\langle .,.\rangle_{\mathcal{X}^*,\mathcal{X}}$ denotes the duality product between $\mathcal{X}$ and the dual space $\mathcal{X}^*$.

Finally, we can state optimality conditions for the penalized problem, without any further assumption:

Theorem 3.5.4. *The solution $(y_\varepsilon, v_\varepsilon, \xi_\varepsilon)$ of problem $(\mathcal{P}_\varepsilon^\alpha)$ satisfies the following optimality system:*

$$\begin{aligned}\forall y \in \tilde{K} \quad & (p_\varepsilon + q_\varepsilon, A_\varepsilon(y - y_\varepsilon))_2 + (A(y - y_\varepsilon), A(y_\varepsilon - y_\alpha))_2 \\ & + r_\varepsilon \left(y - y_\varepsilon, \frac{\alpha}{(y_\varepsilon + \alpha)^2} \right)_2 \geqslant 0,\end{aligned} \tag{3.5.16}$$

$$\forall v \in U_{ad} \quad (\nu(v_\varepsilon - v_d) + v_\varepsilon - v_\alpha - q_\varepsilon, v - v_\varepsilon)_2 \geqslant 0, \tag{3.5.17}$$

$$\forall \xi \in V_{ad} \quad \left(r_\varepsilon \frac{\alpha}{(\xi_\alpha + \alpha)^2} - q_\varepsilon + \xi_\varepsilon - \xi_\alpha, \xi - \xi_\varepsilon \right)_2 \geqslant 0, \tag{3.5.18}$$

$$r_\varepsilon \geqslant 0, \quad r_\varepsilon \left[\left(1, \frac{y_\varepsilon}{y_\varepsilon + \alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon + \alpha} \right)_2 - Area(\Omega) \right] = 0, \tag{3.5.19}$$

where p_ε is given by (3.5.7) and q_ε by (3.5.10), and $\tilde{K} = K \cap (H^2(\Omega) \cap H_0^1(\Omega))$.

Proof. Relation (3.5.11) applied to $y - y_\varepsilon$ gives:

$$\begin{aligned}\forall y \in \mathcal{X} \quad & (y - y_\varepsilon, y_\varepsilon - z_d)_2 + (q_\varepsilon, A_\varepsilon(y - y_\varepsilon))_2 + (A(y - y_\varepsilon), A(y_\varepsilon - y_\alpha))_2 \\ & + r_\epsilon \left(y - y_\varepsilon, \frac{\alpha}{(y_\varepsilon + \alpha)^2} \right)_2 \\ & = \langle s_\varepsilon, y \rangle_{\mathcal{X}^*, \mathcal{X}} - \langle s_\varepsilon, y_\varepsilon \rangle_{\mathcal{X}^*, \mathcal{X}}.\end{aligned}$$

So, with (3.5.15), we obtain

$$\forall y \in \tilde{K} \quad (p_\varepsilon + q_\varepsilon, A_\varepsilon(y - y_\varepsilon))_2 + (A(y - y_\varepsilon), A(y_\varepsilon - y_\alpha))_2 + r_\varepsilon \left(y - y_\varepsilon, \frac{\alpha}{(y_\varepsilon + \alpha)^2} \right)_2 \geqslant 0.$$

□

3.6 Optimality conditions for $(\mathcal{P}^{\alpha})$

3.6.1 Qualification assumptions

Now we would like to study the asymptotic behavior of the previous optimality conditions (3.5.16)-(3.5.19) when ε goes to 0 and we need some estimations on q_ε and r_ε. We have to assume some qualification conditions to pass to the limit in the penalized optimality system. We have

$$A_\varepsilon y_\varepsilon - v_\varepsilon - \xi_\varepsilon = Ay_\varepsilon + g(y_\varepsilon) - v_\varepsilon - \xi_\varepsilon - f + f + g^{'}(y_\varepsilon)y_\varepsilon - g(y_\varepsilon).$$

We set

$$\omega_\varepsilon = g^{'}(y_\varepsilon)y_\varepsilon - g(y_\varepsilon) \quad \text{and} \quad \omega_\alpha = g^{'}(y_\alpha)y_\alpha - g(y_\alpha), \tag{3.6.1}$$

so that

$$A_\varepsilon y_\varepsilon - v_\varepsilon - \xi_\varepsilon = \varepsilon q_\varepsilon + f + \omega_\varepsilon.$$

Let us choose (y, v, ξ) in $\tilde{K} \times U_{ad} \times V_{ad}$, and add relation (3.5.16)-(3.5.18). We have:

$$\begin{aligned}
&(p_\varepsilon, A_\varepsilon(y - y_\varepsilon))_2 + (q_\varepsilon, A_\epsilon(y - y_\varepsilon))_2 + (A(y - y_\varepsilon), A(y_\varepsilon - y_\alpha))_2 + (-q_\varepsilon, v - v_\varepsilon)_2 \\
&\qquad + (\nu(v_\varepsilon - v_d) + v_\varepsilon - v_\alpha, v - v_\varepsilon)_2 + r_\varepsilon \left(y - y_\varepsilon, \frac{\alpha}{(y_\varepsilon + \alpha)^2} \right)_2 \\
&\qquad + r_\varepsilon \left(\frac{\alpha}{(\xi_\alpha + \alpha)^2}, \xi - \xi_\varepsilon \right)_2 + (\xi_\varepsilon - \xi_\alpha, \xi - \xi_\varepsilon)_2 + (-q_\varepsilon, \xi - \xi_\varepsilon)_2 \\
&\qquad \geqslant 0.
\end{aligned}$$

So that:

$$\begin{aligned}
&(q_\varepsilon, f + \omega_\varepsilon + v + \xi - A_\varepsilon y)_2 - r_\varepsilon \left[\left(\frac{\alpha}{(y_\varepsilon + \alpha)^2}, y - y_\varepsilon \right)_2 + \left(\frac{\alpha}{(\xi_\alpha + \alpha)^2}, \xi - \xi_\varepsilon \right)_2 \right] \\
&\leqslant (p_\varepsilon, A_\varepsilon(y - y_\varepsilon)_2 + (A(y - y_\varepsilon), A(y_\varepsilon - y_\alpha))_2 \\
&\quad + (\nu(v_\varepsilon - v_d) + v_\varepsilon - v_\alpha, v - v_\varepsilon)_2 + (\xi_\varepsilon - \xi_\alpha, \xi - \xi_\varepsilon)_2 - \varepsilon \,||\, q_\varepsilon \,||_2^2 \,.
\end{aligned}$$

The right hand side is uniformly bounded with respect to ε by a constant C which only depends of y, v, ξ. Here, we use as well Theorem 3.5.2 when ε goes to 0. Moreover, relation (3.5.19) gives

$$r_\varepsilon \left(1, \frac{y_\varepsilon}{y_\varepsilon + \alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon + \alpha} \right)_2 = r_\varepsilon \text{Area}(\Omega),$$

so that we finally obtain:

$$-\Big(q_\varepsilon, Ay + g^{'}(y_\varepsilon)y - f - v - \xi - \omega_\varepsilon\Big)_2 - r_\varepsilon\left[\left(\frac{\alpha}{(y_\varepsilon+\alpha)^2}, y - y_\varepsilon\right)_2 + \left(\frac{\alpha}{(\xi_\varepsilon+\alpha)^2}, \xi - \xi_\varepsilon\right)_2\right] \leqslant C_{(y,v,\xi)}, \tag{3.6.2}$$

where

$$q_\varepsilon = \frac{Ay_\varepsilon + g(y_\varepsilon) - f - v_\varepsilon - \xi_\varepsilon}{\varepsilon} \in L^2(\Omega) \text{ and } A_\varepsilon = A + g^{'}(y_\varepsilon),\ \omega_\varepsilon = g^{'}(y_\varepsilon)y_\varepsilon - g(y_\varepsilon).$$

We consider two cases:

(i) If

$$\left(1, \frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha}\right)_2 < Area(\Omega),$$

as

$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 \to \left(1, \frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha}\right)_2,$$

there exists $\varepsilon_0 > 0$ such that

$$\forall \varepsilon \ \leqslant \varepsilon_0, \qquad \left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 < Area(\Omega),$$

and relation (3.5.19) implies that $r_\varepsilon = 0$, so the limit value is $r_\alpha = 0$.

(ii) If

$$\left(1, \frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha}\right)_2 = Area(\Omega),$$

as

$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 \to \left(1, \frac{y_\alpha}{y_\alpha+\alpha} + \frac{\xi_\alpha}{\xi_\alpha+\alpha}\right)_2,$$

there exists $\varepsilon_0 > 0$ such that

$$\forall \varepsilon \ \leqslant \varepsilon_0, \qquad \left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 = Area(\Omega),$$

we cannot conclude immediately, so we assume the following condition $(\mathcal{H}_1)$:

$$(\mathcal{H}_1) \qquad \forall \alpha > 0, \qquad \text{such that,} \qquad \left(1, \frac{y_\varepsilon}{y_\epsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 = Area(\Omega),$$

$g^{'}$ is locally Lipschitz continuous, and U_{ad} has a non empty L^∞-interior (denoted $\text{Int}_\infty(U_{ad})$) and

$-(f+\omega_\alpha) \in \mathrm{Int}_\infty(U_{ad})$.

Theorem 3.6.1. *Assume $(\mathcal{H}_1)$, then r_ε is bounded by a constant independent of ε and we may extract a subsequence that converges to r_α.*

Proof. We have already mentioned that $r_\alpha = 0$ when

$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 < Area(\Omega).$$

In the other case, as $g^{'}$ is locally Lipschitz continuous, then ω_ε uniformly converges to ω_α on $\bar{\Omega}$. Indeed, we prove that y_ε uniformly converges to y_α. Therefore, there exists $\varepsilon_0 > 0$ such that $y_\varepsilon - y_\alpha$ remains in a bounded subset of $\mathbb{R}^n$ independently of $\varepsilon \in]0, \varepsilon_0[$. The local Lipschitz continuity of $g^{'}$ yields

$$|g^{'}(y_\varepsilon(x)) - g^{'}(y_\alpha(x))| \leqslant M|y_\varepsilon(x) - y_\alpha(x)| \leqslant M \mid\mid y_\varepsilon - y_\alpha \mid\mid_\infty, \quad \forall x \in \Omega,$$

where M is a constant that does not depend of ε. Thus $\mid\mid g^{'}(y_\varepsilon) - g^{'}(y_\alpha) \mid\mid_\infty \to 0$. As

$$|g^{'}(y_\varepsilon)y_\varepsilon - g^{'}(y_\alpha)y_\alpha| \leqslant |g^{'}(y_\varepsilon)||y_\varepsilon - y_\alpha| + |g^{'}(y_\varepsilon) - g^{'}(y_\alpha)||y_\alpha|,$$

we get

$$\mid\mid g^{'}(y_\varepsilon)y_\varepsilon - g^{'}(y_\alpha)y_\alpha \mid\mid_\infty \leqslant M \mid\mid y_\varepsilon - y_\alpha \mid\mid_\infty + \mid\mid g^{'}(y_\epsilon) - g^{'}(y_\alpha) \mid\mid_\infty \mid\mid y_\alpha \mid\mid_\infty \to 0.$$

Similarly $\mid\mid g(y_\varepsilon) - g(y_\alpha) \mid\mid_\infty \to 0$. As we supposed $-(f+\omega_\alpha) \in \mathrm{Int}_\infty(U_{ad})$, then $-(f+\omega_\varepsilon) \in U_{ad}$ for ε smaller than some $\varepsilon_0 > 0$.
Now, we choose $y = 0,\ v = -(f+\omega_\varepsilon)$ and $\xi = 0$ in relation (3.6.2). We obtain

$$\forall \varepsilon \leqslant \varepsilon_0, \qquad r_\varepsilon \left[\left(\frac{\alpha}{(y_\varepsilon+\alpha)^2}, y_\varepsilon\right)_2 + \left(\frac{\alpha}{(\xi_\varepsilon+\alpha)^2}, \xi_\varepsilon\right)_2\right] \leqslant C, \tag{3.6.3}$$

where C is independent of ε since ω_ε is uniformly bounded with respect to ε, for $\varepsilon \in]0, \varepsilon_0[$.
We still need to proof that:

$$\left(\frac{\alpha}{(y_\varepsilon+\alpha)^2}, y_\varepsilon\right)_2 + \left(\frac{\alpha}{(\xi_\varepsilon+\alpha)^2}, \xi_\varepsilon\right)_2 \neq 0, \quad \forall \alpha > 0 \qquad \text{and} \qquad \varepsilon \to 0,$$

as

$$\left(1, \frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}\right)_2 = Area(\Omega).$$

So, we have:

$$\frac{y_\varepsilon}{y_\varepsilon+\alpha} + \frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha} = 1, \text{ a.e. in } \Omega,$$

in view of section 2.5 we obtain

$$\frac{y_\varepsilon}{y_\varepsilon+\alpha}+\frac{\xi_\varepsilon}{\xi_\varepsilon+\alpha}=1 \Longleftrightarrow y_\varepsilon\xi_\varepsilon=\alpha^2 \Longrightarrow (y_\varepsilon,\xi_\varepsilon)_2=\alpha^2 Area(\Omega), \quad \text{a.e. in } \Omega.$$

Therefore, the set $\{x\in\Omega, /y_\varepsilon(x)\neq 0,\ \xi_\varepsilon(x)\neq 0\}$ is not empty, and the set $\{x\in\Omega, /y_\varepsilon(x)=0,\ \xi_\varepsilon(x)=0\}$ is empty when ε goes to 0 since α is fixed. Hence we obtain:

$$\left(\frac{\alpha}{(y_\varepsilon+\alpha)^2},y_\varepsilon\right)_2+\left(\frac{\alpha}{(\xi_\varepsilon+\alpha)^2},\xi_\varepsilon\right)_2\neq 0, \quad \forall\alpha>0.$$

Passing to limit as $\varepsilon\to 0$, we obtain

$$\left(\frac{\alpha}{(y_\alpha+\alpha)^2},y_\alpha\right)+\left(\frac{\alpha}{(\xi_\alpha+\alpha)^2},\xi_\alpha\right)_2\neq 0, \quad \forall\alpha>0.$$

□

From (3.6.3), r_ε is uniformly bounded (independently of ε). So the relation (3.6.2) becomes:

$$\forall(y,v,\xi)\in\tilde{K}\times U_{ad}\times V_{ad} \qquad -\left(q_\varepsilon, Ay+g^{'}(y_\varepsilon)y-f-v-\xi-\omega_\varepsilon\right)_2 \leqslant C_{(y,v,\xi)}. \tag{3.6.4}$$

Then we have to do another assumption to get the estimation of q_ε $(\mathcal{H}_2)$:

$$\exists p\in[1,+\infty], \quad \exists\varepsilon_0>0, \quad \exists\rho>0,$$
$$(\mathcal{H}_2) \qquad \forall\varepsilon\in]0,\varepsilon_0[, \quad \forall\chi\in L^p(\Omega) \text{ such that } ||\ \chi\ ||_{L^p(\Omega)}\leqslant 1,$$
$$\exists(y_\chi^\varepsilon,v_\chi^\varepsilon,\xi_\chi^\varepsilon) \text{ bounded in } \tilde{K}\times U_{ad}\times V_{ad} \text{ (uniformly with respect to } \chi \text{ and } \varepsilon), \text{ such that}$$
$$Ay_\chi^\varepsilon+g^{'}(y_\varepsilon)y_\chi^\varepsilon=f+\omega_\varepsilon+v_\chi^\varepsilon+\xi_\chi^\varepsilon-\rho\chi \text{ in } \Omega.$$

Then we may conclude:

Theorem 3.6.2. *Assume $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$, then q_ε is bounded in $L^{p^{'}}(\Omega)$ by a constant independent of ε (here $\frac{1}{p}+\frac{1}{p'}=1$).*

Proof. $(\mathcal{H}_2)$ and relation (3.6.4) when applied with $(y_\chi^\epsilon,v_\chi^\varepsilon,\xi_\chi^\varepsilon)$ give:

$$\forall\chi\in L^p(\Omega), \quad ||\ \chi\ ||_{L^p(\Omega)}\leqslant 1, \quad \rho\,(q_\varepsilon,\chi)_{L^{p'}\times L^p}\leqslant C_{\chi,\varepsilon}\leqslant C.$$

□

Then passing to the limit in the penalized optimality system, we obtain the following result.

Theorem 3.6.3. *Assume* $(\mathcal{H}_1)$ *and* $(\mathcal{H}_2)$, *if* $(y_\alpha, v_\alpha, \xi_\alpha)$ *is a solution of* $(\mathcal{P}^\alpha)$, *then Lagrange multipliers* $(q_\alpha, r_\alpha) \in L^{p'}(\Omega) \times \mathbb{R}^+$ *exist, such that*

$$\begin{aligned} &\forall y \in \tilde{K}, \quad [A + g'(y_\alpha)](y - y_\alpha) \in L^p(\Omega), \\ &\Big(p_\alpha + q_\alpha, [A + g'(y_\alpha)](y - y_\alpha)\Big)_{L^{p'} \times L^p} + r_\alpha \left(\frac{\alpha}{(y_\alpha + \alpha)^2}, y - y_\alpha\right)_{L^{p'} \times L^p} \geqslant 0, \end{aligned} \tag{3.6.5}$$

$$\forall v \in U_{ad}, \quad v - v_\alpha \in L^p(\Omega) \quad (\nu(v_\alpha - v_d) - q_\alpha, v - v_\alpha)_{L^{p'} \times L^p} \geqslant 0, \tag{3.6.6}$$

$$\forall \xi \in V_{ad}, \quad \xi - \xi_\alpha \in L^p(\Omega) \quad \left(\frac{r_\alpha \alpha}{(\xi_\alpha + \alpha)^2} - q_\alpha, \xi - \xi_\alpha\right)_{L^{p'} \times L^p} \geqslant 0, \tag{3.6.7}$$

$$r_\alpha \left[\left(1, \frac{y_\alpha}{y_\alpha + \alpha} + \frac{\xi_\alpha}{\xi_\alpha + \alpha}\right)_2 - Area(\Omega)\right] = 0, \tag{3.6.8}$$

where p_α *is given by* (3.5.8).

3.6.2 Sufficient condition for $(\mathcal{H}_2)$ with p=2.

In this subsection we give an assumption dealing with $(y_\alpha, v_\alpha, \xi_\alpha)$ independently of ε. We consider $p = 2$ which corresponds to one of the useful cases and for which we established several results in section 3.3. We always assume that g' is locally Lipschitz continuous (for example g is $\mathcal{C}^2$), and we set the following $(\mathcal{H}_3)$

$$\exists \rho > 0, \ \exists v_0 \in \mathrm{Int}_\infty(U_{ad}), \ \forall \mathcal{X} \in L^2(\Omega) \text{ such that } \| \mathcal{X} \|_{L^2(\Omega)} \leqslant 1,$$

$$\exists (y_\mathcal{X}, \xi_\mathcal{X}) \in \tilde{K} \times V_{ad} \text{ (uniformly bounded by a constant M independent of } \mathcal{X}),$$

$$\text{such that } Ay_\mathcal{X} + g'(y_\alpha)y_\mathcal{X} = f + \omega_\alpha + v_0 + \xi_\mathcal{X} - \rho\mathcal{X} \text{ in } \Omega.$$

Proposition 3.6.4. *If* g' *is locally Lipschitz continuous then* $(\mathcal{H}_3) \Longrightarrow (\mathcal{H}_2)$.

Proof. We have seen that $\| y_\varepsilon - y_\alpha \|_\infty \to 0$, $\| g'(y_\varepsilon) - g'(y_\varepsilon) \|_\infty \to 0$ and $\| \omega_\varepsilon - \omega_\alpha \|_\infty \to 0$. Let be $\mathcal{X} \in L^2(\Omega)$ such that $\| \mathcal{X} \|_2 \leqslant 1$ and $(y_\mathcal{X}, v_0, \xi_\mathcal{X}) \in \tilde{K} \times \mathrm{Int}_\infty(U_{ad}) \times V_{ad}$ given by $(\mathcal{H}_3)$. As $v_0 \in \mathrm{Int}_\infty(U_{ad})$, there exists $\rho_0 > 0$ such that $\mathcal{B}_\infty(v_0, \rho) \subset U_{ad}$. As $y_\mathcal{X}$ is bounded by M, then for ε small enough (less than some $\varepsilon_0 > 0$), we get

$$\begin{aligned} &\| \omega_\alpha - \omega_\varepsilon + (g'(y_\varepsilon) - g'(y_\alpha))y_\mathcal{X} \|_\infty \\ &\leqslant \| \omega_\alpha - \omega_\varepsilon \|_\infty + \| g'(y_\varepsilon) - g'(y_\alpha) \|_\infty \| y_\mathcal{X} \|_\infty \\ &\leqslant \rho_0, \end{aligned}$$

therefore $v_\mathcal{X}^\varepsilon = v_0 + (g'(y_\varepsilon) - g'(y_\alpha))y_\mathcal{X} + \omega_\alpha - \omega_\varepsilon$ belongs to U_{ad} and

$$|| v_{\mathcal{X}}^{\varepsilon} ||_2 \leqslant || v_0 ||_2 + || \omega_\alpha - \omega_\varepsilon ||_2 + || (g^{'}(y_\varepsilon) - g^{'}(y_\alpha))y_{\mathcal{X}} ||_2 \leqslant C,$$

$v_{\mathcal{X}}^{\varepsilon}$ is L^2-bounded independently of $\mathcal{X}$ and ε. Now, we set $y_{\mathcal{X}}^{\varepsilon} = y_{\mathcal{X}} \in \tilde{K}$ and $\xi_{\mathcal{X}}^{\varepsilon} = \xi_{\mathcal{X}} \in V_{ad}$ to obtain

$$\begin{aligned} Ay_{\mathcal{X}}^{\varepsilon} + g^{'}(y_\varepsilon)y_{\mathcal{X}}^{\varepsilon} &= Ay_{\mathcal{X}} + g^{'}(y_\alpha))y_{\mathcal{X}} + (g^{'}(y_\varepsilon) - g^{'}(y_\alpha))y_{\mathcal{X}} \\ &= f + \omega_\alpha + v_0 + \xi_{\mathcal{X}} - \rho\mathcal{X} + (g^{'}(y_\varepsilon) - g^{'}(y_\alpha))y_{\mathcal{X}} \\ &= f + \omega_\varepsilon + v_0 + (g^{'}(y_\varepsilon) - g^{'}(y_\alpha))y_{\mathcal{X}} + \omega_\alpha - \omega_\varepsilon + \xi_{\mathcal{X}} - \rho\mathcal{X} \\ &= f + \omega_\varepsilon + v_{\mathcal{X}}^{\varepsilon} + \xi_{\mathcal{X}}^{\varepsilon} - \rho\mathcal{X}. \end{aligned}$$

We can see that $(\mathcal{H}_2)$ is satisfied. □

We obtain, as an immediate consequence, the following result about the existence of Lagrange multipliers.

Theorem 3.6.5. *Let $(y_\alpha, v_\alpha, \xi_\alpha)$ be a solution of $(\mathcal{P}^\alpha)$ and assume $(\mathcal{H}_1, \mathcal{H}_3)$, then Lagrange multipliers $(q_\alpha, r_\alpha) \in L^2(\Omega) \times \mathbb{R}^+$ exist, such that*

$$\forall y \in \tilde{K}, \quad \Big(p_\alpha + q_\alpha, [A + g^{'}(y_\alpha)](y - y_\alpha)\Big)_2 + r_\alpha \left(\frac{\alpha}{(y_\alpha + \alpha)^2}, y - y_\alpha\right)_2 \geqslant 0, \tag{3.6.9}$$

$$\forall v \in U_{ad}, \quad (\nu(v_\alpha - v_d) - q_\alpha, v - v_\alpha)_2 \geqslant 0, \tag{3.6.10}$$

$$\forall \xi \in V_{ad}, \quad \left(\frac{r_\alpha \alpha}{(\xi_\alpha + \alpha)^2} - q_\alpha, \xi - \xi_\alpha\right)_2 \geqslant 0, \tag{3.6.11}$$

$$r_\alpha \left[\left(1, \frac{y_\alpha}{y_\alpha + \alpha} + \frac{\xi_\alpha}{\xi_\alpha + \alpha}\right)_2 - Area(\Omega)\right] = 0, \tag{3.6.12}$$

where p_α is given by (3.5.8).

Proof. We take $v_0 = -(f + \omega_\alpha)$ to ensure $(\mathcal{H}_3)$. Let $\mathcal{X} \in L^2(\Omega)$ such that
$|| \mathcal{X} ||_{L^2(\Omega)} \leqslant 1$.
We set $\xi_{\mathcal{X}} = \mathcal{X}^+ + \mathcal{X}^- = |\mathcal{X}| \geqslant 0$, where $\mathcal{X}^+ = max(0, \mathcal{X})$ and $\mathcal{X}^- = \max(0, -\mathcal{X})$. As $|| \mathcal{X} ||_{L^2(\Omega)} \leqslant 1$, it is clear that $\xi_{\mathcal{X}} \in V_{ad}$. Let $y_{\mathcal{X}}$ be the solution of

$$[A + g^{'}(y_\alpha)]y_{\mathcal{X}} = \xi_{\mathcal{X}} - \mathcal{X} = 2\mathcal{X}^- \geqslant 0 \quad (a.e.), \ y \in H_0^1(\Omega),$$

thanks to the properties of $[A + g^{'}(y_\alpha)]$ and the maximum principale, then $y_{\mathcal{X}} \geqslant 0$ a.e. in Ω. Therefore $y_{\mathcal{X}} \in \tilde{K}$ and $(\mathcal{H}_3)$ is satisfied (with $\rho = 1$). The optimality system follows and we prove that the multiplier q_α is a $L^2(\Omega)$-function. □

Corollary 3.6.6. *If g is linear and $-f \in U_{ad}$, the conclusions of Theorem 3.6.5 are valid.*

Proof. If g is linear, we use the same proof as the one of Theorem 3.6.5 to bound q_ε in $L^2(\Omega)$. It is sufficient that $-f \in \mathrm{U}_{ad}$. □

Remark. *The Lagrange multiplier $(q_\alpha, r_\alpha) \in L^2(\Omega) \times \mathbb{R}^+$ in (3.6.5) works. But for some examples such that complex constraints, it becomes invalid, it is generally called as Lagrange crisis, and there are some methods to overcome the crisis, see for example [57].*

Next we describe numerical experiments that are carried out by means of AMPL languages [1], with the IPOPT solver [106] ("Interior Point OPTimizer"), KNITRO solver [26] ("Nonlinear Interior point Trust Region Optimization") and SNOPT solver [51]("Sequential Quadratic Optimization Technique").

3.7 Numerical results

In this section, we report on some experiments considering 2D-examples. For two different smoothing functions, we present some numerical results using three solvers from different families: the IPOPT nonlinear programming algorithm, the KNITRO and SNOPT solver on AMPL [1] optimization platform. Our aim is just to verify the qualitative numerical efficiency of our approach.
The discretization process is based on finite difference schemes with a $N \times N$ grid and the size of the grid is given by $h = \frac{1}{N}$ on each side of the domain.
We take $\Omega =]0,1[\times]0,1[\subset \mathbb{R}^2$, $A := -\triangle$ the Laplacian operator $\left(\triangle y = \frac{\partial^2 y}{\partial x_1^2} + \frac{\partial^2 y}{\partial x_2^2}\right)$. We fix the tolerance to $\mathrm{tol} = 10^{-3}$.
In our experiments, we use the two following functions

$$\theta_\alpha^1(x) = \frac{x}{x+\alpha},$$
$$\theta_\alpha^{\log}(x) = \frac{\log(1+x)}{\log(1+x+\alpha)}.$$

3.7.1 Example 1

We set $U_{ad} = L^2(\Omega)$, $\nu = 0.1$, $z_d = 1$ and $v_d = 0$, $g(y) = y^3$.
We fixe the smoothing parameter to 10^{-3} and the penalization parameter to 10^{-3}.

$$f(x_1, x_2) = \begin{cases} 200[2x_1(x_1 - 0.5)^2 - x_2(1-x_2)(6x_1 - 2)] & \text{if } x_1 \leqslant 0.5, \\ 200(0.5 - x_1) & \text{else,} \end{cases}$$

and

$$\psi(x_1, x_2) = \begin{cases} 200[x_1 x_2 (x_1 - 0.5)^2 (1 - x_2)] & \text{if } x_1 \leqslant 0.5, \\ 200[(x_1 - 1) x_2 (x_1 - 0.5)^2 (1 - x_2)] & \text{else.} \end{cases}$$

Figure 3.1: Data of the considered example.

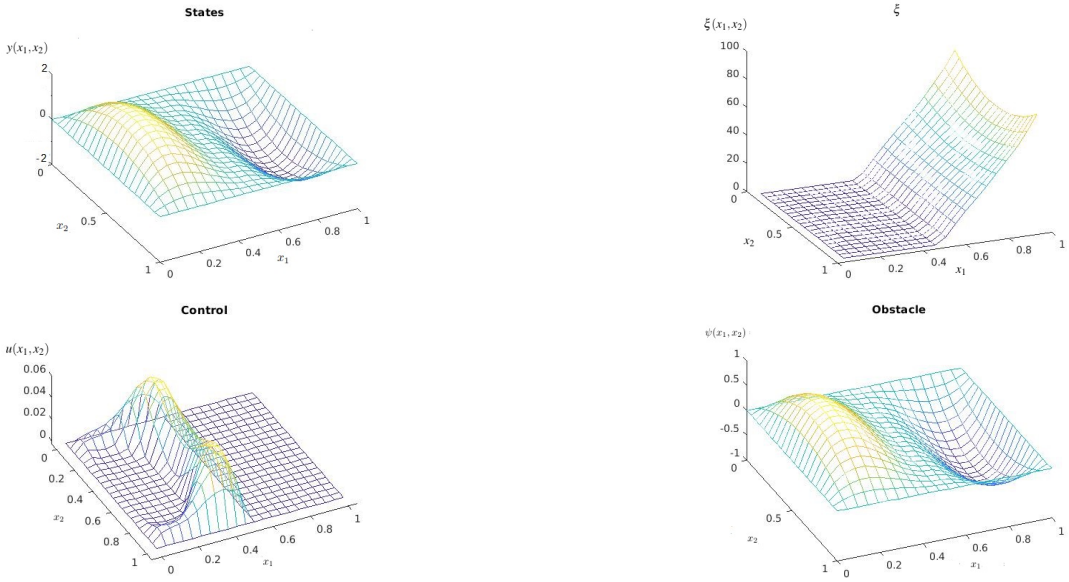

Figure 3.2: Optimal solution with IPOPT solver using the θ_α^1, N=20, $\alpha = 10^{-3}$, and $\varepsilon = 10^{-3}$.

Figure 3.1 presents the data from Example 1. On the right is the function f and on the left is the obstacle ψ. Using these data, we solve the problem $(\mathcal{P}_\alpha^\varepsilon)$ using IPOPT solver and report on the solutions in 3D, in figure 3.2. We find that the state satisfies the obstacle constraint and that the function ξ is greater than or equal to 0. The complementarity constraint is satisfied and its error is equal to 3.15e-7 (we do not need α very close to 0 to have a good approximation of our problem), the equality constraint is satisfied and the obtained objective value is equal to $J(u, v) = 0.6490659$.

