Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Aspects multi-niveaux des problèmes de complémentarité

Malmedy, Vincent

Award date:
2006

Link to publication

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

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

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

Download date: 02. May. 2026

FUNDP
NAMUR

Faculté des Sciences
Département de Mathématique

Rempart de la Vierge, 8
B - 5000 Namur (Belgique)

Aspects multi-niveaux des problèmes de complémentarité



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

Vincent MALMEDY

Promoteur : Philippe TOINT

Année académique 2005-2006

Aspects multi-niveaux des problèmes de complémentarité

Résumé

Ce mémoire traite de la résolution des problèmes de complémentarité non linéaire. Ceux-ci se profilent dans de nombreux domaines, lorsque parmi deux quantités positives, l'une doit être nulle dès lors que l'autre ne l'est pas. Une reformulation de ces problèmes sous forme d'équations semi-lisses puis d'un problème de minimisation différentiable est proposée grâce à une fonction de complémentarité. Le problème ainsi obtenu est résolu par une méthode de région de confiance non monotone dont la convergence locale et globale est établie. Dans un second temps, une approche multi-niveaux des problèmes de complémentarité est envisagée. Nous étudions la façon de définir une version du problème au niveau grossier, soit en discrétisant sa reformulation obtenue au moyen d'une fonction de complémentarité, soit directement, avant de procéder à cette reformulation.

Multilevel aspects of complementarity problems

Abstract

In this paper, we discuss the resolution of nonlinear complementarity problems. They emerge in several fields, when among two nonnegative quantities, one is null as soon as the other is not. A reformulation as nonsmooth equations and then, as a differentiable minimisation problem is proposed thanks to a complementarity function. Resolution is performed by a nonmonotone trust-region method, whose local and global convergence is established. In a second time, a multilevel approach of complementarity problems is considered. We study how to define a coarser version of the problem, either by discretizing the NCP-reformulation or directly, just before applying the NCP-reformulation process.

Remerciements

Je tiens avant tout à remercier particulièrement Philippe Toint pour le soutien qu'il m'a apporté et les conseils qu'il m'a prodigués tout au long de la réalisation de ce mémoire, ainsi que les professeurs et assistants du département de Mathématique des F.U.N.D.P. qui ont contribué à ma formation. Un grand merci aussi à mes parents pour m'avoir soutenu dans mes choix et dans mon travail, ainsi qu'à mes anciens professeurs, et tout spécialement à Raymond Pierret, pour avoir éveillé en moi le goût des mathématiques, pour avoir su l'entretenir et le faire croître. De chaleureux remerciements également à l'organisation de l'olympiade mathématique belge de même qu'à toute l'équipe bénévole de préparation à l'olympiade mathématique internationale, en particulier à Fabienne et Gérald Troessaert ; ils m'ont ouvert à d'autres facettes des mathématiques au travers de la résolution de problèmes.

Merci enfin aux Facultés universitaires Notre-Dame de la Paix à Namur qui m'ont accueilli pendant ces quatre dernières années et tout spécialement à son Conseil d'administration au sein duquel j'ai pu me faire une vision plus juste de la complexité d'une université dans ses missions d'enseignement, de recherche et de services à la communauté.

Table des matières

Introduction 1

Notations 3

1. Quelques notions d'analyse non lisse 4
- 1.1. Généralisation du concept de matrice jacobienne . . . 4
- 1.2. Les fonctions semi-lisses . . . 7
- 1.3. Les fonctions semi-lisse à l'ordre p . . . 12

2. Les problèmes de complémentarité 14
- 2.1. Présentation du problème . . . 14
- 2.2. Exemples de problèmes de complémentarité . . . 15
  - 2.2.1. En optimisation . . . 15
  - 2.2.2. En physique . . . 16
  - 2.2.3. En économie et en finance . . . 17
- 2.3. Piste pour la résolution . . . 18
- 2.4. Quelques fonctions de complémentarité . . . 19

3. Résolution d'équations par méthode de région de confiance non monotone 25
- 3.1. Définition du problème . . . 25
- 3.2. Méthode de type Newton avec projection . . . 27
  - 3.2.1. Présentation de l'algorithme . . . 27
  - 3.2.2. Convergence locale . . . 29
- 3.3. Méthode de région de confiance non monotone . . . 33
  - 3.3.1. Présentation de l'algorithme . . . 33
  - 3.3.2. Convergence globale . . . 39
  - 3.3.3. Convergence locale . . . 50
  - 3.3.4. Implémentation d'un pas satisfaisant la condition de décroissance . . . 56
  - 3.3.5. Application de la méthode aux problèmes de complémentarité . . . 66
  - 3.3.6. Résultats numériques . . . 69

4. Les méthodes multi-niveaux 74

5. Aspects multi-niveaux des problèmes de complémentarité **79**
5.1. Introduction 79
5.2. Formalisation des deux démarches 80
5.2.1. Opérateurs de prolongation et de restriction 80
5.2.2. Explicitation des deux démarches 81
5.3. Quelques exemples 84
5.4. Quelques réflexions 96

Conclusion **97**

Bibliographie **99**

Introduction

Bien que nous n'en ayons pas forcément conscience, une foule de situations font intervenir des paires de quantités (positives) dont chacune est nulle dès lors que l'autre ne l'est pas. La littérature parle à leur propos, de *problèmes de complémentarité*. Ils surgissent dans quasiment tous les domaines. En optimisation, ils comprennent en particulier les conditions de Karush-Kuhn-Tucker des problèmes différentiables et de ce fait, ils interviennent dans bien d'autres disciplines. En physique, ils apparaissent typiquement dans les problèmes de contact en l'absence ou, plus généralement, en présence de frottements ; de là, ils débouchent par exemple, sur la conception des prothèses en ingénierie médicale. Ils permettent aussi de modéliser les phénomènes d'élasticité et de plasticité des solides. Ils sont également à la base de l'économie mathématique, depuis ses principes de bases jusque dans des applications plus complexes telles que les options boursières ; plusieurs prix Nobel d'économie ont d'ailleurs salué les avancées faites en ce sens. La modélisation des transports peut aussi être un terrain propice à l'émergence de tels problèmes. Bref, le nombre et la diversité des applications qui peuvent être formulées par une ou plusieurs relation(s) de complémentarité sont impressionnants.

Cependant, une chose est de modéliser une situation concrète dans un formalisme mathématique de complémentarité, une autre est d'être capable de le résoudre. À ce propos, il faut noter que de nombreux problèmes de ce genre se placent dans le cadre d'espaces de dimension infinie ou en tout cas, de très grande dimension. Il n'est alors pas toujours aisé ou possible de les résoudre, ni analytiquement ni numériquement dans des temps raisonnables. En effet, il est vain de savoir calculer la solution d'un problème en un an ou même en quelques jours, dès lors qu'elle est devenue inutile à ce moment ; un exemple classique de cette préoccupation est donné par les prévisions météorologiques qui perdent tout intérêt dès qu'elles ne concernent plus l'avenir. Des méthodes de résolution rapides sont ainsi généralement indispensables pour que les résultats puissent être exploités en pratique.

Une idée lumineuse est alors apparue durant les dernières dizaines d'années : elle consiste à prendre une version grossière du problème et à utiliser l'information sur la solution qui y est disponible pour résoudre le problème originel. Comment faire pour trouver la solution à ce niveau grossier ? Il est tout à fait possible de le considérer à son tour comme un problème trop fin pour le résoudre directement ; nous pouvons ainsi recourir à un niveau moins détaillé encore

et ainsi de suite. Au final, c'est tout une série de problèmes à différents niveaux de précision, de discrétisation, que nous obtenons. Les techniques qui utilisent ainsi des versions du problème de départ à différentes échelles sont nommées *méthodes multi-niveaux.*

Notre volonté dans ce mémoire est dès lors d'aborder les problèmes de complémentarité dans une perspective multi-niveaux. Nous ne prétendons pas aboutir à ce stade de notre réflexion, à une méthode de résolution multi-niveaux pour les problèmes de complémentarité. Nous avons voulu baliser le terrain au niveau de cette imbrication des techniques multi-niveaux avec les méthodes de résolutions classiques des problèmes de complémentarité.

Pour débuter en la matière, nous avons cru utile de rassembler dans un premier chapitre tout une série de résultats d'analyse non lisse dont nous aurons besoin par la suite et qu'il aurait été inélégant et peu pratique de disséminer tout au long de notre exposé. Ensuite dans un deuxième chapitre, nous verrons plus en détail ce qu'est un problème de complémentarité ; nous développerons plusieurs exemples et indiquerons des pistes pour les résoudre. En particulier, le chapitre suivant sera consacré à l'approfondissement de l'une d'entre elles, proposée par M. Ulbrich dans [29] ; une théorie de convergence locale et globale de cette méthode sera exposée, de même que certaines précisions d'implémentation et résultats numériques. Par après, nous expliquerons davantage dans le chapitre 4 en quoi consistent les méthodes multi-niveaux. Enfin dans un dernier chapitre, nous vous proposerons une réflexion sur la façon d'appliquer ces méthodes multi-niveaux à des problèmes de complémentarité.

Notations

$\mathbb{N}$	l'ensemble des entiers positifs : $\{0, 1, 2, 3, \ldots\}$.
$\mathbb{N}_0$	l'ensemble des entiers strictement positifs : $\{1, 2, 3, \ldots\}$.
$\underline{n}$	$\stackrel{\text{def}}{=} \{1, 2, 3, \ldots, n\}$ lorsque $n \in \mathbb{N}_0$.
$[x]_i$	la i-ième composante du vecteur x ; on la notera aussi x_i si aucune confusion n'est possible avec la version du vecteur x à i-ième itération.
$x \geqslant y$	$\stackrel{\text{def}}{\Leftrightarrow} \forall\, i \in \underline{n} : x_i \geqslant y_i$ lorsque $x, y \in \mathbb{R}^n$ $(n \in \mathbb{N}_0)$.
∇f	le gradient de la fonction $f : \mathbb{R}^n \to \mathbb{R}$ $(n \in \mathbb{N}_0)$: $\left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right) \in \mathbb{R}^n \equiv \mathbb{R}^{n \times 1}$.
$\nabla^2 f$	le hessien de la fonction $f : \mathbb{R}^n \to \mathbb{R}$ $(n \in \mathbb{N}_0)$: $\left(\frac{\partial f}{\partial x_i \partial x_j}\right) \in \mathbb{R}^{n \times n}$.
f'	la matrice jacobienne de $f : \mathbb{R}^n \to \mathbb{R}^m$ $(n, m \in \mathbb{N}_0)$: $\left(\frac{\partial f_i}{\partial x_j}\right)_{\substack{i \in \underline{m} \\ j \in \underline{n}}} \in \mathbb{R}^{m \times n}$.
$\partial_B f$	le B-sous-différentiel de la fonction localement lipschitzienne $f : \mathbb{R}^n \to \mathbb{R}^m$ $(n, m \in \mathbb{N}_0)$ (voir définition 1.2, p. 4).
∂f	la matrice jacobienne généralisée de Clarke de la fonction localement lipschitzienne $f : \mathbb{R}^n \to \mathbb{R}^m$ $(n, m \in \mathbb{N}_0)$ (voir définition 1.3, p. 5).
$\mathcal{P}(X)$	l'ensemble des parties de l'ensemble X.
$\operatorname{sgn}(x)$	le signe de $x \in \mathbb{R}$: $\operatorname{sgn}(x) = \lvert x \rvert / x$ si $x \neq 0$ et $\operatorname{sgn}(x) = 0$ si $x = 0$.
x_+	$\stackrel{\text{def}}{=} \max(0, x)$.
$\langle \cdot, \cdot \rangle$	le produit scalaire usuel sur $\mathbb{R}^n$: $\langle x, y \rangle = x^T y$.
$\lVert \cdot \rVert$	la norme euclidienne : $\lVert x \rVert_2 = \sqrt{\langle x, x \rangle}$.
$\operatorname{conv}(X)$	l'enveloppe convexe de l'ensemble X.
$g(x) = o(f(x)),$ $(x \to x_0)$	fonction g telle que $\lim_{x \to x_0} \frac{g(x)}{f(x)} = 0$.
$g(x) = O(f(x)),$ $(x \to x_0)$	fonction g telle que $\exists\, M \in \mathbb{R} : \limsup_{x \to x_0} \frac{g(x)}{f(x)} \leqslant M$.
K^*	le cône dual du cône $K \subseteq \mathbb{R}^n$: $K^* \stackrel{\text{def}}{=} \left\{ d \in \mathbb{R}^n : v^T d \geqslant 0,\ \forall\, v \in K \right\}$.

Chapitre 1.

Quelques notions d'analyse non lisse

1.1. Généralisation du concept de matrice jacobienne

Considérons un ouvert non vide $U \subset \mathbb{R}^n$ et une fonction $f : U \to \mathbb{R}^m$ localement lipschitzienne sur U, i.e.

$$\forall\, x \in U,\ \exists V_x \subseteq U,\ \exists L_x \geqslant 0,\ \forall y, z \in V_x : \|f(y) - f(z)\| \leqslant L_x \|y - z\| .$$

Nous noterons D_f l'ensemble des points $x \in U$ où f est différentiable (au sens de Fréchet) et admet une matrice jacobienne $f'(x) \in \mathbb{R}^{m \times n}$.

Théorème 1.1 (Rademacher). *Soit une fonction $f : U \to \mathbb{R}^m$ localement lipschitzienne sur l'ouvert $U \subset \mathbb{R}^n$. Alors, l'ensemble $U \setminus D_f$ des points où f n'est pas différentiable est négligeable, i.e. pour tout $\epsilon > 0$, il existe une famille de boîtes $\{B_k\}$ d'hypervolume $\epsilon_k > 0$ telle que $U \setminus D_f \subseteq \bigcup_{k=1}^{\infty} B^k$ et $\sum_{k=1}^{\infty} \epsilon^k < \epsilon$; ainsi, D_f est dense dans U.*

Démonstration. Voir [12], théorème 3.1.1. □

Par le théorème de Rademacher, nous savons que D_f est dense dans U, i.e. $U \subseteq \overline{D_f}$. Par conséquent, tout point $x \in U$ est limite d'au moins une suite $\{x_k\} \subseteq D_f$. Il est donc licite de définir le concept de jacobienne limite (voir [4]) comme suit :

Définition 1.2 (Jacobienne limite). L'ensemble

$$\partial_B f(x) = \left\{ V \in \mathbb{R}^{m \times n} : \exists (x_k) \subseteq D_f : x_k \to x, f'(x_k) \to V \right\}$$

est appelé la *jacobienne limite* (aussi dénommée sous-différentiel de Bouligand, en bref B-sous-différentiel) de f en $x \in U$.

Il n'est pas forcément aisé ni de calculer ni de manipuler la jacobienne limite pour des fonctions non lisses arbitraires. De plus, elle n'est pas particulièrement utile en optimisation non lisse car elle ne permet pas d'obtenir des conditions d'optimalité. Pour remédier (du moins partiellement) à ce problème, nous introduisons la notion suivante (voir [4]) :

Définition 1.3 (Jacobienne généralisée de Clarke). La *jacobienne généralisée de Clarke* de f en $x \in U$ est l'enveloppe convexe de la jacobienne limite de f en x : $\partial f(x) = \text{conv}(\partial_B f(x))$.

Par facilité, nous parlerons souvent de jacobienne généralisée en raccourcissant la terminologie complète.

Pour mieux visualiser ce que représentent ces notions, prenons l'exemple de la fonction

$$g = \|\cdot\|_1 : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto |x| + |y| .$$

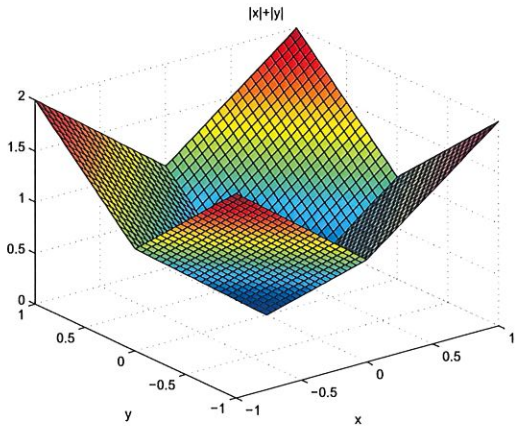


FIG. 1.1. — $g : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto |x| + |y|$.

Cette fonction est différentiable aux points (x, y) où $xy \neq 0$ ($D_g = (\mathbb{R}_0)^2$). Comme le montrera la proposition suivante, la jacobienne généralisée est simplement la jacobienne

$$g'(x, y) = \begin{pmatrix} \operatorname{sgn} x & \operatorname{sgn} y \end{pmatrix} = \partial_B g(x, y) = \partial g(x, y)$$

pour les points $(x, y) \in D_g$ où g est différentiable. Il reste donc à traiter l'origine et les points dont une seule des deux coordonnées est nulle. Commençons par ce dernier cas ; soit par exemple le point $(0, \eta)$ ($\eta > 0$). La jacobienne limite de g en $(0, \eta)$ est composée des limites de suites de matrices jacobiennes de g prises en des points $x_k \in D_g$ convergeant vers $(0, \eta)$. En conséquence, comme $\eta > 0$, à partir d'un moment les x_k vont rester dans $\mathbb{R}_0 \times \mathbb{R}_0^+$.

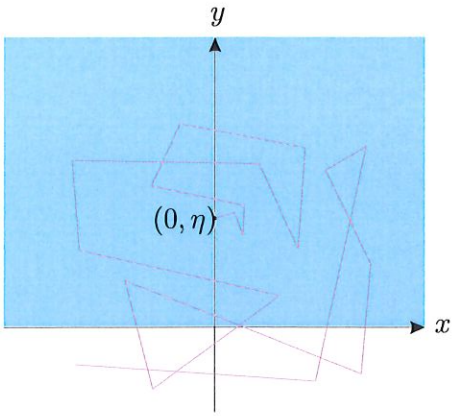


FIG. 1.2. — *La suite* (x_k) *finit par rester dans* $\mathbb{R}_0 \times \mathbb{R}_0^+$.

Or, dans cette partie du plan réel, les matrices jacobiennes de g sont soit $\begin{pmatrix} 1 & 1 \end{pmatrix}$, soit $\begin{pmatrix} -1 & 1 \end{pmatrix}$. Il est alors évident que des suites à valeurs parmi ces deux matrices ne peuvent converger que vers celles-ci ; en outre, ces deux éléments sont effectivement atteint pour $x_k = (\frac{1}{k}, \eta)$ et $x_k = (-\frac{1}{k}, \eta)$ respectivement. Nous concluons donc que la jacobienne limite de g en $(0, \eta)$ est

$$\partial_B g(0,\eta) = \left\{ \begin{pmatrix} 1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \end{pmatrix} \right\}$$

et que la jacobienne généralisée de g en ce même point est l'ensemble

$$\partial g(0,\eta) = [-1,1] \times \{1\}.$$

Par analogie, nous trouvons que

$$\partial_B g(\xi,\eta) = \begin{cases} \left\{ \begin{pmatrix} 1 & \operatorname{sgn}(\eta) \end{pmatrix}, \begin{pmatrix} -1 & \operatorname{sgn}(\eta) \end{pmatrix} \right\} & \text{si } \xi = 0 \text{ et } \eta \neq 0\,; \\ \left\{ \begin{pmatrix} \operatorname{sgn}(\xi) & 1 \end{pmatrix}, \begin{pmatrix} \operatorname{sgn}(\xi) & -1 \end{pmatrix} \right\} & \text{si } \xi \neq 0 \text{ et } \eta = 0\,; \end{cases}$$

de même, la jacobienne généralisée, qui est l'enveloppe convexe de $\partial_B g(\xi, \eta)$, est donnée par

$$\partial g(\xi,\eta) = \begin{cases} [-1,1] \times \{\operatorname{sgn}(\eta)\} & \text{si } \xi = 0 \text{ et } \eta \neq 0\,; \\ \{\operatorname{sgn}(\xi)\} \times [-1,1] & \text{si } \xi \neq 0 \text{ et } \eta = 0. \end{cases}$$

Le cas de l'origine est similaire mais au lieu d'avoir des suites pouvant prendre deux valeurs, quatre possibilités sont ici offertes : $\begin{pmatrix} \pm 1 & \pm 1 \end{pmatrix}$ (les $\pm$ sont indépendants). La jacobienne limite de g à l'origine est donc

$$\partial_B g(0,0) = \left\{ \begin{pmatrix} 1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & -1 \end{pmatrix} \right\}$$

et la jacobienne généralisée est

$$\partial g(0,0) = [-1,1]^2.$$

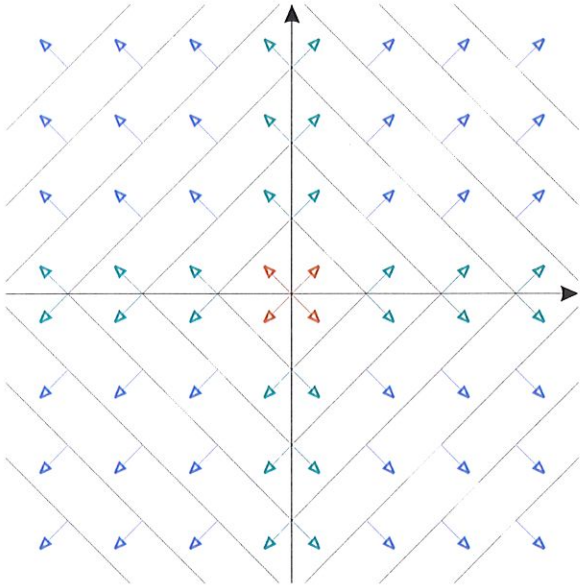


FIG. 1.3. — *Représentation de la jacobienne limite de g.*

Nous donnons à présent quelques propriétés pour pouvoir travailler avec $\partial_B f$ et ∂f :

Proposition 1.4. *Quel que soit $x \in U$, nous avons que :*

(a) *$\partial_B f(x)$ est compact et non vide ;*

(b) *$\partial f(x)$ est compact, convexe et non vide ;*

(c) *l'application $\partial_B f : U \to \mathcal{P}(\mathbb{R}^{m\times n})$ est localement bornée, i.e. pour tout $x \in U$, il existe un voisinage ouvert $\mathcal{N}$ de x tel que $\bigcup_{x\in\mathcal{N}\cap U} \partial_B f(x)$ est borné [12, p. 139] ;*

(d) *l'application d'ensembles $\partial f : U \to \mathcal{P}(\mathbb{R}^{m\times n})$ est supérieurement semi-continue (et donc fermée) i.e. pour tout $x \in U$ et pour tout voisinage ouvert $\mathcal{V}$ de $\partial f(x)$, il existe un voisinage ouvert $\mathcal{N}$ de x tel que $\forall y \in \mathcal{N} : \partial f(y) \subseteq \mathcal{V}$ [12, p. 139] ;*

(e) *si f est continûment différentiable dans un voisinage de $x \in U$, alors $\partial_B f(x) = \partial f(x) = \{f'(x)\}$.*

Démonstration. Voir [4], proposition 2.6.2. □

Les deux premières propriétés permettront de parler de limite, minimum ou maximum sur $\partial_B f(x)$ ou sur $\partial f(x)$. La cinquième montre que la jacobienne généralisée est bien une extension de la simple matrice jacobienne des fonctions de classe C^1, puisque dans le cas de ces fonctions, il n'y a pas de différence entre les concepts de jacobienne simple, limite et généralisée.

1.2. Les fonctions semi-lisses

Nous souhaitons pouvoir travailler avec des fonctions qui ne sont pas forcément continûment différentiables. Deux concepts plus faibles sont abordés pour ce faire : la différentiabilité directionnelle et le caractère semi-lisse (voir [23] et [27]).

Définition 1.5 (Différentiabilité directionnelle). La fonction f est directionnellement différentiable en $x \in U$ si les dérivées directionnelles

$$f'(x;s) = \lim_{\tau \to 0^+} \frac{f(x+\tau s) - f(x)}{\tau}$$

existent dans toutes les directions $s \in \mathbb{R}^n$.

Définition 1.6 (Semi-lisse). La fonction f est semi-lisse en $x \in U$ si elle est localement lipschitzienne en x et que la limite

$$\lim_{\substack{V \in \partial f(x+\tau d) \\ d \to s,\ \tau \to 0^+}} Vd$$

existe pour tous les $s \in \mathbb{R}^n$. Nous notons $S^0(U, \mathbb{R}^m)$, l'ensemble des fonctions $f : U \to \mathbb{R}^m$ semi-lisse sur U.

Remarquez que cette notation signifie que la limite est indépendante du choix de $V \in \partial f(x+\tau d)$ tout au long du processus de convergence.

Si nous reprenons notre exemple précédent, g est directionnellement différentiable au point $x = (0,1)$ et

$$\begin{aligned} g'((0,1)\,;s) &= \lim_{\tau \to 0^+} \frac{g(x+\tau s) - g(x)}{\tau} = \lim_{\tau \to 0^+} \frac{|\tau s_1| + |1 + \tau s_2| - 1}{\tau} \\ &= \lim_{\tau \to 0^+} |s_1| + \frac{|1+\tau s_2| - 1}{\tau} = \lim_{\tau \to 0^+} |s_1| + \frac{\tau s_2}{\tau} \\ &= |s_1| + s_2 \end{aligned}$$

où $s = (s_1, s_2)$. En effet, $1 + \tau s_2 \geqslant 0$ pour τ suffisamment petit (s_2 est fixé) et donc $|1+\tau s_2| - 1 = 1 + \tau s_2 - 1 = \tau s_2$.

La fonction g est aussi semi-lisse en $x = (0,1)$. En effet, supposons d'abord que $s_1 \neq 0$. Comme d tend vers s, il s'en suit qu'à partir d'un moment $d_1 \neq 0$ et $\operatorname{sgn} d_1 = \operatorname{sgn} s_1$; de même quand $\tau \to 0^+$, $\tau > 0$ devient suffisamment petit pour que $x + \tau d \in \mathbb{R}_0 \times \mathbb{R}_0^+ \subseteq D_f$, d'où $\partial g(x+\tau d) = \left\{ \begin{pmatrix} \operatorname{sgn}(x_1 + \tau d_1) & \operatorname{sgn}(x_2 + \tau d_2) \end{pmatrix} \right\} = \left\{ \begin{pmatrix} \operatorname{sgn} d_1 & 1 \end{pmatrix} \right\}$; la limite recherchée vaut ainsi

$$\lim Vd = \lim_{d \to s} \begin{pmatrix} \operatorname{sgn} d_1 & 1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = (\operatorname{sgn} s_1) s_1 + s_2 = |s_1| + s_2.$$

Si en revanche $s_1 = 0$, alors $d_1 \to 0$ d'où $Vs \to V_2 s_2$ avec $V = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \in \partial g(x+\tau d) \subseteq [-1,1]^2$. Comme $\tau \to 0^+$, nous avons $x_2 + \tau d_2 = 1 + \tau d_2 > 0$ pour $\tau > 0$ suffisamment petit ; de la sorte, $V_2 = \operatorname{sgn}(x_2 + \tau d_2) = 1$ et nous trouvons s_2 comme valeur de la limite. Finalement en regroupant les deux cas, nous trouvons que

$$\lim_{\substack{V \in \partial g(x+\tau d) \\ d \to s,\ \tau \to 0^+}} Vd = |s_1| + s_2$$

et donc, g est effectivement semi-lisse en $(0,1)$.

Notre attention est ici éveillée par le constat que cette limite correspond à la dérivée directionnelle. La proposition suivante et son corollaire vont nous indiquer quel rapport existe entre ces deux expressions.

Proposition 1.7. *Pour $x \in U$, les assertions suivantes sont équivalentes :*

(a) *f est semi-lisse en x ;*

(b) *f est directionnellement différentiable et*

$$\sup_{V \in \partial f(x+s)} \|Vs - f'(x;s)\| = o(\|s\|) \quad \textit{lorsque } s \to 0.$$

Démonstration. Voir [27], théorème 2.3. □

Le corollaire suivant fait le parallèle avec la propriété connue pour les fonctions de classe C^1, à savoir que $f'(x;s) = f'(x)s$. Mais dans le cadre des fonctions semi-lisses, la jacobienne généralisée n'est pas forcément réduite à un singleton et il faut donc adapter l'énoncé de cette propriété :

Corollaire 1.8. *Si f est semi-lisse en x, alors pour tout $s \in \mathbb{R}^n$,*

$$f'(x;s) = \lim_{\substack{V \in \partial f(x+\tau s) \\ \tau \to 0^+}} Vs.$$

Rappelons que cette notation signifie que la limite est indépendante du choix de $V \in \partial f(x+\tau s)$, c'est-à-dire que V peut prendre n'importe quelle valeur dans $\partial f(x+\tau s)$ à chaque instant du processus de convergence de τ vers 0^+.