3.7.1.1 Details of the numerical tests

Numerical simulation results using IPOPT solver

In our experiments we make a logarithmic scaling for these two functions to bound their gradients. Each constraint

$$\theta_\alpha((y-\psi)_{i,j}) + \theta_\alpha(\xi_{i,j}) \leqslant 1,$$

is in fact replaced by the following inequality

$$\alpha^2 \ln\left(\frac{\alpha}{(y-\psi)_{i,j}+\alpha} + \frac{\alpha}{\xi_{i,j}+\alpha}\right) \geqslant 0, \qquad 0 \leqslant i,j \leqslant N+1,$$

in the case of the θ_α^1 function and

$$\alpha \ln\left(2 - \left(\frac{\log(1+(y-\psi)_{i,j})}{\log(1+(y-\psi)_{i,j}+\alpha)} + \frac{\log(1+\xi_{i,j})}{\log(1+\xi_{i,j}+\alpha)}\right)\right) \geqslant 0, \qquad 0 \leqslant i,j \leqslant N+1,$$

in the case of the $\theta_\alpha^{\log}$.

This scaling technique is proposed and used in [19] to avoid numerical issues.

We fix the penalization parameter to $\varepsilon = 10^{-3}$, Tables 3.1 and 3.2 give in view of Example 1 and for different values of the parameter α, the complementarity error, the state equation error, and the value of the objective function when using each of the two smoothing functions.

Table 3.1: Using the θ_α^1 smoothing function -Example 1- N=20.

α	$\|\| Ay - g(y) - f - v - \xi \|\|_2$	$(y-\psi,\xi)_2$	J
1.e-1	4.19213e-06	0.00554001	6.4520064e-01
1.e-2	9.73059e-06	3.90692e-05	6.4808311e-01
1.e-3	1.66885e-05	3.15736e-07	6.4906597e-01
1.e-4	7.20318e-06	3.45708e-09	6.4906596e-01

Table 3.2: Using the $\theta_\alpha^{\log}$ smoothing function -Example 1- N=20.

α	$\|\| Ay - g(y) - f - v - \xi \|\|_2$	$(y-\psi,\xi)_2$	J
1.e-1	4.77128e-06	0.00275542	6.4585951e-01
1.e-2	1.00013e-05	2.29846e-05	6.4810490e-01
1.e-3	1.67039e-05	2.54095e-07	6.4906099e-01
1.e-4	3.05604e-07	3.97519e-09	6.4906099e-01

In Table 3.1 and Table 3.2, for different values of α, using the two smooth functions θ_α^1 and $\theta_\alpha^{\log}$, we notice that when α goes to 0 our approach gives consistent results and we also notice that the complementarity error goes to 0, the state equation goes to 0 and objective value remains almost the same for $\alpha \leqslant$ 1.e-3.

We present now, the effect of the penalization parameter ε, where α is fixed. Table 3.3 gives in view of Example 1 and for different values of the parameter ε, the complementarity error, the state equation error, and the obtained value.

Table 3.3: Using the θ_α^1 smoothing function -Example 1- N=20 where $\alpha = 10^{-2}$ is fixed.

ε	$\| Ay - g(y) - f - v - \xi \|_2$	$(y - \psi, \xi)_2$	J
1.e-1	0.000190242	6.60632e-05	6.4772424e-01
1.e-2	2.49746e-05	5.1126e-05	6.4776774e-01
1.e-3	9.73059e-06	3.90692e-05	6.4808311e-01
1.e-4	2.50301e-06	3.77642e-05	6.4810870e-01

Numerical test using different NLP solvers

We prove that all the variables and multipliers involved in our optimality conditions exist and remain bounded. So, one can at least in theory use any standard NLP solver to try to tackle discrete versions of the relaxed problem.

In this subsection, we consider 3 solvers from different families

- a free Interior point solver IPOPT,
- a commercial one KNITRO,
- and a sequential quadratic programming solver SNOPT.

Our main concern is to analyse their efficiency on our problems. Even, we provide information about the number of iterations, it is difficult to make any comparison: iteration of different solvers corresponds to different numerical efforts. We only want to check if these solvers are able to solve the problems.

Table 3.4 gives in view of Example 1 and for the value of the parameter $\alpha = 10^{-2}$, and $\varepsilon = 10^{-3}$, the complementarity error, the state equation error and the obtained value, and the number of iterations.

Table 3.4: Using the θ_α^1 smoothing function -Example 1- N=20.

Solver	$\| Ay - g(y) - f - v - \xi \|_2$	$(y - \psi, \xi)_2$	J	Nb.Iter
SNOPT	3.73674e-09	9.04096e-07	6.47795123e-1	46346
KNITRO	7.72611e-13	9.05794e-05	6.4779048e-01	64
IPOPT	9.73059e-06	3.90692e-05	6.4808311e-01	478

In view of Table 3.4, we remark that, the 3 algorithms obtain the same solution and almost the same objective value. This suggests that our approach can be implemented using any standard NLP solver.

3.7.2 Example 2

We present now, some numerical results an example given in [17]. We take $\Omega =]0,1[\times]0,1[\subset \mathbb{R}^2$, $A := -\triangle$ the Laplacian operator. The discretization is done via finite-differences and the size of the grid is given by $\frac{1}{N}$ on each side of the domain.
We set $U_{ad} = L^2(\Omega)$, $\nu = 100$, $v_d = 0$, and $g(y) = 0$, $\psi(x_1, x_2) = 0$

$$z_d(x_1,x_2) = \begin{cases} 200[x_1x_2(x_1-\frac{1}{2})^2(1-x_2)] & \text{if } 0 < x_1 \leqslant 1/2, \\ 200[(x_1-1)x_2(x_1-\frac{1}{2})^2(1-x_2)] & \text{if } 1/2 < x_1 \leqslant 1, \end{cases}$$

and

$$f(x_1,x_2) = \begin{cases} 200[2x_1(x_1-\frac{1}{2})^2 - x_2(1-x_2)(6x_1-2)] & \text{if } 0 < x_1 \leqslant 1/2, \\ 200(\frac{1}{2}-x_1) & \text{if } 1/2 < x_1 \leqslant 1. \end{cases}$$

Moreover, we put $\alpha = 10^{-3}$ and $\varepsilon = 10^{-3}$. This example is constructed such that the null control $v^* \equiv 0$ is the optimal control for the original problem $(\mathcal{P})$ and

$$y^* = \begin{cases} z_d & \text{if } 0 < x_1 \leqslant 1/2, \\ 0 & \text{if } 1/2 < x_1 \leqslant 1, \end{cases}$$

and $J^* = J(y^*, v^*) = \frac{25}{504} \simeq 0.0496$.

Figure 3.3: Example 2 using θ_α^1, $N = 15$ and $\alpha = 10^{-3}$.

In figure 3.3, using IPOPT solver we draw the approximate and exact solutions. Using the smooth function θ_α^1, and for $\alpha = 10^{-3}$ our approach gives consistent results.

Table 3.5: Using the θ_α^1 smoothing function -Example 2- N=15.

α	$\| Ay - g(y) - f - v - \xi \|_2$	$(y-\psi, \xi)_2$	$\| y - y^* \|_2$	$\| v - v^* \|_2$	$\lvert J - J^* \rvert$
1.e-1	3.63549e-08	0.00420714	0.00029353	3.63549e-07	2.24511e-05
1.e-2	4.61188e-08	4.35654e-05	3.21283e-06	4.60966e-07	3.93248e-05
1.e-3	3.51106e-13	4.39172e-07	2.54572e-08	5.07363e-07	3.94924e-05
1.e-4	3.93201e-13	1.19934e-09	2.01061e-08	1.72508e-06	3.94944e-05

Table 3.6: Using the θ_α^1 smoothing function -Example 2-$\alpha = 10^{-3}$.

N	$\lvert J - J^* \rvert$
3	0.009068814
6	0.000722747
9	0.000270778
12	5.37292e-05
15	3.94924e-05
18	9.16493e-06
21	8.82412e-06
24	8.85534e-07

In Table 3.5, for different values of α, using the smooth function θ_α^1, we notice that when α goes to 0 our approach gives consistent results. The complementarity error goes to 0 and the error on the state equation goes to 0. We remark that when $\alpha \leqslant 10^{-2}$ the largest error concerns the cost function. We decided to use this criterion to analyse the mesh dependence when the value of alpha α is fixed. In Table 3.6, for different values of N, using the smooth function θ_α^1, we found that for N large enough ($N \geqslant 15$) the approximation is good. We also remark that we do not need to take N extremely large ($N \leqslant 30$).

Remark. *(**Constrained control case**) Up to now, we investigate optimal control problems governed by semilinear elliptic variational inequalities involving constraints on the state i.e.* $y \in K = \{y \mid y \in H_0^1(\Omega),\ y \geqslant \psi$ *a.e. in* $\Omega\}$, *and for unconstrained control i.e.* $v \in U_{ad} = L^2(\Omega)$.
It is possible to add constraints on the control. In the following, we give two examples for which the assumption of Theorem 3.6.5 is satisfied.

- *We can consider an admissible control of the form*

$$U_{ad} = \{v \in L^2(\Omega) \quad | \quad v \geqslant \Lambda \text{ a.e. in } \Omega\},$$

where $\Lambda \in L^\infty$. If one can find a real number $\rho > 0$ such that $\Lambda + \rho \leqslant -f$ then the assumption of Theorem 3.6.5 is satisfied.

- *We can choose $g(x) = -\frac{1}{1+x^2}$: it is globally Lipschitz-continuous and C^2. Then*

$$0 \leqslant \omega_\alpha = g^{'}(y_\alpha)y_\alpha - g(y_\alpha) = \frac{3y_\alpha^2 + 1}{(y_\alpha^2 + 1)^2} \leqslant 3.$$

So $-f - 3 \leqslant -f - \omega_\alpha \leqslant -f$ and one can consider a set of admissible controls defined as

$$U_{ad} = \{v \in L^2(\Omega) \quad | \quad \Lambda \leqslant v \leqslant \Gamma \qquad a.e. \ \ in\ \Omega\},$$

where $\Lambda, \Gamma \in L^\infty(\Omega)$ such that $\Lambda + \rho \leqslant -f - 3$ and $-f \leqslant \Gamma - \rho$, (with $\rho > 0$ is a real number) then the assumption of Theorem 3.6.5 is satisfied since $-(f + \omega_\alpha) \subset Int_{L^\infty}(U_{ad})$.

3.8 Conclusion

In this chapter, we introduce a new regularization schema for optimal control of semilinear elliptic variational inequalities with complementarity constraints. We prove that Lagrange multipliers exist. The existence of Lagrange multipliers is an important tool to describe and study algorithms to compute the solutions(s) of $(\mathcal{P}^\alpha)$ (that are "good approximations" of the original problem $(\mathcal{P})$).
In our numerical experiments, we use several standard NLP solvers and obtained promising results. The next step will be to develop an approach based on our optimality conditions.

4 A smooth approach to the solution of nonlinear complementarity problems involving $\mathcal{P}_0$-function

This chapter is a paper accepted in Statistics, Optimization & Information Computing [87].
In this chapter, we present a family of smoothing methods to solve nonlinear complementarity problems (NCPs) involving $\mathcal{P}_0$-function.
Several regularization or approximation techniques like Fisher-Burmeister's method, interior point methods approaches, or smoothing methods already exist. All the corresponding methods solve a sequence of nonlinear systems of equations and depend on parameters that are difficult to drive to zero. The main novelty of our approach is to consider the smoothing parameters as variables that converge by themselves to zero. We do not need any complicated updating strategy, and then obtain nonparametric algorithms. We prove some global and local convergence results and present several numerical experiments, comparisons, and applications that show the efficiency of our approach.

Contents

4.1 Introduction

The nonlinear complementarity problem (NCP) consists in finding $x \in \mathbb{R}^n$ satisfying

$$x \geqslant 0 \quad F(x) \geqslant 0 \quad x^T F(x) = 0, \tag{4.1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$. When F is linear, problem (4.1.1) reduces to a linear complementarity problem (LCP).

NCPs arise in many practical applications, for example, the Karush-Kuhn-Tucker (KKT) systems of mathematical programming problem, economic equilibria, and engineering design problems can be formulated as NCPs (see, for instance, [42, 56, 82]).

Different concepts have been developed to study and solve these problems: reformulation as a system of nonlinear equations or a minimization problem (see [44, 47, 60, 72, 84, 88, 90]). Recently, there have been strong interests in equation reformulation methods for solving the NCPs. One of the most effective methods is to transform the NCP into semi-smooth equation (NCP functions) and solve using semi-smooth Newton methods. The most well-known NCP functions are the Fisher-Burmeister's function introduced by Fisher Burmeister in [43] and the min function studied by Kanzow, Yamashita and Fukushima [62]. Another well-known class of algorithms corresponds to the smoothing methods. The main idea of smoothing approaches is to approximate or regularize the NCP to obtain smooth equations depending on some parameter (see, for example, [27, 61, 69, 92, 109]).

In this chapter, we present a smoothing approximation scheme to solve (4.1.1). We replace

$$0 \leqslant x \perp F(x) \geqslant 0, \tag{4.1.2}$$

by a sequence of smoothed systems of the form

$$G_r(x, F(x)) := (G_r(x_i, F_i(x)))_{i=1,\dots,n} := \left(r\psi^{-1}\left[\psi\left(\frac{x_i}{r}\right) + \psi\left(\frac{F_i(x)}{r}\right)\right]\right)_{i=1,\dots,n} = 0. \tag{4.1.3}$$

All the functions and parameters involved in (4.1.3) will be explicit later. Depending on the context, we apply several functions $(\theta, \psi, G_r, ...)$ on reals or vectors. When applied to vectors, we consider that they apply component-wise. The novelty of our approach is that we do not need any complicated strategy to update the regularization parameter r since we will consider it as a new variable. To solve the smoothed equations system we will use the standard Newton-like method. Without requiring a strict complementarity assumption at the solution of equation (4.1.1), we prove that the proposed algorithm is well defined, globally and superlinearly convergent. At the end of the chapter, we present numerical results to prove the effectiveness of the algorithm.

This chapter is organized as follows: some definitions are introduced in section 4.2. We present our approximation and formulation in section 4.3. In section 4.4, we discuss our approach and scheme to solve (4.1.1). The convergence properties of the algorithm are given in section 4.5. The section 4.6 is devoted to the numerical results with a comparison of our method with other approaches. Finally, we conclude our chapter.

4.2 Preliminaries and problem setting

Consider the NCP, which is to find a solution of the system:

$$x \geqslant 0, \quad F(x) \geqslant 0 \quad \text{and} \quad x^T F(x) = 0 \quad \text{or} \quad 0 \leqslant x \perp F(x) \geqslant 0, \tag{4.2.1}$$

where $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a continuous function satisfying some additional assumptions to be precised later.
From (4.2.1), we obtain the equivalent formulation for component-wise products

$$x \geqslant 0,\ F(x) \geqslant 0 \qquad \text{and} \qquad x_i F_i(x) = 0, \quad i = 1, 2, ..., n.$$

Or equivalently

$$x.F(x) = 0, \qquad x \geqslant 0,\ F(x) \geqslant 0,$$

where "." stands for the Hadamard product. It provides an explanation for the term "complementarity", namely, for all $i = 1, 2, ..., n,\ x_i$ and $F_i(x)$ are complementary in the sense that if one of them is positive then the other term must be zero.
A particular and important class of NCP is the LCP class defined below.

Definition 4.2.1. *When F is affine function:*

$$F(x) = Mx + q, \quad x \in \mathbb{R}^n,\ q \in \mathbb{R}^n,\ M \in \mathbb{R}^{n \times n}.$$

The corresponding NCP is called an LCP. So an problem is to find $x \in \mathbb{R}^n$ such that

$$x \geqslant 0, \quad Mx + q \geqslant 0 \quad \textit{and} \quad x^T(Mx + q) = 0.$$

To solve NCP, there are essentially three different classes of methods: equation-based methods (smoothing), merit functions, and projection-type methods. Our goal in this chapter is to present new and very simple smoothing and approximation schemes to solve NCP and to produce efficient numerical methods. In our approach, we do not need any complicated strategy to update the smoothing

parameter since we will consider it as a new variable.
First, let us introduce the usual assumptions on F and the ones that will be used in this chapter. A well-known and studied situation corresponds to monotone functions F and several methods and algorithms have been developed in this case.
Almost all the solution methods consider at least the following important and standard condition on the mapping F (monotonicity): We recall that F is said to be monotone if $F : \mathbb{R}^n \to \mathbb{R}^n$ satisfies for any $(x, y) \in \mathbb{R}^n$,

$$(x - y)^T (F(x) - F(y)) \geqslant 0.$$

In this work, we will consider a weaker assumption on F:

$$F \quad \text{is a} \quad \mathcal{P}_0\text{-function}, \tag{H_0}$$

to prove the convergence of our approach. We recall the following definitions of $\mathcal{P}_0$ and $\mathcal{P}$ functions. We say that $F : \mathbb{R}^n \to \mathbb{R}^n$ is a $\mathcal{P}_0$-function (respectively $\mathcal{P}$-function) if $\forall x, y \in \mathbb{R}^n$ with $x \neq y$, there exists an index $i_0 \in \{1, 2, ..., n\}$ such that

$$(x_{i_0} - y_{i_0})[F_{i_0}(x) - F_{i_0}(y)] \geqslant 0,$$

$$(\text{respectively} \quad (x_{i_0} - y_{i_0})[F_{i_0}(x) - F_{i_0}(y)] > 0).$$

It is important to notice that the index i_0 can depend on x and y.
A matrix is called a $\mathcal{P}_0$-matrix (resp. $\mathcal{P}$-matrix) if all its principals minors are nonnegative (resp. positive). Note that F is a $\mathcal{P}_0$-function if and only if $\nabla F(x)$ is a $\mathcal{P}_0$-matrix for all $x \in \mathbb{R}^n$. If $\nabla F(x)$ is a $\mathcal{P}$-matrix for all $x \in \mathbb{R}^n$, then F is a $\mathcal{P}$-function. However, the converse is not necessarily true.

4.3 Smoothing approximation functions

In the first part of this section, we introduce the smoothing functions and establish different properties that will be useful for the presentation and the convergence of our algorithm. In the second part of this section, we present a new smoothing function for NCP.
A function $\phi : \mathbb{R}^2 \to \mathbb{R}$ is said to be a NCP function if ϕ satisfies

$$\phi(a, b) = 0 \iff a \geqslant 0,\ b \geqslant 0,\ ab = 0. \tag{4.3.1}$$

An example of such function is $\phi_{\min} : \mathbb{R}^2 \to \mathbb{R}$,

$$\phi_{\min}(a, b) = \min\{a, b\}.$$

Then, $\phi_{\min}$ is a NCP function. Problem (4.2.1) is then equivalent to the following system of nonlinear equations:

$$H(x) = \begin{bmatrix} \phi_{\min}(x_1, F_1(x)) \\ \phi_{\min}(x_2, F_2(x)) \\ \vdots \\ \phi_{\min}(x_n, F_n(x)) \end{bmatrix} = 0. \tag{4.3.2}$$

This system is clearly nonsmooth; classical Newton-like methods can not be used to try to solve it. To overcome this difficulty, there exist several semi-smooth approaches. These techniques may present difficulties to converge. An efficient approach is to approximate (4.3.2) by a smooth one. The following subsection introduces some smoothing functions and establishes different properties that will be useful for our study.

4.3.1 Definition and properties of the smoothing functions

We start our discussion by introducing the function θ with the following properties (these funtions were used in [2, 4, 14, 52]). Let $\theta : \mathbb{R} \to]-\infty, 1[$ be a non-decreasing continuous function such that

$$\theta(t) < 0 \ \text{ if } \ t < 0, \ \ \theta(0) = 0 \ \text{ and } \ \lim_{t \to +\infty} \theta(t) = 1.$$

For instance,

$$\theta^1(t) = \begin{cases} \dfrac{t}{t+1} & \text{if } t \geqslant 0, \\ t & \text{if } t < 0, \end{cases}$$

and

$$\theta^2(t) = 1 - e^{-t}, \ t \in \mathbb{R}.$$

We will often return to these two examples very different from each other. We will also use these two functions in the numerical section.
In order to "detect" if $t = 0$ or $t > 0$ in a "continuous way", we introduce $\theta_r(t) = \theta(\frac{t}{r})$ for $r > 0$ and $\lim_{r \to 0} \theta(t) = 1$ for all $t > 0$.

4.3.1.1 θ-smoothing of a complementarity condition

Let $x, z \in \mathbb{R}$ be two scalars such that

$$0 \leqslant x \perp z \geqslant 0, \tag{4.3.3}$$

that is,

$$x \geqslant 0, \quad z \geqslant 0, \quad xz = 0.$$

In the (x, z)-plane, the set of points obeying (4.3.3) is the union of the two semi-axes $\{x \geqslant 0,\ z = 0\}$ and $\{x = 0,\ z \geqslant 0\}$. Visually, the nonsmoothness of (4.3.3) is manifested by the "kink" at the corner $(x, z) = (0, 0)$.
We consider two possible smooth approximations of (4.3.3), depending how it is rewritten in terms of θ-function.

Lemma 4.3.1. *[52] Given $x, z \in \mathbb{R}_+$ and the parameter $r > 0$, we have the equivalence*

$$xz = 0 \iff \lim_{r \searrow 0} (\theta_r(x) + \theta_r(z)) \leqslant 1.$$

Lemma 4.3.2. *[52] θ_r is sub-additive for non-negative values, i.e. given $x,\ z \geqslant 0$ it holds that*

$$\theta_r(x) + \theta_r(z) \ \geqslant \ \theta_r(x + z),$$

and with equality if and only if $x = 0$ or $z = 0$,

$$xz = 0 \iff \theta_r(x) + \theta_r(z) = \theta_r(x + z).$$

Now, let us discuss the following equation in the one-dimensional case. Let $s, t \in \mathbb{R}_+$, be such that

$$\theta_r(t) + \theta_r(s) = 1. \tag{4.3.4}$$

For instance, let us take θ^1. The equality (4.3.4) is then equivalent to $st = r^2$.
So, when r goes to 0, we simply get $st = 0$. The equation (4.3.4) applied with $s = x \in \mathbb{R}_+$ and $t = F(x) \in \mathbb{R}_+$ is then an approximation of the relation $xF(x) = 0$.

4.3.2 A new smoothing function using θ-function

Our aim is to propose a large class of θ-functions for which the problems

$$x^{(r)} \geqslant 0, \quad F(x^{(r)}) \geqslant 0 \qquad \text{and} \qquad \theta_r(x^{(r)}) + \theta_r(F(x^{(r)})) = 1, \tag{4.3.5}$$

are well-posed and any limit point of $(x^{(r)})$ when r goes to 0, is a solution of NCP. In the multidimensional case, the equation just above has to be interpreted as a system of n equations,

$$\theta_r(x_i^{(r)}) + \theta_r(F_i(x^{(r)})) = 1, \quad i = 1, ..., n.$$

Note that the relation (4.3.5) is symmetric in x and $F(x)$. Thus, our problem can be seen as a fixed point problem for the function $F_{r,\theta}(x)$ defined just below. Indeed, the equation (4.3.5) is equivalent to

$$x = \theta_r^{-1}(1 - \theta_r(F(x))) = r\theta^{-1}(1 - \theta(F(x)/r)) =: F_{r,\theta}(x).$$

By symmetry of the equation (4.3.5), we also have the relations:

$$F(x) = \theta_r^{-1}(1 - \theta_r(x)) = r\theta^{-1}(1 - \theta(x/r)),$$

but we shall not go that direction. As in [53] we propose another way to approximate a solution of the NCP problem as follows. Let $\psi_r(t) = 1 - \theta_r(t)$, the relation (4.3.5) is equivalent to the three following equalities

$$\psi_r(x) + \psi_r(F(x)) = 1 = \psi_r(0),$$

$$\psi_r^{-1}\left[\psi_r(x) + \psi_r(F(x))\right] = 0 \quad \text{and}$$

$$r\psi^{-1}\left[\psi\left(\tfrac{x}{r}\right) + \psi\left(\tfrac{F(x)}{r}\right)\right] = 0.$$

For the sequel, we set for any $x, y \in \mathbb{R}^n$ and any $r > 0$

$$G_r(x, y) := (G_r(x_i, y_i))_{i=1,...,n} := \left(r\psi^{-1}\left[\psi\left(\frac{x_i}{r}\right) + \psi\left(\frac{y_i}{r}\right)\right]\right)_{i=1,...,n}, \tag{4.3.6}$$

where $\psi : \mathbb{R} \to]0, +\infty[$.
First, we characterize the solutions (x, y) of $G_r(x, y) = 0$ when ψ satisfies some conditions independent of F. Let $0 < a < 1$. We say that ψ satisfies condition (H_a) if there exists $s_a > 0$ such that

$$\psi(s) \leqslant \frac{1}{2}\psi(as) \quad \forall s \geqslant s_a \quad \text{or equivalently} \quad \frac{1}{2} + \frac{1}{2}\theta(as) \leqslant \theta(s) \quad \forall s \geqslant s_a. \tag{H_a}$$

The condition (H_a) imposes that the decay of $\psi(s)$ is under some uniform control for large s or in terms of θ that $\theta(s)$ should grow enough quickly with some uniformity for large s. Since ψ and θ are monotone, it is interesting to take a as large as possible in the condition (H_a) since $(H_a) \Longrightarrow (H_b)$ for $b < a$.

Note that we can never take $a = 1$ because $\theta \leqslant 1$ unless θ is constant and equal to one for large s. But in some cases, a can be chosen as close to 1, see for instance θ^2.
One can obtain by simple calculations that:

1. For θ^1, we have
$$\psi^1(t) = \begin{cases} \dfrac{1}{t+1} & \text{if} \quad t \geqslant 0, \\ 1 - t & \text{if} \quad t < 0, \end{cases}$$
and the condition (H_a) is only satisfied for $0 < a < 1/2$ with $s_a \geqslant \dfrac{1}{1-2a}$.

2. For θ^2, we have $\psi(t) = e^{-t}$ and the condition (H_a) is satisfied for any $0 < a < 1$ with $s_a = \dfrac{\ln 2}{1-a}$.

From now on, all the results use the function ψ. Obviously, everything can be easily transposed on θ. The following lemma compare the function G_r defined in (4.3.6) to the min function and will be useful for the rest of our analysis.

Lemma 4.3.3. *If $\psi : \mathbb{R} \to]0, +\infty[$ is an invertible non-increasing function, then for any $(s,t) \in \mathbb{R}^2$ and any $r > 0$*
$$G_r(s,t) \leqslant \min(s,t).$$

Proof. Let $s, t \in \mathbb{R}$ be fixed. By symmetry, we can assume that $s = \min(s,t)$. Since $\psi \geqslant 0$, we obviously have
$$\psi(s/r) \leqslant \psi(s/r) + \psi(t/r).$$
By the fact that ψ is invertible and non-increasing, we get
$$\psi^{-1}(\psi(s/r) + \psi(t/r)) \leqslant s/r.$$
Thus, from the definition of G_r we conclude that
$$G_r(s,t) = r\psi^{-1}[\psi(s/r) + \psi(t/r)] \leqslant s = \min(s,t).$$
□

The next theorem shows how the condition (H_a) gives information about the behavior of G_r.

Theorem 4.3.4. *Let $\psi : \mathbb{R} \to]0, +\infty[$ be an invertible non-increasing function such that*
$$\lim_{t \to -\infty} \psi(t) = +\infty, \psi(0) = 1, \text{and} \lim_{t \to +\infty} \psi(t) = 0.$$

If ψ satisfies the condition (H_a) for some $a \in]0,1[$, then for all $s,t \in \mathbb{R}$,

$$\lim_{r \searrow 0} G_r(s,t) = 0 \iff \min(s,t) = 0.$$

Proof. We start by the direct implication:
Let $s,t \in \mathbb{R}$ be fixed. By Lemma 4.3.3, for any $r > 0$ we have $G_r(s,t) \leqslant \min(s,t)$ and, then $\min(s,t) \geqslant 0$. We finish the proof by contradiction as follows. Assume that $s = \min(s,t) > 0$.
Since ψ is nonincreasing and $s \leqslant t$, we have

$$\psi(s/r) + \psi(t/r) \leqslant 2\psi(s/r).$$

By assumption (H_a) and for r small enough, $2\psi(s/r) \leqslant \psi(as/r)$. Indeed, the ratio s/r goes to infinity as r goes to 0 because $s > 0$. Hence

$$\psi(s/r) + \psi(t/r) \leqslant \psi(as/r).$$

Now since ψ^{-1} is nonincreasing,

$$as/r \leqslant \psi^{-1}(\psi(s/r) + \psi(t/r)),$$

or equivalently with r small enough, $s \leqslant a^{-1}G_r(s,t)$.
Passing to the limit, $\lim_{r \searrow 0} G_r(s,t) = 0$ and, then $s \leqslant 0$ in contradiction with $s > 0$.
Now, we prove the converse ($\Leftarrow$):
Assume $s = \min(s,t)$. Hence, $s = 0$. Since $\psi(0) = 1$, we have

$$G_r(s,t) = r\psi^{-1}(1 + \psi(t/r)).$$

If $t = 0$ then $\lim_{r \searrow 0} G_r(s,t) = \lim_{r \searrow 0} r\psi^{-1}(2) = 0$.
If $t > 0$ then $\lim_{r \searrow 0} \psi(t/r) = 0$. Thus $\lim_{r \searrow 0} G_r(s,t) = 0$ by continuity of ψ^{-1}.
In both cases, we have $\lim_{r \searrow 0} G_r(s,t) = 0$. □

For both θ_r^1 and θ_r^2 examples, the assertion of Theorem 4.3.4 is clearly satisfied. Indeed direct computations lead to

1. For $s > 0$ and $t > 0$ such that $\frac{1}{s} + \frac{1}{t} \leqslant \frac{1}{r}$, we have the following explicit expression

$$G_r^1(s,t) = \frac{st - r^2}{s + t + 2r}. \tag{4.3.7}$$

Note that the denominator is not zero when s, t are nonnegative even when $s = t = 0$. In addition, when $\min(s,t) > 0$ we have $\lim_{r \searrow 0} G_r^1(s,t) = \dfrac{st}{s+t} < \min(s,t)$.

2. For any $s, t \in \mathbb{R}$, we have the following explicit expression

$$G_r^2(s,t) = -r \log(e^{-s/r} + e^{-t/r}). \tag{4.3.8}$$

Assume $s = \min(s,t)$. Then we have $s - r\log(2) \leqslant G_r^2(s,t)$ because

$$e^{-s/r} + e^{-t/r} \leqslant 2e^{-s/r}.$$

Thus, $\min(s,t) - r\log(2) \leqslant G_r^2(s,t) \leqslant \min(s,t)$. Passing to the limit as r goes to 0, we conclude that $\lim_{r \searrow 0} G_r^2(s,t) = \min(s,t)$.

Figure 4.1 illustrate the behaviour of $G_r(x,-x)$.

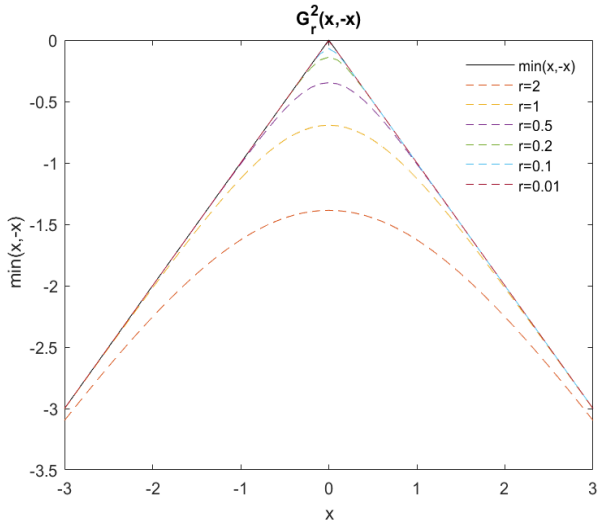

Figure 4.1: Comparison of $G_r^2(x,-x)$ and $\min(x,-x)$.