Proposition 1.9. *Si f est semi-lisse en $x \in U$, alors*

$$\|f(x+s) - f(x) - f'(x;s)\| = o(\|s\|) \quad \textit{lorsque } s \to 0.$$

Démonstration. Voir [27]. □

Nous proposons à présent quelques moyens pour pouvoir déterminer assez facilement si une fonction est semi-lisse. Tout d'abord, la composée de fonctions semi-lisses l'est aussi.

Proposition 1.10. *Soient U_1 un ouvert de $\mathbb{R}^n$ et U_2 un ouvert de $\mathbb{R}^l$. Soient également les applications localement lipschitziennes $f_1 : U_1 \to U_2$ et $f_2 : U_2 \to \mathbb{R}^m$. Si f_1 est semi-lisse en $x \in U_1$ et f_2 est semi-lisse en $f_1(x) \in U_2$, alors l'application composée $f = f_2 \circ f_1 : U_1 \to \mathbb{R}^m$ est semi-lisse en x. De plus,*

$$f'(x;\cdot) = f_2'(f_1(x); f_1'(x;\cdot)).$$

Démonstration. Voir [14], lemme 18. □

Par ailleurs, une fonction est semi-lisse si et seulement s'il en est de même pour ses composantes :

Proposition 1.11 (Composantes semi-lisses). *Une fonction $f : U \to \mathbb{R}^n : x \mapsto f(x) = (f_i(x))_{i \in \underline{n}}$ est semi-lisse sur l'ouvert $U \subseteq \mathbb{R}^m$ si et seulement si ses fonctions composantes f_i ($i \in \underline{n}$) sont semi-lisses sur U. De plus, pour tout $i \in \underline{n}$, $\partial f_i(x) = e_i^T \partial f(x)$ où e_i est le i-ième vecteur de la base canonique de $\mathbb{R}^n$.*

Démonstration. Voir [30], proposition 2.10. □

Ensuite, pour les fonctions continûment différentiables, nous avons la proposition suivante :

Proposition 1.12 (Fonction continûment différentiable). *Si $f : U \to \mathbb{R}^m$ est continûment différentiable dans un voisinage de $x \in U$, alors f est semi-lisse en x.*

Démonstration. Voir [13], proposition 7.4.5 (a). □

Le résultat est identique pour les fonctions convexes :

Proposition 1.13 (Fonction convexe). *Si $f : U \to \mathbb{R}^m$ est convexe sur un voisinage de $x \in U$, alors f est semi-lisse en x.*

Démonstration. Voir [13], proposition 7.4.5 (c). □

Nous voyons ainsi que la fonction de notre exemple précédent $g(x, y) = |x| + |y|$ est semi-lisse.

Une autre classe importante de fonctions semi-lisses est celle des fonctions continûment différentiables par morceaux. Rappelons en quoi consiste cette notion (voir [12], définition 4.5.1) :

Définition 1.14 (C^k par morceaux). Soit $k \in \mathbb{N}_0$. La fonction continue $f : U \to \mathbb{R}^m$ définie sur l'ouvert $U \subseteq \mathbb{R}^n$ est k fois continûment différentiable par morceaux à proximité de $x \in U$ s'il existe un voisinage $V_x \subseteq U$ de x et une collection finie de fonctions k fois continûment différentiables $\{f_i\}_{i \in \underline{N_x}}$ allant de U dans $\mathbb{R}^m$ telles que

$$\forall y \in V_x : f(y) \in \{f_1(y), \ldots, f_{N_x}(y)\}.$$

Si f est k fois continûment différentiable par morceaux à proximité de tout $x \in U$, la fonction f est dite k fois continûment différentiable par morceaux sur U.

Notre fonction $g(x, y) = |x| + |y|$ remplit cette définition puisque

$$g(x, y) = \begin{cases} x + y & \text{si } x \geqslant 0 \text{ et } y \geqslant 0, \\ x - y & \text{si } x \geqslant 0 \text{ et } y \leqslant 0, \\ y - x & \text{si } x \leqslant 0 \text{ et } y \geqslant 0, \\ -x - y & \text{si } x \leqslant 0 \text{ et } y \leqslant 0. \end{cases}$$

Elle est donc de classe C^∞ par morceaux.

De la même façon, nous définissons les fonctions semi-lisses par morceaux ([13], p. 683) comme :

Définition 1.15 (Semi-lisse par morceaux). La fonction continue $f : U \to \mathbb{R}^m$ définie sur l'ouvert $U \subseteq \mathbb{R}^n$ est semi-lisse par morceaux à proximité de $x \in U$ s'il existe un voisinage $V_x \subseteq U$ de x et une collection finie de fonctions semi-lisses $\{f_i\}_{i \in \underline{N_x}}$ allant de U dans $\mathbb{R}^m$ telles que

$$\forall y \in V_x : f(y) \in \{f_1(y), \ldots, f_{N_x}(y)\}.$$

Si f est semi-lisse par morceaux à proximité de tout $x \in U$, la fonction f est dite semi-lisse par morceaux sur U

La proposition suivante va nous permettre de faire le lien entre les caractères continûment différentiable par morceaux et semi-lisse.

Proposition 1.16. *Si $f : U \to \mathbb{R}^m$ est semi-lisse par morceaux à proximité de $x \in U$, alors f est semi-lisse en x.*

Démonstration. Voir [13], proposition 7.4.6. □

Corollaire 1.17 (Fonction continûment différentiable par morceaux). *Si $f : U \to \mathbb{R}^m$ est continûment différentiable par morceaux sur U, alors f est semi-lisse sur U.*

Démonstration. Si nous nous plaçons en un point particulier $x_0 \in U$, il existe une collection finie de fonctions continûment différentiables $\{f_i\}_{i \in \underline{N}}$ tel que cela est affirmé par la définition 1.14. Or par la proposition 1.12, ces fonctions sont aussi semi-lisses, ce qui permet de satisfaire la définition de fonction semi-lisse par morceaux à proximité de x_0. Enfin, par la proposition précédente, nous obtenons que f est semi-lisse en x_0. Dès lors, comme x_0 était choisi arbitrairement dans U, f est semi-lisse sur U tout entier. □

En application de ce résultat, nous retrouvons le fait que $g(x, y) = |x| + |y|$ est semi-lisse en tant que fonction continûment différentiable par morceaux.

La propriété suivante de régularité est essentielle pour assurer une convergence locale rapide des méthodes de type Newton :

Définition 1.18 (BD-régulier). Le point $x \in U$ est dit *BD-régulier* pour f si tous les éléments de la jacobienne limite $\partial_B f(x)$ sont non singuliers (voir [23]).

Le résultat suivant exprime le fait que si un point est BD-régulier, tous les points suffisamment proches le sont aussi et qu'en ces points, l'inverse des éléments de la jacobienne limite est bornée en norme. En outre, si la fonction est semi-lisse, la distance entre les images de deux points est localement bornée inférieurement — et supérieurement, du fait du caractère localement lipschitzien — par un multiple de la distance séparant ces points.

Proposition 1.19. *Soit $x \in U$ un point BD-régulier pour f. Alors il existe $\epsilon > 0$ et $C > 0$ tels que pour tout y tel que $\|y - x\| \leqslant \epsilon$, tout $V \in \partial_B f(y)$ est non singulier et vérifie $\|V^{-1}\| \leqslant C$. Si de plus, f est semi-lisse en x, alors il existe $\delta > 0$ et $\zeta > 0$ tels que*

$$\|f(x) - f(y)\| \geqslant \zeta \|y - x\|$$

pour tout $y \in \mathbb{R}^n$ tel que $\|y - x\| \leqslant \delta$.

Démonstration. Voir [23], proposition 3. □

1.3. Les fonctions semi-lisse à l'ordre p

Nous parvenir à une vitesse de convergence quadratique des algorithmes qui seront présentés dans la suite de ce mémoire, nous aurons besoin d'imposer une condition un peu plus forte que le caractère semi-lisse : le caractère semi-lisse à l'ordre p où $0 < p \leqslant 1$. Il s'agit d'une relaxation semi-lisse de la différentiabilité continue au sens de Hölder, de façon semblable à ce qui a été fait précédemment pour les foncions continûment différentiable. Voyons donc auparavant ce qu'est la continuité au sens de Hölder (voir [10], notation 5.1) :

Définition 1.20 (Continuité au sens de Hölder). Soit $0 < p \leqslant 1$. La fonction continue bornée $f : U \to \mathbb{R}^m$ est continue à l'ordre p au sens de Hölder si

$$\sup_{x,y \in U,\ x \neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|^p} < +\infty.$$

Notez que le cas $p = 1$ est la continuité de Lipschitz habituelle. Nous pouvons alors définir le caractère semi-lisse à l'ordre p ($0 < p \leqslant 1$), de façon analogue au caractère semi-lisse fourni par la proposition 1.7 :

Définition 1.21 (Semi-lisse à l'ordre p). Soit $0 < p \leqslant 1$. La fonction $f : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$ est dite *semi-lisse à l'ordre p* en un point $x \in \mathbb{R}^m$ si elle est localement lipschitzienne en x, directionnellement différentiable en x et satisfait

$$\sup_{V \in \partial f(x+s)} \|Vs - f'(x,s)\| = O(\|s\|^{1+p}) \quad \text{lorsque } s \to 0.$$

Nous noterons $S^p(U, \mathbb{R}^m)$, l'ensemble des fonctions semi-lisses à l'ordre p sur U (voir [27]).

Notez la différence entre le petit « o » pour les fonctions semi-lisses, tandis qu'il s'agit d'un grand « O » pour les fonctions semi-lisses à l'ordre p.

Une fonction semi-lisse à l'ordre $p = 1$ sera dite *fortement semi-lisse.*

Proposition 1.22. *Si $0 < p \leqslant 1$ et que f est semi-lisse à l'ordre p en $x \in U$, alors*

$$\|f(x + s) - f(x) - f'(x; s)\| = O(\|s\|^{1+p}) \quad \textit{lorsque } s \to 0.$$

Démonstration. Voir [27]. □

De même que pour les fonctions semi-lisse (voir proposition 1.10), la composée de fonctions semi-lisses à l'ordre p l'est aussi.

Proposition 1.23. *Soient U_1 un ouvert de $\mathbb{R}^n$ et U_2 un ouvert de $\mathbb{R}^l$. Soient également les applications localement lipschitziennes $f_1 : U_1 \to U_2$ et $f_2 : U_2 \to \mathbb{R}^m$. Si f_1 est semi-lisse à l'ordre p avec $0 < p \leqslant 1$ en $x \in U_1$ et f_2 est semi-lisse à l'ordre p en $f_1(x) \in U_2$, alors l'application composée $f = f_2 \circ f_1 : U_1 \to \mathbb{R}^m$ est semi-lisse à l'ordre p en x. De plus,*

$$f'(x;\cdot) = f_2'(f_1(x); f_1'(x;\cdot)).$$

Démonstration. Voir [14], théorème 21. □

Nous avons la même relation entre la fonction et ses composantes dans le cas semi-lisse à l'ordre p que dans le cas semi-lisse traité par la proposition 1.11 :

Proposition 1.24 (Composantes semi-lisses à l'ordre p). *Une fonction $f : U \to \mathbb{R}^n : x \mapsto f(x) = (f_i(x))_{i\in\underline{n}}$ est semi-lisse à l'ordre p $(0 < p \leqslant 1)$ sur l'ouvert $U \subseteq \mathbb{R}^m$ si et seulement si ses fonctions composantes f_i $(i \in \underline{n})$ sont semi-lisses à l'ordre p sur U. De plus, pour tout $i \in \underline{n}$, $\partial f_i(x) = e_i^T \partial f(x)$ où e_i est le i-ième vecteur de la base canonique de $\mathbb{R}^n$.*

Le résultat suivant est l'analogue de la proposition 1.12 mais cette fois, pour les fonctions semi-lisses à l'ordre p :

Proposition 1.25. *Si f admet en x une dérivée continue à l'ordre p $(0 < p \leqslant 1)$ au sens de Hölder, alors f est semi-lisse à l'ordre p en x.*

Démonstration. Voir [13], proposition 7.4.5 (b). □

En particulier, la continuité au sens de Lipschitz étant synonyme de la continuité à l'ordre 1 au sens de Hölder, nous pouvons en conclure que toute fonction continûment différentiable possédant une dérivée lipschitzienne est semi-lisse à l'ordre 1, c'est à dire fortement semi-lisse.

Chapitre 2.

Les problèmes de complémentarité

2.1. Présentation du problème

Différents problèmes. Nous allons ici vous présenter le concept central sur lequel ce mémoire est basé : les problèmes de complémentarité. En toute généralité [12], ils se présentent comme la recherche de vecteurs $x \in \mathbb{R}^n$ vérifiant

$$\langle f(x), g(x) \rangle = 0, \quad f(x) \in K \quad \text{et} \quad g(x) \in K^*$$

où K est un cône et K^* le cône dual, tandis que f et g sont deux fonctions lisses de $\mathbb{R}^n$ dans $\mathbb{R}^n$. Dans de nombreux cas, $K = (\mathbb{R}^+)^n = K^*$. Le problème se récrit alors

$$\langle f(x), g(x) \rangle = 0, \quad f(x) \geqslant 0 \quad \text{et} \quad g(x) \geqslant 0.$$

Lorsque la seconde fonction est l'identité $g = id : x \mapsto x$, nous arrivons aux conditions

$$\langle f(x), x \rangle = 0, \quad f(x) \geqslant 0 \quad \text{et} \quad x \geqslant 0$$

qui constituent le *problème de complémentarité non linéaire.*

Dans le cas où, en plus, f est affine, i.e. $f : x \mapsto a + Mx$ ($a \in \mathbb{R}^n, M \in \mathbb{R}^{n \times n}$), nous parlerons alors d'un *problème de complémentarité linéaire* :

$$\langle a + Mx, x \rangle = 0, \quad a + Mx \geqslant 0 \quad \text{et} \quad x \geqslant 0.$$

Il existe une extension de ce cas linéaire aux problèmes formulés comme suit :

$$\langle y, x \rangle = 0, \quad Ny - Mx = a, \quad y \geqslant 0 \quad \text{et} \quad x \geqslant 0 \ ;$$

on lui donne le nom de *problème de complémentarité linéaire étendu.* Dans le cas où $N = I$, nous avons $y = a + Mx$ et donc, nous retrouvons le problème de complémentarité linéaire.

Parfois, un système d'équations supplémentaire est aussi ajouté aux conditions du problème

de complémentarité non linéaire :

$$\langle f(x,y), x\rangle = 0, \quad e(x,y) = 0, \quad f(x,y) \geqslant 0 \quad \text{et} \quad x \geqslant 0$$

où $x \in \mathbb{R}^{n_1}$, $y \in \mathbb{R}^{n_2}$ et $f, e : \mathbb{R}^{n_1} \times (\mathbb{R}^+)^{n_2} \to \mathbb{R}^{n_1}$. Nous dirons qu'il s'agit de *problèmes de complémentarité mixtes* car ils mélangent la condition de complémentarité à un système d'équations non lié par les contraintes de positivité sur la seconde variable y.

Adéquation de la dénomination. Le terme *complémentarité* se réfère à l'équation

$$\langle f(x), g(x)\rangle = 0$$

qui se développe, à cause des contraintes de positivités, en

$$f_i(x)g_i(x) = 0, \quad \forall i \in \underline{n}.$$

Il y a complémentarité dans le sens que si l'un des deux facteurs est strictement positif, l'autre doit être obligatoirement nul. On retrouve cela dans les conditions de Karush-Kuhn-Tucker développée ci-après.

2.2. Exemples de problèmes de complémentarité

Il est intéressant de savoir résoudre les problèmes de complémentarité car ils forment une classe de problèmes très large et dont on peut trouver des applications dans de nombreux domaines. Nous n'en présenterons ici que quelques uns.

2.2.1. En optimisation

De tels problèmes surgissent fréquemment en optimisation ; songeons par exemple aux conditions d'optimalité de Karush-Kuhn-Tucker pour la recherche d'extrema d'une fonction objectif non linéaire avec contraintes de positivité :

$$\begin{cases} \min & f(x) \\ \text{s.c.} & c(x) \geqslant 0 \end{cases}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ et $c : \mathbb{R}^m \to \mathbb{R}^n : x \mapsto (c_i(x))_{i\in\underline{n}}$. Les conditions de Karush-Kuhn-Tucker sont alors :

$$\begin{cases} \nabla f(x) &= \sum_{i=1}^m \lambda_i \nabla c_i(x) = C^T(x)\lambda \\ \langle c(x), \lambda\rangle &= 0, \quad c(x) \geqslant 0, \quad \lambda \geqslant 0 \end{cases}$$

avec $\lambda \in \mathbb{R}^m$ et $C(x) = c'(x)$ est la matrice jacobienne de la fonction c. Il s'agit d'un problème d'un problème de complémentarité mixte où le couple (x, y) de la définition est ici remplacé par

(λ, x) et où les fonctions $f(x, y)$ et $e(x, y)$ sont devenues $c(x)$ et $\nabla f(x) - C^T\lambda$ respectivement.

2.2.2. En physique

Comme toujours, la physique regorge d'exemples pour les concepts mathématiques. Les problèmes de complémentarité permettent ainsi de modéliser notamment des problèmes de contact avec friction ou encore des problèmes d'élasticité/plasticité de matériaux (voir [12] pour un développement de ce second exemple).

Problèmes de contact avec friction. Les problèmes de contact consistent à prédire le mouvement de plusieurs corps qui peuvent entrer en contact. Ce problème a de nombreuses répercussions dans l'imagerie virtuelle ou la robotique. Dans ce dernier domaine, le phénomène de friction joue un rôle particulièrement important : c'est par exemple grâce à lui qu'un bras-robot peut saisir une charge à déplacer.

Nous trouvons de façon naturelle une relation de complémentarité (les *conditions de contact de Signorini*) entre d'une part la force de contact normale et d'autre part, l'éloignement des objets considérés. Elles représentent en effet toujours des quantités positives et dont chacune est nulle si l'autre ne l'est pas.

Physiquement, cela veut dire que si les objets ne se touchent pas, il n'y a pas de force de contact et inversement, que si la force de contact est non nulle, les objets sont obligatoirement en contact. La positivité de la force de contact normale correspond au fait qu'il ne peut s'agir que d'une force de compression sur l'autre objet en contact. De même, la positivité de l'éloignement des objets signifie qu'ils ne peuvent s'interpénétrer. Il ne s'agit pas forcément d'une distance au sens classique mais simplement d'une fonction qui est négative si les objets s'interpénètrent, nulle s'ils se touchent seulement et positive s'ils ne sont pas en contact.

Si nous prenons l'exemple d'une sphère (repérée dans l'espace par son centre (x_0, y_0, z_0)) et d'un plan (supposé fixe avec pour équation $z = 0$), l'éloignement entre ces deux corps est alors $|z_0| - R$ où R est le rayon de la sphère.

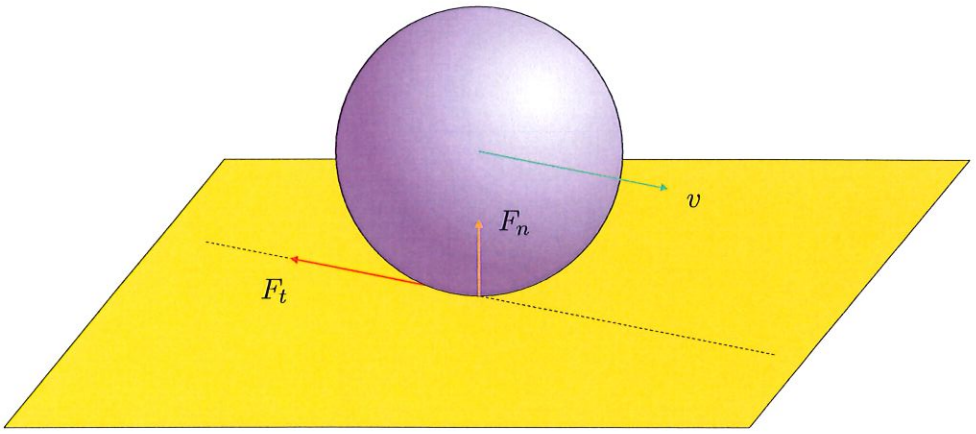


FIG. 2.1. — *Illustration d'un problème de contact : une sphère roulant sur un plan.*

Les phénomènes de friction font intervenir, en sus des forces normales F_n, des forces de contacts tangentielles F_t. Ensemble, elles doivent satisfaire le principe de dissipation maximale de l'énergie. Le problème d'optimisation qui en résulte (développé dans [2, 12]) peut aussi être reformulé comme un problème de complémentarité mixte. Un cas particulier de ce principe est la loi de friction de Coulomb qui indique que la force de frottement est en sens opposé du glissement (v).

Les deux problèmes de complémentarité ainsi obtenus doivent encore être complétés par la troisième loi de Newton qui indique que la somme des forces (extérieures et de contact) agissant sur un corps est égale au produit de la masse de celui-ci et de son accélération. Au final, le système formés par ces corps éventuellement en contact se traduit mathématiquement par un problème de complémentarité linéaire mixte [1].

2.2.3. En économie et en finance

Équilibre walrasien. En 1874, l'économiste français Léon Walras a mis au jour un problème de base en économie mathématique dans un contexte de pur échange et qui porte aujourd'hui son nom. Dans les années 1950, Arrow et Debreu y ont inclu les composantes de production et de consommation. Les travaux de Debreu furent même récompensés par le prix Nobel d'économie en 1983, en remerciement de son rôle de pionnier dans l'économie mathématique. La présentation que nous donnons ici du problème d'équilibre walrasien est celle de L. Mathiesen (reprise dans [12]).

Considérons m activités économiques et n biens de consommation. Le coût unitaire de réalisation de l'activité $i \in \underline{m}$ est c_i et la quantité initiale du bien $j \in \underline{n}$ est b_j. Les inconnues sont les volume d'activité y_i et les prix de vente des biens p_j. La fonction de demande du bien j notée $d_j(p)$ est une fonction des prix des différents biens. On note encore par $A(p)$ la matrice de production : $a_{ij}(p)$ représente la quantité de bien j produite par l'activité i lorsque le vecteur de prix est égal à p. Ainsi, $A(p)^T y$ donne la quantité de chaque bien produite par l'ensemble des activités économiques et $A(p)p$ indique le chiffre d'affaire unitaire de chacune des activités. Nous dirons qu'une paire d'activités et de prix (y, p) est un équilibre si les deux conditions suivantes sont vérifiées :

$$\langle y, c - A(p)p \rangle\,, \quad c - A(p)p \geqslant 0 \quad \text{et} \quad y \geqslant 0\,;$$

$$\big\langle p, b + A(p)^T y - d(p) \big\rangle\,, \quad b + A(p)^T y - d(p) \geqslant 0 \quad \text{et} \quad p \geqslant 0.$$

La première condition signifie que le volume des activités économiques est positif et que la perte (coûts − recettes $= c - A(p)p$) est positive. Toutefois, les activités non rentable (i.e. qui génère une perte strictement positive) ne sont pas effectuée ($y_i = 0$) ; inversement, si le volume d'activité est strictement positif, alors la perte générée par cette activité est nulle, c'est-à-dire qu'il y a équilibre. La seconde condition signifie que les prix sont positifs et que l'offre $(b + A(p)^T y)$ doit être suffisante pour satisfaire la demande $d(p)$. De plus, si le prix est strictement positif, alors l'offre correspond exactement à la demande ; inversement, si l'offre est

strictement supérieure à la demande, cela implique que le bien considéré est gratuit.

Nous avons ainsi abouti à un problème de complémentarité non linéaire avec

$$x = (y, p) \quad \text{et} \quad f(x) = (c - A(p)p, b + A(p)^T y - d(p)).$$

Pour facilité la compréhension de ce problème, nous avons explicité ce que peuvent représenter les différentes variables dans un cas concret. Par exemple, nous pouvons prendre comme activités économiques, m exploitations agricoles et comme biens de consommation, n produits agricoles. Le volume d'activité peut prendre la forme du nombre d'hectares cultivés y_i, avec un coût de culture à l'hectare constant qui vaut c_i. Les inconnues sont ainsi la superficie cultivée par chaque agriculteur i ainsi que les prix p_j des productions. Toutefois, chaque exploitant ne produit pas forcément de tout ni dans les mêmes quantités ; la matrice $A(p)$ reprend cette information : $a_{ij}(p)$ indique la production de l'agriculteur i pour le bien j lorsque les prix sont fixés à p.

Modélisation des options en bourse. En bourse, il est intéressant de pouvoir décrire les variations du cours des produits financiers ; en particulier, les options[1] sont des placements très risqués pour lequel il est évidemment utile de connaître la façon dont ils fluctuent. En 1973, Black et Scholes ont ainsi modélisé le comportement des options européennes, ce qui a valu le prix Nobel d'économie à Scholes en 1997.

Dans [12], Facchinei et Pang formulent le cas (plus compliqué) des options américaines sous forme d'un problème de complémentarité. Une méthode de résolution multi-grilles du problème des options américaines est en outre présentée dans [22] par Oosterlee.

2.3. Piste pour la résolution

Il faut distinguer les problèmes linéaires et ceux qui ne le sont pas. En effet, pour les premiers, les méthodes de résolution connues sont plus nombreuses : outre les techniques non linéaires qui leur sont évidemment applicables, il est possible d'utiliser des algorithmes tels que celui de Lemke (qui converge en un nombre fini d'itérations ; voir [7]) ou le logiciel PATH de Dirkse et Ferris [9].

En ce qui concerne les problèmes non linéaires, la plupart des méthodes existante (voir par exemple [6]) passent par une *fonction (mérite) de complémentarité* $\phi : \mathbb{R}^2 \to \mathbb{R}$ qui satisfait la condition :

$$\phi(x, y) = 0 \quad \Leftrightarrow \quad x \geqslant 0,\ y \geqslant 0,\ xy = 0$$

Admettons pour l'instant qu'une telle fonction existe et utilisons-la pour reformuler le problème

[1] Une option est un instrument financier octroyant à son propriétaire le droit (et non l'obligation) de vendre (*put*) ou d'acheter (*call*) une action à un cours déterminé endéans une période déterminée (option américaine) ou bien à une date fixée (option européenne).

de complémentarité. Avec $f = (f_i)_{i\in\underline{n}}$ et $g = (g_i)_{i\in\underline{n}}$, nous construisons la fonction

$$\Phi = (\Phi_i)_{i\in\underline{n}} : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto (\phi(f_i(x), g_i(x)))_{i\in\underline{n}},$$

de sorte que le problème se récrit de façon équivalente comme

$$\begin{aligned} & \Phi(x) = 0 \\ \Leftrightarrow \quad & \phi(f_i(x), g_i(x)) = 0, \quad \forall\, i \in \underline{n} \\ \Leftrightarrow \quad & f_i(x)g_i(x) = 0,\ f_i(x) \geqslant 0,\ g_i(x) \geqslant 0, \quad \forall\, i \in \underline{n} \\ \Leftrightarrow \quad & \langle f(x), g(x)\rangle = 0,\ f(x) \geqslant 0,\ g(x) \geqslant 0. \end{aligned}$$

Ainsi, il sera possible d'appliquer à ce problème les techniques connues pour la résolution des systèmes d'équations. Mais encore faut-il que la fonction Φ remplisse certaines caractéristiques de régularité pour satisfaire les hypothèses imposées par ces méthodes de résolution. Le choix de la fonction de complémentarité a ainsi toute son importance car elle va déterminer (en partie) les propriétés de régularité de la fonction Φ dont nous cherchons les zéros.

2.4. Quelques fonctions de complémentarité

Une première possibilité pour ϕ est la fonction minimum :

$$\min : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto \min(x, y).$$

Elle remplit bien les conditions de la définition puisque si $\min(x, y) = 0$, alors $x = 0$ ou $y = 0$ donc $xy = 0$, et de plus $0 = \min(x, y) \leqslant x, y$; si réciproquement $x, y \geqslant 0$ et $xy = 0$, c'est-à-dire que $x = 0$ ou $y = 0$, alors $\min(x, y) = 0$. Cette fonction est en outre concave, continûment différentiable sauf sur la première bissectrice du plan $\{(x, x) : x \in \mathbb{R}\} \subset \mathbb{R}^2$ et dont le carré $\min^2$ est continûment différentiable sauf sur cette même droite.

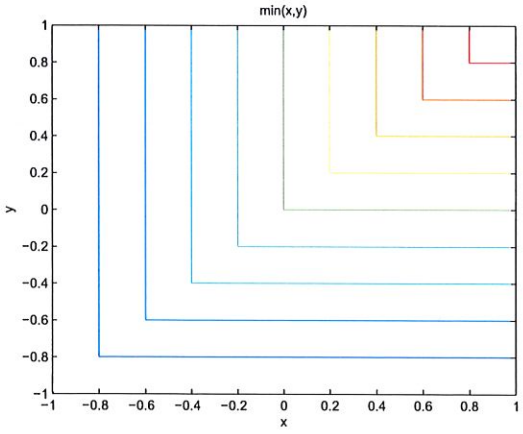


FIG. 2.2. — *La fonction minimum.*

Le fait que le carré de la fonction soit de classe C^1 serait pourtant fort utile car cela permettrait que le carré de la norme (euclidienne) de Φ soit continûment différentiable pour peu qu'il en soit de même pour f et g (puisque $\|\Phi\|^2 = \sum \phi^2(f,g)$). Nous souhaiterions garantir cette propriété afin de pouvoir minimiser le carré de la norme de Φ par des méthodes avec dérivées connues (du type Newton) et ainsi trouver des candidats-solutions de l'équation $\Phi(x) = 0$.

En 1976, Mangasarian [19] a proposé un large choix de fonctions de complémentarité, en démontrant la proposition suivante :

Proposition 2.1. *Soit $\zeta : \mathbb{R} \to \mathbb{R}$ une fonction strictement croissante avec $\zeta(0) = 0$. Alors,*

$$\phi_{Man} : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto \zeta(|a-b|) - \zeta(b) - \zeta(a)$$

est une fonction de complémentarité.

Démonstration. Voir [12], page 74. □

En particulier, lorsque ζ est l'identité sur $\mathbb{R}$, nous trouvons $-2\min(x,y)$ qui est à un facteur près la fonction de complémentarité proposée précédemment. L'avantage de ce résultat est qu'en choisissant correctement la fonction ζ, nous pouvons obtenir des fonctions ϕ_{Man} aussi différentiables que souhaité. Par exemple, en prenant $\zeta : t \mapsto t\,|t|$, la fonction ϕ_{Man} est de classe C^1 ; ou encore, avec $\zeta : t \mapsto t^3$, cette fonction est deux fois continûment différentiable.

Néanmoins, les approches les plus modernes préfèrent utiliser des reformulations non continûment différentiables, mais seulement semi-lisses [30]. La raison en est que lorsque l'on fait le choix d'une reformulation continûment différentiable, alors le gradient de la fonction mérite est donné par :

$$\nabla\Phi(x) = \operatorname{diag}\left(\frac{\partial\phi}{\partial g}(g_i(x),h_i(x))\right)\nabla g(x) + \operatorname{diag}\left(\frac{\partial\phi}{\partial h}(g_i(x),h_i(x))\right)\nabla h(x).$$

Or, comme $\phi(t,0) = 0 = \phi(0,t)$ pour tout $t \geqslant 0$, nous avons que $\nabla\phi(0,0) = (0,0)$. Par conséquent, si la condition de complémentarité stricte est violée pour la i-ième composante, i.e. $g_i(x) = 0 = h_i(x)$, alors la i-ième composante du gradient de Φ est nulle et donc la méthode de Newton n'est pas applicable dans ce cas. Au contraire, avec des reformulations semi-lisses non continûment différentiables, il sera possible de garantir que tout élément du gradient généralisé soit en dehors d'un voisinage de 0 en chaque point du plan $\mathbb{R}^2$. Il n'y aura alors pas besoin d'une hypothèse de complémentarité stricte comme dans le cas de reformulations de classe C^1.

Une fonction couramment utilisée dans ce cadre est celle de Fischer-Burmeister :

$$\phi_{FB} : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto \sqrt{x^2+y^2} - x - y.$$

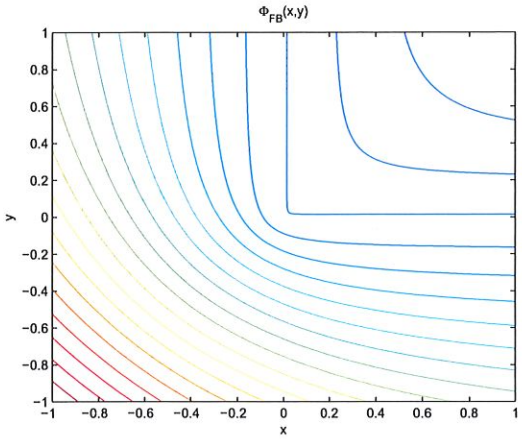


FIG. 2.3. — *La fonction de Fischer-Burmeister.*

Proposition 2.2. *La fonction de Fischer-Burmeister ϕ_{FB} est une fonction de complémentarité continûment différentiable sauf à l'origine, convexe et dont le carré ϕ_{FB}^2 est partout continûment différentiable. Cette fonction est en outre fortement semi-lisse avec $\|V\| \geqslant \sqrt{2} - 1$ pour tout $V \in \partial\phi_{FB}(x)$ et pour tout $x \in \mathbb{R}^2$*

Démonstration.

(i) Si $x, y \geqslant 0$ et $xy = 0$, alors x ou y est nul. Disons qu'il s'agit de x ; alors $\phi_{FB}(x, y) = \sqrt{y^2} - y = 0$ car $y \geqslant 0$. Réciproquement, si $\phi_{FB}(x, y) = 0$, $x + y = \sqrt{x^2 + y^2}$ et en prenant le carré des deux côtés, nous trouvons $x^2 + y^2 + 2xy = x^2 + y^2$, d'où $xy = 0$. Il suit que soit x soit y est nul ; admettons qu'il s'agît de x. Dès lors, $y = \sqrt{y^2} = |y|$ et par conséquent $x, y \geqslant 0$.

(ii) Par la définition de ϕ_{FB}, nous trouvons directement que pour $(x, y) \neq (0, 0)$,

$$\phi'_{FB}(x, y) = \left(\frac{x}{\sqrt{x^2+y^2}} - 1 \quad \frac{y}{\sqrt{x^2+y^2}} - 1\right)$$

et que cette dérivée est continue sur le plan $\mathbb{R}^2$ privé de l'origine.

(iii) Pour montrer que la fonction de Fischer-Burmeister est convexe, il faut prouver que

$$\forall\, t \in [0, 1],\ \forall\, a, b \in \mathbb{R}^2 : t\phi_{FB}(a) + (1 - t)\phi_{FB}(b) \geqslant \phi_{FB}(ta + (1 - t)b).$$

Toutefois, comme ϕ_{FB} est continue sur le plan réel, il suffit de prouver cette inégalité

pour $t = \frac{1}{2}$ [17]. L'expression se récrit alors :

$$\frac{1}{2}\left(\sqrt{a_1^2+a_2^2} - a_1 - a_2 + \sqrt{b_1^2+b_2^2} - b_1 - b_2\right) \geqslant \sqrt{\left(\tfrac{a_1+b_1}{2}\right)^2 + \left(\tfrac{a_2+b_2}{2}\right)^2} - \tfrac{a_1+b_1}{2} - \tfrac{a_2+b_2}{2}$$

$$\Leftrightarrow \quad \sqrt{a_1^2+a_2^2} + \sqrt{b_1^2+b_2^2} \geqslant \sqrt{(a_1+b_1)^2+(a_2+b_2)^2}$$

$$\Leftrightarrow \quad \|a\| + \|b\| \geqslant \|a+b\|$$

qui est vraie par l'inégalité triangulaire avec $a = (a_1, a_2)$ et $b = (b_1, b_2)$.

(iv) En ce qui concerne ϕ_{FB}^2, elle est continûment différentiable avec comme dérivée

$$(\phi_{FB}^2)'(x,y) = 2\phi_{FB}(x,y)\left(\tfrac{x}{\sqrt{x^2+y^2}} - 1 \quad \tfrac{y}{\sqrt{x^2+y^2}} - 1\right)$$

sauf en $(x,y) = (0,0)$. À l'origine, cette dérivée $(\phi_{FB}^2)'(0,0)$ vaut $\begin{pmatrix} 0 & 0 \end{pmatrix}$, c'est-à-dire que

$$\lim_{s\to(0,0)} \frac{\phi_{FB}^2(0+s_1, 0+s_2) - \phi_{FB}^2(0,0) - \langle 0, s\rangle}{\|s\|} = 0$$

où $s = (s_1, s_2)$. Cette limite vaut en effet :

$$\begin{aligned}
\lim_{s\to(0,0)} \frac{\phi_{FB}^2(s_1,s_2)}{\|s\|} &= \lim_{s\to(0,0)} \frac{(\sqrt{s_1^2+s_2^2} - s_1 - s_2)^2}{\sqrt{s_1^2+s_2^2}} \\
&= \lim_{s\to(0,0)} \sqrt{s_1^2+s_2^2} - 2(s_1+s_2) + \frac{(s_1+s_2)^2}{\sqrt{s_1^2+s_2^2}} \\
&= \lim_{s\to(0,0)} \sqrt{s_1^2+s_2^2} - 2(s_1+s_2) + \frac{s_1^2+s_2^2}{\sqrt{s_1^2+s_2^2}} + \frac{2s_1s_2}{\sqrt{s_1^2+s_2^2}} \\
&= \lim_{s\to(0,0)} 2\phi_{FB}(s_1,s_2) + \frac{2s_1s_2}{\sqrt{s_1^2+s_2^2}} \\
&= \lim_{s\to(0,0)} \frac{2s_1s_2}{\sqrt{s_1^2+s_2^2}} = 0,
\end{aligned}$$

car par la positivité du carré $(|s_1| - |s_2|)^2 = s_1^2 + s_2^2 - 2|s_1s_2| \geqslant 0$, nous avons

$$0 \leqslant \left|\frac{2s_1s_2}{\sqrt{s_1^2+s_2^2}}\right| \leqslant \frac{2|s_1s_2|}{\sqrt{2|s_1s_2|}} = \sqrt{2|s_1s_2|} \to 0$$

lorsque s tend vers 0.

Posons $r = \sqrt{x^2+y^2}$ et vérifions que $(\phi_{FB}^2)'(x,y)$ est continue en 0 :

$$\begin{aligned}
\lim_{(x,y)\to(0,0)} [(\phi_{FB}^2)'(x,y)]_1 &= \lim_{(x,y)\to(0,0)} 2\left(\sqrt{x^2+y^2} - x - y\right)\left(\frac{x}{\sqrt{x^2+y^2}} - 1\right) \\
&= \lim_{(x,y)\to(0,0)} 2r^{-1}(r-x-y)(x-r) \\
&= 2 \cdot \lim_{(x,y)\to(0,0)} \frac{rx - xx - yx}{r} - (r-x-y) \\
&= 2 \cdot \lim_{(x,y)\to(0,0)} -x\frac{x+y}{\sqrt{x^2+y^2}} = 0.
\end{aligned}$$

En effet, par l'inégalité des moyennes[2] (cas $1 < 2$),

$$\frac{|x|+|y|}{2} \leqslant \sqrt{\frac{x^2+y^2}{2}} \quad \Leftrightarrow \quad \frac{|x|+|y|}{\sqrt{x^2+y^2}} \leqslant \sqrt{2}.$$

et par conséquent,

$$0 \leqslant \left| -x\frac{x+y}{\sqrt{x^2+y^2}} \right| \leqslant |x| \frac{|x|+|y|}{\sqrt{x^2+y^2}} \leqslant \sqrt{2}\,|x| \to 0$$

lorsque $(x,y) \to (0,0)$. Le même raisonnement est applicable pour la seconde composante (il suffit de permuter x et y), d'où finalement $(\phi_{FB}^2)'(x,y) \to (0 \;\; 0)$ lorsque $(x,y) \to (0,0)$. Dès lors, ϕ_{FB}^2 est partout continûment différentiable.

(v) La fonction de Fischer-Burmeister ϕ_{FB} est fortement semi-lisse sur $\mathbb{R}_0^2$ en tant que fonction différentiable (proposition 1.12) avec un gradient lipschitzien (car de classe C^∞) sur cet ensemble. Il reste à voir que ϕ_{FB} est aussi fortement semi-lisse à l'origine. Nous pouvons calculer que :

$$\phi'_{FB}(0;d) = \lim_{t\to 0^+} \frac{\phi_{FB}(td) - \phi_{FB}(0)}{t} = \lim_{t\to 0^+} \frac{\phi_{FB}(td)}{t} = \lim_{t\to 0^+} \frac{t\phi_{FB}(d)}{t} = \phi_{FB}(d).$$

Alors par la proposition 1.22, ϕ_{FB} est fortement semi-lisse à l'origine puisque

$$\left\|\phi_{FB}(0+d) - \phi_{FB}(0) - \phi'_{FB}(0;d)\right\| = \|\phi_{FB}(d) - 0 - \phi_{FB}(d)\| = 0 = O(\|d\|^2)$$

quand $d \to 0$. Nous en concluons que ϕ_{FB} est fortement semi-lisse sur le plan réel.

(vi) Vu la définition du gradient de ϕ_{FB} sur $\mathbb{R}_0^2$, celui-ci appartient au cercle $\mathcal{C}$ de centre $(-1,-1)$ et de rayon 1. Nous constatons que le gradient de ϕ_{FB} a la même valeur sur

[2] **Inégalités des moyennes** ([17], p. 19) : Soient $r < s$ deux réels non nuls et $\{x_i\}_{i\in\underline{n}} \subset \mathbb{R}^+$, alors la moyenne d'ordre r de ces nombres x_i est inférieure à leur moyenne d'ordre s, i.e. $\left(\frac{1}{n}\sum_{i\in\underline{n}} x_i^r\right)^{1/r} \leqslant \left(\frac{1}{n}\sum_{i\in\underline{n}} x_i^s\right)^{1/s}$.

n'importe quelle demi-droite partant de l'origine, car

$$\frac{tx}{\sqrt{(tx)^2+(ty)^2}} - 1 = \frac{x}{\sqrt{x^2+y^2}} - 1$$

pour tout $t > 0$. Par conséquent, la jacobienne limite à l'origine comprend le cercle $\mathcal{C}$. Mais, elle ne peut comprendre davantage de points car elle est incluse dans la fermeture de l'ensemble des gradients sur $\mathbb{R}^2_0$ qui est ledit cercle. Ainsi, $\partial_B \phi_{FB}(0) = \mathcal{C}$ et $\partial \phi_{FB}(0)$ est la boule de centre $(-1,-1)$ et de rayon 1. Il suit que $\|V\| \geqslant \sqrt{2} - 1$ pour tout $V \in \partial \phi_{FB}(x)$ et pour tout $x \in \mathbb{R}^2$.

□

Plusieurs variantes de cette fonction de Fischer-Burmeister ont été développées durant les dernières décennies. Ainsi, Luo et Tseng [18] ont notamment utilisé la fonction :

$$\phi_{LT} : (x,y) \mapsto \|(x,y)\|_q - x - y, \quad (q > 1),$$

où $\|\cdot\|_q$ est la norme q. Pour $q = 2$, nous retombons évidemment sur ϕ_{FB}. Nous trouvons aussi chez Kanzow et Kleinmichel [16], la fonction de complémentarité

$$\phi_{KK} : (x,y) \mapsto \frac{\sqrt{(x-y)^2 + q \cdot 2xy} - x - y}{2-q}, \quad (0 \leqslant q \leqslant 2).$$

Celle-ci redonne la fonction ϕ_{FB} pour $q = 1$ et la fonction $-\min$ pour $q = 0$. Chen, Chen et Kanzow [3] ont, pour leur part, proposé plusieurs versions pénalisées de la fonction de Fisher-Burmeister :

$$\phi_{CCK1} : (x,y) \mapsto \phi_{FB}(x,y) - q x_+ y_+, \quad (q \geqslant 0),$$

avec $x_+ \overset{\text{def}}{=} \max(0,x)$; ou encore de façon équivalente,

$$\phi_{CCK2} : (x,y) \mapsto (1-q)\phi_{FB}(x,y) - q x_+ y_+, \quad (0 \leqslant q < 1),$$

qui sont toutes deux égales à ϕ_{FB} pour $q = 0$

Chacune de ces fonctions est fortement semi-lisse. Les fonctions ϕ_{LT} et ϕ_{KK} sont continûment différentiable sauf à l'origine, tandis que ϕ_{CCK} l'est sauf sur la frontière du premier quadrant. Comme pour la fonction de Fisher-Burmeister, on peut montrer que le carré de ces différentes fonctions est partout continûment différentiable. Pour de plus amples informations sur ces fonctions de complémentarité, le lecteur est invité à consulter [13], pages 859 et suivantes, ainsi que les trois articles susmentionnés.

Chapitre 3.

Résolution d'équations par une méthode de région de confiance non monotone

Comme indiqué précédemment, nous avons besoin d'une méthode pour résoudre la reformulation équivalente du problème de complémentarité : $\Phi(x) = 0$. Nous allons ici développer l'algorithme que Michael Ulbrich propose dans son article [29].

3.1. Définition du problème

Nous envisageons un système d'équations non linéaires et non lisses sous contraintes simples :

$$H(x) = 0,\ x \in X \tag{3.1.1}$$

où H est une fonction à valeurs dans $\mathbb{R}^n$ et définie sur un ouvert $U \subset \mathbb{R}^n$ contenant le domaine admissible $X = \prod_{i \in \underline{n}} [l_i, u_i]$ (avec $l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$) ; nous supposons que $l_i < u_i$ pour tout $i \in \underline{n}$, sinon la valeur x_i est complètement déterminée et il n'y a plus de raison de la chercher. Nous faisons deux hypothèses :

(H1) la fonction H est semi-lisse ou mieux, semi-lisse à l'ordre p $(0 < p \leqslant 1)$;

(H2) les composantes H_i de la fonction H sont continûment différentiable sur $U \setminus H_i^{-1}(0)$.

Reformulation. Nous reformulons (3.1.1) en un problème de minimisation :

$$\begin{cases} \min & h(x) \\ \text{s.c.} & x \in X \end{cases} \tag{3.1.2}$$

où

$$h : U \to \mathbb{R}^+ : x \mapsto \frac{1}{2} \|H(x)\|^2 .$$

Équivalence. Lorsque le problème initial possède une solution, il est équivalent à sa reformulation. En effet, si x^* est solution de (3.1.1), i.e. $H(x^*) = 0$, alors $h(x^*) = \frac{1}{2}\|H(x^*)\|^2 = 0$ et x^* est donc un minimum (global) pour (3.1.2). Réciproquement, si x_h^* est solution de (3.1.2) et qu'il existe une solution x_H^* de (3.1.1), cette dernière valeur est aussi une solution du problème de minimisation ; par conséquent, $h(x_h^*) = h(x_H^*) = 0$ et x_h^* est de ce fait, une solution du problème initial. Il faut toutefois remarquer que cette seconde implication ne vaut que si le minimum trouvé est global, ce qui est difficile à garantir dans la pratique algorithmique en l'absence de convexité. En résolvant numériquement le problème de minimisation, nous parvenons par conséquent à des candidats-solutions qu'il faut vérifier en les substituant dans le système d'équations initial.

Différentiabilité de l'objectif. Comme cela sera démontré au lemme 3.6, les deux hypothèses que nous avons posées garantissent que h est continûment différentiable sur U, ce qui va nous permettre de remplir un certain nombre d'hypothèses des techniques de région de confiance que nous allons utiliser par la suite pour résoudre le problème de minimisation.

Comme nous souhaitons que cette propriété soit vérifiée par la suite, nous la prenons comme hypothèse à la place de (H2) :

(H2*bis*) h est continûment différentiable sur U.

Notez qu'en conjonction avec l'hypothèse (H1), l'hypothèse (H2*bis*) est plus faible que (H2).

Conditions de Karush-Kuhn-Tucker. Le problème de minimisation s'écrit encore comme

$$\begin{cases} \min & h(x) \\ \text{s.c.} & x - l \geqslant 0,\ u - x \geqslant 0. \end{cases}$$

Les conditions de Karush-Kuhn-Tucker de ce problème sont donc :

$$\begin{cases} \nabla h(x) = \sum_{i=1}^n \lambda_i \nabla_x (x_i - l_i) + \mu_i \nabla_x (u_i - x_i) = \lambda - \mu\ ; \\ l \leqslant x \leqslant u,\ \lambda \geqslant 0,\ \mu \geqslant 0\ ; \\ (x_i - l_i)\lambda_i = 0,\ (u_i - x_i)\mu_i = 0,\ \forall\ i \in \underline{n}. \end{cases}$$

où $\lambda, \mu \in \mathbb{R}^n$ sont les multiplicateurs de Lagrange du problème. Nous pouvons aussi réexprimer ses conditions en fonction de la valeur des x_i ($i \in \underline{n}$). En effet, si $l_i < x_i < u_i$, alors $\lambda_i = 0 = \mu_i$ et donc il faut que $[\nabla h(x)]_i = 0$. Si $x_i = l_i < u_i$, alors $\mu_i = 0$ et donc $[\nabla h(x)]_i = \lambda_i \geqslant 0$. Si enfin, $x_i = u_i > l_i$, alors $\lambda_i = 0$ et donc $[\nabla h(x)]_i = -\mu_i \leqslant 0$. En rassemblant les trois cas, nous trouvons donc que les conditions de Karush-Kuhn-Tucker sont équivalentes à

$$\begin{cases} [\nabla h(x)]_i & = & 0 & \text{lorsque } l_i < x_i < u_i \\ [\nabla h(x)]_i & \geqslant & 0 & \text{lorsque } x_i = l_i \\ [\nabla h(x)]_i & \leqslant & 0 & \text{lorsque } x_i = u_i \end{cases} \quad , \quad \forall\ i \in \underline{n}. \tag{3.1.3}$$

Pistes pour la résolution. Pour obtenir une convergence locale superlinéaire voire quadratique de l'algorithme, nous utiliserons tout d'abord une méthode de type Newton accompagnée d'une projection sur X afin de sauvegarder la réalisabilité des itérées (section 3.2.). Toutefois, ces méthodes de type Newton ne sont pas globalement convergentes. Pour garantir la convergence globale, il faut utiliser en plus d'autres techniques comme les recherches linéaires ou les régions de confiances (voir [13], chapitre 8). Suivant le choix de M. Ulbrich, nous avons opté pour les techniques de région de confiance non monotone pour globaliser la convergence de notre premier algorithme (section 3.3.).

3.2. Méthode de type Newton avec projection

Dans cette section, nous développerons un algorithme de type Newton avec projection pour résoudre le problème de minimisation (3.1.2). Une étude de sa convergence locale sera ensuite réalisée. La convergence globale n'est pas développée car nous savons que les méthodes de type Newton n'en sont pas pourvue.

3.2.1. Présentation de l'algorithme

Soit C une partie fermée convexe non vide de $\mathbb{R}^n$. L'opérateur de projection sur C est défini comme

$$P_C : \mathbb{R}^n \to C : x \mapsto \operatorname{argmin}_{y \in C} \|y - x\| .$$

En ce qui concerne notre problème, nous allons projeter chaque nouvelle itérée sur X afin de maintenir sa réalisabilité ; le coût de cette projection est minime puisque la projection sur $X = \prod_{i \in \underline{n}} [l_i, u_i]$ est donné par

$$P_X(x) = \max(l, \min(x, u))$$

où les minima et maxima sont pris composante par composante :

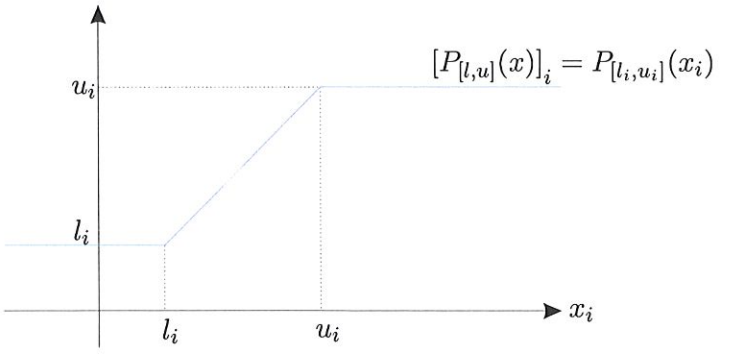


FIG. 3.1. — *Graphe de la projection agissant sur une seule composante.*

Nous sommes maintenant prêt pour dévoiler notre premier algorithme :

Algorithme 3.1 (Méthode de type Newton avec projection).

Étape 1. *Choisir* x_0 *et poser* $k = 0$.

Étape 2. *Si* $H(x_k) = 0$, *alors* STOP.

Étape 3. *Choisir une matrice non singulière* $M_k \in \mathbb{R}^{n\times n}$ *et calculer le pas de type Newton* s_k^N *en résolvant le système* $M_k s_k^N = -H(x_k)$.

Étape 4. *Calculer la projection de* s_k^N *sur* $X - x_k$: $s_k^{PN} = P_X(x_k + s_k^N) - x_k$.

Étape 5. *Mettre à jour :* $x_{k+1} = x_k + s_k^{PN}$; $k = k + 1$; *aller à l'étape 2.*

Ce premier algorithme présente deux différences par rapport à la méthode de Newton classique. La première réside dans le calcul du pas s_k^N à l'étape 3 où nous tolérons de prendre une matrice M_k légèrement différente de la jacobienne généralisée (celle-ci étant réduite à la simple matrice jacobienne dans le cas de fonction de classe C^1) ; la latitude permise à cet égard sera explicitée dans la théorie de la convergence développée *infra*. L'idée d'une telle méthode de Newton semi-lisse fut introduite en 1993 par Qi et Sun [27], puis en 1995, Martínez et Qi [20] en développèrent une version avec dérivée inexacte. La seconde modification est la projection sur X à l'étape 4, qui permet comme nous l'avons déjà mentionné, de maintenir la réalisabilité des itérées.

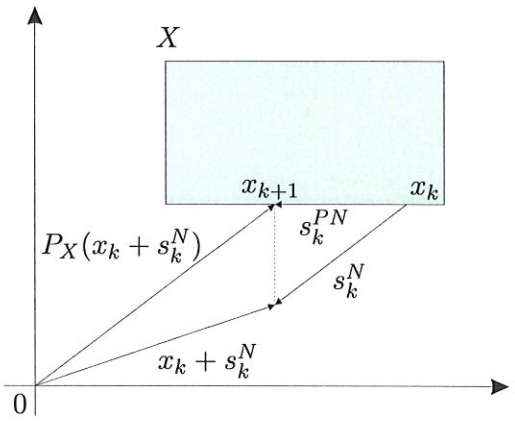


FIG. 3.2. — *Projection du pas de type Newton* s_k^N *sur le domaine admissible* X.

3.2.2. Convergence locale

La projection supplémentaire n'affectera pas la vitesse de convergence de l'algorithme par rapport à une simple méthode de type Newton. En effet, cette projection est lipschitzienne avec une constante de 1, à la suite de la proposition ci-dessous.

Proposition 3.2. *Soit C une partie fermée non vide et convexe de $\mathbb{R}^n$, alors l'opérateur de projection $P_C : \mathbb{R}^n \to C : x \mapsto \operatorname{argmin}_{y \in C} \|y - x\|$ est lipschitzien avec une constante de 1, c'est-à-dire*

$$\|P_C(x) - P_C(y)\| \leqslant \|x - y\|$$

pour tout $x, y \in \mathbb{R}^n$.