Now, we focus on the case where ψ satisfies (H_a) for all $a \in]0,1[$ and prove a stronger result.

Theorem 4.3.5. *Let $\psi : \mathbb{R} \to]0,+\infty[$ be an invertible non-increasing function such that*

$$\lim_{t \to -\infty} \psi(t) = +\infty, \psi(0) = 1, \text{and} \lim_{t \to +\infty} \psi(t) = 0.$$

If ψ satisfies (H_a) for all $a \in]0,1[$, then for any $s,t>0$,

$$\lim_{r\searrow 0} G_r(s,t) = \min(s,t).$$

Proof. By Lemma 4.3.3, we have

$$\forall r>0, \quad \forall s,t\in\mathbb{R}, \quad G_r(s,t) \leqslant \min(s,t).$$

Thus, we have to concentrate on the lower bound of G_r.
Let $s,t>0$ such that $s=\min(s,t)>0$. For each $a\in]0,1[$ and when r is sufficiently small (i.e. $s/r \geqslant s_a > 0$) we can apply the assumption (H_a) to get

$$\psi(s/r)+\psi(t/r) \leqslant 2\psi(s/r) \leqslant \psi(as/r).$$

Since ψ^{-1} is nonincreasing, we deduce

$$as/r \leqslant \psi^{-1}(\psi(s/r)+\psi(t/r)).$$

Thus, for any $a\in]0,1[$ and any $0<r<s/s_a$, we have $as \leqslant G_r(s,t)$,
Hence,

$$a\min(s,t) = as \leqslant \liminf_{r\searrow 0} G_r(s,t) \leqslant \limsup_{r\searrow 0} G_r(s,t) \leqslant \min(s,t).$$

By taking $a\nearrow 1$, we obtain the desired result. □

4.3.3 An approximate formulation

In this section, we present two new reformulations of the complementarity problem (4.2.1) by corresponding to two approximation schemes.

4.3.3.1 Approximation of NCP using θ_r^1-function

Using θ_r^1-function, we regularize each complementarity constraint by considering

$$x_iz_i=0,\ x_i\geqslant 0, z_i\geqslant 0 \quad \text{by} \qquad G_r^1(x_i,z_i) := \frac{x_iz_i-r^2}{x_i+z_i+2r} = 0, \qquad i=1,...,n.$$

This approximation yields the following formulation

$$(\tilde{P}_{\theta_1}) \qquad \begin{cases} F(x) = z, \\ x \geqslant 0, \quad z \geqslant 0, \quad r \searrow 0, \\ G_r^1(x,z) = 0. \end{cases} \tag{4.3.9}$$

We consider the family $\{H_{\theta_1}^r(.),\ r > 0\}$, where

$$H_{\theta_1}^r(x,z) = \begin{bmatrix} F(x) - z \\ G_r^1(x,z) \end{bmatrix}, \tag{4.3.10}$$

is a regularized function of H defined in (4.3.2).

Lemma 4.3.6. *Let $H_{\theta_1}^r(x,z)$ be defined by (4.3.10). Then, for any $(x,z) \in \mathbb{R}_+^{2n}$ the Jacobian matrix of $H_{\theta_1}^r(x,z)$ is*

$$\nabla H_{\theta_1}^r(x,z) = \begin{pmatrix} \nabla F(x) & -I \\ D_a(x,z) & D_b(x,z) \end{pmatrix},$$

where $D_a(x,z) = diag\{a_1(x,z), ..., a_n(x,z)\}$ and $D_b(x,z) = diag\{b_1(x,z), ..., b_n(x,z)\}$ are two diagonal matrices, given by

$$a_i(x,z) = \left(\frac{z_i + r}{x_i + z_i + 2r}\right)^2, \qquad b_i(x,z) = \left(\frac{x_i + r}{x_i + z_i + 2r}\right)^2, \qquad i = 1, ..., n.$$

4.3.3.2 Approximation of NCP using θ_r^2-function

Using the θ_r^2-function defined above, we obtain an approximate formulation for NCP

$$(\tilde{P}_{\theta_2}) \qquad \begin{cases} F(x) = z, \\ x \geqslant 0, \quad z \geqslant 0, \quad r \searrow 0, \\ G_r^2(x,z) = 0. \end{cases} \tag{4.3.11}$$

Where

$$G_r^2(x,z) := \left(G_r^2(x_i,z_i)\right)_{i=1,...,n} := \left(-r\log(e^{-x_i/r} + e^{-z_i/r})\right)_{i=1,...,n}.$$

We consider the family $\{H^r_{\theta_2}(.),\ r>0\}$, where

$$H^r_{\theta_2}(x,z)=\begin{bmatrix}F(x)-z\\ G^2_r(x,z)\end{bmatrix}, \tag{4.3.12}$$

is a regularized function of H defined in (4.3.2).

Lemma 4.3.7. *Let $H^r_{\theta_2}(x,z)$ be defined by (4.3.12). Then, the Jacobian matrix of $H^r_{\theta_2}(x,z)$ is*

$$\nabla H^r_{\theta_2}(x,z)=\begin{pmatrix}\nabla F(x) & -I\\ Q_k(x,z) & Q_l(x,z)\end{pmatrix},$$

where $Q_k(x,z)=diag\{k_1(x,z),...,k_n(x,z)\}$ and $Q_l(x,z)=diag\{l_1(x,z),...,l_n(x,z)\}$ are two diagonal matrices, given by

$$k_i(x,z)=\frac{e^{-x_i/r}}{e^{-x_i/r}+e^{-z_i/r}},\qquad l_i(x,z)=\frac{e^{-z_i/r}}{e^{-x_i/r}+e^{-z_i/r}},\qquad i=1,...,n.$$

In order to study the nonsingularity of the Jacobian matrix of $H^r_{\theta_1}(x,z)$ (resp. $H^r_{\theta_2}(x,z)$) , we state first a basic but essential lemma.

Lemma 4.3.8. *Let $M\in\mathbb{R}^{n\times n}$ be a $\mathcal{P}_0$-matrix. Then any matrix in the following form is nonsingular:*

$$N_s+N_tM,$$

where $N_s\in\mathbb{R}^{n\times n}$ is a positive (negative) diagonal matrix, and $N_t\in\mathbb{R}^{n\times n}$ is a nonnegative (non-positive) diagonal matrix.

Proof. Let $N_s=\mathrm{diag}(s_1,s_2,...,s_n)$ and $N_t=\mathrm{diag}(t_1,t_2,...,t_n)$. If N_s is positive, and N_t is nonnegative, then $s_i>0$ and $t_i\geqslant 0$ for all $i=1,2,...,n$.
Let $v\in\mathbb{R}^n$ be a vector such that $(N_s+N_tM)v=0$. Then, we have $v_i=-\dfrac{t_i}{s_i}(Mv)_i$.
It yields $v_i^2=-\dfrac{t_i}{s_i}v_i(Mv)_i$. If $t_i=0$, then $v_i=0,\quad\forall i=1,...,n$.
If $v_i\neq 0$, we have $\frac{t_i}{s_i}>0$. Owing to $v_i^2\geqslant 0$, we have $v_i(Mv)_i\leqslant 0$. If $v_i(Mv)_i=0$, then $v_i=0$. Otherwise, $v_i(Mv)_i<0$ contradicts the property of M. Based on the above discussion, it is concluded that $v=0$, then N_s+N_tM is a nonsingular matrix. □

By Lemma 4.3.8, we can obtain a property of $H^r_{\theta_1}$ and $H^r_{\theta_2}$ if F is a $\mathcal{P}_0$-matrix.

Theorem 4.3.9. *Let F be a $\mathcal{P}_0$-function. Then, for any $r>0$, and any $(x,z)\in\mathbb{R}^{2n}_+$ the Jacobian matrix $\nabla H^r_{\theta_1}(x,z)$ (resp. $\nabla H^r_{\theta_2}(x,z)$) is nonsingular.*

Proof. For all $r > 0$ and from Lemma 4.3.6 and Lemma 4.3.7, it follows that the diagonal matrix $D_a(x,z)$ (resp. $Q_k(x,z)$) is non-negative, and $D_b(x,z)$ (resp. $Q_l(x,z)$) is non-negative diagonal matrix. Since F is a $\mathcal{P}_0$-function, the Jacobian matrix $\nabla F(x)$ is a $\mathcal{P}_0$-matrix.
We have

$$\det(\nabla H^r_{\theta_1}(x,z)) = \det(D_a(x,z) + \nabla F(x)D_b(x,z)),$$

and

$$\det(\nabla H^r_{\theta_2}(x,z)) = \det(Q_k(x,z) + \nabla F(x)Q_l(x,z)).$$

Since $\nabla F(x)$ is a $\mathcal{P}_0$-matrix and from Lemma 4.3.8, it follows that $D_a(x,z) + \nabla F(x)D_b(x,z)$ (resp. $Q_k(x,z) + \nabla F(x)Q_l(x,z)$) is nonsingular. Hence $\nabla H^r_{\theta_1}(x,z)$ (resp. $\nabla H^r_{\theta_2}(x,z)$) is nonsingular. □

4.4 New approach for solving nonlinear complementarity problems

In this section, we present the idea of our algorithms, we take inspiration from the well-known interior-point methods (IPMs) usually used in nonlinear programming. Even though we don't have any objective function to minimize, the regularization idea behind IPM can be used to tackle NCP.
One can replace the original nonsmooth problems NCPs by a sequence of regularized problems

$$H_r(\mathbf{X}) = 0, \quad \Longleftrightarrow \quad \begin{cases} F(x) = z, \\ x.z = r\mathbf{e}, \end{cases} \tag{4.4.1}$$

where

$$\mathbf{X} = \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{2n}_+, \quad H_r(\mathbf{X}) = \begin{bmatrix} F(x) - z \\ x.z - r\mathbf{e} \end{bmatrix}, \tag{4.4.2}$$

and $r \geqslant 0$ is the smoothing parameter, $\mathbf{e} \in \mathbb{R}^n$ is the vector whose components are all equal to 1.
The Jacobian matrix of H_r with respect to $\mathbf{X}$, does not depend on r and can be denoted by

$$\nabla H_r(\mathbf{X}) = \begin{pmatrix} \nabla F(x) & -I \\ Z & X \end{pmatrix}, \tag{4.4.3}$$

where Z and (resp. X) the diagonal matrix of z (resp. x).
The main difficulty in this approach, is to drive r to 0. In Haddou et al [104] the authors propose a new technique where r is considered as a new variable.

4.4.1 When the parameter becomes a variable

In the system (4.4.1), the status of the parameter r is very distinct from that of the variable $\mathbf{X}$. While $\mathbf{X}$ is computed "automatically" by a Newton iteration, r has to be updated "manually" in an ad-hoc manner.
Our goal is to find a strategy that decreases r during iterations and ensures the nonnegative of variables. However, we must adjust the strategy when the model or its parameters are changed. To avoid this trouble, we consider r as an unknown of the system instead of a parameter as in [104].
We feel that it would be judicious to incorporate the parameter r into the variables. Let us, therefore, consider the enlarged vector of unknowns

$$\mathbb{X} = \begin{bmatrix} \mathbf{X} \\ r \end{bmatrix} \in \mathbb{R}^{2n}_{+} \times \mathbb{R}_{+}, \tag{4.4.4}$$

and then consider a system of $2n + 1$ equations

$$\mathbb{H}_{\theta}(\mathbb{X}) = 0, \tag{4.4.5}$$

to be on $\mathbb{X}$. To this end, let us remind ourselves that our ultimate goal is to solve $H^0_{\theta_1}(\mathbf{X})$, together with the inequalities $x \geqslant 0,\ z \geqslant 0$. We restrict our choice of θ-function to $\theta_r(t) = \theta^1_r(t)$.
Thus, it is really natural to first consider

$$\mathbb{H}_{\theta}(\mathbb{X}) = \begin{bmatrix} H^r_{\theta_1}(\mathbf{X}) \\ r \end{bmatrix}. \tag{4.4.6}$$

where

$$H^r_{\theta_1}(\mathbf{X}) = \begin{bmatrix} F(x) - z \\ G^1_r(\mathbf{X}) \end{bmatrix}.$$

This construction turns out to be to naive. Indeed, if we start from some r^0 and solve the smooth system (4.4.6) by the smooth Newton method, since the last equation is linear, we end up with $r^1 = 0$ at the first iteration. Once the boundary of the interior region is reached, we are "stuck" there.
To prevent r from rushing to zero in just one iteration, we could set

$$\mathbb{H}_{\theta}(\mathbb{X}) = \begin{bmatrix} H^r_{\theta_1}(\mathbf{X}) \\ r^2 \end{bmatrix}. \tag{4.4.7}$$

At this stage, system (4.4.7) is not yet fully adequate. Indeed, the last equation is totally decoupled from the others. Everything happens as if r follows a prefixed sequence, generated by the Newton iterates of the scalar equation $r^2 = 0$, regardless of $\mathbf{X}$. It is desirable to couple r and $\mathbf{X}$ in a tighter way. In this respect, we advocate

$$\mathbb{H}_\theta(\mathbb{X}) = \begin{bmatrix} H^r_{\theta_1}(\mathbf{X}) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 \end{bmatrix}, \tag{4.4.8}$$

where

$$\|x^-\|^2 = \sum_{i=1}^{n} \min^2(x_i, 0), \quad \|z^-\|^2 = \sum_{i=1}^{n} \min^2(z_i, 0).$$

This choice has the benefit of taking into account the nonnegativity condition $x \geqslant 0$ and $z \geqslant 0$. Indeed, the last equation of (4.4.8) implies that, as long as $r \geqslant 0$, we are ascertained that $x^- = z^- = 0$. This amounts to saying that $x \geqslant 0$ and $z \geqslant 0$. Should a component of x or z become negative during the iteration, this equation would contribute to "penalize" it.
Since r is now considered as a variable and the scalar function $t \mapsto \frac{1}{2}|\min(t, 0)|^2$ is differentiable and its derivative is equal to $\min(t, 0)$. From this observation, the Jacobian matrix of $\mathbb{H}_\theta$ is:

$$\nabla_{\mathbb{X}}\mathbb{H}_\theta(\mathbb{X}) = \begin{pmatrix} \nabla_x H^r_{\theta_1} & \nabla_z H^r_{\theta_1} & \partial_r H^r_{\theta_1} \\ & & \\ (x^-)^{\mathrm{T}} & (z^-)^{\mathrm{T}} & 2r \end{pmatrix}, \tag{4.4.9}$$

where x^- is the vector of components $x_i^- = \min(x_i, 0)$ and similarly for z^-,

$$\nabla_x H^r_{\theta_1} = \begin{pmatrix} \nabla_x F(x) \\ D_a(x, z) \end{pmatrix}_{2n\times n}, \quad \nabla_z H^r_{\theta_1} = \begin{pmatrix} -I \\ D_b(x, z) \end{pmatrix}_{2n\times n},$$

$$\partial_r H^r_{\theta_1} = \begin{bmatrix} 0_{n\times 1} \\ \operatorname{diag}\left(\left(\dfrac{-2r}{x_i + z_i + 2r} + \dfrac{2(r^2 - x_i z_i)}{(x_i + z_i + 2r)^2}\right)_{1\leqslant i\leqslant n}\right)\mathbf{e} \end{bmatrix}_{2n\times 1},$$

and $\mathbf{e}$ is a n-dimensional vector whose entries are equal to 1.
If $\mathbb{H}_\theta(\mathbb{X}) = 0$ where $\mathbb{X} \in \mathbb{R}^{2n}_+ \times \mathbb{R}_+$ we obtain $r = 0$ and $x^- = z^- = 0$. Hence in this case, $\nabla_{\mathbb{X}}\mathbb{H}_\theta(\mathbb{X})$ becomes singular, since $\det(\nabla_{\mathbb{X}}\mathbb{H}_\theta(\mathbb{X})) = 0$. To solve this issue, we add a small enough positive

parameter ε in the last equation. We get

$$\frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r = 0. \tag{4.4.10}$$

Hence, we define the following systems

$$\mathbb{H}_\theta(\mathbb{X}) = \begin{bmatrix} H^r_{\theta_1}(\mathbf{X}) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0. \tag{4.4.11}$$

Lemma 4.4.1. *Let $\mathbf{X} \in \bar{\Xi}$ (the closure of Ξ), where Ξ is the interior region defined in*

$$\Xi = \{\mathbf{X} = (x, z) \in \mathbb{R}^{2n} \quad | \quad x > 0, \quad z > 0\}. \tag{4.4.12}$$

Let $r \in \mathbb{R}$ and $\mathbb{X} = [\mathbf{X}; r]^T$. Then,

$$det\ \nabla\mathbb{H}_\theta(\mathbb{X}) = (\varepsilon + 2r)\ det\ \nabla H^r_{\theta_1}(\mathbf{X}).$$

If $\varepsilon + 2r > 0$, the two Jacobian matrices $\nabla\mathbb{H}_\theta$ and $\nabla H^r_{\theta_1}$ are singular or nonsingular at the same time.

Proof. Thanks to the assumption $\mathbf{X} \in \bar{\Xi}$, we have $x \geqslant 0$ and $z \geqslant 0$, so that $x^- = z^- = 0$. Expanding the determinant of (4.4.11) with respect to the last row yields the desired result. □

4.5 Convergence

In this section, we propose a generic algorithm to solve NCP and prove some convergence results. From now on, the enlarged equation (4.4.11) is selected as the reference system in the design of our new algorithm. The idea is simply to apply the standard Newton method to the smooth system (4.4.11). To enforce a global convergence behavior, we also recommend using α line search like Armijo back-tracking technique.
Now, we present our algorithm for our method described above:

Algorithm 4.1 Nonparametric method with Armijo line search

1. Chose $\mathbb{X}^0 = (\mathbf{X}^0, r^0),\ \mathbf{X}^0 \in \Xi,\ r^0 =< x^0, z^0 > /n,\ \tau \in (0, 1/2),\ \varrho \in (0, 1)$. Set $k = 0$.
2. If $\mathbb{H}_\theta(\mathbb{X}^k) = 0$, stop.
3. Find a direction $\mathbf{d}^k \in \mathbb{R}^{2n+1}$ such that

$$\mathbb{H}_\theta(\mathbb{X}^k) + \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbb{X}^k)\mathbf{d}^k = 0.$$

4. Choose $\zeta^k = \varrho^{j_k} \in (0, 1)$, where $j_k \in \mathbb{N}$ is the smallest integer such that

$$\Theta(\mathbb{X}^k + \varrho^{j_k}\mathbf{d}^k) - \Theta(\mathbb{X}^k) \leqslant \tau\varrho^{j_k}\ \nabla\Theta(\mathbb{X}^k)^T\mathbf{d}^k.$$

5. Set $\mathbb{X}^{k+1} = \mathbb{X}^k + \zeta^k\mathbf{d}^k$ and $k \leftarrow k + 1$. Go to step 2.

The merit function used in the line search is:

$$\Theta(\mathbb{X}) = \frac{1}{2}\|\mathbb{H}_\theta(\mathbb{X})\|^2.$$

A detailed description of nonparametric method is given in Algorithm 4.1. A few comments are in order:

- The initial point $\mathbb{X}^0 = (\mathbf{X}^0, r^0)$ must be an interior point, namely, $\mathbf{X}^0 > 0$ and the initial parameter $r^0 =< x^0, z^0 > /n$ has the correct order of magnitude.
- If $\mathbf{X}^k \in \Xi$, then $(x^k)^- = (z^k)^- = 0$ and

$$\mathbf{d}^k = \begin{bmatrix} d\mathbf{X}^k \\ dr^k \end{bmatrix} = -\begin{pmatrix} \nabla_x H^r_{\theta_1} & \nabla_z H^r_{\theta_1} & \partial_r H^r_{\theta_1} \\ 0^{\mathrm{T}} & 0^{\mathrm{T}} & \varepsilon + 2r^k \end{pmatrix}^{-1} \begin{bmatrix} H^r_{\theta_1}(\mathbf{X}^k) \\ \varepsilon r^k + (r^k)^2 \end{bmatrix},$$

  provided that the Jacobian matrix is invertible. The increment for the parameter is then

$$dr^k = -\frac{\varepsilon r^k + (r^k)^2}{\varepsilon + 2r^k}.$$

- There is no need to truncate the Newton direction $\mathbf{d}^k$ to preserve positivity for x^{k+1} and z^{k+1},

since nonnegativity is "guaranteed" at convergence. However, if we want all the iterates to be nonnegative, so we need to carry out additional damping after Step 4 (Armijo's line search).

Proposition 4.5.1. *Let F be a continuous differentiable $\mathcal{P}_0$-function. Then, step 3 in Algorithm 4.1 is well-defined.*

Proof. We know that for all $k \geqslant 0,\ r^k > 0, \mathbf{X}^k > 0$, and $\varepsilon > 0$,

$$\det \nabla \mathbb{H}_\theta(\mathbb{X}^k) = (\varepsilon + 2r^k) \det \nabla H_{\theta_1}^{r^k}(\mathbf{X}^k).$$

By Theorem 4.3.9, $\nabla H_{\theta_1}^{r^k}(\mathbf{X}^k)$ is nonsingular, so Step 3 of Algorithm 4.1 is well-defined. □

4.5.1 Global convergence analysis

Definition 4.5.2. *(Regular zero). Let $\mathbb{X}^* \in \mathbb{R}^{2n+1}$ be a zero of $\mathbb{H}_\theta$, that is, $\mathbb{H}_\theta(\mathbb{X}^*) = 0$. If the Jacobian matrix $\nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbb{X}^*)$ is nonsingular, $\mathbb{X}^*$ is said to be a regular zero of $\mathbb{H}_\theta$.*

The main interest of Algorithm 4.1 lies in the prospect of global convergence, as envisioned by the theory that we are developing now. This global convergence theory, is primarily based on the regularity of zeros [Definition (4.5.2)]. We reproduce a concise result that can be found in the book of Bonnans [20], in view of its importance to our algorithm.
We will prove the global convergence of Algorithm 4.1. First, we show that every $\mathbf{d} \in \triangle$ is a descent direction of Θ at $\mathbb{X}$, where

$$\triangle(\mathbb{X}) = \{\mathbf{d} \in \mathbb{R}^n \quad | \quad \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbb{X})\mathbf{d} = -\mathbb{H}_\theta(\mathbb{X})\}. \tag{4.5.1}$$

Lemma 4.5.3. *(see [107]) If $\mathbb{X}$ is not a solution of NCP, i.e. $\Theta(\mathbb{X}) > 0$, then every $\mathbf{d} \in \triangle(\mathbb{X})$ satisfies the descent condition for $\mathbb{X}$, i.e., $\nabla\Theta(\mathbb{X})^T\mathbf{d} < 0$.*

Using the preceding results, we can prove the following global convergence theorem.

Theorem 4.5.4. *Every limit point $\mathbb{X}^* = (\mathbf{X}^*, r^*)$ of a sequence $\{\mathbb{X}^k\}$ generated by Algorithm 4.1 corresponds to a solution of NCP.*

Proof. Since the sequence $\{\Theta(\mathbb{X}^k)\}$ is nonnegative and decreases monotonically, it converges to some $\Theta^* \geqslant 0$. We assume $\Theta^* > 0$. Let $\mathbb{X}^*$ be an accumulation point of $\{\mathbb{X}^k\}$ and $\{\mathbb{X}^k\}_{k\in K}$ be a subsequence converging to $\{\mathbb{X}^*\}$. Taking a further subsequence if necessary, we can assume without loss of generality that $\lim_{k\to\infty} d^k = d^*$, because Δ is uniformly compact near and closed at $\mathbb{X}^*$ (see [107]). Furthemore, by the closedness of Δ, we have

$$d^* \in \Delta(\mathbb{X}^*). \tag{4.5.2}$$

Since $\Theta(\mathbb{X}^k + \zeta^k \mathbf{d}^k) - \Theta(\mathbb{X}^k) \leqslant \zeta^k \tau \ \nabla\Theta(\mathbb{X}^k)^T \mathbf{d}^k \leqslant 0$. It is obvious that $\{\zeta^k \tau \ \nabla\Theta(\mathbb{X}^k)^T \mathbf{d}^k\}$ converges to 0. In order to prove $\nabla\Theta(\mathbb{X}^k)^T \mathbf{d}^k \to 0$, we show that $\{\zeta^k\}$ is bounded away from 0. Now suppose that there exists a subsequence such that $\zeta^k \to 0$. By the line search rule, we have

$$\frac{\Theta(\mathbb{X}^k + \sigma^k \mathbf{d}^k) - \Theta(\mathbb{X}^k)}{\sigma^k} > \tau \ \nabla\Theta(\mathbb{X}^k)^T \mathbf{d}^k, \tag{4.5.3}$$

where $\sigma^k = \frac{\zeta^k}{\rho^{jk}}$. Since $\sigma^k \to 0$, taking the limit of both sides of (4.5.3) yields

$$\nabla\Theta(\mathbb{X}^*)^T \mathbf{d}^* > \tau \nabla\Theta(\mathbb{X}^*)^T \mathbf{d}^*. \tag{4.5.4}$$

Since $\Theta(\mathbb{X}^*) = \Theta^* > 0$ by assumption, it follows from (4.5.2) and Lemma 4.5.3 that $\nabla\Theta(\mathbb{X}^*)^T \mathbf{d}^* < 0$. Since $\tau < 1$, this contradicts (4.5.4). This implies that $\{\zeta^k\}$ is bounded away from 0, and hence, $\{\nabla\Theta(\mathbb{X}^k)^T \mathbf{d}^k\}$ converge to 0. That is,

$$\lim_{k\to\infty} \nabla\Theta(\mathbb{X}^k)^T d^k = \nabla\Theta(\mathbb{X}^*)^T d^* = 0. \tag{4.5.5}$$

It then follows from (4.5.2) and Lemma 4.5.3 that $\Theta(\mathbb{X}^*) = 0$. This is contradictory to $\Theta^* > 0$. Therefore, we must have $\Theta(\mathbb{X}^k) \to 0$, which implies that any accumulaton point $\mathbb{X}^*$ of $\{\mathbb{X}^k\}$ satisfies $\Theta(\mathbb{X}^*) = 0$ and hence is a solution of NCP. The proof is complete.

□

Below is a result about the Jacobian matrix of $\mathbb{H}_\theta(\mathbb{X})$, when r goes to 0.

Lemma 4.5.5. *Suppose that $\mathbf{X}^* = (x^*, z^*)$ is a solution of NCP, then we have the following equality*

$$\liminf_{r\to 0} \quad \max\left((\nabla_x G_r^1(x^*, z^*))_{ii}, \ (\nabla_z G_r^1(x^*, z^*))_{ii}\right) \geqslant \frac{1}{4}, \quad \forall i = 1, ..., n,$$

where

$$\nabla_x G_r^1(x^*, z^*) = diag\left(\left(\left(\frac{z_i^* + r}{x_i^* + z_i^* + 2r}\right)^2\right)_{1\leqslant i\leqslant n}\right), \quad and \quad \nabla_z G_r^1(x^*, z^*) = diag\left(\left(\left(\frac{x_i^* + r}{x_i^* + z_i^* + 2r}\right)^2\right)_{1\leqslant i\leqslant n}\right).$$

Proof. By considering the two possible situations ($x_i^* \geqslant z_i^*$) and ($x_i^* < z_i^*$), very simple calculation give:

1. $x_i^* \geqslant z_i^*$

$$\max\left(\left(\frac{z_i^* + r}{x_i^* + z_i^* + 2r}\right)^2, \ \left(\frac{x_i^* + r}{x_i^* + z_i^* + 2r}\right)^2\right) = \left(\frac{x_i^* + r}{x_i^* + z_i^* + 2r}\right)^2 \geqslant \left(\frac{x_i^* + r}{2x_i^* + 2r}\right)^2 = \frac{1}{4}.$$

2. $z_i^* > x_i^*$
$$\max\left(\left(\frac{z_i^*+r}{x_i^*+z_i^*+2r}\right)^2,\ \left(\frac{x_i^*+r}{x_i^*+z_i^*+2r}\right)^2\right) = \left(\frac{z_i^*+r}{x_i^*+z_i^*+2r}\right)^2 \geqslant \left(\frac{z_i^*+r}{2z_i^*+2r}\right)^2 = \frac{1}{4}.$$

Since $\nabla_x G$ and $\nabla_z G$ are bounded, the proof is complete by passing to the limit. □

Lemma 4.5.6. *Let $\mathbf{X}^* = (x^*, z^*)$ be a solution of NCP satisfying the strict complementarity condition (i.e. $x_i^* + z_i^* > 0$, $\forall i \in \{1, ..., n\}$). We have*

$$\lim_{r\to 0} \left(\nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)\right) = \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} & 0_{n\times 1} \\ \phi(Z^*) & \phi(X^*) & 0_{n\times 1} \\ 0_{1\times n} & 0_{1\times n} & \varepsilon \end{pmatrix},$$

where

$$\phi(Z^*)_{ii} = \begin{cases} 1 & \text{if} \quad z_i^* \neq 0 \quad \text{and} \quad x_i^* = 0 \\ 0 & \text{if} \quad z_i^* = 0 \quad \text{and} \quad x_i^* \neq 0, \end{cases} \quad \text{and} \quad \phi(X^*)_{ii} = \begin{cases} 1 & \text{if} \quad x_i^* \neq 0 \quad \text{and} \quad z_i^* = 0 \\ 0 & \text{if} \quad x_i^* = 0 \quad \text{and} \quad z_i^* \neq 0. \end{cases}$$

Proof. By definition

$$\mathbb{H}_\theta(\mathbb{X}) = \begin{bmatrix} F(x) - z \\ G_r^1(x, z) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix}.$$

The Jacobian matrix of $\mathbb{H}_\theta$ is:

$$\nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbb{X}) = \begin{pmatrix} \nabla_x F(x) & -I_{n\times n} & 0_{n\times 1} \\ \nabla_x G_r^1(x, z) & \nabla_z G_r^1(x, z) & \partial_r G_r^1(x, z) \\ (x^-)^{\mathrm{T}} & (z^-)^{\mathrm{T}} & 2r + \varepsilon \end{pmatrix}.$$

1. The derivative of $G_r^1(x, z)$ with respect to x is:
$$\nabla_x G_r^1(x^*, z^*) = \operatorname{diag}\left(\left(\left(\frac{z_i^*+r}{x_i^*+z_i^*+2r}\right)^2\right)_{1\leqslant i\leqslant n}\right),$$
when r goes to 0 the only two cases to consider are:
   - $x_i^* \to 0$, and $z_i^* > 0$ $\forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}} (\nabla_x G_r^1(x^*,z^*))_{ii} = \lim_{r\to 0}\left(\frac{z_i^*+r}{z_i^*+2r}\right)^2 = 1.$$

- $x_i^* > 0$, and $z_i^* \to 0\ \forall i \in \{1,...,n\}$ then
$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}} (\nabla_x G_r^1(x^*,z^*))_{ii} = \lim_{r\to 0}\left(\frac{r}{x_i^*+2r}\right)^2 = 0.$$

2. The derivative of $G_r^1(x,z)$ with respect to z is:

$$\nabla_z G_r^1(x^*,z^*) = \operatorname{diag}\left(\left(\left(\frac{x_i^*+r}{x_i^*+z_i^*+2r}\right)^2\right)_{1\leqslant i\leqslant n}\right),$$

as below, the only two cases to consider are:

- $x_i^* \to 0$, and $z_i^* > 0\ \forall i \in \{1,...,n\}$ then
$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}} (\nabla_z G_r^1(x^*,z^*))_{ii} = \lim_{r\to 0}\left(\frac{r}{z_i^*+2r}\right)^2 = 0.$$

- $x_i^* > 0$, and $z_i^* \to 0\ \forall i \in \{1,...,n\}$ then
$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}} (\nabla_z G_r^1(x^*,z^*))_{ii} = \lim_{r\to 0}\left(\frac{x_i^*+r}{x_i^*+2r}\right)^2 = 1.$$

3. The derivative of $G_r^1(x,z)$ with respect to r is:

$$\partial_r G_r^1(x^*,z^*) = \left(\frac{-2r}{x_i^*+z_i^*+2r} + \frac{2(r^2-x_i^*z_i^*)}{(x_i^*+z_i^*+2r)^2}\right)_{i=1,...,n},$$

when r goes to 0 the only two cases to consider are:

- $x_i^* \to 0$, and $z_i^* > 0\ \forall i \in \{1,...,n\}$ then
$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}} (\partial_r G_r^1(x^*,z^*))_i = \lim_{r\to 0}\left(\frac{-2r}{z_i^*+2r} + \frac{2r^2}{(z_i^*+2r)^2}\right) = 0.$$

- $x_i^* > 0$, and $z_i^* \to 0$ $\forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}} (\partial_r G_r^1(x^*, z^*))_i = \lim_{r\to 0} \left(\frac{-2r}{x_i^* + 2r} + \frac{2r^2}{(x_i^* + 2r)^2} \right) = 0.$$

□

Finally, since $\mathbf{X}^* = (x^*, z^*)$ is a solution of NCP, we have $x^* \geqslant 0$ and $z^* \geqslant 0$, so that $x^- = z^- = 0$. Hence

$$\lim_{r\to 0} (\nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)) = \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} & 0_{n\times 1} \\ \phi(Z^*) & \phi(X^*) & 0_{n\times 1} \\ 0_{1\times n} & 0_{1\times n} & \varepsilon \end{pmatrix}.$$

We present now, two situations where we can conclude about the nonsingularity of $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)$.