Démonstration. Soit $u \in \mathbb{R}^n$. Par définition, $P_C(u)$ est l'unique solution du problème (en la variable ξ)

$$\begin{cases} \min & \|\xi - u\|^2 \\ \text{s.c.} & \xi \in C. \end{cases}$$

Étant donné que C est fermé convexe et que $\|\xi - u\|^2$ est de classe C^1, nous savons qu'une condition nécessaire d'optimalité (voir [12], p.13) pour $P_C(u)$ est l'appartenance de l'opposé du gradient de l'objectif au cône normal, c'est-à-dire

$$\begin{aligned} & (v-u)^T \nabla_\xi (\|\xi - u\|^2)\big|_{\xi = P_C(u)} \geqslant 0, \ \forall\, v \in C \\ \Leftrightarrow \quad & (v-u)^T (P_C(u) - u) \geqslant 0, \ \forall\, v \in C. \end{aligned}$$

Soient $x, y \in \mathbb{R}^n$ et prenons $u = x$ et $v = P_C(y) \in C$, puis $u = y$ et $v = P_C(x) \in C$ dans cette équation ; nous trouvons alors que

$$\begin{aligned} (P_C(x) - x)^T (P_C(x) - P_C(y)) &\leqslant 0 \\ \text{et} \quad (P_C(y) - y)^T (P_C(y) - P_C(x)) &\leqslant 0 \end{aligned}$$

et en sommant,

$$\begin{aligned} & (P_C(x) - x - P_C(y) + y)^T (P_C(x) - P_C(y)) \leqslant 0 \\ \Leftrightarrow \quad & (P_C(x) - P_C(y))^T (P_C(x) - P_C(y)) \leqslant (x-y)^T (P_C(x) - P_C(y)) \ ; \end{aligned}$$

d'où finalement,

$$\begin{aligned} & \|P_C(x) - P_C(y)\|^2 \leqslant (x-y)^T (P_C(x) - P_C(y)) \leqslant \|x - y\| \, \|P_C(x) - P_C(y)\| \\ \Rightarrow \quad & \|P_C(x) - P_C(y)\| \leqslant \|x - y\| \, . \end{aligned}$$

par l'inégalité de Cauchy-Schwarz. □

Dans le théorème suivant, nous démontrons la convergence locale de l'algorithme 3.1

Théorème 3.3 (Convergence locale de la méthode de type Newton avec projection). *Supposons que $H : U \to \mathbb{R}^n$ soit localement lipschitzien. Soit $x^* \in X$ un zéro BD-régulier de H auquel H est semi-lisse. Nous prenons $\|x_0 - x^*\|$ et $\delta > 0$ suffisamment petits. Si l'algorithme 3.1 génère une suite infinie d'itérées et que pour tout k, on a*

$$\mu_k \stackrel{\text{def}}{=} \min_{V \in \partial_B H(x_k)} \left\|(M_k - V)s_k^N\right\| \leqslant \delta \left\|s_k^N\right\|, \tag{3.2.1}$$

alors $(x_k)_{k\in\mathbb{N}}$ converge vers x^*.
Si de plus,

$$\lim_{k\to+\infty} \frac{\mu_k}{\left\|s_k^N\right\|} = 0, \tag{3.2.2}$$

alors la suite $(x_k)_{k\in\mathbb{N}}$ converge superlinéairement vers x^*.
Par ailleurs, si H est semi-lisse à l'ordre p $(0 < p \leqslant 1)$ en x^* et que

$$\limsup_{k\to+\infty} \frac{\mu_k}{\left\|s_k^N\right\|^{1+p}} < +\infty, \tag{3.2.3}$$

alors la suite $(x_k)_{k\in\mathbb{N}}$ converge à l'ordre $(1+p)$ vers x^*.

Démonstration.

(i) Comme H est localement lipschitzien, il existe un voisinage de x^* :

$$B_{\epsilon_L}(x^*) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : \|x - x^*\| \leqslant \epsilon_L\}$$

sur lequel H est lipschitzien, i.e.

$$\forall\, x, y \in B_{\epsilon_L}(x^*) : \|H(x) - H(y)\| \leqslant L\|x - y\| \tag{3.2.4}$$

pour un certain $L > 0$. Par ailleurs, x^* est BD-régulier pour H et donc par la proposition 1.19, il existe $\epsilon_{BD} > 0$ et $C > 0$ tels que

$$\forall\, x \in B_{\epsilon_{BD}}(x^*),\ \forall\, V \in \partial_B H(x) : \left\|V^{-1}\right\| \leqslant C. \tag{3.2.5}$$

En prenant $0 < \epsilon \leqslant \min(\epsilon_L, \epsilon_{BD})$, nous obtenons à la fois (3.2.4) et (3.2.5).
Soit maintenant un point arbitraire $x_k \in B_\epsilon(x^*)$ et prenons $V_k \in \partial_B H(x_k)$ tel que le minimum est atteint pour V_k dans l'expression (3.2.1), i.e.

$$\mu_k = \left\|(M_k - V_k)s_k^N\right\|, \tag{3.2.6}$$

ce qui est possible puisque que $\partial_B H(x_k)$ est compact (par la proposition 1.4 (a)) et que $V \mapsto \left\|(M_k - V)s_k^N\right\|$ est une fonction continue. Nous définissons alors

$$d_k \overset{\text{def}}{=} x_k - x^*; \tag{3.2.7}$$

$$e_k \overset{\text{def}}{=} x_k + s_k^N - x^*. \tag{3.2.8}$$

Comme nous avons supposé que δ était suffisamment petit, nous pouvons imposer que $\delta \leqslant \frac{1}{2C}$. Par conséquent,

$$\begin{aligned}
\left\|s_k^N\right\| &= \left\|V_k^{-1}[(M_k + (V_k - M_k))s_k^N]\right\| \\
&\leqslant \left\|V_k^{-1}\right\| \left[\left\|M_k s_k^N\right\| + \left\|(M_k - V_k)s_k^N\right\|\right] \\
&\leqslant \left\|V_k^{-1}\right\| \left[\left\|-H(x_k)\right\| + \left\|(M_k - V_k)s_k^N\right\|\right] && \text{car } M_k s_k^N = -H(x_k) \\
&\leqslant C\left[\left\|H(x_k)\right\| + \mu_k\right] && \text{par (3.2.5) et (3.2.6)} \\
&\leqslant C\left\|H(x_k)\right\| + C\delta\left\|s_k^N\right\| && \text{par (3.2.1)} \\
&\leqslant C\left\|H(x_k)\right\| + \tfrac{1}{2}\left\|s_k^N\right\| \\
\Rightarrow \left\|s_k^N\right\| &\leqslant 2C\left\|H(x_k)\right\|.
\end{aligned} \tag{3.2.9}$$

Il vient ainsi que

$$\begin{aligned}
\left\|s_k^N\right\| &\leqslant 2C\left\|H(x_k)\right\| = 2C\left\|H(x_k) - H(x^*)\right\| \\
&\overset{(3.2.4)}{\leqslant} 2CL\left\|x_k - x^*\right\| \overset{(3.2.7)}{=} 2CL\left\|d_k\right\|.
\end{aligned} \tag{3.2.10}$$

De plus, comme $H(x^*) = 0$, nous avons que

$$\begin{aligned}
V_k e_k &= V_k s_k^N + V_k d_k = [M_k s_k^N + (V_k - M_k)s_k^N] + V_k d_k \quad \text{par (3.2.8)} \\
&= -H(x_k) + (V_k - M_k)s_k^N + V_k d_k \\
&= \underbrace{[H(x^*) - H(x_k) + H'(x^*; d_k)]}_{I_1} + \underbrace{[(V_k - M_k)s_k^N]}_{I_2} + \underbrace{[V_k d_k - H'(x^*; d_k)]}_{I_3}
\end{aligned} \tag{3.2.11}$$

Par la proposition 1.9, nous savons que $\|I_1\| = o(\|d_k\|)$; la proposition 1.7(b) permet quant à elle, d'affirmer que $\|I_3\| \leqslant o(\|d_k\|)$. Nous avons en outre que

$$\|I_2\| \overset{(3.2.6)}{=} \mu_k \overset{(3.2.1)}{\leqslant} \delta\left\|s_k^N\right\| \overset{(3.2.10)}{\leqslant} \delta(2CL\left\|d_k\right\|).$$

Ainsi, pour des ϵ, δ suffisamment petits,

$$\begin{aligned}\|e_k\| &\overset{(3.2.11)}{=} \left\|V_k^{-1}(I_1+I_2+I_3)\right\| \leqslant \left\|V_k^{-1}\right\| \left(\|I_1\|+\|I_2\|+\|I_3\|\right) \\ &\overset{(3.2.5)}{\leqslant} C\left[o(\|d_k\|)+\delta(2CL)\|d_k\|+o(\|d_k\|)\right] \\ &= \left(C\tfrac{o(\|d_k\|)}{\|d_k\|}+\delta(2C^2L)\right)\|d_k\| \\ &\leqslant \tfrac{1}{2}\|d_k\| \end{aligned} \tag{3.2.12}$$

car $\frac{o(\|d_k\|)}{\|d_k\|} \to 0$ quand $\epsilon \geqslant \|d_k\| \to 0$. Maintenant, comme $x^* \in X$ et que P_X a une constante de Lipschitz égale à 1, nous avons que

$$\begin{aligned}\|d_{k+1}\| &= \|x_{k+1}-x^*\| = \left\|P_X(x_k+s_k^N)-x^*\right\| = \left\|P_X(x_k+s_k^N)-P_X(x^*)\right\| \\ &\leqslant 1\cdot\left\|x_k+s_k^N-x^*\right\| \overset{(3.2.8)}{=} \|e_k\| \overset{(3.2.12)}{\leqslant} \tfrac{1}{2}\|d_k\|. \end{aligned} \tag{3.2.13}$$

En conséquence, il suit par induction que $\|d_k\| \to 0$ et qu'ainsi, la suite $(x_k)_{k\in\mathbb{N}}$ converge vers x^* si δ et $\|x_0-x^*\|$ sont suffisamment petits (notamment pour permettre $x_0 \in B_\epsilon(x^*)$).

(ii) Si l'équation (3.2.2) est vérifiée, cela signifie que $I_2 = \mu_k = o(\|s_k^N\|) \leqslant o(\|d_k\|)$ par l'équation (3.2.10) et donc l'équation (3.2.12) devient

$$\begin{aligned}\|e_k\| &= \left\|V_k^{-1}(I_1+I_2+I_3)\right\| \leqslant \left\|V_k^{-1}\right\|\left(\|I_1\|+\|I_2\|+\|I_3\|\right) \\ &\leqslant C\left[o(\|d_k\|)+o(\|d_k\|)+o(\|d_k\|)\right] \\ &= o(\|d_k\|) \end{aligned} \tag{3.2.14}$$

ce qui permet de trouver en lieu et place de l'équation (3.2.13) que

$$\|d_{k+1}\| \leqslant \|e_k\| \overset{(3.2.14)}{\leqslant} o(\|d_k\|). \tag{3.2.15}$$

Dès lors, la suite x_k converge superlinéairement vers x^* car

$$\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} = \frac{\|d_{k+1}\|}{\|d_k\|} \leqslant \frac{o(\|d_k\|)}{\|d_k\|} \to 0$$

quand $k \to \infty$ et donc que $\|d_k\| \to 0$.

(iii) Dans l'hypothèse où l'équation (3.2.3) est satisfaite, $I_2 = \mu_k = O(\|s_k\|^{1+p}) \leqslant O(\|d_k\|^{1+p})$ par l'équation (3.2.10). Si en outre, H est semi-lisse à l'ordre p, nous trouvons par la proposition 1.22 que $\|I_1\| = O(\|d_k\|^{1+p})$ et par la définition 1.21, que $\|I_3\| \leqslant O(\|d_k\|^{1+p})$.

Il suit que cette fois

$$\begin{aligned} \|e_k\| &= \left\|V_k^{-1}(I_1+I_2+I_3)\right\| \leqslant \left\|V_k^{-1}\right\| (\|I_1\|+\|I_2\|+\|I_3\|) \\ &\leqslant C\left[O(\|d_k\|^{1+p})+O(\|d_k\|^{1+p})+O(\|d_k\|^{1+p})\right] \\ &= O(\|d_k\|^{1+p}) \end{aligned} \tag{3.2.16}$$

ce qui permet de remplacer l'équation (3.2.13) par

$$\|d_{k+1}\| \leqslant \|e_k\| \overset{(3.2.16)}{\leqslant} O(\|d_k\|^{1+p}). \tag{3.2.17}$$

Dès lors, la suite x_k converge à l'ordre $1+p$ vers x^* car

$$\frac{\left\|x^{k+1}-x^*\right\|}{\|x^k-x^*\|^{1+p}} = \frac{\|d_{k+1}\|}{\|d_k\|^{1+p}} \leqslant \frac{O(\|d_k\|^{1+p})}{\|d_k\|^{1+p}} \leqslant Q$$

pour un certain $Q>0$ et pour tout $k \in \mathbb{N}$.

□

3.3. Méthode de région de confiance non monotone

3.3.1. Présentation de l'algorithme

Introduction. Ainsi que nous l'avons signalé antérieurement, les méthodes de type Newton n'apportent qu'une convergence locale. Pour remédier à ce défaut, nous utiliserons des techniques de régions de confiance. Nous espérons dès lors pouvoir prouver la convergence globale de notre méthode de résolution pour le problème (3.1.1).

Grâce à l'hypothèse (H2*bis*) : h est continûment différentiable sur U, il est possible de calculer le gradient ∇h de la fonction mérite. Nous posons $g_k = \nabla h(x_k)$.

La première étape des méthodes de région de confiance consiste en la définition à chaque itération, d'un modèle de la fonction relativement facile à minimiser dans une région déterminée, classiquement une boule ; il faut donc également choisir une norme pour définir la région de confiance (voir [6]).

Définition du modèle. Notre choix se porte ici sur un modèle quadratique

$$q_k : \mathbb{R}^n \to \mathbb{R} : s \mapsto q_k(s) \simeq h(x_k+s) - h(x_k)$$

cohérent au moins au premier ordre avec $h(x_k+s)-h(x_k)$ autour de x_k, i.e. vérifiant $h(x_k+0)-h(x_k)=0=q_k(0)$ et $h'(x_k+0)=g_k=q_k'(0)$. Ainsi, nous construisons :

$$q_k(s) = g_k^T s + \tfrac{1}{2}\|M_k s\|^2 . \tag{3.3.1}$$

Ce modèle est inspiré de la méthode de Gauss-Newton. En effet, dans le cas où H est continûment différentiable, le développement de Taylor de $H(x_k + s)$ au premier ordre autour de x_k existe et vaut $H(x_k) + M_k s$ où $M_k = H'(x_k)$. La fonction mérite associée est alors

$$\begin{aligned} \tfrac{1}{2}\left\|H(x_k) + M_k s\right\|^2 &= \tfrac{1}{2}\left\|H(x_k)\right\|^2 + \langle H(x_k), M_k s\rangle + \tfrac{1}{2}\left\|M_k s\right\|^2 \\ &= h(x_k) + \underbrace{\left(M_k^T H(x_k)\right)^T}_{g_k} s + \tfrac{1}{2}\left\|M_k s\right\|^2 \end{aligned}$$

car $g_k = \nabla\left(\frac{1}{2}\left\|H(x_k)\right\|^2\right) = H'(x_k)^T H(x_k) = M_k^T H(x_k)$. En soustrayant $h(x_k)$ de chaque membre, nous trouvons ainsi le modèle $q_k(s)$ proposé.

Lorsque l'on cherche un modèle quadratique d'une fonction, une autre approche possible est celle de la méthode de Newton. On prend alors le développement à l'ordre 2 du demi-carré de la norme de la fonction :

$$\begin{aligned} & h(x_k) + \langle \nabla h(x_k), s\rangle + \frac{1}{2}\left\langle s, (\nabla^2 h(x_k))s\right\rangle \\ = \; & h(x_k) + \left\langle M_k^T H(x_k), s\right\rangle + \frac{1}{2}\left\langle s, \left[J^T J + \sum_{i\in\underline{n}} H_i \nabla^2 H_i\right] s\right\rangle \end{aligned}$$

où l'on suppose cette fois que H est de classe C^2. Nous ne la retiendrons pas car elle demande une connaissance de H plus approfondie et dont nous ne pourrons probablement pas facilement disposer.

Par contre, pour le modèle quadratique de type Gauss-Newton, sa construction à partir de la norme du développement de Taylor à l'ordre 1, nous fournit une indication sur la matrice M_k à utiliser : une approximation de la jacobienne de H en x_k Dans le cas général, nous prendrons une matrice $M_k \in \mathbb{R}^{n\times n}$ avec les mêmes propriétés que précédemment, à savoir qu'elle est non singulière et satisfait la condition

$$\min_{V\in\partial_B H(x_k)} \left\|(M_k - V)s_k^N\right\| \;\leqslant\; \delta\left\|s_k^N\right\| \tag{3.3.2}$$

pour un certain $\delta > 0$ relativement petit, c'est-à-dire que dans la direction s_k, M_k est assez proche de l'un des éléments du B-sous-différentiel $\partial_B H(x_k)$. Ainsi, notre théorie de convergence prendra en considération un choix plus vaste de matrices M_k que celles de la jacobienne limite, permettant ainsi d'utiliser des méthodes quasi-Newton et/ou de tolérer certaines erreurs d'arrondis dans les calculs.

Le calcul explicite du gradient g_k ne sera pas nécessaire car par le lemme 3.6, nous verrons qu'en droite ligne de la relation existant pour les fonction de classe C^1, le gradient g_k peut s'exprimer comme $V_k^T H(x_k)$ où V_k est choisi arbitrairement dans $\partial H(x_k)$. Il ne faut donc

calculer que l'action de V_k^T sur le vecteur $H(x_k)$; cette opération est généralement moins onéreuse que le calcul complet de la matrice V_k.

Pour la résolution des problèmes de complémentarité non linéaire

$$\langle f(x), x\rangle = 0,\ f(x) \geqslant 0,\ x \geqslant 0.$$

(avec $f : \mathbb{R}^n \to \mathbb{R}^n$) reformulés avec une fonction de complémentarité telle que celle de Fischer-Burmeister, les $V \in \partial H(x_k)$ possèdent une forme particulière $D_a + D_b f'(x)$ (avec D_a et D_b diagonales) qui sera développée plus en détail au lemme 3.18 (voir aussi [13], p. 808 pour une méthode explicite de calcul dans le cas de la fonction de Fischer-Burmeister).

Choix d'une norme. Nous optons par ailleurs pour la norme infinie $\|\cdot\|_\infty$ par cohérence avec la forme géométrique du domaine admissible $X = [l, u] = \prod_{i\in\underline{n}}[l_i, u_i]$. De la sorte, il sera aisé de prendre l'intersection de la région de confiance $B_\infty(x_k, \Delta_k) = \{x \in \mathbb{R}^n : \|x - x_k\|_\infty \leqslant \Delta_k\}$ et du domaine admissible X ; celle-ci gardera en outre la forme d'une boîte rectangulaire.

Choix d'une mesure de criticalité. Nous définissons également une mesure de criticalité continue $\chi : X \to \mathbb{R}_+$ satisfaisant

$$\chi(x) = 0 \Leftrightarrow\ x \text{ vérifie (KKT) pour le problème (3.1.2)}$$

où (KKT) sont les conditions de Karush-Kuhn-Tucker du problème de minimisation. Un choix classique [5] pour χ est la norme du gradient projeté :

$$\chi(x) = \|x - P_X(x - \nabla h(x))\|\,.$$

Définition du sous-problème. À chaque itération, la méthode de région de confiance va résoudre approximativement le sous-problème défini comme la minimisation du modèle q_k dans région de confiance intersectée par le domaine admissible :

$$\left\{\begin{array}{ll}\min & q_k(s)\\ \text{s.c.} & l \leqslant x_k + s \leqslant u\\ & \|s\|_\infty \leqslant \Delta_k\end{array}\right.$$
$$\Leftrightarrow \left\{\begin{array}{ll}\min & q_k(s)\\ \text{s.c.} & s \in X_k\end{array}\right.$$

où $X_k \stackrel{\text{def}}{=} [l - x_k, u - x_k] \cap [-\Delta_k, \Delta_k]^n$. Il s'agit donc d'un problème de programmation quadratique sous contrainte de boîtes avec comme domaine admissible X_k.

Nous supposerons que les itérées vérifient des contraintes légèrement relaxées en ce sens que nous tolérerons de sortir un peu de la région de confiance pourvu que l'itérée reste dans le

domaine admissible, i.e.

$$l \leqslant x_k + s_k \leqslant u \quad \text{et} \quad \|s_k\|_\infty \leqslant \beta_1 \Delta_k \tag{3.3.3}$$

pour une constante $\beta_1 \geqslant 1$ indépendante de k. De plus, nous voulons que la réduction prévue $\varsigma_k^P(s_k) \stackrel{\text{def}}{=} q_k(0) - q_k(s_k) = -q_k(s_k)$ soit suffisante, i.e.

$$\varsigma_k^P(s_k) \quad \geqslant \quad \beta_2 \chi(x_k) \min\left(1, \Delta_k, \chi(x_k)\right) \tag{3.3.4}$$

pour une constante $\beta_2 > 0$ indépendante de k.

Chacune des trois composantes de cette condition de décroissance correspond à un cas de figure bien précis. Le 1 est là pour le cas où le minimum est trouvé sur la frontière de $X - x_k$; le Δ_k, pour celui où ce minimum a lieu sur la frontière de la région de confiance $[-\Delta_k, \Delta_k]^n$; enfin, le $\chi(x_k)$ provient du cas où le minimum global se trouve à l'intérieur de ces deux domaines. Ainsi, cette condition requiert une décroissance suffisante quelle que soit la position du minimum global du modèle par rapport à la région de confiance et au domaine admissible (pour plus de détails, voir la preuve du lemme 3.17).

Calcul du pas. Afin d'obtenir une convergence locale et globale rapide, nous utiliserons algorithme de type Newton 3.1 dans la méthode de région de confiance 3.4 pour calculer le pas-candidat s_k^P. Mais comme celui-ci doit satisfaire la condition de réalisabilité (3.3.3), nous le calculerons de manière légèrement différente de s_k^{PN} : alors que s_k^{PN} est la projection de $s_k^N = M_k^{-1} H(x_k)$ sur $X - x_k$, nous obtenons ici s_k^P en le projetant sur l'ensemble admissible X_k du k-ième sous-problème :

$$s_k^P \stackrel{\text{def}}{=} P_{X_k}(s_k^N).$$

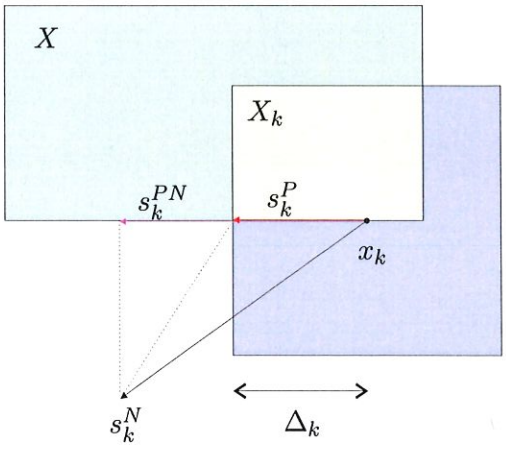


FIG. 3.3. — *Projection du pas de type Newton s_k^N sur le domaine admissible intersecté par la région de confiance : X_k.*

Toutefois, nous montrerons par la suite que $s_k^{PN} = s_k^P$ lorsque l'on s'approche d'une solution BD-régulière et donc que cette modification ne change rien aux vitesses de convergence locale

des méthodes de type Newton.

Si s_k^P satisfait en plus la seconde condition (3.3.4), il devient notre pas-candidat ; autrement, il faudra trouver un autre pas vérifiant les deux conditions (3.3.3) et (3.3.4). Nous indiquons à la section 3.3.4. comment agir pour ce faire.

Acceptation du pas. Classiquement, l'acceptation d'un pas est implémentée en fonction de la réduction effective $\varsigma_k^E(s) = h(x_k) - h(x_k + s)$. Le rapport de réduction ρ_k est alors défini comme le quotient de la réduction effective sur la réduction prévue. Nous n'imposerons pas ici la décroissance monotone de la suite $h(x_k)$ mais seulement une condition de décroissance globale. Pour ce faire, nous considérerons une version relaxée de la réduction effective. De nombreux résultats numériques témoignent d'améliorations significatives de la vitesse de convergence en utilisant ce genre de techniques non monotone ; par ailleurs, cela permet parfois à la méthode d'éviter un minimum local non global puisque la valeur de la fonction peut par moment augmenter.

Ainsi, un pas s_k sera accepté si le rapport de réduction ρ_k est suffisant, c'est-à-dire qu'il est supérieur à une valeur η_1 fixée. Nous noterons $\mathcal{S}$, l'ensemble des itérations réussies :

$$\mathcal{S} \stackrel{\text{def}}{=} \{k \in \mathbb{N} : \text{ le pas } s_k \text{ est accepté}\}.$$

Détermination du rapport de réduction. Au lieu de ne tenir compte que de la dernière itération, nous considérerons (à terme) les m dernières itérées réussies x_k ($k \in \mathcal{S}$) qui seront utilisées pour calculer la décroissance. Cependant, comme l'ensemble $\mathcal{S}$ est vide au départ, nous prendrons tous les indices qu'il contient, en attendant qu'il soit suffisamment rempli pour ne plus prendre que les m derniers.

Dans cet algorithme, nous n'imposons pas une décroissance de la fonction h à chaque itération, mais seulement une décroissance en moyenne pondérée sur les $m_k = \min(m, |\mathcal{S}|)$ dernières itérations réussies :

$$\varsigma_k^{ER}(s) \stackrel{\text{def}}{=} \max\left(h(x_k), \sum_{r=0}^{m_k-1} \lambda_{k,r} h(x_{j_{i-r}})\right) - h(x_k + s).$$

Si $m = 1$, nous retrouvons par ailleurs l'algorithme monotone classique puisque alors $\varsigma^E = \varsigma^{ER}$.

Par ailleurs, la constante λ, choisie au début de la méthode, est là pour empêcher de choisir des poids nuls ; l'algorithme doit ainsi tenir compte un tant soit peu de chacune des m_k itérées réussies précédentes.

Cette façon de procéder est légèrement plus exigeante que l'idée habituelle pour les méthodes non monotone qui prennent comme décroissance de la fonction :

$$\varsigma_k^{ER\infty}(s) \stackrel{\text{def}}{=} \max_{0 \leqslant r < m_k} h(x_{j_{i-r}}) - h(x_k + s).$$

Dans ce cas, la moyenne d'ordre $+\infty$ (i.e. le maximum) est préférée à notre hypothèse de moyenne d'ordre 1 (i.e. arithmétique) pondérée. Notre choix provient du fait que la théorie de convergence globale n'est pas garantie pour notre algorithme en utilisant $\varsigma_k^{ER\infty}$ au lieu de ς_k^{ER}. Nous pouvons toutefois parvenir à des résultats très semblables sans modifier notre implémentation : il suffit de prendre un λ suffisamment petit et de choisir les poids comme suit :

$$\lambda_{k,r} = \begin{cases} \lambda & \text{si } r \neq r_k\,; \\ 1-(m_k-1)\lambda & \text{si } r = r_k\,; \end{cases}$$

où $r_k = \operatorname{argmax}_{0 \leqslant r < m_k} h(x_{j_{i-r}})$.

Une autre possibilité souvent utilisée pour imposer une décroissance non monotone est de définir le rapport de décroissance comme

$$\tilde{\rho}_k = \max\left(\frac{\varsigma_k^E(s_k)}{\varsigma_k^P(s_k)}, \frac{h(x_{r(k)}) - h(x_k+s_k)}{\sum_{i=r(k),i\in\mathcal{S}}^{k} m_k(x_k) - m_k(x_k+s_k)}\right)$$

où $r(k)$ est l'itération de référence pour l'itération k. De cette façon, on considère la réduction effective intervenue depuis l'itération de référence par rapport à la réduction intervenue en travaillant sur chaque modèle depuis l'itération de référence. Diverses manières d'actualiser cette itération de référence existent, notamment celle proposée par [6].

Mise à la jour de la région de confiance. Le rayon de la région de confiance est actualisé à chaque itération de façon classique. Michael Ulbrich introduit cependant un paramètre $\Delta_{\min} \geqslant 0$ pour garantir une taille minimale des régions de confiance après chaque itération réussie : $\Delta_{j+1} \geqslant \Delta_{\min}$ pour tout $j \in \mathcal{S}$. Si $\Delta_{\min} = 0$, il ne joue aucun rôle ; en revanche, si $\Delta_{\min} > 0$ alors cela nous permettra d'utiliser certaines techniques dans les preuves.

Présentation de l'algorithme. Nous sommes à présent en mesure de dévoiler l'algorithme de région de confiance pour notre problème de minimisation :

Algorithme 3.4 (Méthode de région de confiance).

Étape 1. *Initialisation : choisir $\eta_1 \in]0,1[$, $\Delta_{\min} \geqslant 0$ et une mesure de criticalité χ. Choisir $x_0 \in X$, $\Delta_0 \geqslant \Delta_{\min}$ et une matrice non singulière $M_0 \in \mathbb{R}^{n\times n}$. Choisir un entier $m \geqslant 1$ et fixer $\lambda \in]0, \frac{1}{m}]$ pour le calcul de ρ_k. Poser $k = 0$ et $i = -1$.*

Étape 2. *Calculer $\chi_k = \chi(x_k)$. Si $\chi_k = 0$, alors* STOP.

Étape 3. *Calculer un pas-candidat s_k satisfaisant les conditions de réalisabilité* (3.3.3) *et de réduction* (3.3.4) *:*

3.1. *Si l'on vient du point 6, alors poser $s_k^N = s_{k-1}^N$.*
Sinon, résoudre $M_k s_k^N = -H(x_k)$ pour trouver s_k^N.

3.2. *Calculer le pas de Newton projeté $s_k^P = P_{X_k}(s_k^N)$.*

3.3. Si s_k^P satisfait la condition de décroissance (3.3.4), alors poser $s_k = s_k^P$. Sinon, calculer un autre pas vérifiant (3.3.3) et (3.3.4).