Lemma 4.5.7. *Suppose that $\mathbf{X}^* = (x^*, z^*)$ is a solution of NCP, we have two possibilities when computing the determinant of $\mathbb{H}_\theta$ on (x^*, z^*).*

- *If $\mathbf{X}^* = (x^*, z^*)$ satisfies the strict complementarity condition, then by Lemma 4.5.12, $\forall\ \mathbb{I} \subset \{1, ..., n\}$ we have:*

$$\lim_{r\to 0} det\, (\nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)) = \left| \begin{pmatrix} \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} \\ \phi(Z^*) & \phi(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{pmatrix} \right| = \varepsilon \left| \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} \\ \phi(Z^*) & \phi(X^*) \end{pmatrix} \right| = \varepsilon |\nabla F(x^*)_{\mathbb{II}}|,$$

  therefore the matrix $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)$ exists and is invertible if F is $\mathcal{P}$-function.

- *If $\mathbf{X}^* = (x^*, z^*)$ does not satisfy the strict complementarity condition, then by Lemma 4.5.12, $\forall\ \mathbb{I} \subset \{1, ..., n\}$ we have:*

$$\lim_{r\to 0} det\, (\nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)) = \left| \begin{pmatrix} \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} \\ \phi(Z^*) & \phi(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{pmatrix} \right|$$

$$= \varepsilon \left| \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} \\ \phi(Z^*) & \phi(X^*) \end{pmatrix} \right| = \varepsilon \left| \begin{pmatrix} \nabla F(x^*) - I_{n\times n} & -I_{n\times n} \\ \phi(Z^*) + \phi(X^*) & \phi(X^*) \end{pmatrix} \right|,$$

from Lemma 4.5.5 we have:

$$\liminf_{r\to 0} \quad \max\left((\nabla_x G_r^1(x^*, z^*))_{ii},\ (\nabla_z G_r^1(x^*, z^*))_{ii}\right) \geqslant \frac{1}{4} > 0, \quad \forall i = 1, ..., n, \tag{4.5.6}$$

hence $\lim_{r\to 0} \left(\nabla_x G_r^1(x^*, z^*) + \nabla_z G_r^1(x^*, z^*)\right) = \phi(Z^*) + \phi(X^*)$ *is a positive diagonal matrix. From Lemma 4.3.8, we take:*

$$\begin{aligned} M &= \nabla F(x^*) - I_{n\times n}, \\ N_t &= \phi(X^*), \\ N_s &= \phi(Z^*) + \phi(X^*), \end{aligned} \tag{4.5.7}$$

where N_s is a positive diagonal matrix, and N_t is a nonnegative diagonale matrix. Therefore the matrix $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)$ *exists and is invertible if $\nabla F(x^*) - I_{n\times n}$ is $\mathcal{P}_0$-matrix.*

Remark 4.5.8. *The matrix* $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)$ *exists and is invertible if $\nabla F(x^*) - \beta I_{n\times n}$ is $\mathcal{P}_0$-matrix for any $\beta > 0$. Since $\phi(Z^*) + \beta\phi(X^*)$ is a positive diagonal matrix for any $\beta > 0$.*

Remark 4.5.9. *If $\nabla F(x^*) - \beta I_{n\times n}$ is $\mathcal{P}_0$-matrix for any $\beta > 0$ then $\nabla F(x^*)$ is a $\mathcal{P}$-matrix.*

In the following, we focus our attention on the superlinear convergence rate of Algorithm 4.1.

Theorem 4.5.10. *(Theorem 6.9, [20]). Let $\mathbb{H}_\theta : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n+1}$ be a continuously differentiable function.*

(i) (Local analysis) Let $\mathbb{X}^*$ be a regular zero of $\mathbb{H}_\theta$. If $\mathbb{X}^0$ is close enough to $\mathbb{X}^*$, then $\zeta^k = 1$ for all k, and $\mathbb{X}^k$ converge to $\mathbb{X}^*$ super-linearly (and we recover the standard Newton method).

(ii) (Limit point) Let $\mathbb{X}^*$ be a limit point of sequence $\{\mathbb{X}^k\}$. If $\nabla\mathbb{H}_\theta(\mathbb{X}^*)$ is invertible, then $\mathbb{X}^*$ is a regular zero of $\mathbb{H}_\theta$. If $\mathbb{X}^*$ is a regular zero of $\mathbb{H}_\theta$, then $\zeta^k = 1$ for k big enough and $\mathbb{X}^k$ converge to $\mathbb{X}^*$ super-linearly.

Proof. We apply Theorem 4.5.4 and Lemma 4.5.7 under some condition on F. □

The next lemma measures the "additional coercivity" effect of the smoothing.

Lemma 4.5.11. *Assume F is a $\mathcal{P}_0$-function, then*

(i) $\mathbb{H}_\theta$ is a $\mathcal{P}$-function.

(ii) If $\mathbb{H}_\theta(\mathbf{X}, 0)$ exists then it is a $\mathcal{P}_0$-function.

Proof. (i) Let X, Y be two distinct vectors of $\mathbb{R}^{2n}$. Since F is a $\mathcal{P}_0$-function there exists an index $i \in \{1, ..., 2n\}$ such that $X_i \neq Y_i$ and $(X_i - Y_i)(F_i(X) - F_i(Y)) \geqslant 0$. Without loss of generality, we can suppose that $X_i > Y_i$ and $F_i(X) > F_i(Y)$.
Since ψ and ψ^{-1} are decreasing, we obtain consecutively that for any $r > 0$,

$$\psi(X_i/r) + \psi(F_i(X)/r) < \psi(Y_i/r) + \psi(F_i(Y)/r), \tag{4.5.8}$$

so

$$G_r(X_i, F_i(X)) > G_r(Y_i, F_i(Y)).$$

Hence, $\mathbb{H}_\theta$ is a $\mathcal{P}$-function.
We will now deal with the case where $X_i = Y_i, \quad \forall i < 2n + 1$. For $i = 2n + 1$, we can suppose that $X_{2n+1} > Y_{2n+1}$

$$(X_{2n+1} - Y_{2n+1})(\mathbb{H}_\theta(X)_{2n+1} - \mathbb{H}_\theta(Y)_{2n+1}) = (r_1 - r_2)(r_1^2 + \varepsilon r_1 - r_2^2 - \varepsilon r_2) > 0.$$

Hence, $\mathbb{H}_\theta$ is a $\mathcal{P}$ function.
(ii) If $\mathbb{H}_\theta(\mathbb{X}, 0)$ exists, passing to the limit in (4.5.8) as $r \searrow 0$, we obtain that $\mathbb{H}_\theta(\mathbb{X}, 0)$ is a $\mathcal{P}_0$-function. □

Now we would like to study the asymptotic behavior of the Jacobian matrix of our method with the Jacobian matrix of IPM when r goes to 0 and we need a lemma that is used to prove our main result.

Lemma 4.5.12. *We consider the following system*

$$\begin{aligned} &Z.X = 0 \\ &Z \geqslant 0, \quad X \geqslant 0, \end{aligned} \tag{4.5.9}$$

where $Z = diag(z)$ and $X = diag(x)$.
Assume that Z, X are strictly complementary (i.e. $Z + X > 0_{n\times n}$). Then J is singular if and only if T is singular, where

$$J = \begin{pmatrix} \nabla F(x) & -I \\ Z & X \end{pmatrix}, \quad \textit{and} \quad T = \begin{pmatrix} \nabla F(x) & -I \\ \phi(Z) & \phi(X) \end{pmatrix},$$

where $\phi(.)$ is defined in Lemma 4.5.6, here ϕ operates component-wise on Z and X.

Proof. By the strict complementarity hypothesis, we range the rows and the columns of J and T as follows

$$J_\sigma = \begin{pmatrix} \nabla F(x)_\sigma & -I_\sigma \\ \begin{pmatrix} Z_1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & X_2 \end{pmatrix} \end{pmatrix},$$

where $X_2 > 0$ and $Z_1 > 0$, and

$$(T)_\sigma = \begin{pmatrix} \nabla F(x)_\sigma & -I_\sigma \\ \begin{pmatrix} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ & & & \\ & 0 & & 0 \end{pmatrix} & \begin{pmatrix} 0 & & & 0 \\ & \ddots & & \\ & & 1 & 0 \\ 0 & & & \ddots \\ & & 0 & 1 \end{pmatrix} \end{pmatrix}.$$

The determinant of the two matrices J_σ and $(T)_\sigma$ are equal to

$$\det(J_\sigma) = \begin{vmatrix} \nabla F(x)_\sigma & -I_\sigma \\ \begin{pmatrix} Z_1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & X_2 \end{pmatrix} \end{vmatrix} = \pm \prod_{i \in \mathbb{I}_1} x_i \prod_{i \in \mathbb{I}_2} z_i \ \det(C),$$

$$\det(T_\sigma) = \begin{vmatrix} \nabla F(x)_\sigma & -I_\sigma \\ \begin{pmatrix} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ & & & \\ & 0 & & 0 \end{pmatrix} & \begin{pmatrix} 0 & & & 0 \\ & \ddots & & \\ & & 1 & 0 \\ 0 & & & \ddots \\ & & 0 & 1 \end{pmatrix} \end{vmatrix} = \pm \prod_{i \in \mathbb{I}_1} \phi(x_i) \prod_{i \in \mathbb{I}_2} \phi(z_i) \ \det(C),$$

where C is a certain matrix, $\mathbb{I}_1 = \{i \ \mid \ x_i > 0\}$ and $\mathbb{I}_2 = \{i \ \mid \ z_i > 0\}$. Since

$$\pm \prod_{i \in \mathbb{I}_1} x_i \prod_{i \in \mathbb{I}_2} z_i \qquad \text{and} \qquad \prod_{i \in \mathbb{I}_1} \phi(x_i) \prod_{i \in \mathbb{I}_2} \phi(z_i),$$

are nonzeros, then we can conclude that J and T are invertibles and singulars at the same time. □

Theorem 4.5.13. *Suppose that* $\mathbf{X}^* = (x^*, z^*)$ *is a solution of NCP which satisfies the strict complementarity, and* $\nabla H_0(\mathbf{X}^*)$ *be defined by* (4.4.3) *(the Jacobian matrix of the interior-point method) is invertible. Then* $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)$ *is invertible, i.e., the two Jacobian matrices are singular or nonsigular at the same time.*

Proof. In view of Lemma 4.5.12, and thanks to the assumption $\mathbf{X}^* = (x^*, z^*)$ is a solution of NCP, we have $x^* \geqslant 0$ and $z^* \geqslant 0$, so that $x^- = z^- = 0$. Hence

$$\lim_{r\to 0} \det\left(\nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)\right) = \left| \begin{pmatrix} \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} \\ \phi(Z^*) & \phi(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{pmatrix} \right| = \varepsilon \left| \begin{pmatrix} \nabla F(x^*) & -I_{n\times n} \\ \phi(Z^*) & \phi(X^*) \end{pmatrix} \right|,$$

where $\phi(.)$ is defined in Lemma 4.5.6, $Z^* = \mathrm{diag}(z^*)$ and $X^* = \mathrm{diag}(x^*)$.
From Lemma 4.5.12, we conclude that if $\nabla_{\mathbf{X}} H_0(\mathbf{X}^*)$ is invertible then $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{H}_\theta(\mathbf{X}^*, r)$ is invertible. This means, that if the IPM converges our method converges. □

Hypothesis H_0 (F is a $\mathcal{P}_0$-function) assures us that our method is well defined and the Theorem 4.5.13 shows that the domain of convergence of our method is at least as large as that of the IPM.

4.6 Numerical experiments and applications

In this section, we present some numerical experiments for the two smoothing functions. Our aim is just to verify the theoretical assertions for these two "extreme" cases.
First, we study eight test problems with various sizes and characteristics. Then, we present a comparison on some randomly generated problems of our method and other approaches that have been suggested recently in [29, 48]. We also present numerical results for two concrete examples. All the codes are written in Matlab 2020R and run in the system of Windows 10 with PC i5 8-th Gen and 16.00 GB RAM. We take the precision $\varepsilon = 10^{-9}$ (the termination criterion).

Example 4.6.1. *We consider eight test problems (that can be found in [36, 55, 59, 67, 98, 110]) with various sizes and characteristics. In some cases, F is monotone or strongly monotone whereas others can have a non-connected solution set, in this case, F is at most a $\mathcal{P}_0$-function. A precise description of each test problem is given in the appendix.*

Table 4.1: Results for θ_1 and θ_2.

Pb	size	Iter (θ_1, θ_2)	Opt. (θ_1, θ_2)	Feas. (θ_1, θ_2)	cpu time(s) (θ_1, θ_2)	r (θ_1, θ_2)
P1	10	(114, 47)	(7.58e-11, 7.66e-13)	(0, 8.23e-11)	(0.04, 0.0073)	(2.96e-04, 0.0014)
	100	(134, 65)	(9.31e-12, 1.15e-15)	(0, 1.51e-12)	(0.14, 0.07)	(1.06e-04, 9.21e-04)
	500	(148, 66)	(1.87e-12, 6.39e-16)	(0, 4.28e-12)	(5.24, 2.52)	(4.90e-05, 8.95e-04)
	1000	(153, 68)	(1.05e-12, 7.82e-14)	(0, 1.04e-09)	(25.30, 11.41)	(3.70e-05, 0.0012)
P2	10	(116, 47)	(7.53e-11, 7.66e-13)	(0, 8.23e-11)	(0.01, 0.02)	(2.83e-04, 1.41e-03)
	100	(133, 74)	(7.64e-12, 2.64e-13)	(0, 3.47e-10)	(0.19, 0.10)	(8.74e-05, 0.0013)
	500	(147, 84)	(1.60e-12, 1.29e-12)	(0, 8.62e-08)	(5.25, 3.14)	(4.01e-05, 0.0014)
	1000	(153, 115)	(8.22e-13, 2.25e-15)	(0, 3.02e-10)	(24.47, 19.89)	(2.86e-05, 9.53e-04)
P3	10	(14, 16)	(3.46e-10, 1.65e-19)	(0, 3.99e-18)	(0.006, 0.02)	(0.001, 4.49e-04)
	100	(108, 44)	(4.99e-10, 3.63e-20)	(0, 4.75e-21)	(0.22, 0.05)	(0.0014, 0.0042)
	500	(353,140)	(7.58e-10, 3.97e-11)	(9.68e-13, 7.54e-11)	(35.03, 5.14)	(0.0011, 0.002)
	1000	(675, 265)	(8.94e-10, 2.32e-10)	(1.31e-12, 1.58e-11)	(91.41, 24.77)	(0.0011, 0.002)
P4	4	(53, 58)	(2.97e-10, 1.20e-10)	(1.24e-07, 4.19e-08)	(0.008, 0.0242)	(6.06e-04, 1.72e-04)
P5	4	(16, 14)	(3.02e-10, 1.92e-10)	(0, 3.84e-10)	(0.003, 0.009)	(0.0026, 0.018)
P6	7	(10, 13)	(1.06e-10, 7.16e-11)	(0, 0)	(0.1264, 0.0044)	(0.0016, 1.33e-04)
P7	5	(33, 30)	(2.23e-11, 1.16e-10)	(0, 3.44e-14)	(0.011, 0,016)	(0.004, 0.003)
P8	10	(65, 45)	(7.21e-11, 2.27e-11)	(1.34e-12, 5.18e-11)	(0.18, 0.16)	(0.0018, 3.76e-04)

In this table, **Size** *stands for the number of variables,* **Iter** *corresponds to the total number of Jacobian evaluations,* **Opt**. *and* **Feas**. *correspond to the following optimality and feasibility measures*

$$\textit{Opt.} := \max_{1\leqslant i\leqslant n} |x_i F_i(x)| \quad \textit{and} \quad \textit{Feas.} := \|\min(x,\ 0)\|_1 + \|\min(F(x),\ 0)\|_1.$$

The results clearly show that our methods are efficient. We also remark that the second smoothing function is much more efficient and powerful than the first one.
In the next table, we compare the results of θ_2-smoothing approach to three state-of-the-art methods (Namely: Fischer Burmeister (FB-Alg) [29], Newton Min (Min-Alg) and projection method (PM-Alg)[39]. We make a comparison among Algorithm 4.1, FB-Alg, Min-Alg, and PM-Alg by implementing these algorithms to solve the same benchmark test problems available in the literature. Since, we can not compare the iterative numbers, we only present the optimality measures, and cpu time(s).

Table 4.2: Comparison of Algorithm 4.1 (θ_2) with FB-Alg, Min-Alg and PM-Alg.

Pb	size	Algorithm 4.1 (θ_2)		FB-Alg		Min-Alg		PM-Alg	
		Opt.	time(s)	Opt.	time(s)	Opt.	time(s)	Opt.	time(s)
P1	10	7.66e-13	0.0073	4.23e-11	0.0273	3.11e-07	0.0092	2.5e-09	0.63
	100	1.15e-15	0.07	1.55e-10	0.1052	1.55e-14	0.0134	9.1e-11	5.48
	500	6.39e-16	2.52	4.47e-11	10.3936	1.55e-14	0.0282	4.7e-10	96.37
	1000	7.82e-14	11.41	4.47e-11	96.5009	1.55e-14	0.0557	8.8e-11	224.04
P2	10	7.66e-13	0.02	2.47e-11	0.0512	1.76e-12	0.0136	2.2e-10	1.19
	100	2.64e-13	0.10	1.25e-10	1.7285	5.06e-17	0.0193	7.1e-12	5.76
	500	1.29e-12	3.14	6.04e-11	5.6432	2.06e-12	0.6265	5.3e-12	112.41
	1000	2.25e-15	19.89	9.51e-12	25.1059	2.20e-13	3.2308	2.9e-11	336.20
P3	10	1.65e-19	0.02	6.58e-11	0.0272	4.24e-17	0.0078	6.4e-11	1.03
	100	3.63e-20	0.05	5.90e-11	0.0327	4.24e-17	0.0080	1.8e-12	5.19
	500	3.97e-11	5.14	1.30e-11	1.6040	4.24e-17	0.020	5.8e-13	90.22
	1000	2.32e-10	24.77	5.56e-11	11.0915	4.24e-17	0.0314	2.4e-11	350.06
P4	4	1.20e-10	0.0242	2.94e-14	0.0435	6.12e-10	0.0019	3.1e-12	0.19
P5	4	1.92e-10	0.009	2.37e-10	0.2752	2.22e-16	0.0120	1.4e-12	0.34
P6	7	7.16e-11	0.0044	5.70e-10	0.1158	6.05e-17	0.0341	2.3e-11	0.31

The results clearly show that our methods are efficient, competitive, and superior to the Fischer Burmeister method and projection method.

Example 4.6.2. *This example is described in [55, 98]. The corresponding function $F(x)$ is of the form:*

$$F(x) = (AA^T + B + D)x + q,$$

where the matrices A, B and D are randomly generated as: any entry of the square $n \times n$ matrix A and of the $n \times n$ skew-symmetric matrix B is uniformly generated from $]-5,\ 5[$, and any entry of the diagonal matrix D is uniformly generated from $]0,\ 3[$. The vector q is uniformly generated from $]-500,\ 0[$.

The matrix $AA^T + B + D$ is a positive definite and the function F is strongly monotone. We used the M-files proposed in [98] to generate A, B, D and q.

In this example, we will compare our methods already mentioned in sections 4.3 and 4.4, named: Algorithm 4.1 (θ_1), Algorithm 4.1 (θ_2) to some other methods (Newton min method (Min-Alg), Fischer-

Burmeister's method [29] (FB-Alg) and the classical interior-point method [48] (IPM-Alg)). In order to complete this experiment, we propose the performance profiles, developed by E. D. Dolan and J. J. Moré [35], as a tool for the comparative analysis of these methods.
We set "$n_s = 5$" as the number of methods and we have chosen "$n_p = 100$" (problems to be tested). We are interested in the comparison of the computation time and the number of iterations.

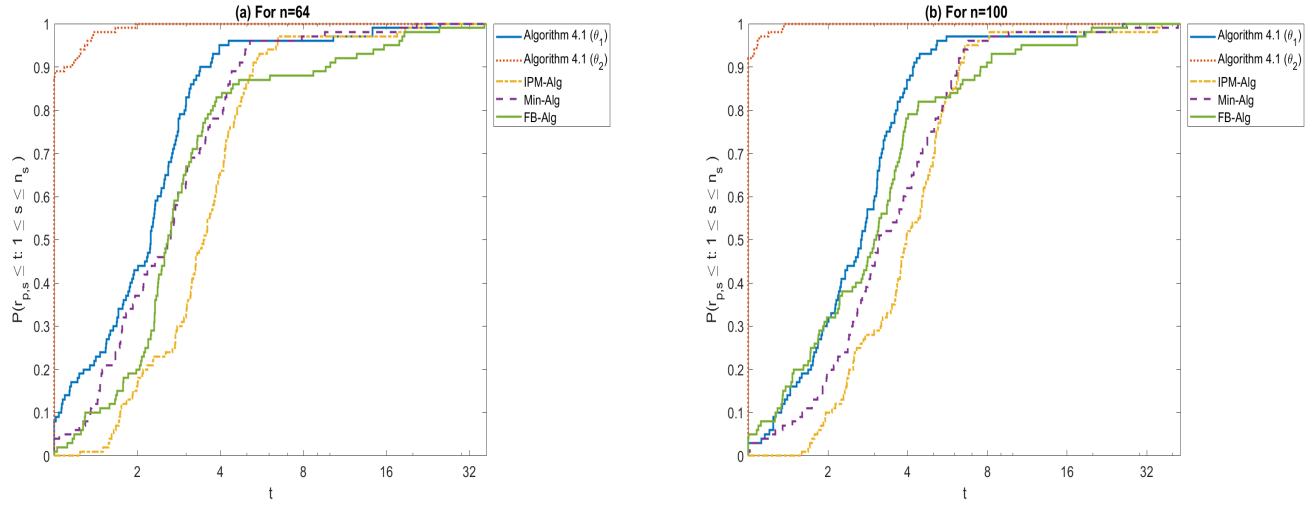

Figure 4.2: Performance profiles where $t_{p,\ s}$ represents the average computation time.

The figure above shows the performance profiles of five methods where the performance measure is execution time. It is clear that our method with the θ_2-function captures our attention (admits the highest probability value). In fact, in the interval [0, 1], our method is able to solve 99% of the problems, while the other methods do not reach 20% and require more time. We also notice that IPM-Alg is the slowest compared to others. However, for $t > 4$, the three algorithms FB-Alg, Min-Alg, and Algorithm 4.1 with θ_1-function confirm their robustness. Figure 3 also indicates that, with respect to the computation time, with the same initial points and under the same stopping criterion, our method with θ_2-function (resp. θ_1-function) is the fastest method, followed respectively by Min-Alg, FB-Algor, and IPM-Alg.

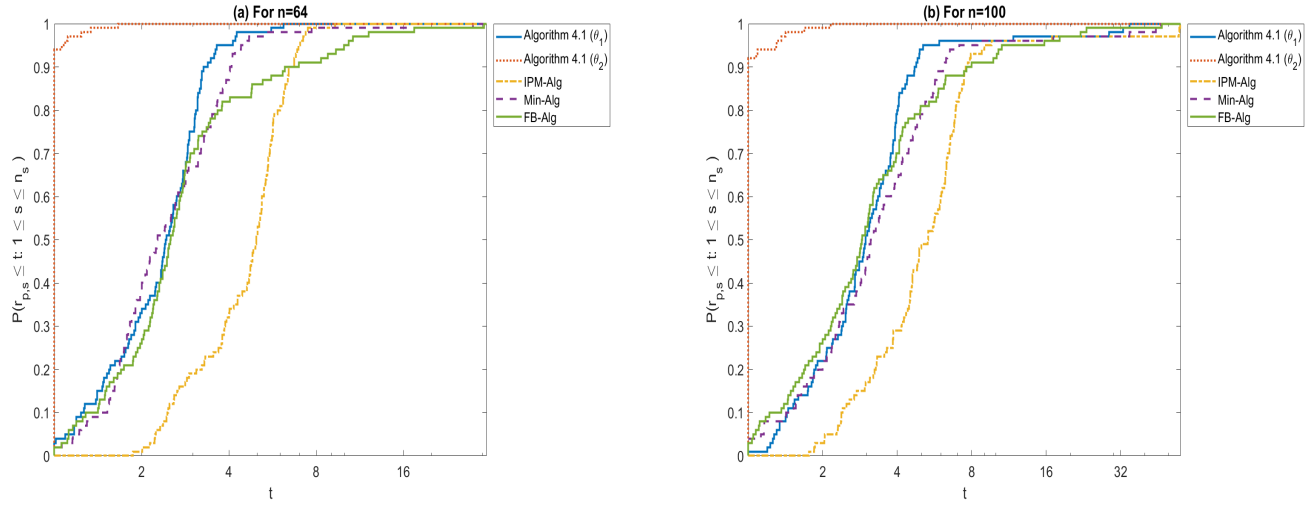

Figure 4.3: Performance profiles where $t_{p,\ s}$ represents the average number of iteratoins.

In Figure 4.3, we illustrate the performance profiles of five methods considering the number of iterations required as a performance measure. We notice that our method with the θ_2-function is the winner (admits the highest probability value) followed by our method with the θ_1-function, Min-Algo, and FB-Alg. We also note that IPM-Alg needs more iterations to resolve problems. The performance of our method with the θ_1-function becomes interesting beyond $t = 3$.

Example 4.6.3. *(Geochemical Models [80]) The problem comes from Geochemistry. We introduce a model which are 2-salts. The main idea is that we need to find a way to reformulate a general problem to a problem which has a form like $G(x) = 0$. We show the numerical results by applying several iteration methods.*
Let T, K are constant vectors which have meaning in chemistry. We define the problem as follows:
Let $x = (x_1, x_2, x_3)$ and $p = (p_1, p_2)$,

$$H : \mathbb{R}^5 \to \mathbb{R}^3 \qquad F : \mathbb{R}^3 \to \mathbb{R}^2 \qquad G : \mathbb{R}^5 \to \mathbb{R}^5$$

$$(x,\ p) \to H(x,p) = \begin{bmatrix} T_1 - x_1 - p_1 \\ T_2 - x_2 - p_2 \\ x_3 - x_2 - x_1 \end{bmatrix}, \quad x \to F(x) = \begin{bmatrix} K_1 - x_1 x_3 \\ K_2 - x_2 x_3 \end{bmatrix}, \quad (x,p) \to G(x,p) = \begin{bmatrix} H(x,p) \\ P^T F(x) \\ p \geqslant 0, F(x) \geqslant 0 \end{bmatrix}.$$

We want to solve the equation

$$G(x,p)=0 \iff \begin{bmatrix} H(x,p) \\ P^T F(x) \\ p \geqslant 0, F(x) \geqslant 0 \end{bmatrix} = 0. \tag{4.6.1}$$

Using our approach with $\theta_r = \theta_r^1$ (resp. $\theta_r = \theta_r^2$), we reformulate (4.6.1) and we get

$$G_{\theta_r^1}(x,p) = \begin{bmatrix} K_1 - x_1x_3 - z_1 \\ K_2 - x_2x_3 - z_2 \\ T_1 - x_1 - x_4 \\ T_2 - x_2 - x_5 \\ x_3 - x_2 - x_1 \\ \dfrac{x_4z_1 - r^2}{x_4 + z_1 + 2r} \\ \dfrac{x_5z_2 - r^2}{x_5 + z_2 + 2r} \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0, \quad G_{\theta_r^2}(x,p) = \begin{bmatrix} K_1 - x_1x_3 - z_1 \\ K_2 - x_2x_3 - z_2 \\ T_1 - x_1 - x_4 \\ T_2 - x_2 - x_5 \\ x_3 - x_2 - x_1 \\ -r\log(e^{-x_4/r} + e^{-z_1/r}) \\ -r\log(e^{-x_5/r} + e^{-z_2/r}) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0,$$

and considering $x_4 = p_1,\ x_5 = p_2$.

Since our focus is on the effect of different smoothing approaches in solving (4.6.1), we replace the complementarity constraint by the Min function, and by the Fischer-Burmeister's function [29]. The corresponding two algorithms are referred to as Min-Alg and FB-Alg, respectively. We make a comparison among Algorithm 4.1, Min-Alg, and FB-Alg, IPM-Alg (The classical interior-point method [48]) by implementing these algorithms to solve problem (4.6.1).

The Table 4.3, and Figure 4.4 show the results with the intial point $x_0 = (3,1,4,5,6)^T,\ T = (2,6)^T$, and $K = (37.5837, 7.6208)^T$.

Table 4.3: Comparison of Algorithm 4.1 (θ_1) and Algorithm 4.1 (θ_2) with Min-Alg, FB-Alg, and IPM-Alg.

Iter k	Algorithm 4.1 (θ_1) $\|G_{\theta_r^1}(x^k)\|$	Algorithm 4.1 (θ_2) $\|G_{\theta_r^2}(x^k)\|$	Min-Alg $\|G_{Min}(x^k)\|$	FB-Alg $\|G_{FB}(x^k)\|$	IPM-Alg $\|G_{IPM}(x^k)\|$
0	24.5837	24.5837	24.5837	24.5837	24.5837
1	8.2104	13.2212	13.3538	8.9798	8.3720
2	2.5518	12.0213	12.1412	4.2253	6.2486
3	4.1723	9.6610	11.0112	1.2774	1.8581
4	1.9497	6.1377	9.9651	0.4355	1.2035
5	0.0382	1.3388	9.0020	0.2001	0.8359
6	0.0063	0.3347	8.1192	0.1049	0.2406
7	0.0012	0.0838	7.1321	0.0537	0.0032
8	2.4936e-04	0.0210	6.0534	0.0272	5.1935e-05
9	5.0533e-05	0.0052	4.6609	0.0137	1.6166e-05
10	1.0487e-05	0.0013	2.9998	0.0068	4.9426e-06
11	2.1474e-06	3.2750e-04	1.2778	0.0034	1.5108e-06
12	4.4258e-07	8.1875e-05	0.0563	0.0017	4.6172e-07
13	9.1435e-08	2.0469e-05	1.2410e-05	8.5782e-04	1.4110e-07
14	1.8834e-08	5.1172e-06	4.4658e-12	4.2899e-04	4.3119e-08
15	4.5031e-09	1.2793e-06		2.1451e-04	1.3176e-08
16	9.6186e-10	3.1982e-07		1.0726e-04	4.0264e-09
17		7.9956e-08		5.3631e-05	1.2304e-09
18		1.9989e-08		2.6816e-05	3.7597e-10
19		4.9973e-09		1.3408e-05	
20		1.2493e-09		6.7041e-06	
21		3.1233e-10		3.3520e-06	
⋮				⋮	
34				4.0918e-10	

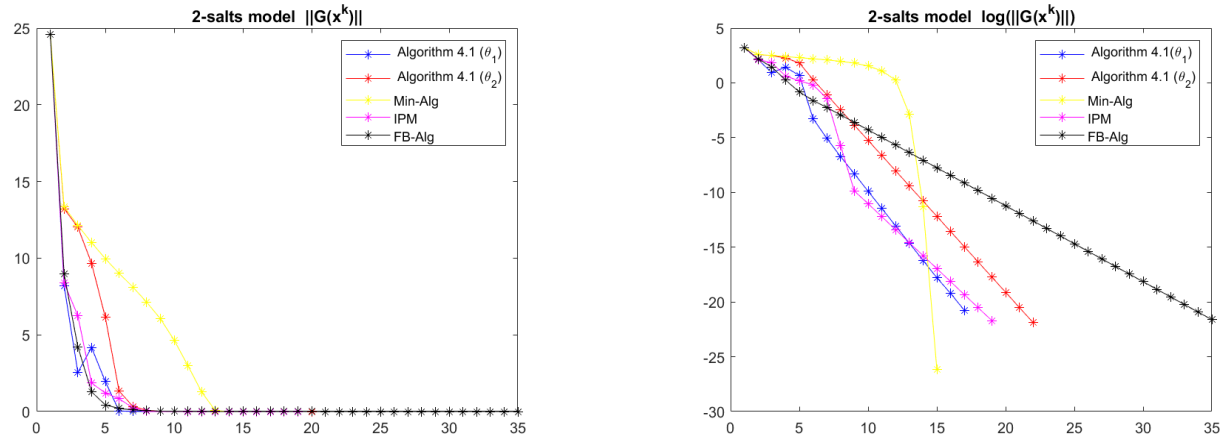

Figure 4.4: 2-salts model of $\|G(x^k)\|_\infty$ and $\log(\|G(x^k\|_\infty)$.

In Figure 4.4, all methods are quadratic convergence. The best method based on the number of iterations is semi-smooth Newton min method (14 iters). Next is our method for θ_1 function with (16 iters). Next is the classical interior-point method and our method for θ_2 function with (18 iters) and (21 iters), respectively. The last is the Fischer-Burmeister's method with (34 iters).

Example 4.6.4. *(An ordinary differential equation) We consider the ordinary differential equation*

$$\begin{cases} x''(t) - |x(t)| = -2 - t, \\ x(0) = -4, \quad x'(0) = 5, \\ t \in [0,\ 5]. \end{cases} \tag{4.6.2}$$

First, we discretize the EDO equation by using the finite difference scheme. We use the second-order centred finite difference to approximate the second order derivative

$$\frac{x_{i-2} - 2x_{i-1} + x_i}{h^2} - |x_i| = (-2 - t)_i. \tag{4.6.3}$$

Equation (4.6.3) was derived with equispace gridpoints $t_i = ih,\ i = 1, ..., N$. In order to approximate the Neumann boundary conditions we use a center difference

$$\frac{x_1 - x_{-1}}{2h} = x'(0) = 1. \tag{4.6.4}$$

Using the classical decomposition of the absolute value [2] we reformulate (4.6.3) as follows

$$\begin{cases} N_1 x^+ - N_2 x^- = q, \\ 0 \leqslant x^+ \perp x^- \geqslant 0, \end{cases} \tag{4.6.5}$$

where

$$N_1 = \frac{1}{h^2}\begin{pmatrix} 2-h^2 & & & & \\ -2 & 1-h^2 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1-h^2 \end{pmatrix}, \quad N_2 = \frac{1}{h^2}\begin{pmatrix} 2+h^2 & & & & \\ -2 & 1+h^2 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1+h^2 \end{pmatrix},$$

and $q = -\frac{1}{h^2}\begin{bmatrix} 8-10h \\ -4 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} 2+h \\ 2+2h \\ \vdots \\ 2+Nh \end{bmatrix}$.

N_1 is invertible, then the problem (4.6.5) is reduced to a standard LCP.
We compare the obtained solution by our methods to the predefined Runge-Kutta ode45 function in Matlab [78]. The domain is $t \in [0,5]$, initial conditions $x(0) = -4$, $x'(0) = 5$ and $N = 100$.

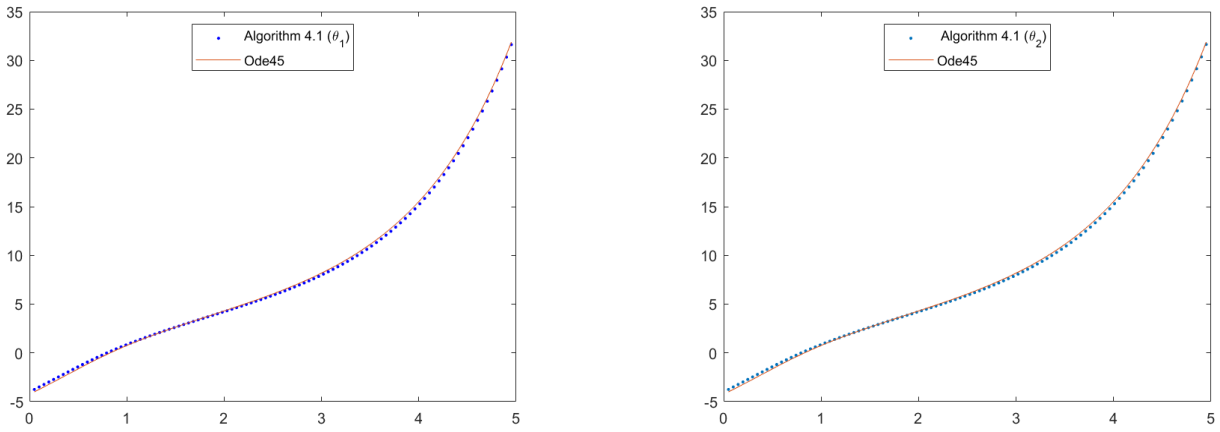

Figure 4.5: Numerical solution of (4.6.2) with ode45 and both methods.

Both methods solve the problem and gives consistent results.

4.7 Conclusion

In this chapter, we have presented a new smoothing approach for solving the nonlinear complementarity problem. For such an approach, some useful properties have been analyzed, which was employed to develop a well-defined and efficient Jacobian Newton algorithm for solving the nonlinear complementarity problem with $\mathcal{P}_0$-function. We have established the global convergence and the super-linear convergence for the developed algorithm. Numerical experiments prove the efficiency of our study in the following aspects:

1. It can find the solution of NCP either with less number of iteration, or with higher precision than the other.

2. It is relatively more robust for the increasing dimension of the test problem. In particular, it seems more suitable to solve large-scale problems.

3. It is more efficient to find the nondegenerate solution of NCP with less iteration number than the others.

4.8 Appendix

We give in this appendix a brief description of each test example.

1. The two first examples **P1** and **P2** [59] correspond to strongly monotone function

$$F(x) = (F_1(x),\ ...,\ F_n(x))^T \quad \text{with} \quad F_i(x) = -x_{i+1} + 2x_i - x_{i-1} + \frac{1}{3}x_i^3 - b_i, \quad i = 1,\ ...,\ n$$

   where $x_0 = x_{n+1} = 0$ and $b_i = (-1)^i$ (resp. $b_i = \frac{(-1)^i}{\sqrt{i}}$), $\quad i = 1,\ ...,\ n,$ for **P1** (resp. **P2**).

2. **P3** is another strongly monotone test problem from [36] where $F(x) = (F_1(x),\ ...,\ F_n(x))^T$ with

$$F_i(x) = -x_{i+1} + 2x_i - x_{i-1} + \arctan(x_i) + \left(i - \frac{\pi}{2}\right), \quad i = 1,\ ...,\ n, \quad (x_0 = x_{n+1} = 0).$$

3. **P4** and **P5** are known as the degenerate and non-degenerate examples of Kojima-Shindo [67]. **P4** and **P5** are respectively defined by

$$F_4(x) = \begin{bmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{bmatrix}, \qquad F_5(x) = \begin{bmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 1 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{bmatrix}.$$

P5 has a unique solution $x^* = \left(\frac{\sqrt{6}}{2},\ 0,\ 0,\ \frac{1}{2}\right)$ with $F(x^*) = \left(0,\ 2 + \frac{\sqrt{6}}{2},\ 5,\ 0\right)$ while **P4** has two optimal solutions $x^* = \left(\frac{\sqrt{6}}{2},\ 0,\ 0,\ \frac{1}{2}\right)$ with $F(x^*) = \left(0,\ 2 + \frac{\sqrt{6}}{2},\ 0,\ 0\right)$ and $x^{**} = (1,\ 0,\ 3,\ 0)$ with $F(x^*) = (0,\ 31,\ 0,\ 4)$.
The first optimal solution of **P4** is degenerate since $x_3^* = F_3(x^*) = 0$.

4. [110] In problem (4.2.1), $x \in \mathbb{R}^7$ and $F(x) : \mathbb{R}^7 \to \mathbb{R}^7$ is given by

$$F_6(x) = \begin{bmatrix} 2x_1 - x_3 + x_5 + 3x_6 - 1 \\ x_2 + 2x_5 + x_6 - x_7 - 3 \\ -x_1 + 2x_3 + x_4 + x_5 + 2x_6 - 4x_7 + 1 \\ x_3 + x_4 + x_5 - x_6 - 1 \\ -x_1 - 2x_2 - x_3 - x_4 + 5 \\ -3x_1 - x_2 - 2x_3 + x_4 + 4 \\ x_2 + 4x_3 - 1.5 \end{bmatrix}.$$

P6, has a non-degenerate solution

$$x^* = (0.2727,\ 2.0909,\ 0,\ 0.54545,\ 0.4545,\ 0,\ 0)^T.$$

5. A complete description of **P7** and **P8** can be found in [55, 98]. These two examples correspond to the Nash-Cournot test problem with $N = 5$ and $N = 10$.
Let $x \in \mathbb{R}^N$, $Q = \sum x_i$ and define the functions $C_i(x_i)$ and $p(Q)$ as follows:

$$P(Q) = 5000^{\frac{1}{\gamma}} Q^{\frac{-1}{\gamma}}, \quad C_i(x_i) = c_i x_i + \frac{b_i}{1 + b_i} L_i^{\frac{1}{b_i}} x_i^{\frac{b_i+1}{b_i}}.$$

The NCP-function is given by

$$F_i(x) = C_i'(x_i) - P(Q) - x_i p'(Q), \quad i = 1, ..., N,$$

with $c_i, L_i, b_i > 0$ and $\gamma \geqslant 1$. For our numerics, we used:

- $N = 5,\ c = [10, 8, 6, 4, 2]^T,\ b = [1.2, 1.1, 1, 0.9, 0.8]^T,\ L = [5, 5, 5, 5, 5]^T, e = [1, 1, 1, 1, 1]^T$ and $\gamma = 1.1$.
- $N = 10,\ c = [5, 3, 8, 5, 1, 3, 7, 4, 6, 3]^T,\ b = [1.2, 1, 0.9, 0.6, 1.5, 1, 0.7, 1.1, 0.95, 0.75]^T,$ $L = [10, 10, 10, 10, 10, 10, 10, 10, 10, 10]^T, e = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]^T$ and $\gamma = 1.2$.

Exact solution of 2-salts model

We compute the exact solution for the problem (4.8.1) in case where the NCP-function is the min-function. We want to compute exact solution of

$$G(x,p) = \begin{bmatrix} T_1 - x_1 - p_1 \\ T_2 - x_2 - P_2 \\ x_3 - x_2 - x_1 \\ \min(p_1,\ K_1 - x_1x_3) \\ \min(p_2,\ K_2 - x_2x_3) \end{bmatrix} = 0. \tag{4.8.1}$$

When having in hand the exact solution of (4.8.1), we choose the initial point in the code and also prove the existence and uniqueness of the solution of (4.8.1). Here we have some conditions as $p_i \geqslant 0,\ K_1 - x_1x_3 \geqslant 0,\ K_2 - x_2x_3 \geqslant 0$. Therefore we have four cases.

1. If $p_1 > 0\ p_2 > 0$.
   In this case, since $K_1,\ K_2 > 0$ then from (4.8.1) we have $x_1,\ x_2,\ x_3 > 0$ and $T_1 - x_1 = p_1 > 0 \Rightarrow 0 < x_1 < T_1,\ T_2 - x_2 = p_2 > 0 \Rightarrow 0 < x_2 < T_2$.

$$\begin{cases} K_1 = x_1x_3 \\ K_2 = x_2 + x_3 \\ x_3 = x_1 + x_2 \end{cases} \Leftrightarrow \begin{cases} K_1 = x_1(x_1 + x_2) \\ K_2 = x_2(x_1 + x_2) \\ x_3 = x_1 + x_2 \end{cases} \Leftrightarrow \begin{cases} x_1 = \dfrac{K_1}{\sqrt{K_1 + K_2}} \\ x_2 = \dfrac{K_2}{\sqrt{K_1 + K_2}} \\ x_3 = \sqrt{K_1 + K_2} \end{cases}$$

   Then $p_1 = T_1 - x_1 = T_1 - \dfrac{K_1}{\sqrt{K_1 + K_2}}$ and we need $T_1 > \dfrac{K_1}{\sqrt{K_1 + K_2}}$ since $p_1 > 0$.
   The same for $p_2 = T_2 - x_2 = T_2 - \dfrac{K_2}{\sqrt{K_1 + K_2}}$ and the conditon $T_2 > \dfrac{K_2}{\sqrt{K_1 + K_2}}$. We get the exact solution of (4.8.1) in this case

$$(x,p) = \left(\frac{K_1}{\sqrt{K_1 + K_2}},\ \frac{K_2}{\sqrt{K_1 + K_2}},\ \sqrt{K_1 + K_2},\ T_1 - \frac{K_1}{\sqrt{K_1 + K_2}},\ T_2 - \frac{K_2}{\sqrt{K_1 + K_2}}\right)^T, \tag{4.8.2}$$

where $T_1 > \dfrac{K_1}{\sqrt{K_1+K_2}}$ and $T_2 > \dfrac{K_2}{\sqrt{K_1+K_2}}$.

2. If $p_1 = 0$ $p_2 > 0$.
   Since $p_1 = 0$ then $x_1 = T_1,\ K_1 \geqslant x_1 x_3$ and $x_3 = x_2 + x_1 = x_2 + T_1$. Since $p_2 > 0$ then $K_2 = x_2 x_3 = x_2(x_2 + T_1) \Rightarrow T_1 = \dfrac{K_2 - x_2^2}{x_2}$ and we get a equation

$$x_2 = \frac{-T_1 + \sqrt{T_1^2 + 4K_2}}{2}. \tag{4.8.3}$$

   Since $T_1,\ T_2 \geqslant 0,\ K_2 > 0,$ then $x_2 > 0$ and $x_3 = T_2 + x_2 > 0$. We also need $T_2 - x_2 > 0$ or $T_2 > x_2$. Then we have $T_2^2 + T_1 T_2 - K_2 > 0$. The last condition to check is that $K_1 \geqslant x_1 x_3 = x_1(x_1 + x_2) = T_1(T_1 + x_2) = \dfrac{K_2 - x_2^2}{x_2}(\dfrac{K_2 - x_2^2}{x_2} + x_2)$. Then $x_2 \geqslant \dfrac{K_2}{\sqrt{K_1+K_2}}$.
   This implies $T_2 > x_2 \geqslant \dfrac{K_2}{\sqrt{K_1+K_2}}$. Then we get the solution $x = [T_1,\ x_2,\ T_1 + x_2,\ 0, T_2 - x_2]^T$ where x_2 is in (4.8.3), $T_2^2 + T_1 T_2 - K_2 > 0$ and $T_2 > \dfrac{K_2}{\sqrt{K_1+K_2}}$.

3. If $p_1 > 0$ $p_2 = 0$.
   Since $p_2 = 0$ then $x_2 = T_2,\ K_2 \geqslant x_2 x_3$ and $x_3 = x_2 + x_1 = T_2 + x_1$. Since $p_1 > 0$ then $K_1 = x_1 x_3 = x_1(x_1 + T_2) \Rightarrow T_2 = \dfrac{K_1 - x_1^2}{x_1}$ and we get a equation

$$x_1^2 + T_2 x_1 - K_1 = 0 \Leftrightarrow x_1 = \frac{-T_2 \pm \sqrt{T_2^2 + 4K_1}}{2}.$$

   Since we want the solution x to be nonnegative, we choose

$$x_1 = \frac{-T_2 + \sqrt{T_2^2 + 4K_1}}{2} > 0. \tag{4.8.4}$$

   If $T_1,\ T_2,\ K_1 > 0$ is then $x_1 > 0$ and $x_3 = T_2 + x_1 > 0$. We also need $T_1 - x_1 > 0$ or $T_1 - \dfrac{-T_2 + \sqrt{T_2^2 + 4K_1}}{2} > 0 \Leftrightarrow T_1^2 + T_1 T_2 - K_1 > 0$ to ensure that $p_1 = T_1 - x_1$ is nonnegative.
   The last one is to check $K_2 \geqslant x_2 x_3 = T_2(T_2 + x_1) = \dfrac{K_1 - x_1^2}{x_1}(\dfrac{K_1 - x_1^2}{x_1} + x_1) \Leftrightarrow x_1 \geqslant \dfrac{K_1}{\sqrt{K_1+K_2}}$ then we get $T_1 > \dfrac{K_1}{\sqrt{K_1+K_2}}$. Then we get the solution $x = [x_1, T_2, T_2 + x_1, T_2 - x_1, 0]^T$ where x_1 is in (4.8.4) and $T_1^2 + T_1 T_2 - K_1 > 0$ and $T_1 > \dfrac{K_1}{\sqrt{K_1+K_2}}$.

4. If $p_1 = 0$ $p_2 = 0$.

We get

$$\begin{cases} x_1 &= & T_1 \\ x_2 &= & T_2 \\ x_3 &= & x_1 + x_2 = T_1 + T_2. \end{cases}$$

We need some conditions that $K_1 - x_1 x_3 \geqslant p_1 = 0 \Rightarrow K_1 \geqslant T_1(T_1 + T_2)$ and the same for $K_2 \geqslant T_2(T_1 + T_2) \geqslant 0$. If $T_1,\ T_2$ is nonnegative then the exact solution of (4.8.1) is

$$x = [T_1,\ T_2,\ T_1 + T_2,\ 0,\ 0]^T,$$

where $K_1 \geqslant T_1(T_1 + T_2) \geqslant 0$ and $K_2 \geqslant T_2(T_1 + T_2) \geqslant 0$.

5 New smoothing methods for solving the linear complementarity problems involving $\mathcal{P}_0$-matrix

This chapter is a paper submitted to RAIRO entitled: New smoothing methods for solving the linear complementarity problems with $\mathcal{P}_0$-matrix. We have chosen to present this chapter this way for a better correspondence with our paper [85].

Based on smoothing techniques, we propose two new methods to solve linear complementarity problems (LCPs) called TLCP and Soft-LCP. The idea of these two new methods takes inspiration from interior-point methods in optimization. The technique that we propose avoids any parameter management while ensuring good theoretical convergence results. In our approach we do not need any complicated strategy to update the smoothing parameter r since we will consider it as a new variable. Our methods are validated by extensive numerical tests, in which we compare our methods to several other classical methods.

Contents

5.1 Introduction

The linear complementarity problem consists in finding a vector in a finite-dimensional real vector space that satisfies a certain system of inequalities. Specifically, given a vector $q \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{n\times n}$, the linear complementarity problem, abbreviated LCP, is to find a vector $x \in \mathbb{R}^n$ such that

$$0 \leqslant x \perp (Mx + q) \geqslant 0. \tag{5.1.1}$$

This problem is known to have a unique solution for any $q \in \mathbb{R}^n$ if and only if M is a $\mathcal{P}$-matrix [34, 100]. The linear complementarity problem has many important applications in engineering and equilibrium modeling [42, 89], and many numerical methods are developed to solve LCPs [18, 28]. Although the effectiveness of complementarity algorithms has improved substantially in recent years, the fact remains that increasingly more difficult problems are being proposed that are exceeding the capabilities of these algorithms. As a result, there is a real need to propose new methods and algorithms to address complicated and difficult situations. To solve LCP, there are essentially three different classes of methods: equation-based methods (smoothing), merit functions, and projection-type methods. Our goal in this chapter is to present new and very simple smoothing and approximation schemes to solve LCP and to produce efficient numerical methods.
Many algorithms have been proposed to solve problem LCP [34, 83]. They may be based on pivoting techniques [33, 71], which often suffer from the combinatorial aspect of the problem, on interior point methods, which originate from an algorithm introduced by Karmarkar in linear optimization [63], see also [66] for one of the first accounts on the use of interior-point methods to solve LCP. Some researchers try to solve LCP by reformulating them as an unconstrained optimization [47], and on nonsmooth Newton approaches [39], and rewrite the complementarity conditions as a system of smooth equations [73], such as the one considered here. See [34, 83] for other iterative methods.
In this work, we propose two new algorithms called TLCP and Soft-LCP for solving the LCP. The principle of these algorithms is as follows: first, we proposed two smoothing techniques to regularize the complementary condition, we replace

$$0 \leqslant x \perp z \geqslant 0,$$

by

$$\theta_r(x) + \theta_r(z) = \mathbf{e}, \quad r \searrow 0,$$

and

$$\forall \rho > 0 \qquad x = r \log \left(\mathbf{e} + e^{\frac{x - \rho z}{r}} \right), \quad r \searrow 0,$$

where θ_r, log, e and max operates component-wise on x and z, and $\mathbf{e} \in \mathbb{R}^n$ is the vector whose entries are all equal to 1, then we give a strategy that decreases r during iterations and ensures the nonnegatives of variables. The main difference in our approach is that we do not need any complicated strategy to update the parameter r since we will consider it as a new variable. Finally, the two new algorithms are solved using the standard Newton method. To enforce a global convergence behavior, we also recommend using Armijo's line search.

This chapter is structured as follows. In section 5.2 of this chapter we gives some definitions and properties of the smoothing functions. In section 5.3, we present our two approximation for the problem LCP and give the new formulation of the problem LCP. In section 5.4, we propose two new methods to solve the LCP. In section 5.5, we propose two generic algorithms to solve LCP and prove some convergence results. In section 5.6, we provide some numerical results where we present a comparison on some randomly generated problems of our two methods with other approaches that have been suggested recently in [29, 48] and we study two concrete examples, the first one is a second-order ordinary differential equation and the second is an obstacle problem also, we tested our algorithms on several absolute value equations problems. Finally, we conclude our chapter.

5.2 Preliminaries and problem setting

In this section, we present some necessary definitions and lemmas. A matrix $M \in \mathbb{R}^{n \times n}$ is said to be positive definite if $\langle x, Mx \rangle > 0$ for all nonzero $x \in \mathbb{R}^n$. $M \in \mathbb{R}^{n \times n}$ is called a $\mathcal{P}$-matrix if all its minors are positive. As a consequence, if M is positive definite, then M is a $\mathcal{P}$-matrix. A matrix $M \in \mathbb{R}^{n \times n}$ is a $\mathcal{P}_0$-matrix if every of its principal minors is nonnegative.

First, we state a result for the unique solution of an LCP, the following result was proved by Cottle, Pang, and Stone [34]. Next, we give the definition of θ-smoothing function and Soft-Max function that will use to approximate the complementarity condition.

Theorem 5.2.1. *(Theorem 3.3.7, [34]). A matrix $M \in \mathbb{R}^{n \times n}$ is a $\mathcal{P}$-matrix if and only if the LCP (5.1.1) has a unique solution for every $q \in \mathbb{R}^n$.*

5.2.1 Definition of θ-smoothing function

We introduce the function θ with the following properties (these functions were used in [53, 52]).
Let $\theta : \mathbb{R} \to]-\infty, 1[$, be a non-decreasing continuous smooth concave function such that

$$\theta(t) < 0 \text{ if } t < 0,\ \theta(0) = 0 \text{ and } \lim_{t\to+\infty} \theta(t) = 1.$$

One possible way to build such function is to consider non-increasing probability density functions $f : \mathbb{R}_+ \to \mathbb{R}_+$ and then take the corresponding cumulative distribution function

$$\theta(t) = \int_0^t f(x)dx.$$

By definition of f we can verify that

$$\lim_{t\to+\infty} \theta(t) = \int_0^{+\infty} f(x)dx = 1,$$

and

$$\theta(0) = \int_0^0 f(x)dx = 0.$$

The non-decreasing hypothesis gives the concavity of θ. We then extend this functions for negative values in a smooth way.
Example of this family are $\theta^1(t) = t/(t+1)$ if $t \geqslant 0$ and $\theta^1(t) = t$ if $t < 0$.
We introduce $\theta_r(t) := \theta(\frac{t}{r})$ for $r > 0$. This definition is similar to the perspective functions in convex analysis. This functions satisfy

$$\theta_r(0) = 0 \ \ \forall r > 0 \text{ and } \lim_{r\searrow 0} \theta(t) = 1 \ \ \forall t > 0.$$

There are some examples of such functions

$$\theta_r^1(t) = \frac{t}{t+r} \text{ if } t \geqslant 0 \quad \text{and} \quad \theta_r^1(t) = t/r \text{ if } t < 0,$$

$$\theta_r^2(t) = 1 - e^{-t/r},\ t \in \mathbb{R}.$$

The function θ_r^1 will be extensively used in this chapter.

5.2.1.1 θ-smoothing of a complementarity condition

Let $(x, z) \in \mathbb{R}^2$ be two scalars such that

$$0 \leqslant x \perp z \geqslant 0, \tag{5.2.1}$$

that is,

$$x \geqslant 0, \qquad z \geqslant 0, \qquad xz = 0.$$

In the (x, z)-plane, the set of points obeying (5.2.1) is the union of the two semi-axes $\{x \geqslant 0,\ z = 0\}$ and $\{x = 0,\ z \geqslant 0\}$. Visually, the nonsmoothness of (5.2.1) is manifested by the "kink" at the corner $(x, z) = (0, 0)$.
We consider two possible smooth approximations of (5.2.1), depending how it is rewritten in terms of θ-function.

Lemma 5.2.2. *[52] Given $x, z \in \mathbb{R}_+$ and the parameter $r > 0$, we have the equivalence*

$$xz = 0 \iff \lim_{r \searrow 0} (\theta_r(x) + \theta_r(z)) \leqslant 1.$$

Lemma 5.2.3. *[52] θ_r is sub-additive for non-negative values, i.e. given $x,\ z \geqslant 0$ it holds that*

$$\theta_r(x) + \theta_r(z) \ \geqslant \ \theta_r(x + z).$$

and with equality if and only if $x = 0$ or $z = 0$,

$$xz = 0 \iff \theta_r(x) + \theta_r(z) = \theta_r(x + z).$$

Our objective is to approximate the complementarity constraints by using these theta functions then we will present the max function which will be the basic idea of our second approximation.

5.2.2 Soft-Max Function

Let f be a function defined as:

$$f(x_1, ..., x_n) = \max(x_1, ..., x_n),$$

obviously, the max function is non-differentiable. We approximate the max function by a smooth function, noted Soft-Max function as introduced in [30]:

$$\forall r > 0, \qquad f_r(x_1, ..., x_n) = r \log \left(\sum_{i=1}^{n} e^{\ x_i/r} \right).$$

Indeed: $\forall r > 0$ and $\forall x \in \mathbb{R}^n$,

$$r \log \left(\sum_{i=1}^{n} e^{\ x_i/r} \right) \leqslant r \log \left(n \max_i e^{\ x_i/r} \right) = \max_i x_i + r \log n,$$

and

$$\max_i x_i \leqslant r \log \left(\sum_{i=1}^{n} e^{\, x_i/r} \right) \leqslant r \log \left(\sum_{i=1}^{n} e^{\, x_i/r} \right) + r \log n.$$

Then

$$|\max_i x_i - f_r(x)| \leqslant r \log n, \quad \forall\ i = 1, ..., n.$$

Thus f_r is a uniformly smoothing approximation function of f. Notice that the accuracy of the Soft-Max approximation depends on scale r. Figure 5.1 illustrate the behaviour of Soft-Max function.

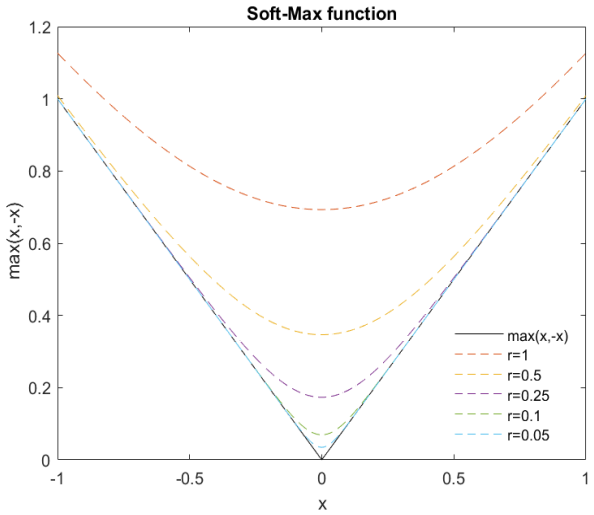

Figure 5.1: Smoothing by Soft-Max function.

5.3 An approximate formulation

In this section, we present our two formulations for LCP (5.1.1), the first with the θ-function and the second with the Soft-Max function.

Consider the linear complementarity problem, which is to find a solution of the system $F(\mathbf{X}) = 0$, with

$$F(\mathbf{X}) = \begin{bmatrix} Mx + q - z \\ x.z \end{bmatrix}, \tag{5.3.1}$$

where $\mathbf{X} = (x,\ z) \in \mathbb{R}^{2n}_{+}$. Recall that the Hadamard product $x.z$ of two vectors x and z is the vector having its ith component equal to $x_i z_i$.

5.3.1 Approximation of LCP using θ-function

We reformulate the problem LCP using θ_r-function, we regularize each complementarity constraint by considering

$$x_i z_i = 0,\ x_i \geqslant 0,\ z_i \geqslant 0 \quad \text{by} \quad \theta_r(x_i) + \theta_r(z_i) = 1,\ x_i \geqslant 0,\ z_i \geqslant 0 \qquad \forall\ i = 1, ..., n.$$

In fact $x_i z_i = 0$ should be approximated by

$$\theta_r(x_i) + \theta_r(z_i) \leqslant 1, \qquad \text{(both can be zeros)}$$

but we use an implicit assumption of strict complementarity. Using this approximation, we obtain the following formulation:

$$(P_\theta^r) \qquad \begin{cases} Mx + q = z, \\ x \geqslant 0, \quad z \geqslant 0, \\ (\theta_r(x) + \theta_r(z) - \mathbf{e}) = 0. \end{cases} \tag{5.3.2}$$

We consider the family $\{F_\theta^r(.),\ r > 0\}$, where

$$F_\theta^r(\mathbf{X}) = \begin{bmatrix} Mx + q - z \\ r\,(\theta_r(x) + \theta_r(z) - \mathbf{e}) \end{bmatrix} \in \mathbb{R}^{2n}, \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{2n}_+, \tag{5.3.3}$$

is a regularized function of F defined in (5.3.1). Here, it is understood that θ_r operates componentwise on x and z, while $\mathbf{e} \in \mathbb{R}^n$ is the vector whose entries are all equal to 1. It is highly recommended that the smoothed complementarity equations in (5.3.3) be premultiplied by r, so as to control the magnitude of their partial derivatives.
Indeed, for all $t \geqslant 0$,

$$\theta_r'(t) = \frac{1}{r}\theta'\left(\frac{t}{r}\right),$$

can be seen to blow up when $r \searrow 0$, while $r\theta_r'(t)$ tends to the finite limit $\theta'(0)$.