Étape 4. Calculer le rapport de réduction $\rho_k = \rho_k(s_k)$:

4.1. Déterminer le nombre de poids $m_k = \min(i+1, m)$ et choisir ces poids $\lambda_{k,r} \in \mathbb{R}$ de sorte que

$$\begin{cases} \lambda_{k,r} \geqslant \lambda, & \forall\, r = 0, \ldots, m_k - 1\,; \\ \sum_{r=0}^{m_k-1} \lambda_{k,r} = 1. \end{cases} \tag{3.3.5}$$

4.2. Calculer la réduction effective relaxée $\varsigma_k^{ER} = \varsigma_k^{ER}(s_k)$ avec

$$\varsigma_k^{ER}(s) \stackrel{\text{def}}{=} \max\left(h(x_k), \sum_{r=0}^{m_k-1} \lambda_{k,r} h(x_{j_{i-r}})\right) - h(x_k + s). \tag{3.3.6}$$

4.3. Calculer le rapport de réduction ρ_k comme étant $\rho_k(s) \stackrel{\text{def}}{=} \varsigma_k^{ER}(s)/\varsigma_k^P(s)$.

Étape 5. Calculer le nouveau rayon de la région de confiance Δ_{k+1} :
fixer $\eta_1 < \eta_2 < 1$ et $0 \leqslant \gamma_0 < \gamma_1 < 1 < \gamma_2$.

5.1. Si $\rho_k \leqslant \eta_1$, alors choisir $\Delta_{k+1} \in]\gamma_0\Delta_k, \gamma_1\Delta_k]$.

5.2. Si $\eta_1 < \rho_k < \eta_2$, alors choisir $\tilde{\Delta}_{k+1} \in [\gamma_1\Delta_k, \Delta_k]$; $\Delta_{k+1} = \max(\Delta_{\min}, \tilde{\Delta}_{k+1})$.

5.3. Si $\rho_k \geqslant \eta_2$, alors choisir $\tilde{\Delta}_{k+1} \in [\Delta_k, \gamma_2\Delta_k]$; $\Delta_{k+1} = \max(\Delta_{\min}, \tilde{\Delta}_{k+1})$

Étape 6. Si $\rho_k \leqslant \eta_1$, alors rejeter le pas s_k et poser $x_{k+1} = x_k$, $M_{k+1} = M_k$, $k = k+1$ puis aller à l'étape 3.

Étape 7. Accepter le pas : $x_{k+1} = x_k + s_k$ et choisir une matrice $M_{k+1} \in \mathbb{R}^{n\times n}$. Poser $j_{i+1} = k$, $k = k+1$ et $i = i+1$ puis aller à l'étape 2.

La suite croissante $(j_i)_{i\geqslant 0} = \mathcal{S}$ énumère les indices k des pas acceptés, de sorte que

$$x_k = x_{j_i} \qquad \forall\, j_{i-1} < k \leqslant j_i,\ \forall\, i \geqslant 1. \tag{3.3.7}$$

Réciproquement, si $k \notin \{j_i\}_{i\geqslant 1}$, alors le pas s_k a été rejeté.

Notez aussi que lorsque $i = -1$, m est nul et donc la somme dans (3.3.6) est prise sur un ensemble d'indices vide ; de ce fait, elle est nulle et $\varsigma_k^{ER}(s) = \varsigma_k^E(s)$ puisque $h(x_k) \geqslant 0$.

3.3.2. Convergence globale

Nous commencerons par établir que la fonction mérite h est continûment différentiable. À cet effet, nous prouverons tout d'abord le lemme suivant :

Lemme 3.5. *Soit la fonction $f : U \to \mathbb{R}$ localement lipschitzienne sur l'ouvert non vide*

$U \subset \mathbb{R}^n$. Si f est continûment différentiable sur $U \setminus f^{-1}(0)$, alors le carré de la fonction : f^2 est continûment différentiable sur U. De plus, $\nabla f^2(x) = 2f(x)v^T$ pour tout $v \in \partial f(x)$ et pour tout $x \in U$.

Démonstration. Par hypothèse, f et donc f^2 sont de classe C^1 sur $U \setminus f^{-1}(0)$ et le gradient de f^2 vaut $\nabla f^2(x) = 2f(x)\nabla f(x)$ sur cet ensemble. Mais, $\partial f(x) = \{\nabla f(x)^T\}$ pour tout $x \in U \setminus f^{-1}(0)$ comme nous l'avons vu à la proposition 1.4 (e), d'où $\nabla f^2(x) = 2f(x)v^T$ avec $v \in \partial f(x)$. De plus, f est localement lipschitzienne, i.e. il existe $L > 0$ et $\delta > 0$ tels que

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

pour tout $y \in U$ tel que $\|x - y\| \leqslant \delta$. Il suit que la fonction f^2 est différentiable en tout point $x \in f^{-1}(0)$ avec $\nabla f^2(x) = 0 \in \mathbb{R}^n$ car alors

$$\begin{aligned} 0 \leqslant \lim_{y \to x} \frac{\left|f^2(y) - f^2(x) - (y-x)\nabla f^2(x)\right|}{\|y - x\|} &= \lim_{y \to x} \frac{|f(y) - f(x)| \cdot |f(y) + f(x)|}{\|y - x\|} \\ &\leqslant \lim_{y \to x} \frac{L\|y - x\| \, |f(y) + f(x)|}{\|y - x\|} \\ &= \lim_{y \to x} L\,|f(y) + f(x)| = 0. \end{aligned}$$

Donc, $\nabla f^2(x) = 0 = 2 \cdot 0 \cdot v^T = 2f(x)v^T$ pour tout $v \in \partial f(x)$. Nous pouvons dès lors affirmer que $\nabla f^2(x) = 2f(x)v^T$ pour tout $v \in \partial f(x)$ et pour tout $x \in U$. Ainsi, f^2 est différentiable sur U tout entier.

Enfin, si $x \in f^{-1}(0)$ et $v(y) \in \partial f(y)$, alors $\nabla f^2(y) = 2f(y)v(y)^T \to 0$ lorsque $y \to x$ comme produit du scalaire $f(y)$ tendant vers $f(x) = 0$ (par continuité de f) et du vecteur borné $v(y) \in \partial f(y)$ (qui est compact et donc borné par la proposition 1.4 (b)). Nous concluons dès lors que f^2 est continûment différentiable sur U tout entier. □

Lemme 3.6. *Sous les hypothèses :*

(H1) la fonction H est semi-lisse ou mieux, semi-lisse à l'ordre p $(0 < p \leqslant 1)$;

(H2) les composantes H_i de la fonction H sont continûment différentiable sur $U \setminus H_i^{-1}(0)$;

la fonction $h : x \mapsto \frac{1}{2}\|H(x)\|^2$ est continûment différentiable sur U avec comme gradient $\nabla h(x) = V^T H(x)$ pour n'importe quel $V \in \partial H(x)$.
En particulier, les hypothèses (H1) et (H2) impliquent (H2bis).

Démonstration. Pour $i \in \underline{n}$, la composante H_i de la fonction H est semi-lisse (par le corollaire 1.11) et donc localement lipschitzienne sur U. De plus, par hypothèse, H_i est aussi continûment différentiable sur $U \setminus f^{-1}(0)$. Dès lors, H_i^2 est de classe C^1 sur U en vertu du lemme 3.5. Il en est de même pour $h = \frac{1}{2}\sum_{i=1}^n H_i^2$ en tant que somme de fonctions continûment différentiables. Par ailleurs, quel que soit $V = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix}^T \in \partial H(x)$, nous avons que $v_i \in \partial H_i(x), \forall\, i \in \underline{n}$

(corollaire 1.11) et il vient ainsi que

$$\nabla h(x) \stackrel{\text{def}}{=} \frac{1}{2}\sum_{i=1}^{n} \nabla(H_i^2(x)) = \sum_{i=1}^{n} H_i(x)v_i = V^T H(x)$$

grâce au lemme 3.5. □

Dans le lemme suivant, nous montrerons une importante propriété de décroissance de la fonction h le long des itérations x_k.

Lemme 3.7. *Soient données les variables x_k, s_k, Δ_k, j_i,... générées par l'algorithme de région de confiance 3.4. Alors pour tous les indices calculés $i \geqslant 1$, nous avons*

$$h(x_{j_i}) < h(x_0) - \eta_1\lambda\left(\sum_{r=0}^{i-2} \varsigma_{j_r}^P(s_{j_r})\right) - \eta_1\varsigma_{j_{i-1}}^P(s_{j_{i-1}}) \tag{3.3.8}$$

Démonstration. Par souci de concision, nous utiliserons les notations raccourcies $\varsigma_k^E = \varsigma_k^E(s_k)$, $\varsigma_k^{ER} = \varsigma_k^{ER}(s_k)$ et $\varsigma_k^P = \varsigma_k^P(s_k)$. Nous pouvons dès à présent noter que $\varsigma_k^P > 0$ à cause de la condition de décroissance (3.3.4). La preuve est réalisée par induction sur i. Pour $i = 1$, nous avons :

$$\begin{aligned} h(x_{j_1}) &\stackrel{(3.3.7)}{=} h(x_{j_0+1}) = h(x_{j_0}) - (h(x_{j_0}) - h(x_{j_0+1})) \\ &= h(x_{j_0}) - \varsigma_{j_0}^E \\ &< h(x_{j_0}) - \eta_1\varsigma_{j_0}^P \\ &= h(x_0) - \eta_1\varsigma_{j_0}^P \end{aligned}$$

car l'itération j_0 est réussie, i.e. $\rho_{j_0} > \eta_1$ d'où $\varsigma_{j_0}^E = \varsigma_{j_0}^{ER} > \eta_1\varsigma_{j_0}^P$ (en effet au départ, $\mathcal{S} = \emptyset \Rightarrow \varsigma_{j_0}^E = \varsigma_{j_0}^{ER}$).

Supposons maintenant que (3.3.8) soit vérifié pour $i = 1, \ldots, k$. Si $\varsigma_{j_k}^E = \varsigma_{j_k}^{ER}$, alors nous aboutissons à

$$\begin{aligned} h(x_{j_{k+1}}) &= h(x_{j_k+1}) && \text{par (3.3.7)} \\ &= h(x_{j_k}) - \varsigma_{j_k}^E = h(x_{j_k}) - \varsigma_{j_k}^{ER} \\ &< \left(h(x_0) - \eta_1\lambda\left(\sum_{r=0}^{k-2} \varsigma_{j_r}^P\right) - \eta_1\varsigma_{j_{k-1}}^P\right) - \eta_1\varsigma_{j_k}^P && \text{par (3.3.8) et car } \rho_{j_k} > \eta_1 \\ &\leqslant h(x_0) - \eta_1\lambda\left(\sum_{r=0}^{k-1} \varsigma_{j_r}^P\right) - \eta_1\varsigma_{j_k}^P && \text{car } \lambda \leqslant 1. \end{aligned}$$

Si en revanche $\varsigma_{j_k}^E \neq \varsigma_{j_k}^{ER}$, alors $\varsigma_{j_k}^{ER} > \varsigma_{j_k}^E$; en posant $q = m_k - 1 = \min(i, m-1)$, nous

obtenons que

$$h(x_{j_{k+1}}) \overset{(3.3.7)}{=} h(x_{j_k+1}) \overset{(3.3.6)}{=} \left(\sum_{p=0}^{q} \lambda_{j_k,p} h(x_{j_{k-p}})\right) - \varsigma_{j_k}^{ER}$$
$$\overset{(3.3.8)}{<} \left(\sum_{p=0}^{q} \lambda_{j_k,p} \left(h(x_0) - \eta_1 \lambda \left(\sum_{r=0}^{k-p-2} \varsigma_{j_r}^{P}\right) - \eta_1 \varsigma_{j_{k-p-1}}^{P}\right)\right) - \eta_1 \varsigma_{j_k}^{P} \quad (3.3.9)$$

car $\rho_{j_k} > \eta_1$. Or $(\underline{q} \times \underline{k-q-2}) \subset \{(p,r) : 0 \leqslant p \leqslant q,\ 0 \leqslant r \leqslant k-p-2\}$ (puisque le $-p \geqslant -q$ permet d'obtenir davantage d'éléments dans le second ensemble) et par définition des poids, $\sum_{p=0}^{q} \lambda_{j_i,p} = 1$ et $\lambda_{j_i,p} \geqslant \lambda$. Dès lors, nous pouvons développer l'expression précédente :

$$\begin{aligned}
h(x_{j_{k+1}}) &\overset{(3.3.9)}{<} h(x_0)\left(\sum_{p=0}^{q} \lambda_{j_k,p}\right) - \eta_1\lambda\left(\sum_{p=0}^{q} \lambda_{j_k,p} \sum_{r=0}^{k-p-2} \varsigma_{j_r}^{P}\right) - \eta_1\left(\sum_{p=0}^{q} \underbrace{\lambda_{j_k,p}}_{\geqslant\lambda} \varsigma_{j_{k-p-1}}^{P}\right) - \eta_1\varsigma_{j_k}^{P} \\
&\leqslant h(x_0) - \eta_1\lambda\left(\sum_{p=0}^{q} \lambda_{j_k,p}\right)\left(\sum_{r=0}^{k-q-2} \varsigma_{j_r}^{P}\right) - \eta_1\left(\sum_{p=0}^{q} \lambda \varsigma_{j_{k-p-1}}^{P}\right) - \eta_1\varsigma_{j_k}^{P} \\
&= h(x_0) - \eta_1\lambda\left(\sum_{r=0}^{k-q-2} \varsigma_{j_r}^{P}\right) - \eta_1\lambda\left(\sum_{r=k-q-1}^{k-1} \varsigma_{j_r}^{P}\right) - \eta_1\varsigma_{j_k}^{P} \\
&= h(x_0) - \eta_1\lambda\left(\sum_{r=0}^{k-1} \varsigma_{j_r}^{P}\right) - \eta_1\varsigma_{j_k}^{P}.
\end{aligned}$$

De la sorte, nous venons de prouver que l'inégalité (3.3.8) était vérifiée pour $i = k+1$, ce qui nous permet de conclure par induction qu'elle est satisfaite pour tout $i \geqslant 1$. □

Nous montrons à présent que chaque fois que l'on n'est pas à un point critique mais suffisamment proche de celui-ci, il existe un rayon de confiance suffisamment petit pour que l'itération suivante soit réussie. En conséquence, les itérations ne pourront pas être indéfiniment non réussies ; ou encore dit autrement, tant qu'un point critique n'est pas atteint (mais que l'on en est suffisamment proche), l'algorithme ne reste pas en permanence au même point. De plus, le rayon de la région de confiance ne pourra pas tendre vers zéro lorsque l'algorithme se rapprochera (sans y être encore parvenu) d'un point critique.

Lemme 3.8. *Soient données les variables x_k, s_k, Δ_k, j_i,... générées par l'algorithme de région de confiance 3.4. Alors, pour tout point non critique $x \in X$ (i.e. avec $\chi(x) \neq 0$) et pour n'importe quelle valeur $0 < \eta < 1$, il existe des $\Delta > 0$ et $\delta > 0$ tels que*

$$\rho_k \geqslant \eta$$

chaque fois que $\|x_k - x\| \leqslant \delta$ et que $\Delta_k \leqslant \Delta$.

Démonstration. Puisque $\chi(x) \neq 0$, il existe par continuité de χ, des $\epsilon > 0$ et $\delta > 0$ tels que

$\chi(x_k) \geqslant \epsilon$ pour tous les k avec $\|x_k - x\| \leqslant \delta$. Prenons à présent $\Delta \in]0, \min(1, \epsilon)]$ et n'importe quel k satisfaisant $\|x_k - x\| \leqslant \delta$ et $0 < \Delta_k \leqslant \Delta$; nous obtenons de la condition de décroissance (3.3.4) que

$$\varsigma_k^P(s_k) \overset{\text{def}}{=} -q_k(s_k) \overset{(3.3.4)}{\geqslant} \beta_2 \underbrace{\chi(x_k)}_{\geqslant \epsilon} \underbrace{\min(1, \Delta_k, \chi(x_k))}_{=\Delta_k} \geqslant \beta_2 \epsilon \Delta_k. \tag{3.3.10}$$

De plus, avec la condition de réalisabilité (3.3.3), il vient que

$$\|s_k\|_\infty \overset{(3.3.3)}{\leqslant} \beta_1 \Delta_k \overset{(3.3.10)}{\leqslant} \frac{\beta_1}{\beta_2 \epsilon} \varsigma_k^P(s_k). \tag{3.3.11}$$

Par ailleurs, comme h est continûment différentiable, nous pouvons prendre son développement en série de Taylor au premier ordre. De la sorte, nous avons que pour un certain $\tau_k \in [0, 1]$,

$$\begin{aligned}
\varsigma_k^E(s_k) &\overset{\text{def}}{=} h(x_k) - h(x_k + s_k) \overset{\text{Taylor}}{=} -\nabla h(x_k + \tau_k s_k)^T s_k \\
&\overset{(3.3.1)}{=} \underbrace{\left(-q_k(s_k) + g_k^T s_k + \tfrac{1}{2}\|M_k s_k\|^2\right)}_{=0} - \nabla h(x_k + \tau_k s_k)^T s_k \\
&= -q_k(s_k) + (g_k - \nabla h(x_k + \tau_k s_k))^T s_k + \underbrace{\tfrac{1}{2}\|M_k s_k\|^2}_{\geqslant 0} \\
&\geqslant \varsigma_k^P(s_k) + (g_k - \nabla h(x_k + \tau_k s_k))^T s_k.
\end{aligned} \tag{3.3.12}$$

En outre, ∇h est continu et il existe de ce fait, un $\delta' > 0$ tel que

$$\left\|\nabla h(x') - \nabla h(x)\right\|_1 \leqslant (1 - \eta)\frac{\beta_2 \epsilon}{2\beta_1}$$

pour tout $x' \in X$ tel que $\|x' - x\| < \delta'$. En réduisant Δ et δ si nécessaire, nous pouvons garantir que $\delta + \sqrt{n}\beta_1 \Delta < \delta'$. Alors, pour tout k satisfaisant $\|x_k - x\| \leqslant \delta$ et $0 < \Delta_k \leqslant \Delta$, nous trouvons que

$$\|(x_k + \tau_k s_k) - x\| \leqslant \|x_k - x\| + \underbrace{\tau_k}_{\leqslant 1} \|s_k\| \leqslant \delta + \sqrt{n}\|s_k\|_\infty \leqslant \delta + \sqrt{n}\beta_1 \Delta < \delta'. \tag{3.3.13}$$

D'où, pour de tels indices k,

$$\begin{aligned}
\|g_k - \nabla h(x_k + \tau_k s_k)\|_1 &\leqslant \|g_k - \nabla h(x)\|_1 + \|\nabla h(x_k + \tau_k s_k) - \nabla h(x)\|_1 \\
&= \|\nabla h(x_k) - \nabla h(x)\|_1 + \|\nabla h(x_k + \tau_k s_k) - \nabla h(x)\|_1 \\
&\leqslant (1 - \eta)\frac{\beta_2 \epsilon}{2\beta_1} + (1 - \eta)\frac{\beta_2 \epsilon}{2\beta_1} = (1 - \eta)\frac{\beta_2 \epsilon}{\beta_1}
\end{aligned}$$

car $\|x_k - x\| \leqslant \delta < \delta'$ et $\|(x_k + \tau_k s_k) - x\| < \delta'$ par (3.3.13). Par l'inégalité de Cauchy-Schwarz,

nous obtenons que

$$\begin{aligned}\left|[g_k - \nabla h(x_k + \tau_k s_k)]^T s_k\right| &\leqslant \|(g_k - \nabla h(x_k + \tau_k s_k))\|_1 \|s_k\|_\infty \leqslant (1-\eta)\frac{\beta_2 \epsilon}{\beta_1}\|s_k\|_\infty \\ &\overset{(3.3.11)}{\leqslant} (1-\eta)\varsigma_k^P(s_k).\end{aligned}$$

Cela implique que pour tous ces k, nous avons que

$$\begin{aligned}\varsigma_k^{ER}(s_k) &\geqslant \varsigma_k^E(s_k) \overset{(3.3.12)}{\geqslant} \varsigma_k^P(s_k) + (g_k - \nabla h(x_k + \tau_k s_k))^T s_k \\ &\geqslant \varsigma_k^P(s_k) - \left|(g_k - \nabla h(x_k + \tau_k s_k))^T s_k\right| \\ &\geqslant \varsigma_k^P(s_k) - (1-\eta)\varsigma_k^P(s_k) \\ &= \eta\varsigma_k^P(s_k).\end{aligned}$$

Finalement, nous trouvons ainsi que $\rho_k = \varsigma_k^{ER}(s_k)/\varsigma_k^P(s_k) \geqslant \eta$. □

Lemme 3.9. *L'algorithme de région de confiance 3.4*

- *soit aboutit après un nombre fini d'itérations à un point x_k vérifiant les conditions de Karush-Kuhn-Tucker du problème de minimisation* (3.1.2) ;
- *soit produit une suite infinie $(s_{j_i})_{i\in\mathbb{N}}$ de pas acceptés.*

Démonstration. Supposons un instant que l'algorithme 3.4 ne se termine pas en un nombre fini d'itérations, ni ne produise une suite infinie de pas acceptés. Alors il existe un indice minimal k_0 à partir duquel tous les pas sont rejetés. En particulier, $x_k = x_{k_0}$ pour tout $k \geqslant k_0$ et la suite des rayons des régions de confiance (Δ_k) tend vers 0 lorsque $k \to \infty$ puisque par définition de la mise à jour de ce rayon,

$$\Delta_{k_0+j} \leqslant \gamma_1^j \Delta_{k_0}$$

avec $0 < \gamma_1 < 1$ et $j \in \mathbb{N}$. Mais comme l'algorithme ne se termine pas, nous pouvons en déduire que $\chi(x_{k_0}) \neq 0$ et qu'en application du lemme 3.8, le pas s_k devra être accepté dès que Δ_k deviendra suffisamment petit (l'hypothèse $\|x_k - x_{k_0}\| = 0 \leqslant \delta$ est d'office vérifiée). Nous arrivons ainsi à une contradiction avec notre hypothèse, prouvant ainsi que l'assertion du lemme est vérifiée. □

Lemme 3.10. *Supposons que l'algorithme de région de confiance 3.4 produise une suite infinie $(s_{j_i})_{i\in\mathbb{N}}$ de pas acceptés avec $\mathcal{S} = \{j_i\}_{i\geqslant 1}$. S'il existe un sous-ensemble $\mathcal{S}' \subseteq \mathcal{S}$ tel que*

$$\sum_{k\in\mathcal{S}'} \Delta_k = +\infty, \tag{3.3.14}$$

alors

$$\liminf_{k\to\infty,\ k\in\mathcal{S}'} \chi(x_k) = 0.$$

Démonstration. Nous supposons par l'absurde, que la thèse soit fausse et qu'en conséquence, il existe $\epsilon > 0$ tel que $\chi(x_k) \geqslant \epsilon$ pour tout $k \in \mathcal{S}' \subset \mathcal{S}$. Comme une somme finie de nombres finis

est finie, il suit de (3.3.14) que $\mathcal{S}'$ est infinie. Pour chaque indice $k \in \mathcal{S}'$, l'itération est réussie et satisfait donc la condition de décroissance suffisante (3.3.4) :

$$\varsigma_k^P(s_k) \geqslant \beta_2 \chi(x_k) \min(1, \Delta_k, \chi(x_k)) \geqslant \beta_2 \chi(x_k) \min(1, \Delta_k, \epsilon).$$

Alors, en application de la positivité de $h = \frac{1}{2}\|H\|^2$ et du lemme 3.7,

$$\begin{aligned} h(x_0) \geqslant h(x_0) - h(x_j) &> \eta_1 \lambda \left(\sum_{r=0}^{R-2} \varsigma_{j_r}^P \right) + \eta_1 \varsigma_{j_{R-1}}^P \geqslant \eta_1 \lambda \sum_{r=0}^{R-1} \varsigma_{j_r}^P \\ &= \eta_1 \lambda \sum_{k \in \mathcal{S}, k<j} \varsigma_k^P \geqslant \eta_1 \lambda \sum_{k \in \mathcal{S}', k<j} \varsigma_k^P \\ &\geqslant \eta_1 \lambda \sum_{k \in \mathcal{S}, k<j} \beta_2 \chi(x_k) \min(1, \Delta_k, \epsilon) \\ &\geqslant \eta_1 \lambda \beta_2 \epsilon \sum_{k \in \mathcal{S}, k<j} \min(1, \Delta_k, \epsilon) \end{aligned}$$

pour tout $j = j_R \in \mathcal{S}'$ et car $\lambda \leqslant 1$. Nous avons ainsi construit une suite bornée (par $h(x_0)$) et qui tend vers l'infini lorsque $j \to +\infty$. Menant à une contradiction, notre hypothèse est absurde et de ce fait, la thèse du lemme doit être vérifiée. □

Grâce à cette série de lemmes, nous allons à présent pouvoir montrer notre premier théorème de convergence globale. Notez que nous imposons dans ce premier résultat, qu'à la fois $\gamma_0 > 0$ et $\Delta_{\min} > 0$, i.e. il existe, lors de toutes les itérations, une bornes inférieure (relative ou absolue) pour le rayon de la région de confiance. Nous affaibliront cette exigence par la suite. Sous cette hypothèse, tout point d'accumulation de la suite des itérées (x_k) vérifie les conditions de Karush-Kuhn-Tucker du problème de minimisation (3.1.2).

Théorème 3.11. *Supposons que $\gamma_0 > 0$ **et** que $\Delta_{\min} > 0$ et qu'ainsi, l'algorithme de région de confiance 3.4 ne se termine pas après un nombre fini d'itérations avec un point vérifiant les conditions de Karush-Kuhn-Tucker du problème de minimisation* (3.1.2). *Alors il génère une infinité de pas acceptés (s_{j_i}) tels que tout point d'accumulation de la suite (x_k) est un point satisfaisant les conditions de Karush-Kuhn-Tucker de ce problème.*

Démonstration. Supposons que l'algorithme de région de confiance 3.4 ne se termine pas après un nombre fini d'itérations. Alors par le lemme 3.9, une infinité de pas acceptés (s_{j_i}) est générée. Admettons par l'absurde, qu'il existe un point d'accumulation x^* de la suite (x_k) qui ne vérifie pas les conditions de Karush-Kuhn-Tucker du problème de minimisation (3.1.2). Alors par définition de la mesure de criticalité, $\chi(x^*) \neq 0$. Il suit du lemme 3.8 qu'il existe des $\Delta > 0$ et $\delta > 0$ pour lesquels $k \in \mathcal{S}$ dès que $\|x_k - x^*\| \leqslant \delta$ et $\Delta_k \leqslant \Delta$. Mais comme x^* est un point d'accumulation de (x_k) (i.e. il existe des points de la suite (x_k) aussi proches que l'on veut de

x^* sans y être égaux), il existe un suite croissante d'indices $(j'_i)_{i\in\mathbb{N}} \subseteq \mathcal{S}$ tels que $\|x_{j'_i} - x^*\| \leqslant \delta$ et $x_{j'_i} \to x^*$, $\forall\, i \in \underline{n}$.

Maintenant, si $(j'_i - 1) \in \mathcal{S}$ (i.e. l'itération précédente est réussie), alors le rayon de la région de confiance a un diamètre minimum : $\Delta_{j'_i} \geqslant \Delta_{\min} > 0$. Sinon $s_{j'_i-1}$ a été rejeté, ce qui n'a pu se produire que si $\Delta_{j'_i-1} \geqslant \Delta$; en effet, dans ce cas, $x_{j'_i}$ et $x_{j'_i-1}$ seraient égaux et ainsi, le lemme 3.9 trouverait à nouveau à s'appliquer si cette inégalité n'était pas vérifiée, menant à l'acceptation du pas. Il suit que dans cette seconde hypothèse, $\Delta_{j'_i} \geqslant \gamma_0 \Delta_{j'_i-1} \geqslant \gamma_0 \Delta$. Dans tous les cas, $\Delta_{j'_i} \geqslant \min(\Delta_{\min}, \gamma_0 \Delta) > 0$, ce qui permet d'affirmer que $\sum_{i\in\mathbb{N}} \Delta_{j'_i} = +\infty$. En conséquence, le lemme 3.10 ($\mathcal{S}' = \{j'_i\}_{i\in\mathbb{N}}$) nous apprend que

$$\lim_{i\to\infty} \chi(x_{j'_i}) = \liminf_{i\to\infty} \chi(x_{j'_i}) = 0.$$

Or, cette limite est précisément $\chi(x^*) \neq 0$ par continuité de la mesure de criticalité ; nous arrivons ainsi à une contradiction. Dès lors, notre hypothèse de l'absurde est fausse et nous concluons finalement que tout point d'accumulation x^* de la suite des itérées (x_k) vérifie les conditions de Karush-Kuhn-Tucker du problème de minimisation (3.1.2). □

Le théorème suivant traite du cas plus large où (au moins) l'une des hypothèses $\gamma_0 > 0$ ou $\Delta_{\min} = 0$ est remplie. Il est basé sur les résultats de Powell dans les années 1970 [24, 25, 26], mais y inclut le cas de matrices M_k différant légèrement de la jacobienne.