5.3.2 Approximation of LCP using Soft-Max

It is obvious that the vectors x and z satisfy complementarity condition if and only if

$$\forall \rho > 0, \quad \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \max(0, x_1 - \rho z_1) \\ \vdots \\ \max(0, x_n - \rho z_n) \end{pmatrix}.$$

Using the Soft-Max function defined below, we approximate

$$\max(0, x_i - \rho z_i) \quad \text{by} \quad r \log\left(1 + e^{\frac{x_i - \rho z_i}{r}}\right), \quad \forall i = 1, ..., n, \tag{5.3.4}$$

we obtain

$$(P_s^r) \qquad \begin{cases} Mx + q = z, \quad \forall \rho > 0, \\ x - r \log\left(\mathbf{e} + e^{\frac{x - \rho z}{r}}\right) = 0. \end{cases} \tag{5.3.5}$$

We consider the family $\{F_s^r(.),\ r > 0\}$, where

$$F_s^r(\mathbf{X}) = \begin{bmatrix} Mx + q - z \\ \left(x - r \log\left(\mathbf{e} + e^{\frac{x - \rho z}{r}}\right)\right) \end{bmatrix} \in \mathbb{R}^{2n}, \text{ and } \mathbf{X} = \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{2n}_+, \tag{5.3.6}$$

is a regularized function of F defined in (5.3.1). By the same way as for (5.3.3), $\log(.)$, $e^{(.)}$ operates componentwise on x and z.

Lemma 5.3.1. *Let $F_s^r(\mathbf{X})$ be defined by (5.3.6). Then, for any $(x, z) \in \mathbb{R}^{2n}_+$ the Jacobian matrix of $F_s^r(\mathbf{X})$ is*

$$\nabla F_s^r(\mathbf{X}) = \begin{pmatrix} M & -I \\ D_a(\mathbf{X}) & D_b(\mathbf{X}) \end{pmatrix},$$

where $D_a(\mathbf{X}) = diag\{a_1(\mathbf{X}), ..., a_n(\mathbf{X})\}$ and $D_b(\mathbf{X}) = diag\{b_1(\mathbf{X}), ..., b_n(\mathbf{X})\}$ are two diagonal matrices, and

$$a_i(\mathbf{X}) = \frac{1}{1 + e^{\frac{x_i - \rho z_i}{r}}}, \qquad b_i(\mathbf{X}) = \frac{\rho e^{\frac{x_i - \rho z_i}{r}}}{1 + e^{\frac{x_i - \rho z_i}{r}}}, \qquad i = 1, ...n.$$

Let $F_\theta^r(\mathbf{X})$ be defined by (5.3.3). We restrict our choice of θ-fucntion to $\theta_r(t) = \theta_r^1(t)$. Then, the Jacobian matrix of $F_\theta^r(\mathbf{X})$ is

$$\nabla F_s^r(\mathbf{X}) = \begin{pmatrix} M & -I \\ Q_k(\mathbf{X}) & Q_l(\mathbf{X}) \end{pmatrix},$$

where $Q_k(\mathbf{X}) = diag\{k_1(\mathbf{X}), ..., k_n(\mathbf{X})\}$ and $Q_l(\mathbf{X}) = diag\{l_1(\mathbf{X}), ..., l_n(\mathbf{X})\}$ are two diagonal matrices,

and

$$k_i(\mathbf{X}) = \frac{r^2}{(x_i + r)^2}, \qquad l_i(\mathbf{X}) = \frac{r^2}{(z_i + r)^2}, \qquad i = 1, ..., n.$$

Lemma 5.3.2. *Let $M \in \mathbb{R}^{n\times n}$ be a $\mathcal{P}_0-$matrix. Then any matrix in the following form is nonsingular:*

$$N_s + N_t M,$$

where $N_s \in \mathbb{R}^{n\times n}$ is a positive (negative) diagonal matrix, and $N_t \in \mathbb{R}^{n\times n}$ is a nonnegative (non-positive) diagonal matrix.

Proof. Let $N_s = \text{diag}(s_1, s_2, ..., s_n)$ and $N_t = \text{diag}(t_1, t_2, ..., t_n)$. If N_s is positive, and N_t is nonnegative, then $s_i > 0$ and $t_i \geqslant 0$ for all $i = 1, 2, ..., n$.
Let $v \in \mathbb{R}^n$ be a vector such that $(N_s + N_t M)v = 0$. Then, we have $v_i = -\frac{t_i}{s_i}(Mv)_i \quad \forall i = 1, ..., n$.
It yields $v_i^2 = -\frac{t_i}{s_i} v_i (Mv)_i$. If $t_i = 0$, then $v_i = 0$.
If $v_i \neq 0$, we have $\frac{t_i}{s_i} > 0$. Owing to $v_i^2 \geqslant 0$, we have $v_i(Mv)_i \leqslant 0$. If $v_i(Mv)_i = 0$, then $v_i = 0$. Otherwise, $v_i(Mv)_i < 0$ contradicts the property of M. Based on the above discussion, it is concluded that $v = 0$, then $N_s + N_t M$ is a nonsingular matrix. □

By Lemma 5.3.2, we can obtain a property of F_s^r and F_θ^r if M is a $\mathcal{P}_0$-matrix.

Theorem 5.3.3. *Let M be a $\mathcal{P}_0$-matrix. Then, for any $r > 0$, and any $(x, z) \in \mathbb{R}^{2n}_+$ the Jacobian matrix $\nabla F_s^r(\mathbf{X})$ (resp. $\nabla F_\theta^r(\mathbf{X})$) is nonsingular.*

Proof. For all $r > 0$ and from Lemma 5.3.1, it follows that the diagonal matrix $D_a(\mathbf{X})$ (resp. $Q_k(\mathbf{X})$) is non-negative, and $D_b(\mathbf{X})$ (resp. $Q_l(\mathbf{X})$) is non-negative diagonal matrix.
We have $\det(\nabla F_s^r(\mathbf{X})) = \det(D_a(\mathbf{X}) + M D_b(\mathbf{X}))$ (resp. $\det(\nabla F_\theta^r(\mathbf{X})) = \det(Q_k(\mathbf{X}) + M Q_l(\mathbf{X}))$), since M is a $\mathcal{P}_0$-matrix and from Lemma 5.3.2, it follows that $D_a(\mathbf{X}) + M D_b(\mathbf{X})$ (resp. $Q_k(\mathbf{X}) + M Q_l(\mathbf{X})$) is nonsingular. Hence $\nabla F_s^r(\mathbf{X})$ (resp. $\nabla F_\theta^r(\mathbf{X})$) is nonsingular. □

5.4 Solving LCP via new algorithm

In this section, we present the idea of our algorithms for optimization problems to solve the LCP, but here we don't have any objective function to minimize. Our methods take inspiration from IPMs.
We recall that the IPMs have replaced the original nonsmooth problem LCP by a sequence of regularized problems

$$F_r(\mathbf{X}) = 0, \tag{5.4.1}$$

where

$$\mathbf{X} = \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{2n}_+, \quad F_r(\mathbf{X}) = \begin{bmatrix} Mx + q - z \\ x.z - r\mathbf{e} \end{bmatrix}, \tag{5.4.2}$$

and $r \geqslant 0$ is the smoothing parameter. The Jacobian matrix of F_r with respect to $\mathbf{X}$, does not depend on r and can be denoted by

$$\nabla_{\mathbf{X}} F_r(\mathbf{X}) = \begin{pmatrix} M & -I \\ Z & X \end{pmatrix}, \tag{5.4.3}$$

where $Z = \text{diag}(z)$ and $X = \text{diag}(x)$, i.e., the diagonal matrix of z (resp. x).

5.4.1 When the parameter becomes a variable

In the system (5.4.1), the status of the parameter r is very distinct from that of the variable $\mathbf{X}$. While $\mathbf{X}$ is computed "automatically" by a Newton iteration, r has to be updated "manually" in an ad-hoc manner.

Our goal is to find a strategy that decreases r during iterations and ensures the nonnegative of variables. However, we must adjust the strategy when the model or its parameters are changed. To avoid this trouble, we consider r as an unknown of the system instead of a parameter as in [104].

We feel that it would be judicious to incorporate the parameter r into the variables. Let us therefore consider the enlarged vector of unknowns

$$\mathbb{X} = \begin{bmatrix} \mathbf{X} \\ r \end{bmatrix} \in \mathbb{R}^{2n}_+ \times \mathbb{R}_+, \tag{5.4.4}$$

and then consider a system of $2n + 1$ equations

$$\mathbb{F}_\theta(\mathbb{X}) = 0, \quad (\text{resp. } \mathbb{F}_s(\mathbb{X}) = 0), \tag{5.4.5}$$

to be on $\mathbb{X}$. To this end, let us remind ourselves that our ultimate goal is to solve $F^0_\theta(\mathbf{X}) = 0$ (resp. $F^0_s(\mathbf{X}) = 0$), together with the inequalities $x \geqslant 0,\ z \geqslant 0$. Thus, it is really natural to first consider

$$\mathbb{F}_\theta(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ r\left(\theta^1_r(x) + \theta^1_r(z) - \mathbf{e}\right) \\ r \end{bmatrix}, \tag{5.4.6}$$

and

$$\mathbb{F}_s(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ x - r \log\left(\mathbf{e} + e^{\frac{x - \rho z}{r}}\right) \\ r \end{bmatrix}. \tag{5.4.7}$$

This construction turns out to be to naive. Indeed, if we start from some r^0 and solve the smooth system (5.4.6) and (5.4.7) by the smooth Newton method, since the last equation is linear, we end up with $r^1 = 0$ at the first iteration. Once the boundary of the interior region is reached, we are "stuck" there.

To prevent r from rushing to zero in just one iteration, we could set

$$\mathbb{F}_\theta(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ r\left(\theta_r^1(x) + \theta_r^1(z) - \mathbf{e}\right) \\ r^2 \end{bmatrix}, \tag{5.4.8}$$

and

$$\mathbb{F}_s(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ x - r \log\left(\mathbf{e} + e^{\frac{x - \rho z}{r}}\right) \\ r^2 \end{bmatrix}. \tag{5.4.9}$$

At this stage, system (5.4.8) (resp. (5.4.9)) is not yet fully adequate. Indeed, the last equation is totally decoupled from the others. Everything happens as if r follows a prefixed sequence, generated by the Newton iterates of the scalar equation $r^2 = 0$, regardless of $\mathbf{X}$. It is desirable to couple r and $\mathbf{X}$ in a tighter way. In this respect, we advocate

$$\mathbb{F}_\theta(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ r\left(\theta_r^1(x) + \theta_r^1(z) - \mathbf{e}\right) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 \end{bmatrix}, \tag{5.4.10}$$

and

$$\mathbb{F}_s(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ x - r \log\left(\mathbf{e} + e^{\frac{x - \rho z}{r}}\right) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 \end{bmatrix}, \tag{5.4.11}$$

where

$$\|x^-\|^2 = \sum_{i=1}^{n} \min^2(x_i, 0), \quad \|z^-\|^2 = \sum_{i=1}^{n} \min^2(z_i, 0).$$

This choice has the benefit of taking into account the nonnegativity condition $x \geqslant 0$ and $z \geqslant 0$. Indeed, the last equation of (5.4.10) and (5.4.11) implies that, as long as $r \geqslant 0$, we are ascertained that $x^- = z^- = 0$. This amounts to saying that $x \geqslant 0$ and $z \geqslant 0$. Should a component of x or z become negative during the iteration, this equation would contribute to "penalize" it.
Since r is now considered as a variable and the scalar function $t \mapsto \frac{1}{2}|\min(t, 0)|^2$ is differentiable and its derivative is equal to $\min(t, 0)$. From this observation, the two Jacobian matrices of $\mathbb{F}_\theta$ and $\mathbb{F}_s$ are:

$$\nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbb{X}) = \begin{pmatrix} M_{n\times n} & -I_{n\times n} & 0_{n\times 1} \\ Q_k(\mathbf{X}) & Q_l(\mathbf{X}) & W\mathbf{e} \\ (x^-)^{\mathrm{T}} & (z^-)^{\mathrm{T}} & 2r \end{pmatrix}, \tag{5.4.12}$$

and

$$\nabla_{\mathbb{X}} \mathbb{F}_s(\mathbb{X}) = \begin{pmatrix} M_{n\times n} & -I_{n\times n} & 0_{n\times 1} \\ D_a(\mathbf{X}) & D_b(\mathbf{X}) & V\mathbf{e} \\ (x^-)^{\mathrm{T}} & (z^-)^{\mathrm{T}} & 2r \end{pmatrix}, \tag{5.4.13}$$

where x^- is the vector of components $x_i^- = \min(x_i, 0)$ and similarly for z^-,

$$V = \operatorname{diag}\left(\left(-\log(1 + e^{\frac{x_i - \rho z_i}{r}}) + \frac{\frac{x_i - \rho z_i}{r} e^{\frac{x_i - \rho z_i}{r}}}{1 + e^{\frac{x_i - \rho z_i}{r}}} \right)_{1\leqslant i\leqslant n} \right),$$

$$W = \operatorname{diag}\left(\left(\frac{x_i^2}{(x_i + r)^2} + \frac{z_i^2}{(z_i + r)^2} - 1 \right)_{1\leqslant i\leqslant n} \right).$$

If $\mathbb{F}_\theta(\mathbb{X}) = 0$ (resp. $\mathbb{F}_s(\mathbb{X}) = 0$) where $\mathbb{X} \in \mathbb{R}_+^{2n} \times \mathbb{R}_+$ we obtain $r = 0$ and $x^- = z^- = 0$. Hence in this case, $\nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbb{X})$ becomes singular (resp. $\nabla_{\mathbb{X}} \mathbb{F}_s(\mathbb{X})$ becomes singular) since $\det \nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbb{X}) = 0$ (resp. $\det \nabla_{\mathbb{X}} \mathbb{F}_s(\mathbb{X}) = 0$). To solve this issue, we add a small enough positive parameter ε in the last equation. We get

$$\frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r = 0. \tag{5.4.14}$$

Hence, we define the following systems

$$\mathbb{F}_\theta(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ r\left(\theta_r^1(x) + \theta_r^1(z) - \mathbf{e}\right) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0, \tag{5.4.15}$$

and

$$\mathbb{F}_s(\mathbb{X}) = \begin{bmatrix} Mx + q - z \\ x - r\log\left(\mathbf{e} + e^{\frac{x-\rho z}{r}}\right) \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix} = 0. \tag{5.4.16}$$

Lemma 5.4.1. *Let $\mathbf{X} \in \bar{\Xi}$, where Ξ is the interior region defined in*

$$\Xi = \{\mathbf{X} = (x, z) \in \mathbb{R}^{2n} \quad | \quad x > 0, \quad z > 0\}. \tag{5.4.17}$$

Let $r \in \mathbb{R}$ and $\mathbb{X} = [\mathbf{X}; r]^T$. Then,

$$det\ \nabla\mathbb{F}_\theta(\mathbb{X}) = (\varepsilon + 2r)\ det\ \nabla F_\theta^r(\mathbf{X}).$$

and

$$det\ \nabla\mathbb{F}_s(\mathbb{X}) = (\varepsilon + 2r)\ det\ \nabla F_s^r(\mathbf{X}).$$

If $\varepsilon + 2r > 0$, the two Jacobian matrices $\nabla\mathbb{F}_\theta$ and ∇F_θ^r (resp. $\nabla\mathbb{F}_s$ and ∇F_s^r) are singular or nonsingular at the same time.

Proof. Thanks to the assumption $\mathbf{X} \in \bar{\Xi}$, we have $x \geqslant 0$ and $z \geqslant 0$, so that $x^- = z^- = 0$. Expanding the determinant of (5.4.15) and (5.4.16) with respect to the last row yields the desired result. □

5.5 Convergence

In this section, we propose two generic algorithms to solve LCP and prove some convergence results. From now on, the enlarged equations (5.4.15) and (5.4.16) are selected as the reference systems in the design of our new algorithms. The idea is simply to apply the standard Newton method to the smooth system (5.4.15) and (5.4.16). To enforce a global convergence behavior, we also recommend using α line search like Armijo back-tracking technique.
Now, we present our algorithms for our methods described above:

Algorithm 5.1 Nonparametric TLCP with Armijo line search

1. Choose $\mathbb{X}^0 = (\mathbf{X}^0, r^0),\ \mathbf{X}^0 > 0,\ r^0 =< x^0, z^0 > /n,\ \tau \in (1, 1/2), \varrho \in (0, 1)$. Set $k = 0$.
2. If $\mathbb{F}_\theta(\mathbb{X}^k) = 0$, stop.
3. Find a direction $d^k \in \mathbb{R}^{2n+1}$ such that

$$\mathbb{F}_\theta(\mathbb{X}^k) + \nabla_{\mathbb{X}}\mathbb{F}_\theta(\mathbb{X}^k)d^k = 0.$$

4. Choose $\zeta^k = \varrho^{j_k} \in (0, 1)$, where $j_k \in \mathbb{N}$ is the smallest integer such that

$$\Theta_\theta(\mathbb{X}^k + \varrho^{j_k} d^k) \leqslant (1 - 2\tau\varrho^{j_k})\ \Theta_\theta(\mathbb{X}^k).$$

5. Set $\mathbb{X}^{k+1} = \mathbb{X}^k + \zeta^k d^k$ and $k \leftarrow k + 1$. Go to step 2.

Algorithm 5.2 Nonparametric Soft-LCP method with Armijo line search

1. Choose $\mathbb{X}^0 = (\mathbf{X}^0, r^0),\ \mathbf{X}^0 > 0,\ r^0 =< x^0, z^0 > /n,\ \tau \in (1, 1/2),\ \varrho \in (0, 1)$. Set $k = 0$.
2. If $\mathbb{F}_s(\mathbb{X}^k) = 0$, stop.
3. Find a direction $d^k \in \mathbb{R}^{2n+1}$ such that

$$\mathbb{F}_s(\mathbb{X}^k) + \nabla_{\mathbb{X}}\mathbb{F}_s(\mathbb{X}^k)d^k = 0.$$

4. Choose $\zeta^k = \varrho^{j_k} \in (0, 1)$, where $j_k \in \mathbb{N}$ is the smallest integer such that

$$\Theta_s(\mathbb{X}^k + \varrho^{j_k} d^k) \leqslant (1 - 2\tau\varrho^{j_k})\ \Theta_s(\mathbb{X}^k).$$

5. Set $\mathbb{X}^{k+1} = \mathbb{X}^k + \zeta^k d^k$ and $k \leftarrow k + 1$. Go to step 2.

The merit function used in the line search is

$$\Theta_\theta(\mathbb{X}) = \frac{1}{2}\|\mathbb{F}_\theta(\mathbb{X})\|^2 \quad (resp. \quad \Theta_s(\mathbb{X}) = \frac{1}{2}\|\mathbb{F}_s(\mathbb{X})\|^2).$$

A detailed description of Nonparametric Soft-LCP is given in Algorithm 5.2. A few comments are in order:

- The initial point $\mathbb{X}^0 = (\mathbf{X}^0, r^0)$ must be an interior point, namely, $\mathbf{X}^0 > 0$ and the initial parameter
  $r^0 =< x^0, z^0 > /n$ has the correct order of magnitude.

- If $\mathbf{X}^k > 0$, then $(x^k)^- = (z^k)^- = 0$ and

$$d^k = \begin{bmatrix} d\mathbf{X}^k \\ dr^k \end{bmatrix} = - \begin{pmatrix} \nabla F_s^r(\mathbf{X}^k) & \partial_r F_s^r(\mathbf{X}^k) \\ 0 & \varepsilon + 2r^k \end{pmatrix}^{-1} \begin{bmatrix} F_s^r(\mathbf{X}^k) \\ \varepsilon r^k + (r^k)^2 \end{bmatrix}.$$

  provided that the Jacobian matrix is invertible. The increment for the parameter is then

$$dr^k = -\frac{\varepsilon r^k + (r^k)^2}{\varepsilon + 2r^k}$$

- There is no need to truncate the Newton direction d^k to preserve positivity for x^{k+1} and z^{k+1}, since nonnegativity is "guaranteed" at convergence. However, if we want all the iterates to be nonnegative, so we need to carry out an additional damping after Step 4 (Armijo's line search).

Proposition 5.5.1. *Let $M \in \mathbb{R}^{n\times n}$ be a $\mathcal{P}_0$-matrix. Then, step 3 in Algorithm 5.1 (resp. Algorithm 5.2) is well-defined.*

Proof. We know that for all $k \geqslant 0,\ r^k > 0, \mathbf{X}^k > 0$, and $\varepsilon > 0$,

$$\det \nabla \mathbb{F}_s(\mathbb{X}^k) = (\varepsilon + 2r^k) \det \nabla F_s^{r^k}(\mathbf{X}^k),$$

and

$$\det \nabla \mathbb{F}_\theta(\mathbb{X}^k) = (\varepsilon + 2r^k) \det \nabla F_\theta^{r^k}(\mathbf{X}^k).$$

By Theorem 5.3.3, we know that $\nabla F_s^{r^k}(\mathbf{X}^k)$ (resp. $\nabla F_\theta^{r^k}(\mathbf{X}^k)$) is nonsingular, so Step 3 of Algorithm 5.1 (resp. Algorithm 5.2) is well-defined. □

5.5.1 Global convergence analysis

Definition 5.5.2. *(Regular zero). Let $\bar{\mathbf{X}} \in \mathbb{R}^{2n}$ be a zero of F, that is, $F(\bar{\mathbf{X}}) = 0$. If the Jacobian matrix $\nabla F(\bar{\mathbf{X}})$ is nonsingular, $\bar{\mathbf{X}}$ is said to be a regular zero of F.*

The main interest of Algorithm 5.1 and Algorithm 5.2 lies in the prospect of global convergence, as envisioned by the theory that we are developing now. This global convergence theory, is primarily based on the regularity of zeros [definition 5.5.2]. We reproduce a concise result that can be found in the book of Bonnans [20], in view of its importance to our algorithm.
We will prove the global convergence of Algorithm 5.1 (resp. Algorithm 5.2). First, we show that every $d \in \triangle_\theta$ (resp. $d \in \triangle_s$) is a descent direction of Θ_θ at $\mathbb{X}$ (resp. Θ_s at $\mathbb{X}$), where

$$\triangle_\theta(\mathbb{X}) = \{d \in \mathbb{R}^n \quad | \quad \nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbb{X})d = -\mathbb{F}_\theta(\mathbb{X})\}. \tag{5.5.1}$$

and

$$\triangle_s(\mathbb{X}) = \{d \in \mathbb{R}^n \quad | \quad \nabla_{\mathbb{X}} \mathbb{F}_s(\mathbb{X})d = -\mathbb{F}_s(\mathbb{X})\}. \tag{5.5.2}$$

Lemma 5.5.3. *(see [107]) If $\mathbb{X}$ is not a solution of LCP, i.e. $\Theta_\theta(\mathbb{X}) > 0$ (resp. $\Theta_s(\mathbb{X}) > 0$), then every $d \in \triangle_\theta(\mathbb{X})$ (resp. $d \in \triangle_s(\mathbb{X})$) satisfies the descent condition for $\mathbb{X}$, i.e., $\nabla\Theta_\theta(\mathbb{X})^T d < 0$ (resp. $\nabla\Theta_s(\mathbb{X})^T d < 0$).*

Similar to the proof in [107], we can prove the following result.

Theorem 5.5.4. *Every limit point $\mathbb{X}^* = (\mathbf{X}^*, r^*)$ of a sequence $\{\mathbb{X}^k\}$ generated by Algorithm 5.1 (resp. Algorithm 5.2) corresponds to a solution of LCP.*

Now we would like to study the asymptotic behavior of the Jacobian matrix $\mathbb{F}_\theta$ (resp. $\mathbb{F}_s$) when r goes to 0.

Lemma 5.5.5. *Let $\mathbf{X}^*$ be a solution of LCP satisfying the strict complementarity condition and $\mathbb{F}_\theta(\mathbb{X})$ be defined by (5.4.15). Then the Jacobian matrix of $\mathbb{F}_\theta(\mathbf{X}^*, r)$ when r goes to 0 is:*

$$\lim_{r\to 0} \left(\nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r)\right) = \begin{pmatrix} M_{n\times n} & -I_{n\times n} & 0_{n\times 1} \\ \phi_\theta(Z^*) & \phi_\theta(X^*) & 0_{n\times 1} \\ 0_{1\times n} & 0_{1\times n} & \varepsilon \end{pmatrix},$$

where

$$\phi_\theta(Z^*)_{ii} = \begin{cases} 0 & \text{if} \quad z_i^* \neq 0 \quad \text{and} \quad x_i^* = 0 \\ 1 & \text{if} \quad z_i^* = 0 \quad \text{and} \quad x_i^* \neq 0, \end{cases} \quad \text{and} \quad \phi_\theta(X^*)_{ii} = \begin{cases} 0 & \text{if} \quad x_i^* \neq 0 \quad \text{and} \quad z_i^* = 0 \\ 1 & \text{if} \quad x_i^* = 0 \quad \text{and} \quad z_i^* \neq 0. \end{cases}$$

Proof. Let S defined as

$$S = \{(x_i, z_i, r)/ \;\; \theta_r^1(x_i) + \theta_r^1(z_i) = 1, \forall i \in \{1, ..., n\}\},$$

by Lemma 5.2.3, we have

$$\theta_r^1(x_i) + \theta_r^1(z_i) = 1 \iff x_i z_i = r^2, \quad \forall i \in \{1, ..., n\}.$$

We can therefore define the set S in the form:

$$S = \{(x_i, z_i, r)/ \;\; x_i z_i - r^2 = 0, \forall i \in \{1, ..., n\}\}.$$

Since $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP, we deduce that (x^*, z^*, r) is near to S, then

$$x_i^* z_i^* - r^2 = o(r)\,,$$

i.e. $x_i^* z_i^* - r^2$ is negligent by r. In view of the assumption of the strict complementary of $\mathbf{X}^* = (x^*, z^*)$ we have to consider two cases if $z_i^* > 0$ then $x_i^* = o(r)$ and if $x_i^* > 0$ then $z_i^* = o(r)$. By definition

$$\mathbb{F}_\theta(\mathbb{X}) = \begin{bmatrix} (\mathbb{F}_\theta)_1(\mathbb{X}) \\ (\mathbb{F}_\theta)_2(\mathbb{X}) \\ (\mathbb{F}_\theta)_3(\mathbb{X}) \end{bmatrix} = \begin{bmatrix} Mx + q - z \\ \left(\dfrac{rx_i}{x_i + r} + \dfrac{rz_i}{z_i + r}\right)_{1 \leqslant i \leqslant n} - r\mathbf{e} \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix}.$$

The jacobian matrix of $\mathbb{F}_\theta$ is:

$$\nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbb{X}) = \begin{pmatrix} M_{n\times n} & -I_{n\times n} & 0_{n\times 1} \\ \nabla_x (\mathbb{F}_\theta)_2(\mathbb{X}) & \nabla_z (\mathbb{F}_\theta)_2(\mathbb{X}) & \partial_r (\mathbb{F}_\theta)_2(\mathbb{X}) \\ (x^-)^{\mathrm{T}} & (z^-)^{\mathrm{T}} & 2r + \varepsilon \end{pmatrix},$$

1. The derivative of $(\mathbb{F}_\theta)_2(\mathbf{X}, r)$ with respect to x is:

$$\nabla_x (\mathbb{F}_\theta)_2(x^*, z^*, r) = \operatorname{diag}\left(\left(\left(\frac{r}{x_i^* + r}\right)^2\right)_{1 \leqslant i \leqslant n}\right),$$

when r goes to 0 and in view of the strict complementary of $\mathbf{X}^* = (x^*, z^*)$, the only two cases to consider are:

- $x_i^* \to 0$, and $z_i^* > 0 \quad \forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}} \left(\nabla_x \left(\mathbb{F}_\theta\right)_2 \left(x_i^*, z_i^*, r\right)\right)_{ii} = \lim_{r\to 0} \left(\frac{r}{o(r)+r}\right)^2 = \lim_{r\to 0} \left(\frac{r}{r}\right)^2 = 1.$$

- $x_i^* > 0$, and $z_i^* \to 0 \quad \forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}} \left(\nabla_x \left(\mathbb{F}_\theta\right)_2 \left(x_i^*, z_i^*, r\right)\right)_{ii} = \lim_{\substack{r\to 0\\ z_i^*\to 0}} \left(\frac{r}{x_i^*+r}\right)^2 = 0.$$

2. The derivative of $\left(\mathbb{F}_\theta\right)_2 (\mathbf{X}, r)$ with respect to z is:

$$\nabla_z \left(\mathbb{F}_\theta\right)_2 \left(x^*, z^*, r\right) = \operatorname{diag}\left(\left(\left(\frac{r}{z_i^*+r}\right)^2\right)_{1\leqslant i\leqslant n}\right),$$

as below, the only two cases to consider are:

- $x_i^* \to 0$, and $z_i^* > 0 \quad \forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}} \left(\nabla_z \left(\mathbb{F}_\theta\right)_2 \left(x_i^*, z_i^*, r\right)\right)_{ii} = \lim_{\substack{r\to 0\\ x_i^*\to 0}} \left(\frac{r}{z_i^*+r}\right)^2 = 0.$$

- $x_i^* > 0$, and $z_i^* \to 0 \quad \forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}} \left(\nabla_z \left(\mathbb{F}_\theta\right)_2 \left(x_i^*, z_i^*, r\right)\right)_{ii} = \lim_{r\to 0} \left(\frac{r}{o(r)+r}\right)^2 = 1.$$

3. The derivative of $\mathbb{F}_{\theta,\,2}(\mathbf{X}, r)$ with respect to r is:

$$\partial_r \left(\mathbb{F}_\theta\right)_2 \left(x^*, z^*, r\right) = \left(\left(\frac{x_i^*}{x_i^*+r}\right)^2 + \left(\frac{z_i^*}{z_i^*+r}\right)^2 - 1\right)_{1\leqslant i\leqslant n}.$$

when r goes to 0 and in view of the strict complementary of $\mathbf{X}^* = (x^*, z^*)$, the only two cases to consider are:

- $x_i^* \to 0$, and $z_i^* > 0 \quad \forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}} \left(\partial_r \left(\mathbb{F}_\theta\right)_2 \left(x_i^*, z_i^*, r\right)\right)_i = \lim_{r\to 0} \left(\left(\frac{o(r)}{o(r)+r}\right)^2 + \left(\frac{z_i^*}{z_i^*+r}\right)^2 - 1\right) = 0.$$

- $x_i^* > 0$, and $z_i^* \to 0 \quad \forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\neq 0\\ z_i^*\to 0}} \left(\partial_r \left(\mathbb{F}_\theta\right)_2 \left(x_i^*, z_i^*, r\right)\right)_i = \lim_{r\to 0} \left(\left(\frac{x_i^*}{x_i^* + r} \right)^2 + \left(\frac{o(r)}{o(r) + r} \right)^2 - 1 \right) = 0.$$

Finally, since $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP, we have $x^* \geqslant 0$ and $z^* \geqslant 0$, so that $x^- = z^- = 0$. Hence

$$\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r) = \begin{pmatrix} \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_\theta(Z^*) & \phi_\theta(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0_{1\times n} & 0_{1\times n} \end{matrix} & \varepsilon \end{pmatrix},$$

□

Here we present the same result but for the system $\mathbb{F}_s(\mathbb{X})$.

Lemma 5.5.6. *Let $\mathbf{X}^*$ be a solution of LCP satisfying the strict complementarity condition and $\mathbb{F}_s(\mathbb{X})$ be defined by (5.4.16). Then the Jacobian matrix of $\mathbb{F}_s(\mathbf{X}^*, r)$ when r goes to 0 is:*

$$\lim_{r\to 0} \left(\nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r)\right) = \begin{pmatrix} M_{n\times n} & -I_{n\times n} & 0_{n\times 1} \\ \phi_s(Z^*) & \phi_s(X^*) & 0_{n\times 1} \\ 0_{1\times n} & 0_{1\times n} & \varepsilon \end{pmatrix},$$

where

$$\phi_s(Z^*)_{ii} = \begin{cases} 1 \quad \textit{if} \quad z_i^* \neq 0 \quad \textit{and} \quad x_i^* = 0 \\ 0 \quad \textit{if} \quad z_i^* = 0 \quad \textit{and} \quad x_i^* \neq 0, \end{cases} \quad \textit{and} \quad \phi_s(X^*)_{ii} = \begin{cases} 1 \quad \textit{if} \quad x_i^* \neq 0 \quad \textit{and} \quad z_i^* = 0 \\ 0 \quad \textit{if} \quad x_i^* = 0 \quad \textit{and} \quad z_i^* \neq 0. \end{cases}$$

Proof. Let

$$\mathbb{F}_s(\mathbb{X}) = \begin{bmatrix} \left(\mathbb{F}_s\right)_1 (\mathbb{X}) \\ \left(\mathbb{F}_s\right)_2 (\mathbb{X}) \\ \left(\mathbb{F}_s\right)_3 (\mathbb{X}) \end{bmatrix} = \begin{bmatrix} Mx + q - z \\ \left(x_i - r \log\left(1 + e^{\frac{x_i - \rho z_i}{r}} \right) \right)_{1\leqslant i\leqslant n} \\ \frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r \end{bmatrix}.$$

The Jacobian matrix of $\mathbb{F}_s$ is:

$$\nabla_{\mathbb{X}}\mathbb{F}_s(\mathbb{X}) = \begin{pmatrix} M_{n\times n} & -I_{n\times n} & 0_{n\times 1} \\ \nabla_x\left(\mathbb{F}_s\right)_2(\mathbb{X}) & \nabla_z\left(\mathbb{F}_s\right)_2(\mathbb{X}) & \partial_r\left(\mathbb{F}_s\right)_2(\mathbb{X}) \\ (x^-)^{\mathrm{T}} & (z^-)^{\mathrm{T}} & 2r+\varepsilon \end{pmatrix}.$$

Let us to calculate $\lim\limits_{r\to 0}\nabla_{\mathbb{X}}\mathbb{F}_s(\mathbf{X}^*,r)$:

1. The derivative of $\left(\mathbb{F}_s\right)_2(\mathbf{X},r)$ with respect to x is:

$$\nabla_x\left(\mathbb{F}_s\right)_2(x^*,z^*,r) = \operatorname{diag}\left(\left(\frac{1}{1+e^{\frac{x_i^*-\rho z_i^*}{r}}}\right)_{1\leqslant i\leqslant n}\right),$$

   when r goes to 0 and in view of the strict complementary of $\mathbf{X}^* = (x^*,z^*)$, the only two cases to consider are:

   - $x_i^* \to 0$, and $z_i^* > 0\ \forall i\in\{1,...,n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}}\left(\nabla_x\left(\mathbb{F}_s\right)_2(x_i^*,z_i^*,r)\right)_{ii} = \lim_{r\to 0}\frac{1}{1+e^{-\frac{\rho z_i^*}{r}}} = 1.$$

   - $x_i^* > 0$, and $z_i^* \to 0\ \forall i\in\{1,...,n\}$ then

$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}}\left(\nabla_x\left(\mathbb{F}_s\right)_2(x_i^*,z_i^*,r)\right)_{ii} = \lim_{r\to 0}\frac{1}{1+e^{\frac{x_i^*}{r}}} = 0.$$

2. The derivative of $\left(\mathbb{F}_s\right)_2(\mathbf{X},r)$ with respect to z is:

$$\nabla_z\left(\mathbb{F}_s\right)_2(x^*,z^*,r) = \operatorname{diag}\left(\left(\frac{\rho e^{\frac{x_i^*-\rho z_i^*}{r}}}{1+e^{\frac{x_i^*-\rho z_i^*}{r}}}\right)_{1\leqslant i\leqslant n}\right),$$

   as below, the only two cases to consider are:

   - $x_i^* \to 0$, and $z_i^* > 0\ \forall i\in\{1,...,n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i^*\neq 0}}\left(\nabla_z\left(\mathbb{F}_s\right)_2(x_i^*,z_i^*,r)\right)_{ii} = \lim_{r\to 0}\rho\frac{e^{-\frac{\rho z_i^*}{r}}}{1+e^{-\frac{\rho z_i^*}{r}}} = 0.$$

- $x_i^* > 0$, and $z_i^* \to 0$ $\forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}} \left(\nabla_z \left(\mathbb{F}_s\right)_2 \left(x_i^*, z_i^*, r\right)\right)_{ii} = \lim_{r\to 0} \rho \frac{e^{\frac{x_i^*}{r}}}{1 + e^{\frac{x_i^*}{r}}} = \rho,$$

we take $\rho = 2$ to ensure the convergence (see Figure 5.4).

3. The derivative of $\left(\mathbb{F}_s\right)_2 (\mathbf{X}, r)$ with respect to r is:

$$\partial_r \left(\mathbb{F}_s\right)_2 \left(x^*, z^*, r\right) = \left(-\log(1 + e^{\frac{x_i^* - \rho z_i^*}{r}}) + \frac{\frac{x_i^* - \rho z_i^*}{r} e^{\frac{x_i^* - \rho z_i^*}{r}}}{1 + e^{\frac{x_i^* - \rho z_i^*}{r}}} \right)_{1 \leqslant i \leqslant n},$$

when r goes to 0 and in view of the strict complementary of $\mathbf{X}^* = (x^*, z^*)$, the only two cases two consider are:

- $x_i^* \to 0$, and $z_i^* > 0$ $\forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ x_i^*\to 0\\ z_i\neq 0}} \left(\partial_r \left(\mathbb{F}_s\right)_2 \left(x_i^*, z_i^*, r\right)\right)_i = \lim_{r\to 0} \left[-\log(1 + e^{-\frac{\rho z_i^*}{r}}) - \frac{\rho z_i^*}{r} \frac{e^{-\frac{\rho z_i^*}{r}}}{1 + e^{-\frac{\rho z_i^*}{r}}} \right] = 0.$$

- $x_i^* > 0$, and $z_i^* \to 0$ $\forall i \in \{1, ..., n\}$ then

$$\lim_{\substack{r\to 0\\ z_i^*\to 0\\ x_i^*\neq 0}} \left(\partial_r \left(\mathbb{F}_s\right)_2 \left(x_i^*, z_i^*, r\right)\right)_i = \lim_{r\to 0} \left[-\log(1 + e^{\frac{x_i^*}{r}}) + \frac{x_i^*}{r} \frac{e^{\frac{x_i^*}{r}}}{1 + e^{\frac{x_i^*}{r}}} \right] = 0.$$

Finally, thanks to the assumption $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP, we have $x^* \geqslant 0$ and $z^* \geqslant 0$, so that $x^- = z^- = 0$. Hence

$$\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r) = \begin{pmatrix} \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_s(Z^*) & \phi_s(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{pmatrix},$$

□

We present now, the situations where we can conclude about the nonsingularity of $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r)$ (resp. $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r)$).

Lemma 5.5.7. *Suppose that $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP, we have one possibilitie when computing the determinant of $\mathbb{F}_\theta$ on (x^*, z^*). (resp. $\mathbb{F}_s$ on (x^*, z^*)).*
If $\mathbf{X}^* = (x^*, z^*)$ satisfies the strict complementarity condition, then from Lemma 5.5.9,
$\forall \ \mathbb{I} \subset \{1, ..., n\}$, we have:

- $$\lim_{r\to 0} det\ (\nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r)) = \left| \begin{pmatrix} \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_\theta(Z^*) & \phi_\theta(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{pmatrix} \right| = \varepsilon \left| \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_\theta(Z^*) & \phi_\theta(X^*) \end{pmatrix} \right| = \varepsilon |M_{\mathbb{II}}|,$$

  therefore the matrix $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r)$ exists and is invertible if M is $\mathcal{P}$-matrix.

- $$\lim_{r\to 0} det\ (\nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r)) = \left| \begin{pmatrix} \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_s(Z^*) & \phi_s(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{pmatrix} \right| = \varepsilon \left| \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_s(Z^*) & \phi_s(X^*) \end{pmatrix} \right| = \varepsilon |M_{\mathbb{II}}|,$$

  therefore the matrix $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r)$ exists and is invertible if M is $\mathcal{P}$-Matrix.

In the following, we focus our attention on the superlinear convergence rate of Algorithm 5.1 and Algorithm 5.2.

Theorem 5.5.8. *(Theorem 6.9, [20]). Let $\mathbb{F}_\theta : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n+1}$ (resp. $\mathbb{F}_s : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n+1}$) be a continuously-differentiable function.*

(i) (Local analysis) Let $\mathbb{X}^*$ be a regular zero of $\mathbb{F}_\theta$ (resp. $\mathbb{F}_s$). If $\mathbb{X}^0$ is close enough to $\bar{\mathbb{X}}$, then $\zeta^k = 1$ for all k, and $\mathbb{X}^k$ converge to $\mathbb{X}^*$ super-linearly (and we recover the standard Newton method).

(ii) (Limit point) Let $\mathbb{X}^*$ be a limit point of sequence $\{\mathbb{X}^k\}$. If $\nabla \mathbb{F}_\theta(\mathbb{X}^*)$ (resp. $\nabla \mathbb{F}_s(\mathbb{X}^*)$) is invertible, then $\mathbb{X}^*$ is a regular zero of $\mathbb{F}_\theta$ (resp. $\mathbb{F}_s$). If $\mathbb{X}^*$ is a regular zero of $\mathbb{F}_\theta$ (resp. $\mathbb{F}_s$), then $\zeta^k = 1$ for k big enough and $\mathbb{X}^k$ converge to $\mathbb{X}^*$ super-linearly.

Proof. We apply Theorem 5.5.4 and Lemma 5.5.7 under some condition on F. □

Now we would like to study the asymptotic behavior of the Jacobian matrix of our methods with the Jacobian matrix of interior-point methods when r goes to 0 and we need a lemma that is used to prove our main result.

Lemma 5.5.9. *We consider the following system*

$$\begin{aligned} &Z.X = 0 \\ &Z \geqslant 0, \quad X \geqslant 0, \end{aligned} \tag{5.5.3}$$

where $Z = diag(z)$ and $X = diag(x)$. Assume that Z, X are strictly complementary. Then J is singular if and only if T is singular, where

$$J = \begin{pmatrix} M & -I \\ Z & X \end{pmatrix}, \quad \text{and} \quad T = \begin{pmatrix} M & -I \\ \phi(Z) & \phi(X) \end{pmatrix},$$

such that

$$\phi(t) = \begin{cases} 1 & \text{if} \quad t \neq 0 \\ 0 & \text{if} \quad t = 0, \end{cases}$$

here ϕ operates componentwise on Z and X , and it verifies the following system

$$\begin{aligned} &\phi(Z).\phi(X) = 0 \\ &\phi(Z) \geqslant 0, \quad \phi(X) \geqslant 0. \end{aligned}$$

Proof. By the strict complementarity hypothesis, we range the rows and the columns of J and T as follows

$$J_\sigma = \begin{pmatrix} M_\sigma & -I_\sigma \\ \begin{pmatrix} Z_1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & X_2 \end{pmatrix} \end{pmatrix},$$

where $X_2 > 0$ and $Z_1 > 0$, and

$$(T)_\sigma = \begin{pmatrix} M_\sigma & -I_\sigma \\ \begin{pmatrix} 1 & & \\ & \ddots & 0 \\ & & 1 \\ & & \\ 0 & & 0 \end{pmatrix} & \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ & & 1 \quad 0 \\ 0 & & \ddots \\ & 0 & 1 \end{pmatrix} \end{pmatrix}.$$

The determinant of the two matrices J_σ and $(T)_\sigma$ are equal to

$$\det(J_\sigma) = \begin{vmatrix} M_\sigma & -I_\sigma \\ \begin{pmatrix} Z_1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & X_2 \end{pmatrix} \end{vmatrix} = \pm \prod_{i \in \mathbb{I}_1} x_i \prod_{i \in \mathbb{I}_2} z_i \quad \det(C),$$

$$\det(T_\sigma) = \begin{vmatrix} M_\sigma & -I_\sigma \\ \begin{pmatrix} 1 & & \\ & \ddots & 0 \\ & & 1 \\ & & \\ 0 & & 0 \end{pmatrix} & \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ & & 1 \quad 0 \\ 0 & & \ddots \\ & 0 & 1 \end{pmatrix} \end{vmatrix} = \pm \prod_{i \in \mathbb{I}_1} \phi(x_i) \prod_{i \in \mathbb{I}_2} \phi(z_i) \quad \det(C),$$

where C is a certain matrix, $\mathbb{I}_1 = \{i \mid x_i > 0\}$ and $\mathbb{I}_2 = \{i \mid z_i > 0\}$. Since

$$\pm \prod_{i \in \mathbb{I}_1} x_i \prod_{i \in \mathbb{I}_2} z_i \qquad \text{and} \qquad \prod_{i \in \mathbb{I}_1} \phi(x_i) \prod_{i \in \mathbb{I}_2} \phi(z_i),$$

are nonzeros, then we can conclude that J and T are invertibles and singulars at the same time. □

Theorem 5.5.10. *Suppose that $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP which satisfies the strict complementarity condition (i.e. $x_i^* + z_i^* > 0$, $\forall i \in \{1, ..., n\}$), and $\nabla_{\mathbf{X}} F_0(\mathbf{X}^*)$ define by (5.4.3) (the Jacobian matrix of the Interior-Point Methods) is invertible. Then $\lim_{r \to 0} \nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r)$ is invertible, i.e. the two Jacobian matrices are singular or nonsigular at the same time.*

Proof. In view of Lemma 5.5.9, and thanks to the assumption $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP, we

have $x^* \geqslant 0$ and $z^* \geqslant 0$, so that $x^- = z^- = 0$. Hence

$$\lim_{r\to 0} \det\ (\nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r)) = \left| \begin{pmatrix} \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_\theta(Z^*) & \phi_\theta(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{pmatrix} \right| = \varepsilon \left| \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_\theta(Z^*) & \phi_\theta(X^*) \end{pmatrix} \right|,$$

where $\phi_\theta(.)$ is defined in Lemma 5.5.5, $Z^* = \text{diag}(z^*)$ and $X^* = \text{diag}(x^*)$.
From Lemma 5.5.9, we conclude that if $\nabla_{\mathbf{X}} F_0(\mathbf{X}^*)$ is invertible then $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_\theta(\mathbf{X}^*, r)$ is invertible. This means, that if the Interior Point Method converges our method converges.

□

Here we present the same result but for the system $\mathbb{F}_s(\mathbb{X})$.

Theorem 5.5.11. *Suppose that $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP which satisfies the strict complementarity condition, and $\nabla_{\mathbf{X}} F_0(\mathbf{X}^*)$ define by (5.4.3) (the Jacobian matrix of the Interior-Point Methods) is invertible. Then $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r)$ is invertible, i.e. the two Jacobian matrices are singular or nonsigular at the same time.*

Proof. In view of Lemma 5.5.9, and thanks to the assumption $\mathbf{X}^* = (x^*, z^*)$ is a solution of LCP, we have $x^* \geqslant 0$ and $z^* \geqslant 0$, so that $x^- = z^- = 0$. Hence

$$\lim_{r\to 0} \det\ (\nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r)) = \left| \begin{bmatrix} \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_s(Z^*) & \phi_s(X^*) \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \varepsilon \end{bmatrix} \right| = \varepsilon \left| \begin{pmatrix} M_{n\times n} & -I_{n\times n} \\ \phi_s(Z^*) & \phi_s(X^*) \end{pmatrix} \right|,$$

where $\phi_s(.)$ is defined in Lemma 5.5.6, $Z^* = \text{diag}(z^*)$ and $X^* = \text{diag}(x^*)$.
From Lemma 5.5.9, we conclude that if $\nabla_{\mathbf{X}} F_0(\mathbf{X}^*)$ is invertible then $\lim_{r\to 0} \nabla_{\mathbb{X}} \mathbb{F}_s(\mathbf{X}^*, r)$ is invertible. This means, that if the Interior Point Method converges our method converges.

□

Hypothesis (M is a $\mathcal{P}_0$-matrix) assures us that our method is well defined and the Theorem 5.5.10 (resp. Theorem 5.5.11) shows that the domain of convergence of our method is at least as large as that of the interior-point methods.

5.6 Numerical results

Through this chapter, we studied two methods Soft-LCP and TLCP to solve the LCP, we present in this section some numerical experiments. First, we present a comparison on some randomly generated problems of our two methods with other approaches that have been suggested recently in [29, 48]. Then, we study two concrete examples, the first one is a second order ordinary differential equation and the second is an obstacle problem that can be formulated as LCP (5.1.1).
Finally We tested our algorithms on several absolute value equations problems. Our results are very promising and outperform standard methods.
For all the numerical tests and all the considered methods, the used codes are simple Matlab codes. We restrict our choice of θ-function to $\theta_r^1(x)$.
Our aim is to validate our approach and run some preliminary comparison with other methods, and not to optimize the performance of the algorithm.

5.6.1 Comparisons of methods for LCPs

We generate for several problem sizes, n=32, 64, 128, 256 the data (M, q) in order to have a solution for LCP as follows

```
R=rand(n, n);
M=R'*R+n*eye(n);
h=rand(n) ;
z=round(h).*rand(n, 1);
t=(1-round(h)).*rand(n, 1);
q=-M*t+z;
```

We compare our two methods denoted Soft-LCP and TLCP with other methods:

- TLCP2 method which is the same algorithm with a different formulation for the complementarity

$$\theta_r(x_i) + \theta_r(z_i) - \theta_r(x_i + z_i) = 0.$$

  In this case we don't necessarily need the constraint

$$\frac{1}{2}\|x^-\|^2 + \frac{1}{2}\|z^-\|^2 + r^2 + \varepsilon r = 0.$$

  since it is a reformulation of the complementarity and not a relaxation (we can use a fixed r).

- The classical interior-point method IPM [48].

- The classical Fischer-Burmeister method [29], we regularize each comlementarity constraint by considering

$$x_i z_i = 0, \quad x_i \geqslant 0,\ z_i \geqslant 0 \quad \text{by} \quad \sqrt{x_i^2 + z_i^2 + r^2} - (x_i + z_i) = 0, \quad i = 1, ..., n.$$

The main idea of all these methods is to regularize the complementarity condition $x^T z = 0$ and solve a system of equations using Newton's method. We use an Infeasible IPM and the classical Fischer-Burmeister method referred to as FB-Algor to compare with our methods. For TLCP2, we have fixed r to 1. We take for all this methods the initial point $(x_0,\ z_0) = 1,$ where $1 \in \mathbb{R}^n$ is the vector whose components are all equal to 1 and $r_0 = \langle x_0,\ z_0 \rangle / n$ and the precision is set as 10^{-6}.
The comparative results are given in the Table 5.1 to 5.5. We are interested in the following aspects: the comp.err, computed as $|x^T z|$, feas.err computed as $\|Mx+q-z\|$ the number of iterations computed as nb-iter and the time.

Table 5.1: Results from Soft-LCP with n=32, 64, 128, 256.

n	comp.err	feas.err	r	nb-iter	time
32	4.8594e-07	3.1676e-14	0.0021	9	0.0081
64	1.6217e-05	9.4826e-13	0.0072	10	0.0292
128	1.1151e-07	5.5129e-12	0.0014	14	0.0364
256	5.1985e-06	9.3893e-12	0.0020	23	0.1900

Table 5.2: Results from FB-Algor with n=32, 64, 128, 256.

n	comp.err	feas.err	r	nb-iter	time
32	4.5973e-08	1.2373e-07	1.324e-06	12	0.0218
64	9.3296e-08	2.8274e-07	9.715e-06	14	0.0305
128	5.1455e-08	1.8937e-07	4.799e-06	15	0.0372
256	2.2314e-07	5.5151e-07	4.050e-06	17	0.1294

Table 5.3: Results from TLCP with n=32, 64, 128, 256.

n	comp.err	feas.err	r	nb-iter	time
32	3.0137e-07	1.0516e-13	4.8821e-04	11	0.0083
64	1.1888e-07	7.1035e-13	4.8813e-04	11	0.0121
128	4.5973e-07	4.8634e-12	4.8793e-04	11	0.0244
256	4.4479e-07	9.9347e-12	2.4369e-04	12	0.0986

Table 5.4: Results from TLCP2 with n=32, 64, 128, 256.

n	comp.err	feas.err	r	nb-iter	time
32	7.0412e-09	8.7429e-09	1	11	0.0154
64	5.7344e-09	2.1192e-09	1	12	0.0429
128	2.3476e-07	1.02336e-08	1	11	0.1993
256	8.0873e-08	8.92116e-07	1	38	1.6799

Table 5.5: Results from IPM with n=32, 64, 128, 256.

n	comp.err	feas.err	r	nb-iter	time
32	9.4531e-07	0	0	198	0.5606
64	9.8840e-07	1.3455e-12	0	212	0.9313
128	9.1254e-07	4.1933e-10	0	248	3.6056
256	9.4437e-07	0	0	238	8.7353

In the above comparisons, we notice that our methods have much better results in terms of iteration numbers and CPU-time than classic interior-point-method IPM and FB-Algor. The TLCP method requires the fewest iteration numbers.

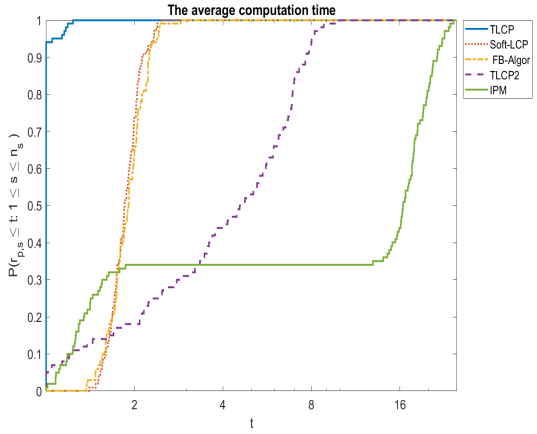

Figure 5.2: Performance profiles where $t_{p,s}$ represents the average computation time.

The figure above shows the performance profiles of five solvers where the performance measure is execution time. It is clear that the TLCP method captures our attention (admits the highest probability value). In fact, in the interval [0, 1], TLCP is able to solve 99% of the problems, while the other solvers do not reach 20% and require more time. We also notice that IPM is the slowest compared to others. However, for $t > 2$, the three algorithms TLCP, Soft-LCP and FB-Algor confirm their robustness. Figure 5.2 also indicates that, with respect to the computation time, with the same initial points and under the same stopping criterion, TLCP is the fastest solver, followed respectively by Soft-LCP, FB-Algor, TLCP2 and IPM.

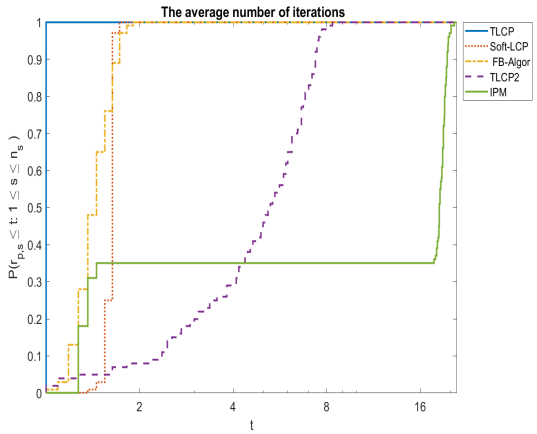

Figure 5.3: Performance profiles where $t_{p,s}$ represents the the average number of iterations.

In Figure 5.3, we illustrate the performance profiles of five solvers considering the number of iterations required as a performance measure. We notice that TLCP is the winner (admits the highest probability value) followed by FB-Algor, Soft-LCP. We also note that IPM and TLCP2 need more iterations to resolve problems. The performance of Soft-LCP becomes interesting beyond $t = 2$.

5.6.1.1 Sensitivity of ρ

We will now study the sensitivity of the parameter ρ. We solve our problem using Soft-LCP with different parameters ρ. We take $\rho = 1, 1.1, 1.2, ..., 100.$ and $n = 128$.

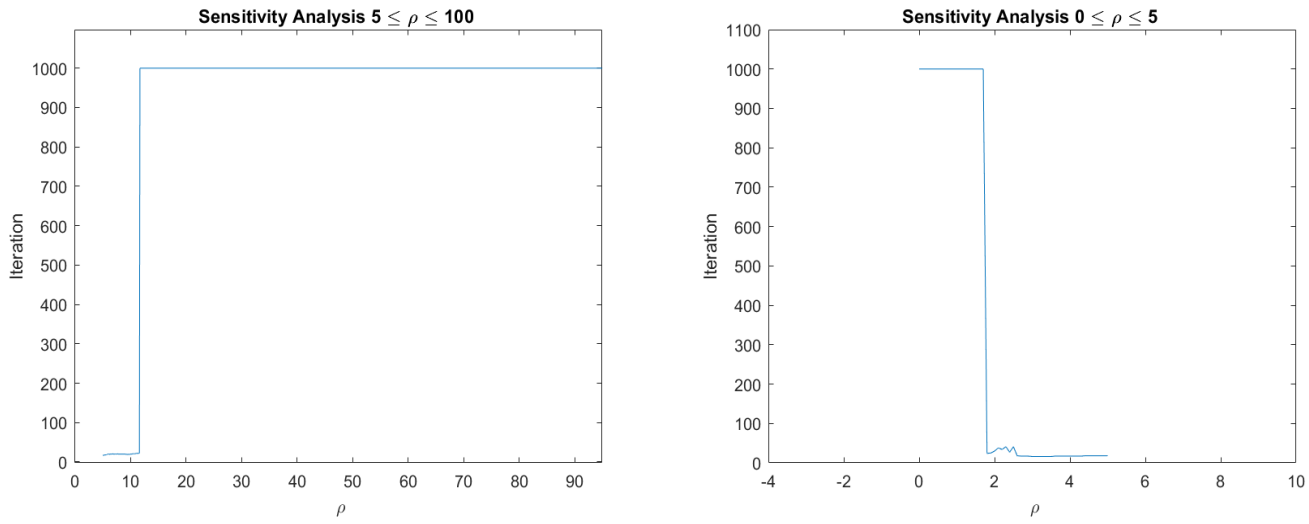

Figure 5.4: Sensitivity Analysis of ρ.

We notice a loss of convergence between 0 and 2, however between 2 and 12 convergence is assured, then from 13, there is a divergence. We conclude that we cannot choose ρ as a random parameter. In all the cases where there is convergence, the number of iterations is almost the same however in the event of divergence the number of iterations exceeds the maximum number fixed in our algorithm. We have fixed $\rho = 3$ in our approach to ensure the convergence.

5.6.2 An obstacle problem

Let f and g two continuous functions defined in $[0,1]$. We want to solve the following obstacle problem: find $u : [0,1] \to \mathbb{R}$ such that:

$$\begin{cases} -u^{"}(x) & \geqslant & f(x), \\ u(x) & \geqslant & g(x) \quad \text{on }]0,1[, \\ (-u^{"}(x) - f(x))(u(x) - g(x)) & = & 0, \end{cases}$$

and $u(0) = u(1) = 0$.
The first equation means a maximum concavity of the function u. In the second equation, we want the solution u to be above g. In the third equation, we have at least equality in one of the two previous equations. In order to get a linear complementarity problem, we set $z = u - g$ and we discretize by using the finite difference. We introduce a uniform subdivision $x_i = i * h, i = 0, \dots N+1$ of $[0,1]$, where $h = \frac{1}{N+1}$.
We use the second-order centered finite difference to approximate the second order derivatives $z^{''}(x)$ and $g^{''}(x)$. We then try to solve the following problem:

$$\begin{cases} \frac{-z_{i-1}+2z_i-z_{i+1}}{h^2} + \frac{-g_{i-1}+2g_i-g_{i+1}}{h^2} - f_i & \geqslant & 0, \\ z_i & \geqslant & 0, \quad \text{for } i = 1, ..., N, \ u_0 = u_{N+1} = 0. \\ \left(\frac{-z_{i-1}+2z_i+z_{i+1}}{h^2} + \frac{-g_{i-1}+2g_i-g_{i+1}}{h^2} - f_i\right)(z_i) & = & 0, \end{cases}$$

Where $g_i = g(x_i)$, $f_i = f(x_i)$, $z_i = z(x_i)$ and $u_i = u(x_i)$. We obtain the following complementarity problem:

$$\begin{aligned} (Mz+q)^T z &= 0, \\ z &\geqslant 0, \\ Mz+q &\geqslant 0, \end{aligned}$$

where

$$M = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}, \text{ and } q = Mg - f.$$

If 1 is not an eigenvalue of M is equivalent to AVE, ([74], Prop. 2),

$$(M-I)^{-1}(M+I)x - |x| = (M-I)^{-1}q.$$

We present in the following figures, the results of our two methods and LPM method from [74]. The obstacle g is chosen here to be

$$g(x) = \max(0.8 - 20 * (x - 0.2)^2, \max(1 - 20(x - 0.75)^2, 1.2 - 30(x - 0.41)^2))$$

$f(x) = 1$ and $N = 50$.

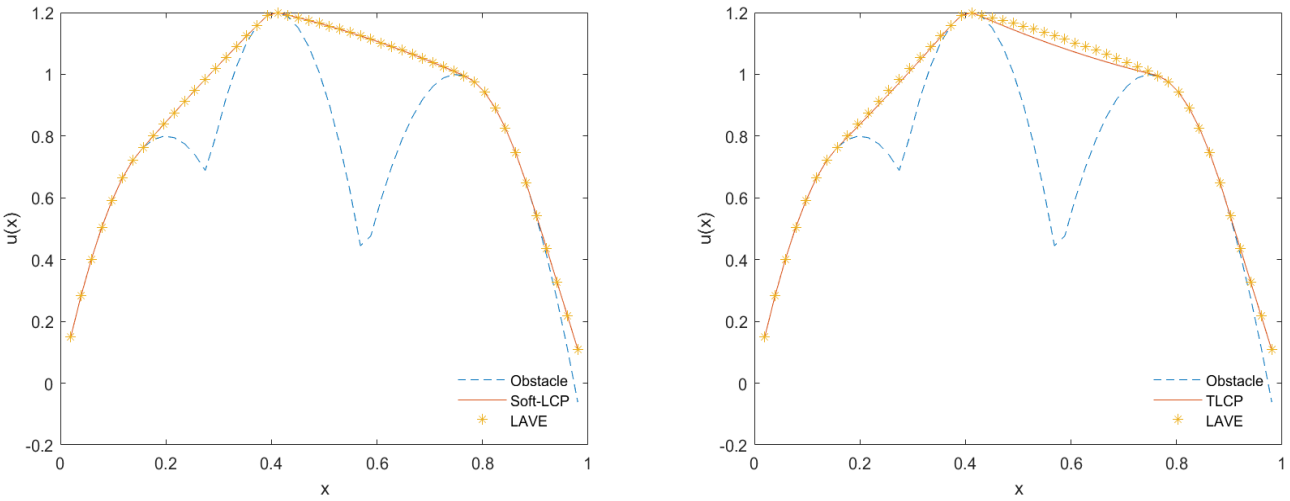

Figure 5.5: Numerical solution of the obstacle problem (5.6.2) with TLCP, Soft-LCP methods, and method from [74].