Théorème 3.12. *Supposons que $\gamma_0 > 0$ **ou** que $\Delta_{\min} = 0$ et qu'ainsi, l'algorithme de région de confiance 3.4 ne se termine pas après un nombre fini d'itérations avec un point vérifiant les conditions de Karush-Kuhn-Tucker du problème de minimisation* (3.1.2)*. Alors il génère une infinité de pas acceptés (s_{j_i}) tels que*

$$\liminf_{k\to\infty} \chi(x_k) = 0. \tag{3.3.15}$$

En particulier, si (x_k) converge vers x^*, alors x^* vérifie les conditions de Karush-Kuhn-Tucker de ce problème.

Démonstration. À nouveau par le lemme 3.9, une infinité de pas acceptés (s_{j_i}) est générée et donc $\mathcal{S}$ est infini. Supposons par l'absurde que l'équation (3.3.15) ne soit pas vérifiée, i.e.

$$\liminf_{k\to\infty} \chi(x_k) > 0. \tag{3.3.16}$$

Alors, en appliquant la contraposée du lemme 3.10, pour tout $\mathcal{S}' \subseteq S$, $\sum_{k\in\mathcal{S}'} \Delta_k < +\infty$; en particulier si $\mathcal{S}' = S$, nous obtenons

$$\sum_{k\in\mathcal{S}} \Delta_k < +\infty. \tag{3.3.17}$$

Mais la série infinie $\sum_{k\in\mathcal{S}} \Delta_k$ est croissante ($\Delta_k \geqslant 0$) ; elle converge donc soit vers un réel, soit vers $+\infty$. Vu que ce second cas est exclu, la série converge dans $\mathbb{R}^+$ et est par conséquent une série de Cauchy, i.e. $\forall\, \epsilon > 0,\ \exists\, N \in \mathbb{N},\ \forall j_{i-1}, j_{k-1} \in \mathcal{S}$ tels que $j_{i-1} > j_{k-1} > N$, nous avons :

$$\left| \sum_{\alpha=0}^{i-1} \Delta_{j_\alpha} - \sum_{\alpha=0}^{k-1} \Delta_{j_\alpha} \right| = \left| \sum_{\alpha=k}^{i-1} \Delta_{j_\alpha} \right| = \sum_{\alpha=k}^{i-1} \Delta_{j_\alpha} < \epsilon$$

En se rappelant l'hypothèse de faisabilité (3.3.3) du sous-problème, nous savons que $\|s_k\|_\infty \leqslant \beta_1 \Delta_k$ pour tout $k \in \mathbb{N}$. Il vient alors que la suite des itérées (x_{j_i}) est de Cauchy puisque $\forall\, \epsilon > 0,\ \exists\, N \in \mathbb{N},\ \forall j_i, j_k \in \mathcal{S}$ tels que $j_i > j_k > N$, nous avons :

$$\begin{aligned} \|x_{j_i} - x_{j_k}\|_\infty &= \left\| \sum_{\alpha=k+1}^{i} (x_{j_\alpha} - x_{j_{\alpha-1}}) \right\|_\infty \\ &\leqslant \sum_{\alpha=k+1}^{i} \|x_{j_\alpha} - x_{j_{\alpha-1}}\|_\infty = \sum_{\alpha=k+1}^{i} \|s_{j_{\alpha-1}}\|_\infty = \sum_{\alpha=k}^{i-1} \|s_{j_\alpha}\|_\infty \\ &\leqslant \sum_{\alpha=k}^{i-1} \beta_1 \Delta_{j_\alpha} < \beta_1 \epsilon \end{aligned}$$

Nous en déduisons que la sous-suite des itérées (x_{j_i}) converge vers un point $x^* \in \mathbb{R}^n$ tel que

$$\chi(x^*) = \lim_{i\to\infty} \chi(x_{j_i}) = \liminf_{i\to\infty} \chi(x_{j_i}) \neq 0$$

par continuité de la mesure de criticalité. Nous distinguons à présent deux cas de figure :

Cas 1. $\Delta_{\min} > 0$:

Par hypothèse, nous avons alors forcément $\gamma_0 > 0$; le théorème 3.11 s'applique pour le point d'accumulation x^* (limite de la sous-suite (x_{j_i})) en donnant pour résultat que x^* satisfait les conditions de Karush-Kuhn-Tucker, autrement dit que $\chi(x^*) = 0$; cela entraîne ainsi une contradiction avec notre supposition initiale.

Cas 2. $\Delta_{\min} = 0$:

En prenant $x = x^*$ et $\eta = \eta_2$ dans le lemme 3.8, nous trouvons qu'il existe des constantes $\Delta > 0$ et $\delta > 0$ telles que $\rho_k \geqslant \eta_2$ chaque fois que $\|x_k - x^*\| \leqslant \Delta$ et $\Delta_k \leqslant \Delta$; en conséquence, l'itération k est réussie (donc $k \in \mathcal{S}$) et $\Delta_{k+1} \geqslant \Delta_k$ lorsque $\|x_k - x^*\| \leqslant \Delta$ et $\Delta_k \leqslant \Delta$. Mais comme $x_k \to x^*$, il est possible de trouver un $K_1 \geqslant 0$ suffisamment grand pour que $\|x_k - x^*\| \leqslant \Delta$ dès que $k \geqslant K_1$. À nouveau, nous traitons deux volets séparément :

2.1. Supposons qu'il existe une constante $K_2 \geqslant K_1$ telle que $\Delta_k \leqslant \Delta$ pour tout $k \geqslant K_2$. Dans ce cas, les hypothèses du lemme 3.8 sont pleinement rencontrées ce qui entraîne que toutes les itérations sont réussies pour $k \geqslant K_2$ et que le rayon de la région de confiance Δ_k augmente à chaque itération ($\Delta_{k+1} \geqslant \Delta_k$ pour $k \geqslant K_2$). Il

est donc *a fortiori* supérieur à Δ_{K_2} puisque par récurrence,

$$\Delta_k \geqslant \Delta_{k-1} \geqslant \cdots \geqslant \Delta_{K_2+1} \geqslant \Delta_{K_2}.$$

Mais cela ne se peut car sinon, nous aurions que

$$\sum_{k\in\mathcal{S}} \Delta_k \geqslant \sum_{k\in\mathcal{S},\ k\geqslant K_2} \Delta_k \geqslant \sum_{k\in\mathcal{S},\ k\geqslant K_2} \Delta_{K_2} = +\infty,$$

ce que nous avons exclu précédemment par l'équation (3.3.17).

2.2. Nous avons donc $\Delta_k > \Delta$ pour des valeurs de k aussi grande que nous le désirons. De ce fait, il existe un nombre infini de telles valeurs de k. Par ailleurs, la série $\sum_{l\in\mathcal{S}} \Delta_l$ étant convergente dans $\mathbb{R}$, son terme général Δ_l doit tendre vers 0 quand $l \to +\infty$ (et $l \in \mathcal{S}$) et donc il existe $K_2 \geqslant K_1$ tel que $\Delta_{j_i} \leqslant \Delta$ pour tout $j_i \geqslant K_2$ ($j_i \in \mathcal{S}$). Comme dit précédemment nous pouvons trouver pour chaque $j_i \geqslant K_2$, un indice $k_i > j_i$ minimal tel que $\Delta_{k_i} > \Delta$.
Cet indice k_i ne peut être dans $\mathcal{S}$ parce sinon avec $k_i > j_i \geqslant K_2$, nous aurions $\Delta_{k_i} \leqslant \Delta$. Posons alors $j_i' = k_i - 1 \geqslant j_i \geqslant K_2 \geqslant K_1$. Par la même application du lemme 3.8, j_i' est dans $\mathcal{S}$ puisque $\Delta_{j_i'} \leqslant \Delta$ du fait du caractère minimal de k_i. De plus, il résulte de la mise à jour du rayon de la région de confiance (avec $\Delta_{\min} = 0$) que $\Delta_{k_i} = \Delta_{j_i'+1} \leqslant \gamma_2 \Delta_{j_i'}$ et donc que $\Delta < \Delta_{k_i} \leqslant \gamma_2 \Delta_{j_i'}$, d'où $\Delta_{j_i'} > \frac{\Delta}{\gamma_2}$. Nous avons ainsi créé une suite croissante d'indices $(j_i') \subseteq \mathcal{S}$ telle que $\Delta_{j_i'} > \frac{\Delta}{\gamma_2}$. Ainsi, en prenant la série correspondante, nous constatons que

$$\sum_{(j_i')\subseteq\mathcal{S}} \Delta_{j_i'} > \sum_{(j_i')\subseteq\mathcal{S}} \frac{\Delta}{\gamma_2} = \infty.$$

Cela nous amène à nouveau à contredire l'équation (3.3.17).

Finalement, toutes nos suppositions étaient absurdes, ce qui prouve que la thèse du théorème était en revanche vraie. □

Sous une hypothèse supplémentaire, le résultat suivant nous permet de remplacer la lim inf du théorème précédent par une simple limite. Il est issu des travaux de Thomas [28] qui souligna en particulier l'importance de l'hypothèse $\eta_1 > 0$ qui passe un peu inaperçue dans la définition de l'algorithme. Un exemple de non-convergence dans le cas $\eta_1 = 0$ fut d'ailleurs donné par Yuan dans [31].

Théorème 3.13. *Supposons que $\gamma_0 > 0$ **ou** que $\Delta_{\min} = 0$ et qu'ainsi, l'algorithme de région de confiance 3.4 ne se termine pas après un nombre fini d'itérations avec un point vérifiant les conditions de Karush-Kuhn-Tucker du problème de minimisation (3.1.2). Alors il génère une infinité de pas acceptés (s_{j_i}). De plus, s'il existe un ensemble Ω contenant la suite (x_k)*

et sur lequel χ est uniformément continu et borné, alors

$$\lim_{k\to\infty} \chi(x_k) = 0. \tag{3.3.18}$$

Démonstration. Comme nous l'avons mentionné en introduction, le seul apport de ce théorème est la transformation de la limite inférieure en une simple limite. Supposons par l'absurde, l'illicéité de ce remplacement, c'est-à-dire que $\lim_{k\to\infty} \chi(x_k)$ soit n'existe pas, soit existe mais n'est pas nulle. Toutefois, si cette limite existe, elle est égale à la limite inférieure dont nous savons par le théorème précédent qu'elle est nulle; il y a déjà ainsi contradiction. Il nous faut donc conjecturer que ladite limite n'existe pas. Cela montre, qu'en plus d'une sous-suite convergeant vers 0, il doit nécessairement exister une sous-suite ne tendant pas vers 0. Il existe donc une constante $\epsilon > 0$ telle que $\chi(x_j) \geqslant 2\epsilon$ pour une infinité de $j \in \mathbb{N}$. Cependant, nous avons indiqué antérieurement (équation 3.3.7) que pour chaque indice $j \in \mathbb{N}$, il existe un $j_i \in \mathcal{S}$ tel que $x_j = x_{j_i}$. Quitte à supprimer des éléments de la suite $(j_i) \subset \mathcal{S}$, nous pouvons sans perte de généralité supposer qu'elle est imbriquée avec une suite $(k_i) \subset \mathcal{S}$ de telle sorte que

$$j_0 < k_0 < j_1 < \cdots < j_i < k_i < j_{i+1} < \cdots .$$

Nous choisissons en outre les k_i de façon à ce que

$$\chi(x_{k_i}) \leqslant \epsilon \quad \text{et} \quad \chi(x_l) > \epsilon \text{ pour tout } l \in \mathcal{S} \text{ tel que } j_i < l < k_i,$$

c'est-à-dire que x_{k_i} est le premier élément de la queue $(x_l)_{l\in\mathcal{S},\ l>j_i}$ à être inférieur ou égal à ϵ; cet élément existe toujours à cause de l'existence d'une sous-suite tendant vers 0 et dont les indices sont pris comme précédemment dans $\mathcal{S}$, sans perte de généralité. Posons $\mathcal{S}' = \bigcup_{i=0}^{+\infty} \mathcal{S}'_i$ avec $\mathcal{S}'_i = \{l \in \mathcal{S} : j_i \leqslant l < k_i\}$. Alors,

$$\liminf_{l\to+\infty,\ l\in\mathcal{S}'} \chi(x_l) \geqslant \epsilon.$$

Il suit dès lors du lemme 3.10 que

$$\sum_{l\in\mathcal{S}'} \Delta_l < +\infty.$$

À cause de cette convergence dans $\mathbb{R}^+$, le terme général Δ_l de cette série doit tendre vers 0 quand $l \to \infty$ (et $l \in \mathcal{S}'$). En prenant seulement quelques uns de ces termes, leur somme (finie) doit aussi tendre vers 0 lorsque leur indice tend vers l'infini. C'est le cas en particulier de la somme $\sum_{l\in\mathcal{S}'_i} \Delta_l$ lorsque $i \to +\infty$, étant donné que $l \in \mathcal{S}'_i \Rightarrow l \geqslant j_i \to +\infty$ quand $i \to +\infty$. En se rappelant alors la condition de faisabilité 3.3.3 imposée, nous trouvons que

$$\|x_{k_i} - x_{j_i}\| = \left\| \sum_{l\in\mathcal{S}'_i} (x_{l+1} - x_l) \right\|_\infty \leqslant \sum_{l\in\mathcal{S}'_i} \|x_{l+1} - x_l\|_\infty \leqslant \sum_{l\in\mathcal{S}'_i} \|s_l\|_\infty \leqslant \beta_1 \sum_{l\in\mathcal{S}'_i} \Delta_l \to 0$$

lorsque $i \to +\infty$; nous pouvons ainsi affirmer que

$$\lim_{i \to +\infty} (x_{k_i} - x_{j_i}) = 0.$$

Par ailleurs, en vertu de la définition des deux suites (j_i) et (k_i), $\chi(x_{j_i}) \geqslant 2\epsilon$ et $\chi(x_{k_i}) \leqslant \epsilon$, d'où

$$|\chi(x_{k_i}) - \chi(x_{j_i})| \geqslant \epsilon, \ \forall\, i \geqslant 0,$$

ce qui est manifestement inconciliable avec la continuité uniforme de la mesure de criticalité χ. Nous en déduisons alors immédiatement la véracité de la thèse. □

3.3.3. Convergence locale

Pour étudier la convergence locale, nous utiliserons les résultats de la méthode de type Newton avec projection décrite précédemment. En effet, nous montrerons que sous des hypothèses assez peu contraignantes, l'algorithme de région de confiance 3.4 se confond avec l'algorithme de type Newton 3.1 à partir du moment où les itérées se rapprochent suffisamment d'une solution x^* BD-régulière de $H(x) = 0$. Mais pour que le pas de Newton projeté s_k^P soit accepté par l'algorithme de région de confiance, il va falloir qu'il satisfasse (3.3.3) et (3.3.4) à proximité de x^* ; nous savons déjà que (3.3.3) est vérifié puisque $s_k^P \in X_k$; reste à garantir que la condition de décroissance (3.3.4) soit aussi remplie. Nous verrons qu'il suffit alors d'exiger que l'implication suivante soit vérifiée :

$$\|x_k - x^*\| < \beta_3 \quad \text{et} \quad \varsigma_k^P(s_k^P) \geqslant \beta_4 h(x_k) \ \Rightarrow \ \varsigma_k^P(s_k^P) \geqslant \beta_2 \chi(x_k) \min(1, \Delta_k, \chi(x_k)) \quad (3.3.19)$$

où $\beta_3 > 0$ et $0 < \beta_4 < 1$ sont des constantes indépendantes de k. Vu que $h(x) = 0$, cette hypothèse signifie que la condition de décroissance (3.3.4) est vérifiée dès que l'on est proche de x^* et que la réduction prédite est au moins égale à une certaine fraction β_4 de la réduction effective maximale qui soit envisageable : $h(x_k) - h(x^*) = h(x_k)$. Cette exigence n'est donc pas déraisonnable et nous en donnerons d'ailleurs des implémentations au moyen du pas de Cauchy dans la sous-section 3.3.4.

Théorème 3.14 (Convergence locale). *Supposons vérifiées les hypothèses*

(H1) la fonction H est semi-lisse ou mieux, semi-lisse à l'ordre p $(0 < p \leqslant 1)$;

(H2) les composantes H_i de la fonction H sont continûment différentiable sur $U \setminus H_i^{-1}(0)$.

Soit $\Delta_{\min} > 0$ et supposons qu'ainsi, l'algorithme de région de confiance 3.4 ne se termine pas après un nombre fini d'itérations avec un point vérifiant les conditions de Karush-Kuhn-Tucker du problème de minimisation (3.1.2). Soit $x^* \in X$ un zéro BD-régulier de H vérifiant (3.3.19). Alors, il existe $\delta > 0$ et $\epsilon > 0$ tels que chaque fois que k' vérifie à la fois $(k'-1) \in \mathcal{S}$,

$\|x_{k'} - x^*\| \leqslant \epsilon$, *et*

$$\mu_k \stackrel{\text{def}}{=} \min_{V \in \partial_B H(x_k)} \left\|(M_k - V)s_k^N\right\| \leqslant \delta \left\|s_k^N\right\|, \quad \forall\, k \geqslant k',$$

nous avons :

(a) *pour tout* $k \geqslant k'$, *le pas* s_k^P *est égal à* $s_k^{PN} \stackrel{\text{def}}{=} P_X(x_k + s_k^N) - x_k$, *est choisi comme candidat et est accepté, i.e.* $k \in \mathcal{S}$;

(b) $x_k \to x^*$;

(c) *si pour tout* $k \geqslant k'$,

$$\lim_{k \to +\infty} \frac{\mu_k}{\left\|s_k^N\right\|} = 0, \tag{3.3.20}$$

alors la suite (x_k) *converge superlinéairement vers* x^* *et de plus,*

$$\lim_{k \to +\infty} \frac{q_k(s_k^P)}{h(x_k)} = \lim_{k \to +\infty} \frac{q_k(s_k^N)}{h(x_k)} = -1; \tag{3.3.21}$$

(d) *si pour tout* $k \geqslant k'$,

$$\limsup_{k \to +\infty} \frac{\mu_k}{\left\|s_k^N\right\|^{1+p}} < +\infty,$$

alors la suite (x_k) *converge à l'ordre* $(1+p)$ *vers* x^* *et de plus,* (3.3.21) *est aussi vérifié.*

Démonstration. Nous allons montrer qu'il existe des ϵ et δ suffisamment petits que pour vérifier les différentes assertions du théorème. En particulier, nous pouvons choisir $\epsilon < \beta_3$ (où β_3 est la constante introduite dans (3.3.19)).

Étant donné que x^* est un zéro BD-régulier de H qui est semi-lisse en x^*, il existe, en vertu de la proposition 1.19, des valeurs $\zeta > 0$, $\epsilon' > 0$ et $C > 0$ telles que

$$\|H(x)\| = \|H(x) - \underbrace{H(x^*)}_{=0}\| \geqslant \zeta \|x - x^*\| \quad \text{et} \quad \left\|V^{-1}\right\| \leqslant C,\ \forall\, V \in \partial_B H(x) \tag{3.3.22}$$

pour tout $x \in B_{\epsilon'}(x^*) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : \|x - x^*\| \leqslant \epsilon'\}$. Sans perte de généralité, $\epsilon \leqslant \epsilon'$ et nous pouvons donc remplacer $B_{\epsilon'}(x^*)$ par $B_\epsilon(x^*) \subseteq B_{\epsilon'}(x^*)$.

Prenons maintenant un k arbitraire tel que $(k-1) \in \mathcal{S}$, $x_k \in B_\epsilon(x^*)$ et $\mu_k \leqslant \delta \left\|s_k^N\right\|$. Si aucune valeur de k ne vérifie ces trois conditions, le théorème est démontré à défaut de pouvoir remplir ses hypothèses.

Comme dans la preuve de la convergence de la méthode de type Newton (théorème 3.3, p. 30), nous définissons $d_k \stackrel{\text{def}}{=} x_k - x^*$ et $e_k \stackrel{\text{def}}{=} x_k + s_k^N - x^*$; nous prenons aussi un $V_k \in \partial_B H(x_k)$ tel que

$$\mu_k = \left\|(M_k - V_k)s_k^N\right\|,$$

ce qui est possible car $\partial_B H(x_k)$ est compact (par la proposition 1.4) et $V \mapsto \|(M_k - V)s_k^N\|$ est une fonction continue. En prenant $\delta > 0$ suffisamment petit, nous pouvons imposer que $\delta \leqslant \frac{1}{2C}$. Par conséquent, comme dans la preuve précitée,

$$\begin{aligned}
\|s_k^N\| &= \|V_k^{-1}[(M_k + (V_k - M_k))s_k^N]\| \\
&\leqslant \|V_k^{-1}\| \left[\|M_k s_k^N\| + \|(M_k - V_k)s_k^N\| \right] \\
&\leqslant C\left[\|H(x_k)\| + \mu_k \right] \\
&\leqslant C\|H(x_k)\| + C\delta \|s_k^N\| \\
&\leqslant C\|H(x_k)\| + \tfrac{1}{2}\|s_k^N\| \\
\Rightarrow \|s_k^N\| &\leqslant 2C\|H(x_k)\| \qquad (3.3.23)
\end{aligned}$$

car le calcul du pas de Newton provient de la résolution de $M_k s_k^N = -H(x_k)$. Il suit ainsi que

$$\begin{aligned}
\|s_k^N\| &\leqslant 2C\|H(x_k)\| = 2C\|H(x_k) - H(x^*)\| \\
&\leqslant 2CL\|x_k - x^*\| = 2CL\|d_k\| \\
&\leqslant 2CL\epsilon
\end{aligned}$$

où L est la constante de Lipschitz de la fonction H localement lipschitzienne sur son domaine U et donc, quitte à réduire ϵ, lipschitzienne sur $B_\epsilon(x^*)$. Dès lors, nous trouvons que

$$\begin{aligned}
\|s_k^N\| &\leqslant 2CL\epsilon \\
&\leqslant \Delta_{\min} && \text{car } \Delta_{\min} > 0 \text{ et } \epsilon \text{ suffisamment petit} \\
&\leqslant \Delta_k && \text{car } (k-1) \in \mathcal{S}. \qquad (3.3.24)
\end{aligned}$$

Il suit que

$$s_k^P = P_{X_k}(s_k^N) = P_{X - x_k}(s_k^N) = P_X(x_k + s_k^N) - x_k = s_k^{PN}$$

car la projection sur $X_k = (X - x_k) \cap [-\Delta_k, \Delta_k]^n$ revient à celle sur $X - x_k$, étant donné que $s_k^N \in [-\Delta_k, \Delta_k]^n$ par (3.3.24). Comme $x_k \in X$ et que la projection P_X est lipschitzienne (avec constante de 1), il vient que

$$\|s_k^P\| = \|P_X(x_k + s_k^N) - \underbrace{P_X(x_k)}_{=x_k}\| \leqslant \|s_k^N\|. \qquad (3.3.25)$$

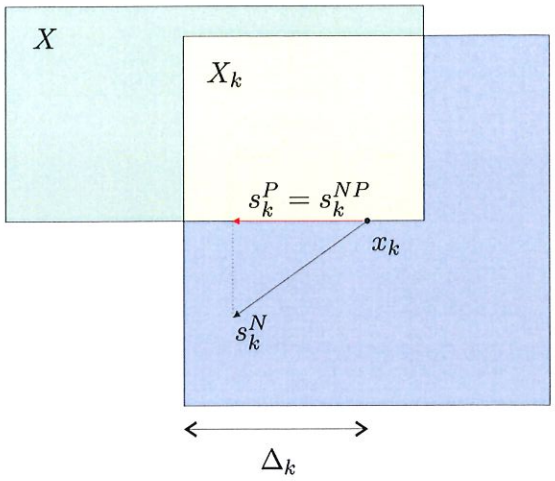


FIG. 3.4. — *Projection du pas de type Newton sur le domaine admissible : cohérence des deux approches.*

En outre, vu que $g_k = V_k^T H(x_k)$, $M_k s_k^N = -H(x_k)$ et par définition de μ_k, nous obtenons que

$$\begin{aligned}
\left|q_k(s_k^N) + h(x_k)\right| &= \left|(g_k^T s_k^N + \tfrac{1}{2}\left\|M_k s_k^N\right\|^2) + h(x_k)\right| \\
&= \left|(H(x_k)^T V_k) s_k^N + h(x_k) + h(x_k)\right| \\
&= \left|H(x_k)^T (M_k + (V_k - M_k)) s_k^N + 2h(x_k)\right| \\
&\leqslant \left|H(x_k)^T M_k s_k^N + 2h(x_k)\right| + \left\|H(x_k)\right\| \mu_k \\
&\leqslant \left|-2h(x_k) + 2h(x_k)\right| + \left\|H(x_k)\right\| \mu_k \\
&= \left\|H(x_k)\right\| \mu_k \qquad (3.3.26) \\
&\leqslant \left\|H(x_k)\right\| (\delta \left\|s_k^N\right\|) \\
&\leqslant \left\|H(x_k)\right\| \delta(2C \left\|H(x_k)\right\|) = 4C\delta h(x_k). \qquad (3.3.27)
\end{aligned}$$

De façon équivalente,

$$\begin{aligned}
& -4C\delta h(x_k) \leqslant q_k(s_k^N) + h(x_k) \leqslant 4C\delta h(x_k) \\
\Leftrightarrow\ & -4C\delta h(x_k) - h(x_k) \leqslant q_k(s_k^N) \leqslant 4C\delta h(x_k) - h(x_k) \\
\Leftrightarrow\ & -(1 + 4C\delta) h(x_k) \leqslant q_k(s_k^N) \leqslant -(1 - 4C\delta) h(x_k) \qquad (3.3.28)
\end{aligned}$$

De plus, comme $x^* \in X$ et que par définition d'une projection, $P_X(x_k + s_k^N)$ est le point de X le plus proche de $x_k + s_k^N$, nous pouvons affirmer que

$$\left\|s_k^P - s_k^N\right\| = \left\|P_X(x_k + s_k^N) - (x_k + s_k^N)\right\| \leqslant \left\|x^* - (x_k + s_k^N)\right\| = \left\|e_k\right\|. \qquad (3.3.29)$$

Alors, nous obtenons que

$$\begin{aligned}
\left|q_k(s_k^P) - q_k(s_k^N)\right| &= \left|g_k^T s_k^P + \tfrac{1}{2}(s_k^P)^T M_k^T M_k s_k^P - g_k^T s_k^N - \tfrac{1}{2}(s_k^N)^T M_k^T M_k s_k^N\right| \\
&= \left|g_k^T(s_k^P - s_k^N) + \tfrac{1}{2}(s_k^P + s_k^N)^T M_k^T M_k (s_k^P - s_k^N)\right| \\
&= \left|\left(g_k^T + \tfrac{1}{2}(s_k^P + s_k^N)^T M_k^T M_k\right)\right| (s_k^P - s_k^N) \\
&\leqslant \left(\|V_k\| \, \|H(x_k)\| + \left\|s_k^N\right\| \|M_k\|^2\right) \left\|s_k^P - s_k^N\right\| \qquad \text{par (3.3.25)} \\
&\leqslant \left(\|V_k\| + 2C \|M_k\|^2\right) \|H(x_k)\| \, \|e_k\| \quad \text{par (3.3.23) et (3.3.29).} \qquad (3.3.30)
\end{aligned}$$

Il est à présent possible de borner $q_k(s_k^P)$ en reprenant (3.3.28) :

$$\begin{array}{rcccl}
 & -\left(\|V_k\|+2C\|M_k\|^2\right)\|H(x_k)\|\|e_k\| & \leqslant q_k(s_k^P)-q_k(s_k^N) & \leqslant \left(\|V_k\|+2C\|M_k\|^2\right)\|H(x_k)\|\|e_k\| & \\
\Rightarrow & -\left(\|V_k\|+2C\|M_k\|^2\right)\|H(x_k)\|\|e_k\|+q_k(s_k^N) & \leqslant \quad q_k(s_k^P) & \leqslant \left(\|V_k\|+2C\|M_k\|^2\right)\|H(x_k)\|\|e_k\|+q_k(s_k^N) & \\
\Rightarrow & -\left(\|V_k\|+2C\|M_k\|^2\right)\|H(x_k)\|\|e_k\|-(1+4C\delta)h(x_k) & \leqslant \quad q_k(s_k^P) & \leqslant \left(\|V_k\|+2C\|M_k\|^2\right)\|H(x_k)\|\|e_k\|-(1-4C\delta)h(x_k) & \\
\Rightarrow & -\left(1+\left[\left(\|V_k\|+2C\|M_k\|^2\right)\frac{2\|e_k\|}{\|H(x_k)\|}+4C\delta\right]\right)h(x_k) & \leqslant \quad q_k(s_k^P) & \leqslant -\left(1-\left[\left(\|V_k\|+2C\|M_k\|^2\right)\frac{2\|e_k\|}{\|H(x_k)\|}+4C\delta\right]\right)h(x_k) &
\end{array}$$