We remark that the both TLCP, Soft-LCP, and LPM method [74] have 19 common points on the curve g and none below g over 50 points. This example also confirms that our approach, TLCP and Soft-LCP method gives consistent results.

5.6.3 An ordinary differential equation

We consider the ordinary differential equation

$$x''(t) - |x(t)| = -2 - t, \quad x(0) = -4, \quad x'(0) = 5, \quad t \in [0,\ 5]. \tag{5.6.1}$$

First, we discretize the EDO equation by using the finite difference scheme. We use the second-order centred finite difference to approximate the second order derivative

$$\frac{x_{i-2} - 2x_{i-1} + x_i}{h^2} - |x_i| = (-2 - t)_i. \tag{5.6.2}$$

Equation (5.6.2) was derived with equispace gridpoints $t_i = ih,\ i = 1, ...N$. In order to approximate the Neumann boundary conditions we use a center difference

$$\frac{x_1 - x_{-1}}{2h} = x^{'}(0) = 1. \tag{5.6.3}$$

Using the classical decomposition of the absolute value [2] we reformulate (5.6.2) as follows

$$\begin{cases} N_1 x^+ - N_2 x^- = q, \\ 0 \leqslant x^+ \perp x^- \geqslant 0, \end{cases} \tag{5.6.4}$$

where

$$N_1 = \frac{1}{h^2}\begin{pmatrix} 2-h^2 & & & & \\ -2 & 1-h^2 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1-h^2 \end{pmatrix}, \quad N_2 = \frac{1}{h^2}\begin{pmatrix} 2+h^2 & & & & \\ -2 & 1+h^2 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1+h^2 \end{pmatrix},$$

and $q = -\frac{1}{h^2}\begin{pmatrix} 8-10h \\ -4 \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} 2+h \\ 2+2h \\ \vdots \\ 2+Nh \end{pmatrix}$.

N_1 is invertible, then the problem (5.6.4) is reduced to a standard LCP.

We compare the obtained solution by Soft-LCP and TLCP to the predefined Runge-Kutta ode45 function in Matlab [78]. The domain is $t \in [0, 5]$, initial conditions $x(0) = -4, x^{'}(0) = 5$ and $N = 100$.

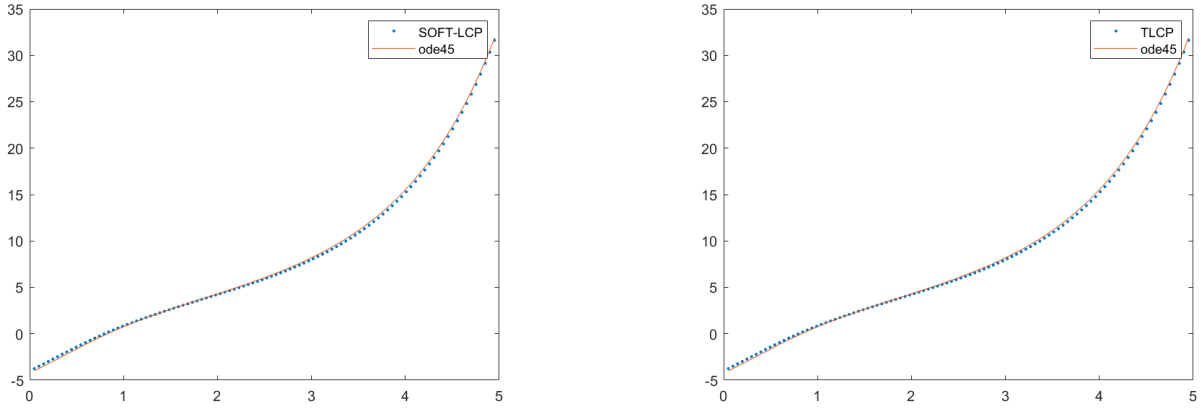

Figure 5.6: Numerical solution of (5.6.3) with ode45 and both methods.

Both methods solve the problem and gives consistent results.

5.6.4 Application to Absolute Value Equation

We consider the absolute value equation (AVE), defined as

$$Ax - |x| = b, \tag{5.6.5}$$

with $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. We studied two cases where AVE has a unique solution and for general AVE. Using the same technique as in [2], (5.6.5) can be cast as the following complementarity problem

$$A(x^+ - x^-) - (x^+ + x^-) = b, \qquad 0 \leqslant x^+ \perp x^- \geqslant 0, \tag{5.6.6}$$

equivalent to

$$(A - I)x^+ = (A + I)x^- + b, \quad 0 \leqslant x^+ \perp x^- \geqslant 0, \tag{5.6.7}$$

where $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. This decompsition guarantes that $|x| = x^+ + x^-$. So AVE can be cast as the following LCP

$$x^+ = Mx^- + q, \ 0 \leqslant x^+ \perp x^- \geqslant 0, \tag{5.6.8}$$

with $M = (A - I)^{-1}(A + I)$ and $q = (A - I)^{-1}b$.

5.6.4.1 Random uniquely solvable generated problem

We consider the special case where AVE is uniquely solvable, to guarantee the convergence of the Newton method. One way to generate such AVE is to generate a matrix A with singular values exceeding 1. We first chose a random A from a uniform distribution on $[-10, 10]$, then we chose a random x from a uniform distribution on $[-1, 1]$. Finally we computed $b = Ax - |x|$. We ensured that the singular values of each A exceeded 1 by actually computing the minimum singular value and rescaling A by dividing it by the minimum singular value multiplied by a random number in the interval $[0, 1]$. We generate like [77] for the several values for $n = 32,\ 64,\ 128,\ 256,\ 512,\ 1024$, the data $(A,\ b)$ by the following Matlab code in order to have a solution for AVE:

```
n=input('dimension of matrix A=');
R=10*(rand(n,n)-rand(n,n));
A=R/(min(svd(R))*rand(1));
x=rand(n,1)-rand(n,1);
b=A*x-abs(x).
```

The required precision for solving AVE is 10^{-6}. For each n we consider 100 instances.
Now, we compare our methods Soft-LCP and TLCP to Generalized Newton method from [75], which is denoted GN. In this method, we solve each iteration a linear system:

$$(A - D(x^i))x^{i+1} = b, \tag{5.6.9}$$

where $D(x^i) = \text{diag}(\text{sign}(x^i))$. Results are summarized in Table 5.6, which gives the number of iterations, the time required to solve all the 100 instances. Our methods solve all 100 AVEs to an accuracy of 10^{-6} and validate our approach. We notice that the GN method is the fastest because at each iteration it solves only one linear system, the TLCP method gives the fewest iterations to solve the 100 instances.

Table 5.6: Comparison of Soft-LCP and TLCP with GN method, in the case with singular values of A exceeds 1 for 100 randomly generated AVE of size n.

n	it-Soft-LCP	Time-Soft-LCP	it-TLCP	Time-TLCP	it-GN	Time-GN
32	201	0.0394	104	0.0238	255	0.0071
64	201	0.1041	107	0.0646	274	0.0182
128	200	0.2844	111	0.1767	274	0.0641
256	212	1.6986	106	0.9727	290	0.2301
512	284	11.0947	110	5.0497	295	1.2925
1024	284	42.1565	111	45.3930	291	14.8541

5.6.4.2 Random generated problem

Now we present results for general AVE, which is the main interest of our algorithm. The data are generated like [74] for several n and for several values of the parameteres, in each situation we solve 100 instances of the problem. We choose a random A from a uniform distributin on $[-10, 10]$, then chose a random x from a uniform distribution on $[-1, 1]$ and set $b = Ax - |x|$. The data (A, b) are generated by Matlab script:

```
n=input('dimension of matrix A=');
rand('state',0);
A=10*(rand(n,n)-rand(n,n));
x=rand(n,1)-rand(n,1);
b=A*x-abs(x);
```

We will compare 4 methods valid for general AVE:

- TLCP method from Algorithm 1;
- Soft-LCP method from Algorithm 2;
- Concave minimization method CMM from [74];
- Successive linear programming method LPM from [76];

In Table 5.7-5.10, "nnztot" gives the number of violated expressions for all problems, "nnzx" gives the maximum violated expressions for one problem, "nb-iter" gives the number of iteration for all the problems. We also provide the time in seconds and the number of problems where we did not manage to solve AVE.

Table 5.7: Results from TLCP on with 100 consecutive random AVEs.

n	nnztot	nnzx	nb-iter	time	nb-failure
32	3	1	1647	0.6274	3
64	5	1	1776	1.2548	5
128	5	1	2359	2.4182	7
256	8	1	2448	22.8817	8

Table 5.8: Results from Soft-LCP on with 100 consecutive random AVEs.

n	nnztot	nnzx	nb-iter	time	nb-failure
32	1	1	960	0.4287	1
64	1	1	1032	0.8351	1
128	3	1	1478	1.6692	3
256	1	1	1996	18.3965	1

Table 5.9: Results from CMM on with 100 consecutive random AVEs.

n	nnztot	nnzx	nb-iter	time	nb-failure
32	13	1	640	4.2832	13
64	11	1	588	7.0034	11
128	13	1	693	19.9940	13
256	15	1	753	143.6931	15

Table 5.10: Results from LPM on with 100 consecutive random AVEs.

n	nnztot	nnzx	nb-iter	time	nb-failure
32	8	1	313	2.2422	8
64	19	4	411	6.0978	18
128	21	3	433	18.2642	20
256	29	5	606	156.4612	22

In every cases our methods manage to reduce the number of unsolved problem, which was our principal aim. In every case it gives the smallest number of unsolved problem in a very reasonable time.

5.7 Conclusion

In this chapter, we propose two methods to solve the LCP. A complete analysis is provided to validate our approach. Furthermore, a numerical study shows that our approach is interesting. Numerical experiments on several LCP problems and a comparison with some existing methods proves the efficiency of our study.
We have presented an application of absolute value equation (AVE) and two examples (an obstacle problem and ODE) and show that our two methods are promising.

6 General conclusion and perspectives

This manuscript presents different regularisation/relaxation methods for complementarity problems, and optimal control problems under complementarity constraints. In particular, in Section 2.5 we introduce a family of smoothing functions that are central in the regularisations proposed through this document.

In Chapter 3, we propose a numerical investigation of optimal control problems governed by semi-linear elliptic variational inequalities with state constraints. To obtain the optimality system of the underlying problem, we first relaxed the feasible domain and then applied some mathematical programming and penalization techniques. We reported on several numerical experiments using various optimization platforms, and solvers such as KNITRO, IPOPT, and SNOPT to illustrate the efficiency of the proposed numerical scheme.
In Chapter 4, we used again the smoothing functions to propose a regularisation technique for NCPs. We have developed a smoothing method to solve NCPs involving $\mathcal{P}_0$-functions, and proposed a "non-parametric" algorithm to solve the nonlinear equations based on the (semi-smooth) Newton method. We performed some global and local convergence analysis of the proposed method. Then, we presented extensive numerical experiments that demonstrate the efficiency of our approach.
Thereafter, still based on smoothing techniques, we have proposed in Chapter 5 two new methods to solve LCPs. We provied a complete analysis to validate our approach. Numerical experiments on several LCP problems and comparisons to other existing methods proved that our approach is promising.

Regarding the work done in chapter 3, we would like to go further: use our optimality conditions to develop our code and conduct more extensive experiments. Concerning the LCP and NCP, we would like in future work to weaken the hypotheses to attack real problems. Several questions remain open. In particular for AVE problems: A good topic of research is the study of the AVE without any condition or assumption on the existence and uniqueness of solutions. Another interesting question is the reformulation of the nonlinear AVE ($F(x) - |x| = b$) as LCP or NCP and deducing new results concerning its solvability.

Bibliography

[1] Ampl modeling language for mathematical programming. available from: http://www.ampl.com.

[2] L. Abdallah, M. Haddou, and T. Migot. Solving absolute value equation using complementarity and smoothing functions. *Journal of Computational and Applied Mathematics*, **327**:196–207, 2018.

[3] A. Akkouche, A. Maidi, and A. Aidene. Optimal control of partial differential equations based on the variational iteration method. *Computers and Mathematics with Applications*, **68**(5):622–631, 2014.

[4] A. Auslender, R. Cominetti, and M. Haddou. Asymptotic analysis for penalty and barrier methods in convex and linear programming. *Mathematics of Operations Research*, **22**(1):43–62, 1997.

[5] A. Auslender and M. Teboulle. *Asymptotic cones and functions in optimization and variational inequalities*, volume 1. Springer Science & Business Media, 2006.

[6] I. B-Gharbia. *Résolution de problème de complémentarité application à un écoulement diphasique dans un milieu poreau.* Theses, University of Paris Dauphine, 2012.

[7] V. Barbu. *Optimal Control of Variational Inequalities.* Boston : Pitman Advanced Pub. Program, 1984.

[8] V. Barbu. *Analysis and Control of Non Linear Infinite Dimensional Systems*, volume **190**. Mathematics in Science and Engineering, Academic Press, 1993.

[9] M. Bergounioux. Optimality conditions for optimal control of elliptic problems governed by variational inequalities, 1995. Rapport de Recherche, Université d'Orléans.

[10] M. Bergounioux. Optimal control of variational inequalities: A mathematical programming approach. *International Federation for Information Processing. Springer, Boston, MA*, **34**(2):123–130, 1996.

[11] M. Bergounioux. Optimal control of an obstacle problem. *Applied Mathematics and Optimization*, **36**:147–172, 1997.

[12] M. Bergounioux. Optimal control of semilinear elliptic obstacle problems. *Journal of Nonlinear and Convex Analysis*, **3**(31):25–39, 2002.

[13] M. Bergounioux and M. Haddou. A SQP-augmented Lagrangian method for optimal control of semilinear elliptic variational inequalities. *ISNM International Series of Numerical Mathematics*, **143**:57–72, 2003.

[14] M. Bergounioux and M. Haddou. A new relaxation method for a discrete image restoration problem. *Journal of Convex Analysis*, **17**:861–883, 2010.

[15] M. Bergounioux and F. Mignot. Control of variational inequalities and Lagrange multipliers. *Control, Optimisation and Calculus of Variations*, **5**:45–70, 2000.

[16] M. Bergounioux and D. Tiba. General optimality conditions for constrained convex control problems. *SIAM Journal on Control and Optimization*, **34**(2):698–711, 1996.

[17] A. Bermudez and C. Saguez. Optimal control of variational inequalities: Optimality conditions and numerical methods, free boundary problems: Applicatons and theory. *Research Notes in Mathematics. Pitman, Boston*, pages 478–487, 1988.

[18] S. C. Billups, P. S. Dirkse, and M. C. Ferris. A comparison of large scale mixed complementarity problem solvers. *Computational Optimization and Applications*, **7**:3–25, 1997.

[19] S. I. Birbil, S. H. Fang, and J. Han. An entropic regularization approach for mathematical programs with equilibrium constraints. *Computers and Operations Research*, **31**(13):2249–2262, 2004.

[20] F. Bonnans. *Optimisation continue: cours et problèmes corrigés, Mathématiques appliquées pour le Master*. Dunod, 2006.

[21] A. P. Bourgeat, S. Granet, and F. Smaï. Compositional two-phase flow in saturated-unsaturated porous media : Benchmarks for phase appearance, disappearance. *De Gruyter*, :81–106, 2013.

[22] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge: Cambridge University Press, 2004.

[23] H. Buchholzer. *The semismooth Newton method for the solution of reactive transport problmes including mineral precipitation-dissolution reactions*. Theses, University of Wurzburg, 2011.

[24] H. Buchholzer, C. Kanzow, P. Knabner, and S. Kräutle. The semismooth Newton method for the solution of reactive transport problems including mineral precipitation-dissolution reactions. *Computational Optimization and Applications*, **50**(2):193–221, 2011.

[25] Q. Bui and H. C. Elman. Semi-smooth Newton methods for nonlinear complementarity formulation of compositional two-phase flow in porous media, 2018. Technical report.

[26] H. R. Byrd, J. Nocedal, and A. R. Waltz. Knitro: An integrated package for nonlinear optimization. *Springer, Boston, MA*, **83**:35–59, 2006.

[27] C. Chen and O. L. Mangasarian. A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications volume*, **5**:97–138, 1996.

[28] X. Chen. Smoothing methods for complementarity problems and their applications a survey. *Journal of The Operations Research Society of Japan*, **43**(1):32–47, 2000.

[29] X. Chen, B. Chen, and C. Kanzow. A penalized Fischer-Burmeister NCP-function: theoretical investigation and numerical results. *Int. fur Angewandte Mathmatik der Univ*, 1997.

[30] Y. Chen. Smoothing for nonsmooth optimization, 2019. ELE 522: Large-Scale Optimization for Data Science, Princeton University.

[31] F. H. Clarke. Generalized gradient and applications. *Transactions of the AMS*, **205**:247–262, 1975.

[32] F. H. Clarke. *Optimization and nonsmooth analysis*. Second edition, classics in applied mathematics, SIAM, Philadelphia, PA, USA, 1990.

[33] R. Cottle and G. Dantzig. Complementarity pivot theory of mathematcal programming. *Linear Algebra and its Applications*, **1**(1):103–125, 1968.

[34] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, 1992.

[35] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, **91**:201–213, 2002.

[36] L. Dong-hui and Z. Jin-ping. A penalty technique for nonlinear problems. *Journal of Computational Mathematics*, **16**(1):40–50, 1998.

[37] K. Erleben. Numerical methods for linear complementarity problems in physics-based animation, Febuary 2013. Siggraph course notes.

[38] F. Facchinei, H. Jiang, and L. Qi. A smoothing method for mathematical programs with equilibrium constraints. *Mathematical Programming*, **85**:81–106, 1995.

[39] F. Facchinei and J. S. Pang. *Finite-dimensional Variational Inequalities and Complementarity Problem*, volume **1-2**. Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[40] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems*, volume 1. Springer, 2003.

[41] S. C. Fang, J. R. Rajasekera, and H. S. Tsao. Entropic optimization and mathematical programming. *Norwell: Kluwer Academic Publishers*, 1997.

[42] M. C. Ferris and J. S. Pang. Engineering and economic applications of complementarity problems. *Mathematical Programmming*, **39**(4):669–713, 1997.

[43] A. Fischer. A special Newton-type optimizaton method. *Optimization*, **3-4**:269–284, 1992.

[44] A. Fischer. Solution of monotone complementarity problems with locally Lipschitz functions. *Mathematical Programming*, **76**(2):513–532, 1997.

[45] A. Friedman. *Variational Principles and Free-Boundray Problems*. Pure and Applied Mathematics, 1982.

[46] W. J. G. Some theoretical aspects of road traffic research. *Proceeding of the Institute of Civil Engineers*, **2**:325–378, 1952.

[47] C. Geiger and C. Kanzow. On the solution of monotone complementarity problems. *Computational Optimization and Applications*, **5**:155–173, 1995.

[48] M. E. Ghami. *New primal-dual interior-point methods based on kernel functions*. Theses, Delft university of technology, Oct 2005.

[49] I. B. Gharbia and J. Jaffré. Gas phase appearance and disappearance as a problem with complementarity constraints. *Mathematics and Computers in Simulation*, **99**:28–36, 2014.

[50] J. C. Gilbert. Mise à jour de la métrique dans les méthodes de quasi-Newton réduites en optimisation avec contraintes d'égalité. *ESAIM: Mathematical Modelling and Numerical Analysis*, **22**(2):251–288, 1988.

[51] P. E. Gill, W. Murray, M. A. Sanders, A. Drud, and E. Kalvelagen. Gams/snopt: An sqp algorithm for large-scale constrained optimization. *Computers and Operations Research*, 2000.

[52] M. Haddou. A new class of smoothing methods for mathematical programs with equilibrium constraints. *Pacific Journal of Optimization*, **5**(1):87–95, 2009.

[53] M. Haddou and P. Maheux. Smoothing methods for nonlinear complementarity problems. *Journal of Optimization Theory and Applications*, **160**:711–729, 2014.

[54] M. Haddou, T. Migot, and J. Omer. A generalized direction in interior point method for monotone linear complementarity problems. *Optimization Letters*, **13**:35–53, 2019.

[55] P. T. Harker. Accelerating the convergence of the diagonalization and projection algorithms for finite-dimensional variational inequalities. *Mathematical Programming*, **41**:29–59, 1988.

[56] P. T. Harker and J. S. Pang. Finite dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programmming*, **48**(3):161–220, 1990.

[57] J. H. He. Lagrange crisis and generalized variational principle for 3D unsteady flow. *International Journal of Numerical Methods for Heat and Fluid Flow*, **30**:1189–1196, 2020.

[58] J. H. He and H. Latifizadeh. A general numerical algorithm for nonlinear differential equations by the variational iteration method. *International Journal of Numerical Methods for Heat and Fluid Flow*, **30**:4797–4810, 2020.

[59] C. Huang and S. Wang. A power penalty approach to a nonlinear complementarity problem. *Operations research letters*, **38**(1):72–76, 2010.

[60] C. Kanzow. Some equation-based methods for the nonlinear complementarity problem. *Optimization Methods and Software*, **3**:327–340, 1994.

[61] C. Kanzow and H. Pieper. Jacobian smoothing methods for nonlinear complementarity problems. *SIAM Journal on Optimization*, **9**(2):342–373, 1999.

[62] C. Kanzow, N. Yamashita, and M. Fukushima. New NCP-functions and their properties. *Journal of Optimization Theory and Applications*, **94**:115–135, 1997.

[63] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, **4**(1):373–395, 1984.

[64] W. Karush. *Minima of Functions of Several Variables with Inequalities as Side Constraints*. Master's thesis, Dept. of mathematics, univ. of Chicago, Chicago, Illinois, 1939.

[65] M. C. Kemp and Y. Kimura. Introduction to mathematical economics. *New York: springer*, pages 38–44, 1978.

[66] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise. A unified approach to interior point algorithms for linear complementarity problems: A summary. *Operations Research Letters*, **10**(5):247–254, 1991.

[67] M. Kojima and S. Shindo. Extensions of Newton and quasi-Newton methods to systems of PC1 equations. *Journal of The Operations Research Society of Japan*, **29**:352–374, 1986.

[68] M. A. Krasnosel'skii and Y. B. Rutickii. Convex functions and Orlicz spaces. *SIAM Journal on Control and Optimization*, **5**(3):290–291, 1996.

[69] N. Krejic and S. Rapaji. Globally convergent Jacobian smoothing inexact Newton methods for NCP. *Computational Optimization and Applications*, **41**:243–261, 2008.

[70] H. W. Kuhn and A. W. Tucker. Nonlinear programming. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, 1951.

[71] C. E. Lemke. Bimatrix equilibrium points and mathematical programming. *Management science*, **11**(3):681–689, 1965.

[72] C. F. Ma, P. Nie, and G. P. Liang. A new smoothing equations approach to the nonlinear complementarity problems. *Journal of Computational Mathematics*, **21**:747–758, 2003.

[73] O. L. Mangasarian. The ill-posed linear complementarity problem. 1995.

[74] O. L. Mangasarian. Absolute value equation solution via concave minimization. *Optimization Letters*, **1**:3–8, 2007.

[75] O. L. Mangasarian. A generalized Newton method for absolute value equations. *Optimization Letters*, **3**:101–108, 2009.

[76] O. L. Mangasarian. Absolute value equation solution via linear programming. *Journal of Optimization Theory and Applications*, **161**:870–876, 2014.

[77] O. L. Mangasarian. Linear complementarity as absolute value equation solution. *Optimization Letters*, **8**:1529–1534, 2014.

[78] Matlab. Version R2020. Natick, Massachusetts: The Math Works Inc. 2019.

[79] F. Mignot and J. P. Puel. Optimal control in some variational inequalities. *SIAM Journal on Control and Optimization*, **22**(3):466–476, 1984.

[80] T. Migot. Analyse methématique de modèles géochimiques, 2014. Rapport de recherche, INRIA Rennes, équope SAGE.

[81] T. Migot. *Contributions aux méthodes numériques pour les problèmes de complémentarité et problèmes d'optimisation sous contraintes de complèmentarité*. Theses, INSA Rennes, 2017.

[82] J. J. More. Global methods for nonlinear complementarity problems. *Mathematics of Operations Research*, **21**(3):589–614, 1996.

[83] K. G. Murty and T. Y. Feng. *Linear Complementarity, Linear and Nonlinear Programming*. Berlin : Heldermann Verlag, 1988.

[84] P. Nie. A null space approach for solving nonlinear complementarity problems. *Acta Mathematicae Applicatae Sinica, English Series*, **22**(1):9–20, 2006.

[85] E. L. Osmani, M. Haddou, L. Abdallah, and N. Bensalem. New smoothing methods for solving the linear complementarity problem with $\mathcal{P}_0$-matrix. *Disponible sous archives-ouvertes.fr/hal-03516404: https://hal.archives-ouvertes.fr/hal-03516404*, 2021.

[86] E. L. Osmani, M. Haddou, and N. Bensalem. A new relaxation method for optimal control of semilinear elliptic variational inequalities obstacle problems. *Numerical Algebra, Control & Optimization*, 2021.

[87] E. L. Osmani, M. Haddou, and N. Bensalem. A new smoothing method for nonlinear complementarity problems involving $\mathcal{P}_0$-function. *Statistics, Optimization & Information Computing*, **10**(4):1267–1292, 2022.

[88] J. S. Pang. A B-differentiable equations based, globally and locally quadratically convergent algorithm for nonlinear programming, complementarity, and variational inequality problems. *Math. Programming*, **51**:101–131, 1991.

[89] J. S. Pang. *Complementarity problems. In: R. Horst, P. Pardalos, (eds.) Handbood of Global Optimization*. Kluwer Academic, Boston, 1994.

[90] J. S. Pang and S. A. Gabriel. NE-SQP: A robust algorithm for nonlinear complementarity problem. *Mathematical Programmming*, **60**:295–337, 1993.

[91] J. S. Pang and L. Qi. Nonsmooth equations: motivation and algorithms. *SIAM Journal on Optimization*, **3**:443–465, 1993.

[92] H. D. Qi and L. Z. Liao. A smoothing Newton method for nonlinear complementarity problems. *Computational Optimization and Applications*, **17**:231–253, 2000.

[93] L. Qi. A convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, **18**:227–244, 1993.

[94] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, **58**:353–368, 1993.

[95] D. Ralph and S. J. Wright. Some properties of regularization and penalization schemes for MPECs. *Optimization Methods and Software*, pages 527–556, 2004.

[96] R. T. Rockafellar. *Convex analysis*. Princeton university press, 1970.

[97] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[98] V. Ruggiero. Parallel algorithms and numerical nonlinear optimization. available from: http://dm.unife.it/pn2o/index.html. *University of Ferrara*, 2001.

[99] A. Ruszczynski. *Nonlinear Optimization*. Princeton, NJ: Princeton university press, 2006.

[100] H. Samelson, R. Thrall, and O. Wesler. A partition theorem for the euclidean n-space. *Proceedings of the American mathematical society*, **9**:805–807, 1958.

[101] S. Scholtes. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, **11**:918–936, 2001.

[102] F. Troltzsch. *Optimality conditions for parabolic control problems and applications*. 1984.

[103] J. Vignes. Implémentation des méthodes d'optimisation : Test d'arrêt optimal, contrôle et précision de la solution. *Revue française d'automatique, d'informatique et de recherche opérationnelle*, **18**(1):1–8, 1984.

[104] D. T. S. Vu, I. B. Gharbia, M. Haddou, and Q. Tran. A new approach for solving nonlinear algebraic systems with complementarity conditions. application to compositional multiphase equilibrium problems. *Mathematics and Computers in Simulation*, **190**:1243–1274, 2021.

[105] G. R. Walsh. Saddle-point property of Lagrangian function. *Methods of optimization. New York: John Wiley & Sons*, pages 39–44, 1975.

[106] A. Wächter and T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical programming*, **106**:25–57, 2006.

[107] N. Yamashita and M. Fukushima. Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems. *Mathematical Programming*, **76**:469–491, 1997.

[108] A. Yassine. *Etudes adaptatives et comparatives de certains algorithmes en optimisation : implémentations effectives et applications*. Theses, Université Joseph-Fourier-Grenoble I, 1989.

[109] L. Zhang, J. Han, and Z. Huang. Superlinear quadratic one step smoothing Newton method for P0 NCP. *Acta Mathematica Sinica*, **21**:117–128, 2005.

[110] W. Zhong, Y. Min, and W. Chang. A partially smoothing Jacobian method for nonlinear complementarity problems with P0 function. *Journal of Computational and Applied Mathematics*, **286**:158–171, 2015.

[111] J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in Banach spaces. *Applied mathematics and Optimization*, **5**:49–62, 1979.

Titre : Méthodes numériques pour les problèmes de complémentarité et les problèmes de contrôle optimal sous contraintes de complémentarité.

Mots clés : Méthode de point intérieur, problème de complémentarité linéaire, problème de complémentarité non linéaire, contrôle optimal, méthodes de régularisation, convergence globale-locale.

Résumé : Les problèmes de complémentarité interviennent dans de nombreux domaines scientifiques : économie, physique, transport, théorie des jeux et mathématiques.
Dans cette thèse, on apporte plusieurs contributions théoriques, algorithmiques et numériques pour résoudre des problèmes de complémentarité et de contrôle optimal sous contraintes de complémentarité.
On s'intéresse plus particulièrement aux méthodes de régularisation pour la résolution numérique de ces deux types de problèmes, où nous avons proposé de nouvelles techniques de régularisation.

En effet, dans la première partie, nous nous sommes intéressés aux problèmes de contrôle optimal sous contraintes de complémentarité.
Nous avons étudié les problèmes de contrôle optimal régis par les inégalités variationnelles elliptiques semi-linéaires impliquant des contraintes sur la variable d'état. Nous avons présenté un nouveau schéma de régularisation pour la contrainte de complémentarité.
Nous avons prouvé l'existence de multiplicateurs de Lagrange.

Ensuite, dans la deuxième partie, nous avons étudié les problèmes de complémentarité linéaire et non linéaire en proposant de nouvelles méthodes de régularisation pour résoudre ce genre de problèmes. L'idée de ces méthodes prend inspiration de la méthode des points intérieurs.

Dans ce travail nous nous sommes concentrés sur les propriétés théoriques des algorithmes et leurs applications numériques.

Titre: Numerical methods for complementarity problems, and optimal control problems under complementarity constraints.

Key words: Interior points methods, linear complementarity problem, nonlinear complementarity problem, optimal control, regularization methods, global-local convergence.

Abstract: Complementarity problems occur in many scientific fields: economics, physics, transport, game theory, and mathematics.
In this thesis, we offer several theoretical, algorithmic, and numerical contributions to solve the complementarity problems and optimal control problems under complementarity constraints.
We are particularly interested in the regularization methods for the numerical resolution of these types of problems, we have proposed new regularization techniques.

Indeed, In the first part, we focused on optimal control problems under complementarity constraints.
We studied optimal control problems governed by semi linear elliptic variational inequalities involving constraints on the state. We presented a new regularisation schema for the complementarity constraint. We proved that Lagrange multipliers exist.

Then, in the second part, we have studied linear complementarity problems (LCPs) and nonlinear complementarity problems (NCPs) by proposing new methods of regularisation to solve these kind of problems. The idea of these methods takes inspiration from interior point methods.

Throughout this manuscript, we have focused on the theoretical properties of algorithms and their digital applications.