De plus,

$$\begin{aligned}
V_k e_k &= V_k s_k^N + V_k d_k \\
&= [M_k s_k^N + (V_k - M_k) s_k^N] + V_k d_k \\
&= -H(x_k) + (V_k - M_k) s_k^N + V_k d_k \\
&= [-H(x_k) + \underbrace{H(x^*)}_{=0} + H'(x^*; d_k)] + [(V_k - M_k) s_k^N] + [V_k d_k - H'(x^*; d_k)],
\end{aligned}$$

ce qui implique que

$$\begin{aligned}
\|e_k\| &\leqslant C\left(\left\|-H(x_k) + H(x^*) + H'(x^*; d_k)\right\| + \left\|(V_k - M_k) s_k^N\right\| + \left\|V_k d_k - H'(x^*; d_k)\right\|\right) \\
&= C\left[o(\|d_k\|) + \mu_k + o(d_k)\right] \\
&= o(\|d_k\|)
\end{aligned}$$

par les propositions 1.9 et 1.7. En combinant cette inégalité avec (3.3.22), nous concluons que

$$\frac{\|e_k\|}{\|H(x_k)\|} \leqslant \frac{1}{\zeta} \frac{\|e_k\|}{\|d_k\|} \to 0 \quad \text{lorsque } \delta, \epsilon \to 0. \qquad (3.3.31)$$

Prenons maintenant $\theta \in]0, 1[$ arbitraire. À cause du caractère localement borné de $\partial_B H$ (proposition 1.4 (c)), du fait que $\|M_k\|$ est borné et de (3.3.31), nous aboutissons, en prenant δ et ϵ suffisamment petits, à :

$$2\left(\|V_k\| + 2C\|M_k\|^2\right) \frac{\|e_k\|}{\|H(x_k)\|} + 4C\delta \leqslant \theta.$$

En combinaison avec les bornes données précédemment sur $q_k(s_k^P)$, cela montre que

$$-(1+\theta)h(x_k) \leqslant q_k(s_k^P) \leqslant -(1-\theta)h(x_k). \tag{3.3.32}$$

Si de plus, θ est choisi inférieur à $1-\beta_4$, alors

$$\varsigma_k^P(s_k^P) \stackrel{\text{def}}{=} -q_k(s_k^P) \stackrel{(3.3.32)}{\geqslant} (1-\theta)h(x_k) \geqslant \beta_4 h(x_k).$$

Dès lors, en vertu de notre hypothèse (3.3.19), $s_k = s_k^P = s_k^{PN}$. Or,

$$\begin{aligned} \|d_{k+1}\| &= \|x_{k+1} - x^*\| = \left\|P_X(x^k + s_k^N) - x^*\right\| = \left\|P_X(x^k + s_k^N) - P_X(x^*)\right\| \\ &\leqslant 1 \cdot \left\|x^k + s_k^N - x^*\right\| = \|e_k\| \end{aligned} \tag{3.3.33}$$

car la projection P_X est lipschitzienne de rang 1. Ainsi, il vient que :

$$\left\|H(x_k + s_k^P)\right\| = \left\|H(x_k + s_k^P) - \underbrace{H(x^*)}_{=0}\right\| \leqslant L\left\|x_k + s_k^P - x^*\right\| = L\|d_{k+1}\| \leqslant L\|e_k\| \tag{3.3.34}$$

où L est comme précédemment, la constante de Lipschitz de H sur $B_\epsilon(x^*)$. Nous savons par ailleurs que

$$\begin{aligned} \varsigma_k^{ER}(s_k^P) &\geqslant \varsigma_k^E(s_k^P) = h(x_k) - h(x_k + s_k^P) = h(x_k) - \tfrac{1}{2}\left\|H(x_k + s_k^P)\right\|^2 \\ &\stackrel{(3.3.34)}{\geqslant} h(x_k) - \tfrac{L^2}{2}\|e_k\|^2 = \left(1 - L^2\frac{\|e_k\|^2}{\|H(x_k)\|^2}\right) h(x_k) \\ &\geqslant (1-\theta)\, h(x_k) \end{aligned} \tag{3.3.35}$$

parce que par (3.3.31), $\frac{\|e_k\|}{\|H(x_k)\|} \to 0$ lorsque $\epsilon, \delta \to 0$. Si maintenant, nous imposons en plus, que θ soit strictement inférieur à $\frac{1-\eta_1}{1+\eta_1}$, alors par (3.3.32) et (3.3.35), nous trouvons que

$$\rho_k(s_k^P) = \frac{\varsigma_k^{ER}(s_k^P)}{\varsigma_k^P(s_k^P)} \geqslant \frac{(1-\theta)h(x_k)}{(1+\theta)h(x_k)} > \frac{1 - \frac{1-\eta_1}{1+\eta_1}}{1 + \frac{1-\eta_1}{1+\eta_1}} = \frac{(1+\eta_1) - (1-\eta_1)}{(1+\eta_1) + (1-\eta_1)} = \eta_1.$$

Dès lors, le pas s_k^P est accepté ; puisque ϵ et δ sont suffisamment petit, nous obtenons aussi que $x_{k+1} \in B_\epsilon(x^*)$ car $\|d_{k+1}\| \leqslant \|e_k\| = o(\|d_k\|) \leqslant o(\epsilon)$ par (3.3.33), (3.3.31) et le fait que $x_k \in B_\epsilon(x^*)$. Il suit inductivement que tous les pas suivant s_k seront eux aussi acceptés puisque le précédent est accepté, qu'il est dans la boule $B_\epsilon(x^*)$ et que la majoration des μ_k est toujours supposée vérifiée par hypothèse. Le point (a) est ainsi démontré.

Du fait que l'algorithme de région de confiance 3.4 se réduit après suffisamment d'itérations à celui de type Newton avec projection, les conclusions sur la convergence locale sont identiques : la suite des itérées x_k converge vers x^* (point (b)) ; la convergence superlinéaire (point (c)) et à l'ordre $(1+p)$ (point (d)) nous sont données par le théorème 3.3. Quant à la limite de l'équation (3.3.21), nous remarquons d'une part, qu'avec (3.3.20), i.e. $\mu_k = o(\|s_k^N\|)$, nous

pouvons renforcer (3.3.27) pour parvenir à

$$\begin{aligned} \left|q_k(s_k^N) + h(x_k)\right| &\overset{(3.3.26)}{\leqslant} \|H(x_k)\|\,\mu_k = \|\|H(x_k\|)\|\,o(\|s_k^N\|) \\ &\leqslant \|H(x_k)\|\,o(H(x_k)) = o(h(x_k)), \end{aligned} \tag{3.3.36}$$

car $\|s_k^N\| \leqslant 2C\,\|H(x_k)\|$ en vertu de (3.3.24). D'autre part, en reprenant (3.3.30) et (3.3.31), nous observons que pour les points (c) et (d),

$$\left|q_k(s_k^P) - q_k(s_k^N)\right| \overset{(3.3.30)}{=} O(\|H(x_k)\|\,\|e_k\|) \overset{(3.3.31)}{=} o(h(x_k)). \tag{3.3.37}$$

Finalement, par l'inégalité triangulaire, nous trouvons

$$\left|q_k(s_k^P) + h(x_k)\right| \leqslant \left|q_k(s_k^P) - q_k(s_k^N)\right| + \left|q_k(s_k^N) + h(x_k)\right| \leqslant o(h(x_k)),$$

d'où,

$$\lim_{k\to\infty} \frac{\left|q_k(s_k^P) + h(x_k)\right|}{h(x_k)} = 0 \quad \Leftrightarrow \quad \lim_{k\to\infty} \frac{\left|q_k(s_k^P)\right|}{h(x_k)} = -1.$$

□

3.3.4. Implémentation d'un pas satisfaisant la condition de décroissance

Introduction. La théorie de convergence présentée ci-dessus repose sur la condition de décroissance abstraite (3.3.4) :

$$\varsigma_k^P(s_k) \geqslant \beta_2 \chi(x_k) \min(1, \Delta_k, \chi(x_k))$$

impliquant une mesure de criticalité χ. Nous avons aussi fait l'hypothèse que

$$\|x_k - x^*\| < \beta_3 \quad \text{et} \quad \varsigma_k^P(s_k^P) \geqslant \beta_4 h(x_k) \quad \Rightarrow \quad \varsigma_k^P(s_k^P) \geqslant \beta_2 \chi(x_k) \min(1, \Delta_k, \chi(x_k))$$

(avec $\beta_3 > 0$ et $0 < \beta_4 < 1$, indépendants de k), pour obtenir une convergence locale rapide de notre méthode.

Pas de Cauchy. Nous donnons à présent une implémentation concrète du choix du pas s_k permettant de vérifier ces deux conditions, dans le cas où le pas de Newton projeté ne pourrait satisfaire la condition de décroissance précitée. Il s'agit du pas de Cauchy préconditionné s_k^C défini comme solution du problème d'optimisation :

$$\begin{cases} \min & q_k(s) \\ \text{s.c.} & s = -tD(x_k)^{2\gamma} g_k \\ & s \in X,\ t \geqslant 0 \end{cases}$$

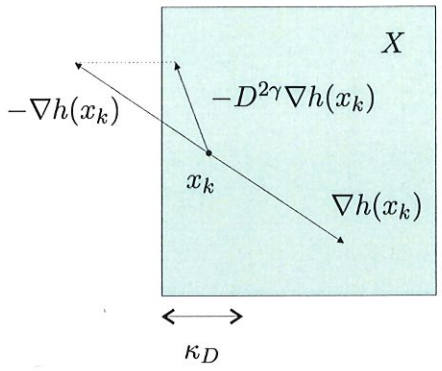


FIG. 3.5. — *Interprétation du préconditionneur pour le pas de Cauchy.*

où $\gamma \geqslant 1$ est fixé et où la matrice diagonale (semi-définie positive) de préconditionnement $D(x) \in \mathbb{R}^{n\times n}$ est définie par

$$[D(x)]_{ii} = \begin{cases} \min(\kappa_D, x_i - l_i) & \text{si } [\nabla h(x)]_i > 0\,; \\ \min(\kappa_D, u_i - x_i) & \text{si } [\nabla h(x)]_i < 0\,; \\ \min(\kappa_D, x_i - l_i, u_i - x_i) & \text{si } [\nabla h(x)]_i = 0\,; \end{cases}$$

pour une constante $\kappa_D > 0$.

Voyons concrètement ce que signifie cette matrice $D(x)$. Supposons que la i-ième composante du gradient de h (que nous cherchons à minimiser sur $X = [l, u]$) soit strictement positive et donc que $[s]_i$ soit négatif ; alors l'itérée suivante $[x]_i$ sera plus proche de la frontière l_i du domaine admissible X. Or, comme nous ne pouvons pas sortir de ce domaine (pour la résolution du sous-problème), il est intéressant de diminuer l'avancée selon $[x]_i$ lorsque que l'itérée est proche du bord l_i ($[x]_i - l_i < \kappa_D$) afin de pouvoir progresser de manière significative dans les autres directions. Il en va de même avec les bornes u_i lorsque la i-ième composante du gradient de h est strictement négative. Lorsque la i-ième composante du gradient de h est nulle, $[s]_i$ est forcément nul. La valeur de $[D(x)]_{ii}$ peut alors est choisie arbitrairement ; notre choix sera justifié par l'utilisation qui en sera faite par la suite.

Par exemple, si $g_k = \nabla h(x_k) = (2, -1)$ et $\gamma = 1 = \kappa_D$, alors nous trouvons

$$-D(x)^{2\gamma} g_k = -\begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 1 \end{pmatrix}^2 \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}$$

Ceci est illustré à la figure 3.5.

Nouvelle condition de décroissance. Considérons maintenant une mesure de criticalité χ telle que

$$\beta_5 \chi(x) \quad \leqslant \quad \chi_{AS}(x) \stackrel{\text{def}}{=} \|D(x)^\gamma \nabla h(x)\|\,, \quad \forall\, x \in X \tag{3.3.38}$$

pour un certain $\beta_5 > 0$. Nous allons montrer que si la décroissance avec le pas s_k est égale à une fraction de la décroissance due au pas de Cauchy préconditionné, i.e.

$$q_k(s_k) \leqslant \alpha q_k(s_k^C) \tag{3.3.39}$$

avec $\alpha \in]0,1[$ constant, alors il existe $\beta_2 > 0$ tel que pour cette mesure de criticalité χ, la condition de décroissance (3.3.4) est satisfaite :

$$\varsigma_k^P(s_k) \geqslant \beta_2 \chi(x_k) \min(1, \Delta_k, \chi(x_k))$$

Un choix naturel pour χ est alors χ_{AS}/β_5 ; mais encore faut-il que χ_{AS} soit une mesure de criticalité, c'est-à-dire qu'elle soit continue et s'annule aux et seulement aux points vérifiant les conditions d'optimalité de Karush-Kuhn-Tucker du problème de minimisation (3.1.2).

Lemme 3.15. *La fonction* $\chi_{AS} : \mathbb{R}^n \to \mathbb{R}^+ : x \mapsto \|D(x)^\gamma \nabla h(x)\|$ *est une mesure de criticalité.*

Démonstration.

(i) Nous avons que $\chi_{AS}(x)$ est nul si et seulement si $D(x)^\gamma \nabla h(x) = 0$, i.e.

$$[D(x)]_{ii}^\gamma [\nabla h(x)]_i = 0, \quad \forall\, i \in \underline{n}.$$

En vertu de la définition de $D(x)$, si $[x]_i = l_i$, alors $[D(x)]_{ii} = 0$ si et seulement si $[\nabla h(x)]_i \geqslant 0$; si $[x]_i = u_i$, alors $[D(x)]_{ii} = 0$ si et seulement si $[\nabla h(x)]_i \leqslant 0$; enfin si $l_i < [x]_i < u_i$, alors $[D(x)]_{ii} = 0$ si et seulement si $[\nabla h(x)]_i = 0$. Il y a donc annulation de $D(x)^\gamma \nabla h(x) = 0$ si et seulement si

$$\begin{cases} [\nabla h(x)]_i \geqslant 0 & \text{lorsque } [x]_i = l_i\,; \\ [\nabla h(x)]_i \leqslant 0 & \text{lorsque } [x]_i = u_i\,; \\ [\nabla h(x)]_i = 0 & \text{lorsque } l_i < [x]_i < u_i. \end{cases}$$

Nous constatons alors qu'il s'agit des conditions de Karush-Kuhn-Tucker du problème de minimisation de (3.1.2) comme souhaité.

(ii) Il nous reste encore à prouver la continuité de χ_{AS}. Soit x un point arbitraire de X et posons

$$I_+ \stackrel{\text{def}}{=} \{i \in \underline{n} : [\nabla h(x)]_i > 0\} \quad \text{et} \quad I_- \stackrel{\text{def}}{=} \{i \in \underline{n} : [\nabla h(x)]_i < 0\}.$$

Puisque h est continûment différentiable, ∇h est continu et par conséquent, $[\nabla h(\cdot)]_i$ ne change pas de signe sur un petit voisinage de x lorsque $[\nabla h(x)]_i \neq 0$; en d'autres termes, il existe $\delta > 0$ tel que pour tout $y \in B_\delta(x)$ et pour tout $i \in I_+ \cup I_-$, nous avons $[\nabla h(x)]_i [\nabla h(y)]_i > 0$. Prenons dès lors $y \in B_\delta(x)$ et définissons

$$r(x,y) = D(y)^\gamma \nabla h(y) - D(x)^\gamma \nabla h(x).$$

Alors pour tout $i \in I_+$, nous trouvons

$$\begin{aligned}
&|[r(x,y)]_i| = \\
&\quad \Big|\Big([D(y)]_{ii}^{\gamma}[\nabla h(y)]_i - [D(y)]_{ii}^{\gamma}[\nabla h(x)]_i\Big) - \Big([D(x)]_{ii}^{\gamma}[\nabla h(x)]_i - [D(y)]_{ii}^{\gamma}[\nabla h(x)]_i\Big)\Big| \\
&\leqslant [D(y)]_{ii}^{\gamma}\big|[\nabla h(y)]_i - [\nabla h(x)]_i\big| + [\nabla h(x)]_i\big|[D(y)]_{ii}^{\gamma} - [D(x)]_{ii}^{\gamma}\big| \\
&\leqslant \kappa_D^{\gamma}\big|[\nabla h(y) - \nabla h(x)]_i\big| + [\nabla h(x)]_i\big|\min(\kappa_D, y_i - l_i)^{\gamma} - \min(\kappa_D, x_i - l_i)^{\gamma}\big|,
\end{aligned}$$

par définition de $D(x)$ et $D(y)$. Lorsque $i \in I_-$, nous obtenons de même que

$$|[r(x,y)]_i| \leqslant \kappa_D^{\gamma}\big|[\nabla h(y) - \nabla h(x)]_i\big| + \big|[\nabla h(x)]_i\big|\big|\min(\kappa_D, u_i - y_i)^{\gamma} - \min(\kappa_D, u_i - x_i)^{\gamma}\big|.$$

Enfin si $i \notin I_+ \cup I_-$, $[\nabla h(x)]_i = 0$ et il s'en suit que

$$\begin{aligned}
|[r(x,y)]_i| &= \big|[D(y)]_{ii}^{\gamma}[\nabla h(y)]_i\big| = [D(y)]_{ii}^{\gamma}\big|[\nabla h(y)]_i\big| \\
&= [D(y)]_{ii}^{\gamma}\big|[\nabla h(y)]_i - \underbrace{[\nabla h(x)]_i}_{=0}\big| \\
&\leqslant \kappa_D^{\gamma}\big|[\nabla h(y) - \nabla h(x)]_i\big|
\end{aligned}$$

Ainsi, dans tous les cas, $|[r(x,y)]_i| \to 0$ lorsque $y \to x$ par continuité des fonctions min et ∇h. En conséquence, $r(x,y) \to 0$ quand $y \to x$ et donc

$$\begin{aligned}
|\chi_{AS}(y) - \chi_{AS}(x)| &= \big|\,\|D(y)^{\gamma}\nabla h(y)\| - \|D(x)^{\gamma}\nabla h(x)\|\,\big| \\
&\leqslant \|D(y)^{\gamma}\nabla h(y) - D(x)^{\gamma}\nabla h(x)\| \\
&= \|r(x,y)\| \to 0
\end{aligned}$$

lorsque $y \to x$, i.e. χ_{AS} est continue.

Avec ces deux éléments, nous avons ainsi prouvé que χ_{AS} était une mesure de criticalité. □

Comme nous l'avons vu dans le théorème 3.13, il est intéressant d'avoir la propriété de continuité uniforme pour la mesure de criticalité car à ce moment, nous savons que la limite inférieure du théorème 3.12 se transforme en une véritable limite pourvu que cette mesure soit aussi bornée.

Lemme 3.16. *La mesure de criticalité χ_{AS} est uniformément continue sur toute partie $\Omega \subseteq U$ sur laquelle ∇h est borné et uniformément continu.*

Démonstration. Les trois inégalités du lemme précédent pour les cas où $i \in I_+$, I_- et $(I_+ \cup I_-)^c$ sont toujours valables pour des $x, y \in \Omega \subseteq U$ suffisamment proches. Pour montrer la continuité uniforme sur Ω, il suffit de montrer que lorsque x et y se rapprochent, $[r(x,y)]_i \to 0$. Cette convergence se déduit facilement du fait que d'une part, ∇h est uniformément continu sur Ω, ce qui implique que $[\nabla h(y) - \nabla h(x)]_i \to 0$ lorsque $\|y - x\| \to 0$ avec $x, y \in \Omega$, et que d'autre part, ∇h est borné sur Ω et que la fonction minimum est aussi uniformément continue sur Ω,

ce qui entraîne que $\left|[\nabla h(x)]_i\right| \cdot \left|\min(\kappa_D, u_i - y_i)^\gamma - \min(\kappa_D, u_i - x_i)^\gamma\right|$ tend également vers zéro. □

Notez que les hypothèses du lemme précédent sont immédiatement vérifiées si Ω est compact et que $h \in C^1$. Or, nous avons supposé le caractère continûment différentiable de h, de sorte que la compacité de Ω suffit pour que χ_{AS} soit uniformément continue sur Ω. Lorsque $-\infty < l < u < +\infty$, le domaine admissible X est compact et contient les itérées x_k, ce qui permet de vérifier directement l'hypothèse supplémentaire du théorème 3.13 pour la mesure de criticalité χ_{AS}.

Nous disposons à présent des outils nécessaires pour montrer que la condition de décroissance tirée du pas de Cauchy suffit pour obtenir la condition (3.3.4) :

Proposition 3.17. *Soit χ une mesure de criticalité satisfaisant*

$$\beta_5 \chi(x) \leqslant \chi_{AS}(x), \quad \forall\, x \in X \tag{3.3.40}$$

pour un certain $\beta_5 > 0$. Supposons que la suite $(\|M_k\|)_{k\in\mathbb{N}}$ soit bornée supérieurement par une constante C_M. Alors, il existe une constante β_2 dépendant uniquement de α, β_5, γ, κ_D et C_M, telle que si $\chi(x_k) \neq 0$ et si le pas-candidat s_k vérifie $q_k(s_k) \leqslant \alpha q_k(s_k^C)$, alors

$$\varsigma_k^P(s_k) \geqslant \beta_2 \chi(x_k) \min(1, \Delta_k, \chi(x_k)).$$

Démonstration. Pour faciliter l'écriture, nous noterons $D_k = D(x_k)$, $d_k = -D_k^{2\gamma} g_k$ et $\hat{g}_k = D_k^\gamma g_k$, de sorte que $\|\hat{g}_k\| = \chi_{AS}(x_k)$. La preuve se décompose en deux étapes : d'abord, nous donnerons une borne supérieure pour

$$q_k(s_k^C) = q_k(t^* d_k) = \min\{q_k(t d_k) : t \geqslant 0,\ t d_k \in X_k\},$$

puis nous l'appliquerons à $q_k(s_k) \leqslant \alpha q_k(t^* d_k)$.

Nous observons en premier lieu que d_k est une direction de descente de q_k en 0 car

$$\nabla q_k(0)^T d_k = g_k^T d_k = -g_k^T D_k^{2\gamma} g_k = -\left\|D_k^\gamma g_k\right\|^2 < 0.$$

Comme il faut rester dans la région de confiance, le pas t doit être inférieur à

$$\begin{aligned} t_1 &\overset{\text{def}}{=} \min_{i\in\underline{n}} \left\{ \frac{\Delta_k}{|[d_k]_i|} : [d_k]_i \neq 0 \right\} = \min_{i\in\underline{n}} \left\{ \frac{\Delta_k}{[D_k]_{ii}^\gamma\, |[\hat{g}_k]_i|} : [d_k]_i \neq 0 \right\} \\ &\geqslant \frac{\Delta_k}{\kappa_D^\gamma \|\hat{g}_k\|} = \frac{\Delta_k}{\kappa_D^\gamma \cdot \chi_{AS}(x_k)}. \end{aligned} \tag{3.3.41}$$

La longueur du pas t est également limitée supérieurement par l'ensemble admissible $X - x_k$. Nous distinguons d'une part les contraintes $x_i \geqslant l_i$ et d'autre part, $x_i \leqslant u_i$. Pour les premières,

le pas maximum autorisé est

$$\begin{aligned} t_2 &\overset{\text{def}}{=} \min_{i\in\underline{n}} \left\{ \frac{[x_k-l]_i}{|[d_k]_i|} : [d_k]_i < 0 \right\} = \min_{i\in\underline{n}} \left\{ \frac{[x_k-l]_i}{[D_k]_{ii}^{\gamma}\,|[\hat{g}_k]_i|} : [g_k]_i > 0,\ [x_k]_i > l_i \right\} \\ &\geqslant \min_{i\in\underline{n}} \left\{ \frac{[x_k-l]_i}{\min(\kappa_D,[x_k-l]_i)^{\gamma}\,\|\hat{g}_k\|} : [g_k]_i > 0,\ [x_k]_i > l_i \right\} \\ &\geqslant \min_{i\in\underline{n}} \left\{ \frac{\min(\kappa_D,[x_k-l]_i)}{\min(\kappa_D,[x_k-l]_i)^{\gamma}\,\|\hat{g}_k\|} : [g_k]_i > 0,\ [x_k]_i > l_i \right\} \\ &\geqslant \frac{\min(\kappa_D,[x_k-l]_i)^{1-\gamma}}{\chi_{AS}(x_k)} \geqslant \frac{\kappa_D^{1-\gamma}}{\chi_{AS}(x_k)} \end{aligned} \tag{3.3.42}$$

car $\gamma \geqslant 1$ et donc $(1-\gamma) \leqslant 0$. Il en est de même pour les contraintes de borne supérieures : le pas maximum est alors

$$\begin{aligned} t_3 &\overset{\text{def}}{=} \min_{i\in\underline{n}} \left\{ \frac{[u-x_k]_i}{|[d_k]_i|} : [d_k]_i > 0 \right\} = \min_{i\in\underline{n}} \left\{ \frac{[u-x_k]_i}{[D_k]_{ii}^{\gamma}\,|[\hat{g}_k]_i|} : [g_k]_i < 0,\ [x_k]_i < u_i \right\} \\ &\geqslant \min_{i\in\underline{n}} \left\{ \frac{[u-x_k]_i}{\min(\kappa_D,[u-x_k]_i)^{\gamma}\,\|\hat{g}_k\|} : [g_k]_i < 0,\ [x_k]_i < u_i \right\} \\ &\geqslant \min_{i\in\underline{n}} \left\{ \frac{\min(\kappa_D,[u-x_k]_i)}{\min(\kappa_D,[u-x_k]_i)^{\gamma}\,\|\hat{g}_k\|} : [g_k]_i < 0,\ [x_k]_i < u_i \right\} \\ &\geqslant \frac{\min(\kappa_D,[u-x_k]_i)^{1-\gamma}}{\chi_{AS}(x_k)} \geqslant \frac{\kappa_D^{1-\gamma}}{\chi_{AS}(x_k)} \end{aligned} \tag{3.3.43}$$

car $\gamma \geqslant 1$ et donc $(1-\gamma) \leqslant 0$. Si $M_k d_k = 0$ (le modèle est affine), nous posons $t_4 = +\infty$. Dans le cas contraire, le modèle

$$q_k(td_k) = g_k^T(td_k) + \tfrac{1}{2}\,\|M_k(td_k)\|^2 = t\cdot g_k^T d_k + t^2\cdot \tfrac{1}{2}\,\|M_k d_k\|^2 \quad (t \geqslant 0)$$

atteint son minimum global sur $\mathbb{R}^+$ en $t = t_4$ défini comme

$$\begin{aligned} t_4 &= \frac{-g_k^T d_k}{\|M_k d_k\|^2} = \frac{g_k^T D_k^{2\gamma} g_k}{\|M_k d_k\|^2} = \frac{\|\hat{g}_k\|^2}{\left\|M_k D_k^{\gamma}\hat{g}_k\right\|^2} \\ &\geqslant \frac{\|\hat{g}_k\|^2}{\|M_k\|^2\left\|D_k^{\gamma}\right\|^2\|\hat{g}_k\|^2} = \frac{1}{\|M_k\|^2\left\|D_k^{\gamma}\right\|^2} \\ &\geqslant \frac{1}{C_M^2\kappa_D^{2\gamma}}. \end{aligned} \tag{3.3.44}$$

Nous obtenons alors que $t^* = \min(t_1,t_2,t_3,t_4)$. Si $t^* < t_4$, nous avons par définition de t_4 que $t^*\,\|M_k d_k\|^2 \leqslant \|\hat{g}_k\|^2$ et donc que

$$\begin{aligned} q_k(t^* d_k) &= -t^*\,\|\hat{g}_k\|^2 + \tfrac{1}{2}\,(t^*)^2\,\|M_k d_k\|^2 \\ &< -t^*\,\|\hat{g}_k\|^2 + \tfrac{1}{2}t^*\,\|\hat{g}_k\|^2 = -\tfrac{1}{2}t^*\,\|\hat{g}_k\|^2 \\ &= -\tfrac{1}{2}t^*\chi_{AS}(x_k)^2 = -\tfrac{1}{2}\min(t_1,t_2,t_3)\,\chi_{AS}(x_k)^2. \end{aligned} \tag{3.3.45}$$

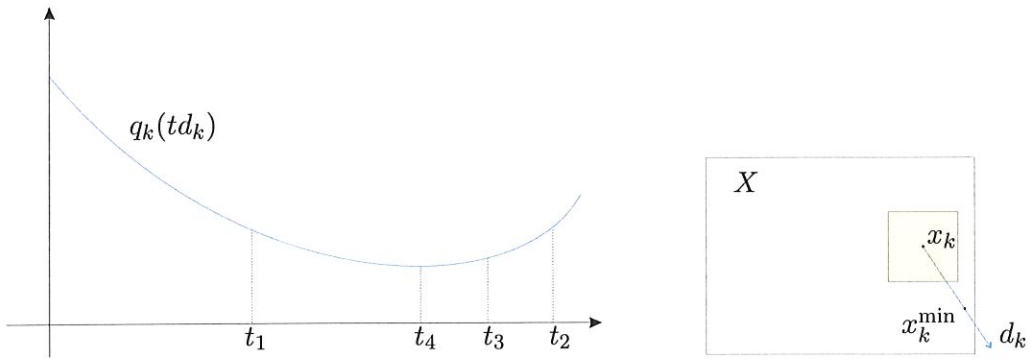


FIG. 3.6. — *Cas où $t^* = t_1$: on s'arrête sur la frontière de la région de confiance.*

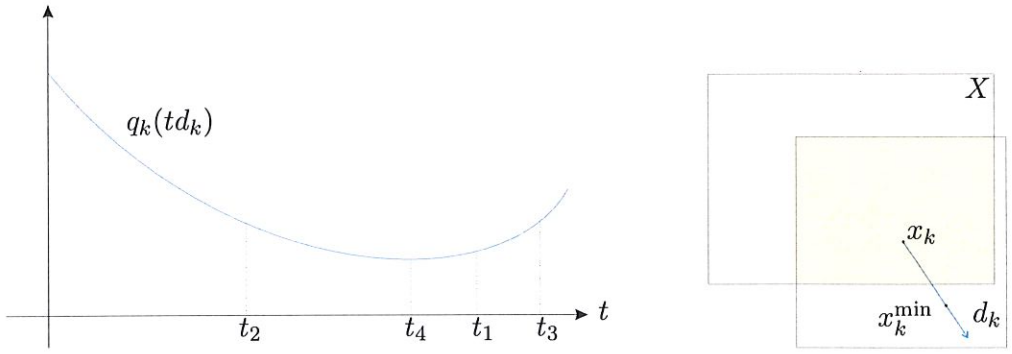


FIG. 3.7. — *Cas où $t^* = t_2$ (similaire à $t^* = t_3$) : on s'arrête sur la frontière du domaine admissible.*

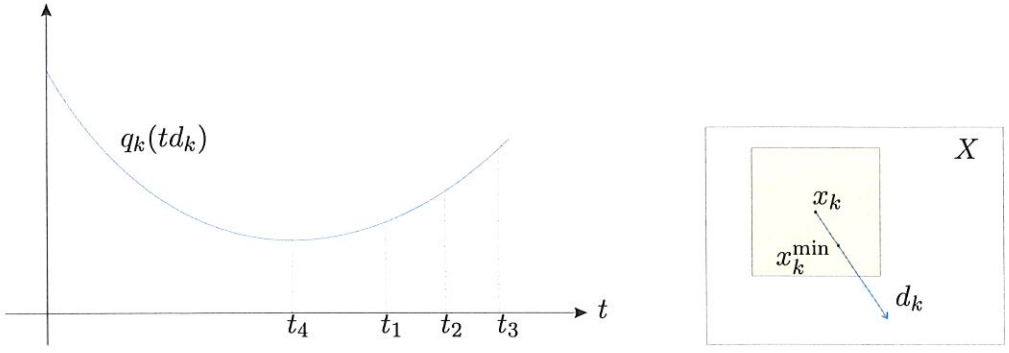


FIG. 3.8. — *Cas où $t^* = t_4$: le minimum global du modèle (sur $\mathbb{R}$) se trouve dans la région de confiance et dans le domaine admissible.*

Grâce aux minorations (3.3.41), (3.3.42) et (3.3.43) trouvée pour t_1, t_2 et t_3, nous pouvons affirmer que

$$\min(t_1, t_2, t_3) \geqslant \frac{\min(\Delta_k, \kappa_D)}{\kappa_D^{\gamma} \chi_{AS}(x_k)} ;$$

il s'en suit que

$$\begin{aligned} q_k(t^* d_k) &< -\tfrac{1}{2} \min(t_1, t_2, t_3) \chi_{AS}(x_k)^2 \leqslant -\tfrac{1}{2} \frac{\min(\Delta_k, \kappa_D)}{\kappa_D^{\gamma} \chi_{AS}(x_k)} \chi_{AS}(x_k)^2 \\ &= -\frac{\min(\Delta_k, \kappa_D)}{2\kappa_D^{\gamma}} \chi_{AS}(x_k) \overset{(3.3.40)}{\leqslant} -\frac{\min(\Delta_k, \kappa_D)}{2\kappa_D^{\gamma}} \beta_5 \chi(x_k). \end{aligned} \tag{3.3.46}$$

En revanche, si $t^* = t_4$, nous trouvons

$$\begin{aligned} q_k(t^* d_k) &= -t^* \|\hat{g}_k\|^2 + \tfrac{1}{2} (t^*)^2 \|M_k d_k\|^2 \\ &= -t^* \|\hat{g}_k\|^2 + \tfrac{1}{2} t^* \|\hat{g}_k\|^2 = -\tfrac{1}{2} t^* \|\hat{g}_k\|^2 = -\tfrac{1}{2} t_4 \chi_{AS}(x_k)^2 \\ &\leqslant -\frac{1}{2C_M^2 \kappa_D^{2\gamma}} \chi_{AS}(x_k)^2 \leqslant -\frac{1}{2C_M^2 \kappa_D^{2\gamma}} \beta_5^2 \chi(x_k)^2. \end{aligned} \tag{3.3.47}$$

En rassemblant les deux cas de figure, il vient que

$$\begin{aligned} q_k(s_k) \leqslant \alpha q_k(t^* d_k) &\leqslant \alpha \max\left(-\frac{\beta_5 \min(\Delta_k, \kappa_D)}{2\kappa_D^{\gamma}} \chi(x_k), -\frac{\beta_5^2}{2C_M^2 \kappa_D^{2\gamma}} \chi(x_k)^2 \right) \\ &= -\chi(x_k) \min\left(\frac{\alpha\beta_5}{2\kappa_D^{\gamma}} \min(\Delta_k, \kappa_D), \frac{\alpha\beta_5^2}{2C_M^2 \kappa_D^{2\gamma}} \chi(x_k) \right). \end{aligned} \tag{3.3.48}$$

Mais puisque $\varsigma_k^P(s_k) = -q_k(s_k)$, nous concluons que

$$\begin{aligned} \varsigma_k^P(s_k) &\geqslant \chi(x_k) \min\left(\frac{\alpha\beta_5 \kappa_D^{1-\gamma}}{2}, \frac{\alpha\beta_5}{2\kappa_D^{\gamma}} \cdot \Delta_k, \frac{\alpha\beta_5^2}{2C_M^2 \kappa_D^{2\gamma}} \chi(x_k) \right) \\ &\geqslant \underbrace{\min\left(\frac{\alpha\beta_5 \kappa_D^{1-\gamma}}{2}, \frac{\alpha\beta_5}{2\kappa_D^{\gamma}}, \frac{\alpha\beta_5^2}{2C_M^2 \kappa_D^{2\gamma}} \right)}_{\beta_2 > 0} \chi(x_k) \min(1, \Delta_k, \chi(x_k)). \end{aligned} \tag{3.3.49}$$

□

Notez que par le biais de cette preuve, nous avons retrouvé l'interprétation des trois composantes de cette condition de décroissance : le 1 est là pour le cas où le minimum est trouvé sur la frontière du domaine admissible X ; le Δ_k, pour celui où ce minimum a lieu sur la frontière de la région de confiance ; enfin, le $\chi(x_k)$ provient du cas où le minimum se trouve à l'intérieur de ces deux domaines.

Ainsi, pour toutes les mesures de criticalité qui satisfont $\beta_5 \chi(x) \leqslant \chi_{AS}(x)$, $\forall\, x \in X$, nous

pouvons remplacer la condition de décroissance (3.3.4) par la condition (3.3.39) :

$$q_k(s_k) \leqslant \alpha q_k(s_k^C)$$

pour une certaine constante $\alpha > 0$.

En particulier, le pas de Cauchy $s_k = s_k^C$ vérifie toujours cette condition et est donc toujours un pas-candidat envisageable ; nous pouvons donc l'utiliser lorsque le pas de Newton projeté ne satisfait pas la condition de décroissance (3.3.4).

Avant d'obtenir un algorithme implémentable, il nous reste encore à définir une mesure de criticalité χ

- qui soit uniformément continue et bornée (pour garantir les résultats du théorème 3.13) ;
- qui permette de vérifier rapidement la condition de décroissance (3.3.4) ;
- qui satisfasse la condition (3.3.19), à savoir que la condition de décroissance (3.3.4) doit être vérifiée pour $s_k = s_k^P$ chaque fois que

$$\|x_k - x^*\| < \beta_3 \quad \text{et} \quad \varsigma_k^P(s_k^P) \geqslant \beta_4 h(x_k)$$

pour des constantes $\beta_3 > 0$ et $0 < \beta_4 < 1$ indépendantes de k.

Plusieurs approches sont possibles pour définir une telle mesure.

I. Approche universelle. Nous débuterons par une approche applicable dans tous les cas. Définissons

$$\chi_I : \mathbb{R}^n \to \mathbb{R}^+ : x \mapsto \min\left(\chi_{AS}(x), \sqrt{\beta_4 h(x)}\right)$$

avec $\beta_4 \in]0, 1[$. Cette fonction χ_I est une mesure de criticalité puisqu'elle est continue et qu'elle n'est nulle que si $h(x) = 0$ ou $\chi_{AS}(x) = 0$, ce qui implique dans les deux cas que x vérifie les conditions (nécessaires) d'optimalité de Karush-Kuhn-Tucker du problème de minimisation (3.1.2). Choisissons à présent une mesure de criticalité vérifiant

$$\beta_5 \chi(x) \leqslant \chi_I(x), \quad x \in X, \tag{3.3.50}$$

pour un certain $\beta_5 > 0$. Alors d'office, la majoration $\beta_5 \chi(x) \leqslant \chi_{AS}(x)$, $\forall\, x \in X$ est satisfaite. Nous pouvons dès lors changer le point 3.3. de l'algorithme de région de confiance comme suit :

> *3.3. Si s_k^P satisfait la condition*
>
> $$q_k(s_k) \leqslant \max\left(\alpha q_k(s_k^C), -\beta_4 h(x_k)\right), \tag{3.3.51}$$
>
> *alors poser $s_k = s_k^P$.*
> *Sinon, calculer un autre pas vérifiant* (3.3.3) *et* (3.3.51).

Notez que cette nouvelle condition (3.3.51) est une relaxation de la condition (3.3.39) qui

elle-même était moins exigeante que la condition de décroissance (3.3.4) comme mentionné ci-dessus.

En effet, d'une part, si $q_k(s_k) \leqslant -\beta_4 h(x_k)$, alors

$$\varsigma_k^P(s_k) \geqslant \left(\sqrt{\beta_4 h(x_k)}\right)^2 \geqslant \chi_I^2(x_k) \geqslant \beta_5^2 \chi^2(x_k) \geqslant \underbrace{\beta_5^2}_{\beta_2} \chi(x_k) \min\left(1, \Delta_k, \chi(x_k)\right)$$

et donc, la condition (3.3.4) est immédiatement vérifiée. D'autre part, si $q_k(s_k) \leqslant \alpha q_k(s_k^C)$, alors la proposition 3.17 est applicable. Finalement, il est facile de voir que le choix de $\chi = \chi_I$ est uniformément continue si χ_{AS} et H sont uniformément continus sur Ω ; la continuité uniforme de χ_{AS} peut être obtenu à cet effet par le lemme 3.16 qui requiert la continuité uniforme et le caractère borné de ∇h.

II. Nous discutons maintenant d'une situation dans laquelle il suffit que χ remplisse la condition (3.3.38). Supposons que $\min_{V \in \partial_B H(x)} \|V\| \leqslant C_V < \infty$ sur X. Alors, avec $V(x) = \operatorname{argmin}_{V \in \partial_B H(x)} \|V\|$, nous avons que

$$\chi_{AS}(x) \stackrel{\text{def}}{=} \|D(x)^\gamma \nabla h(x)\| \leqslant \kappa_D^\gamma \|V(x)\| \, \|H(x)\| \leqslant \kappa_D^\gamma C_V \|H(x)\|$$

puisque $\nabla h(x) = V(x)^T H(x)$ par le lemme 3.6. Par conséquent, si $\beta_4 \in]0,1[$ est fixé et que la mesure de criticalité χ vérifie $\beta_5' \chi(x) \leqslant \chi_{AS}(x)$, $\forall\, x \in X$, alors nous trouvons $\beta_5 \chi(x) \leqslant \chi_I(x)$, $x \in X$ pour tout $\beta_5 > 0$ inférieur ou égal à

$$\beta_5' \min\left(1, \frac{\sqrt{\beta_4/2}}{\kappa_D^\gamma C_V}\right).$$

En effet,

$$\begin{aligned}
\beta_5 \chi(x) &\leqslant \beta_5' \min\left(1, \frac{\sqrt{\beta_4/2}}{\kappa_D^\gamma C_V}\right) \chi(x) \leqslant \min\left(1, \frac{\sqrt{\beta_4/2}}{\kappa_D^\gamma C_V}\right) \chi_{AS}(x) \\
&= \min\left(\chi_{AS}(x), \frac{\sqrt{\beta_4/2 \cdot 2h(x_k)}}{\kappa_D^\gamma C_V \|H(x_k)\|} \chi_{AS}(x)\right) = \min\left(\chi_{AS}(x), \sqrt{\beta_4 h(x)}\right) = \chi_I(x)
\end{aligned}$$

De la sorte, le paragraphe I est applicable. Nous pouvons en particulier choisir $\chi = \chi_{AS}$.

III. Décroissance d'une fraction du pas de Cauchy. Enfin, nous établissons une condition suffisante pour que l'algorithme 3.4 puisse être implémenté au moyen de la simple condition de décroissance de Cauchy :

3.3. Si s_k^P satisfait la condition

$$q_k(s_k) \leqslant \alpha q_k(s_k^C) \tag{3.3.52}$$

alors poser $s_k = s_k^P$.
Sinon, calculer un autre pas vérifiant (3.3.3) *et* (3.3.52).

Supposons que la mesure de criticalité χ remplisse la condition (3.3.38), que les suites $(\|M_k\|)_{k\in\mathbb{N}}$ et $\left(\left\|M_k^{-1}\right\|\right)_{k\in\mathbb{N}}$ soient bornées et que

$$\left\|M_k^T H(x_k) - g_k\right\| = o(\|H(x_k)\|) \quad \text{lorsque } k \to +\infty.$$

Notez que cette dernière affirmation est vraie pour tout $M_k \in \partial H(x_k)$ car alors le membre de gauche est nul. Aussi, sous ces hypothèses, nous obtenons à cause du fait que la valeur au point de Newton est inférieure à celle au point de Cauchy,

$$\begin{aligned}
q_k(s_k^C) &\geqslant q_k(-(M_k^T M_k)^{-1} g_k) = -\tfrac{1}{2} g_k^T (M_k^T M_k)^{-1} g_k \\
&= -h(x_k) + \tfrac{1}{2}(M_k^T H(x_k) - g_k)^T (M_k^T M_k)^{-1}(M_k^T H(x_k) + g_k) \\
&= -h(x_k) - \tfrac{1}{2}(M_k^T H(x_k) - g_k)^T (M_k^T M_k)^{-1}\left((M_k^T H(x_k) - g_k) - 2M_k^T H(x_k)\right) \\
&\geqslant -h(x_k) - \tfrac{1}{2} o\left(\|H(x_k)\|\right) \left\|(M_k^T M_k)^{-1}(M_k^T H(x_k) - g_k) - (M_k^T M_k)^{-1} \cdot 2M_k^T H(x_k)\right\| \\
&= -h(x_k) + o\left(\|H(x_k)\|^2\right) = -h(x_k) + o(h(x_k)).
\end{aligned}$$

où la dernière inégalité est obtenue en prenant l'inégalité de Cauchy-Schwarz. En conséquence, pour tout β_4 pris arbitrairement dans $]\alpha, 1[$, nous pouvons trouver $\beta_3 > 0$ tel que pour tous les indices k satisfaisant $\|x_k - x^*\| < \beta_3$ et $\varsigma_k^P(s_k^P) \geqslant \beta_4 h(x_k)$,

$$q_k(s_k^P) = -\varsigma_k^P(s_k^P) \leqslant -\beta_4 h(x_k) \leqslant \alpha q_k(s_k^C),$$

ce qui implique finalement, que la condition de décroissance (3.3.4) est vérifiée.

3.3.5. Application de la méthode aux problèmes de complémentarité

Considérons le problème de complémentarité non linéaire mixte

$$\begin{cases} x_i F_i(x) = 0, & F_i(x) \geqslant 0, \quad x_i \geqslant 0, & \text{pour } i = 1, \ldots, m; \\ & F_i(x) = 0 & \text{pour } i = m+1, \ldots, n; \end{cases} \tag{3.3.53}$$

où $F : U \to \mathbb{R}^n$ est définie sur l'ouvert $U \subseteq \mathbb{R}^n$ contenant $X \stackrel{\text{def}}{=} [0, +\infty[^m \times \mathbb{R}^{n-m}$ (avec $0 \leqslant m \leqslant n$). Nous reformulons ce problème grâce à une fonction de complémentarité ϕ, de sorte que

$$\begin{cases} H_i(x) \stackrel{\text{def}}{=} \phi(x_i, F_i(x)) & \text{pour } i = 1, \ldots, m\,; \\ H_i(x) \stackrel{\text{def}}{=} F_i(x) & \text{pour } i = m+1, \ldots, n. \end{cases} \tag{3.3.54}$$

Nous obtenons ainsi la fonction $H : U \to \mathbb{R}^n$ définie sur l'ouvert U contenant le domaine admissible X.

Nous supposons que les hypothèses suivantes sont vérifiées :

(H3) la fonction $F : U \to \mathbb{R}^n$ est continûment différentiable avec une dérivée lipschitzienne ;

(H4) la fonction de complémentarité $\phi : \mathbb{R}^2 \to \mathbb{R}$ est continûment différentiable sur $\mathbb{R}^2 \setminus \phi^{-1}(0)$ et semi-lisse à l'ordre 1 sur $\mathbb{R}^2$; de plus, chaque fois que $\phi(a, b) = 0$ et que $v = (v_1, v_2)$ est pris dans $\partial\phi(a, b)$, on a $v_1 \geqslant 0$, $v_2 \geqslant 0$, $v \neq 0$ ainsi que $v_1 = 0$ si $a > 0$ et $b = 0$, tandis que $v_2 = 0$ si $a = 0$ et $b > 0$.

Cette dernière hypothèse est par exemple vérifiée pour la fonction de Fisher-Burmeister (pénalisée ou non).

Lemme 3.18. *Sous les hypothèses (H3) et (H4), la fonction H définie par* (3.3.54) *satisfait les hypothèses (H1) avec $p = 1$ et (H2). De plus, pour tout $x \in U$ et tout $V \in \partial H(x)$, il existe des matrices diagonales $D_a, D_b \in \mathbb{R}^{n \times n}$ telles que*

$$V = D_a + D_b F'(x) \tag{3.3.55}$$

$$[D_a]_{ii} = 0 \quad \text{et} \quad [D_b]_{ii} = 1 \qquad \text{pour } i = m+1, \ldots, n. \tag{3.3.56}$$

En outre, si x est une solution du problème de complémentarité (3.3.53), *alors $D_a + D_b$ est définie positive et pour tout $i \in \underline{m}$,*

$$\begin{aligned} [D_a]_{ii} &= 0 \quad \text{si } x_i > 0 \text{ et } F_i(x) = 0\,; \\ [D_b]_{ii} &= 0 \quad \text{si } x_i = 0 \text{ et } F_i(x) > 0. \end{aligned}$$

Démonstration. Par la proposition 1.25, la fonction F est semi-lisse à l'ordre 1, puisque différentiable à dérivée lipschitzienne. Il résulte aussi des propositions 1.23 et 1.24 que H est semi-lisse à l'ordre 1, puisque ses composantes sont semi-lisses à l'ordre 1 soit directement, soit en tant que composée de fonctions présentant cette propriété. Nous avons ainsi vérifié l'assertion (H1) avec $p = 1$.

Comme $F_i = H_i$ pour $i = m+1, \ldots, n$, il y a égalité entre les $(n-m)$ dernières lignes de V et de F' (i.e. les gradients transposés des F_i pour $i = m+1, \ldots, n$). De la sorte, $[D_a]_{ii} = 0$ et $[D_b]_{ii} = 1$ (pour la même plage d'indices) conduit à l'égalité des $(n-m)$ dernières lignes de V et de $D_a + D_b F'(x)$. Fixons à présent $i \in \underline{m}$ et prenons un x arbitraire dans U tel que $H_i(x) = \phi(x_i, F(x_i)) \neq 0$. Alors, en vertu de l'hypothèse (H4), H_i est de classe C^1 dans un

voisinage de x (en tant que composée de fonctions continûment différentiable en x) ; l'assertion $(H2)$ est ainsi vérifiée.

Prenons cette fois un $x \in U$ tout à fait arbitraire et désignons $(x_i, F_i(x))$ par $f_i(x) \in \mathbb{R}^n$, de sorte que (par le théorème 2.6.6 dans [4])

$$\begin{aligned} \partial H_i(x) &= \operatorname{conv}\left(\partial\phi(f_i(x))\partial f_i(x)\right) \\ &= \partial\phi(f_i(x))\nabla f_i(x)^T \\ &= \left\{v_1 e_i^T + v_2 \nabla F_i(x)^T : v = (v_1, v_2) \in \partial\phi(x_i, F_i(x))\right\}. \end{aligned}$$

Il suffit donc de prendre $([D_a]_{ii}, [D_b]_{ii}) \in e_i^T \partial H \subseteq \partial\phi(x_i, F_i(x)))$ puisque (par la proposition 2.6.2 (e) dans [4]), nous savons que

$$\partial H(x) \subseteq \partial_C H(x) \stackrel{\text{def}}{=} \partial H_1(x) \times \cdots \times \partial H_n(x).$$

Lorsque x est solution du problème de complémentarité non linéaire mixte, $\phi(x_i, F_i(x)) = 0$ pour tout $i \in \underline{m}$. Il s'en suit que $v \geqslant 0$ et $v \neq 0$ pour tout $v \in \partial\phi(x_i, F_i(x))$, ce qui implique que $v_1 + v_2 > 0$ et donc tous les éléments de la matrice diagonale $D_a + D_b$ sont strictement positifs, i.e. $D_a + D_b$ est définie positive. Si $x_i > 0$ et $F_i(x) = 0$, alors par (H4), v_1 est nul et il en est donc de même pour $[D_a]_{ii}$. Si cette fois, $x_i = 0$ et $F_i(x) > 0$, alors par (H4), $v_2 = 0$ d'où, $[D_b]_{ii} = 0$. □

Définition 3.19 (Régularité forte). Une solution x^* du problème de complémentarité non linéaire mixte (3.3.54) est dite fortement régulière s'il existe un voisinage $U_{x^*} \subseteq \mathbb{R}^n$ de x^* et un voisinage $U_0 \subseteq \mathbb{R}^n$ de 0 tels que pour tout $y \in U_0$, le problème de complémentarité mixte linéarisé

$$\begin{cases} x_i F_i^y(x) = 0, & F_i^y(x) \geqslant 0, \quad x_i \geqslant 0, & \text{pour } i = 1, \ldots, m\,; \\ & F_i^y(x) = 0 & \text{pour } i = m+1, \ldots, n\,; \end{cases} \tag{3.3.57}$$

avec $F^y(x) \stackrel{\text{def}}{=} F(x^*) + F'(x^*)(x - x^*) + y$, admet exactement une solution $x(y)$ vérifiant $x(y) \in U_{x^*}$ ainsi que le caractère lipschitzien de l'application $\varpi : U_0 \to \mathbb{R}^n : y \mapsto x^*(y)$.

Un autre critère nous est donné par le lemme suivant pour déterminer si une solution est fortement régulière ou non :

Lemme 3.20. *Une solution x^* du problème de complémentarité non linéaire mixte* (3.3.54) *est fortement régulière si et seulement si la sous-matrice*

$$F'(x^*)_{I^*I^*} \stackrel{\text{def}}{=} \left(\frac{\partial F_i}{\partial x_j}(x^*)\right)_{i,j \in I^*}$$

est non singulière et que son complément de Schur dans $F'(x^*)$, à savoir

$$F'(x^*)_{N^*N^*} - F'(x^*)_{N^*I^*} F'(x^*)^{-1}_{I^*I^*} F'(x^*)_{I^*N^*}$$

est une P-matrice, c'est-à-dire que ses mineurs principaux sont tous strictement positifs, avec $I^* = \{i \in \underline{n} : i > m \text{ ou } x_i^* > 0\}$ *et* $N^* = \{i \in \underline{m} : x_i = 0 \text{ et } F_i(x^*) = 0\}$.

Il peut alors être prouvé sans trop de difficulté que

Théorème 3.21. *Si* x^* *est une solution fortement régulière du problème de complémentarité non linéaire mixte* (3.3.54)*, alors tous les éléments de* $\partial H(x^*)$ *sont non singuliers. En particulier,* x^* *est un point BD-régulier de* H.

Démonstration. Voir [11], théorème 3.4. □

3.3.6. Résultats numériques

Nous présentons maintenant les résultats numériques obtenus par M. Ulbrich avec cet algorithme. Ses tests sont basés sur 40 problèmes de complémentarité mixte issus de la librairie MCPLIB [8], c'est-à-dire des problèmes qui peuvent s'écrire sous la forme (3.3.54). Ces problème ont été reformulé grâce à la fonction de Fischer-Burmeister pénalisée

$$\phi_{CCK2}(x, y) = (1 - q)\phi_{FB}(x, y) - qx_+y_+, \quad (0 \leqslant q < 1),$$

de sorte que

$$\begin{cases} H_i(x) \overset{\text{def}}{=} \phi_{CCK2}(x_i, F_i(x)) & \text{pour } i = 1, \ldots, m\,; \\ H_i(x) \overset{\text{def}}{=} F_i(x) & \text{pour } i = m+1, \ldots, n. \end{cases}$$

Les problèmes se posait alors sous la forme

$$H(x) = 0,\ x \in X.$$

Toutefois, nous pouvons nous interroger sur la nécessité de garder la contrainte $x \in X$, étant donné que $H(x) \neq 0$ sur X^c. Les tests furent effectués sur les deux versions (avec ou sans contraintes) pour voir si l'algorithme s'en sortait mieux avec l'une ou l'autre. Deux valeurs du paramètre q ont aussi été prises en compte pour savoir s'il vaut mieux rester proche de la fonction de Fischer-Burmeister ($q = 0.05$) ou s'il est préférable de prendre davantage en compte la partie x_+y_+ de la fonction de complémentarité ϕ_{CCK2}.

Quatre variantes furent ainsi mises en compétition :

- Algo. 1 : l'algorithme 3.4 avec $q = 0,3$ appliqué à la version avec contraintes ;
- Algo. 2 : l'algorithme 3.4 avec $q = 0,05$ appliqué à la version avec contraintes ;
- Algo. 3 : l'algorithme 3.4 avec $q = 0,3$ appliqué à la version sans contraintes ;
- Algo. 4 : l'algorithme 3.4 avec $q = 0,05$ appliqué à la version sans contraintes.

Comme la jacobienne F' était disponible, les matrices M_k ont été choisies parmi les éléments de $\partial_B H(x_k)$ avec la précision machine. Néanmoins, lorsque M_k était mal conditionnée, un petit multiple de l'identité fut ajouté au produit $M_k^T M_k$. L'algorithme reprend le pas 3.3. dans la

version du paragraphe (III) de la sous-section précédente. Au cas où s_k^P ne satisfaisait pas les conditions, on a procédé à une résolution complète du sous-problème grâce au solveur BPMPD [21]. Le critère de terminaison avec réussite qui fut choisi est

$$\max\left(\{|\min\left(x_i, F_i(x)\right)| : i \in \underline{m}\} \cup \{|F_i(x)| : m < i \leqslant n\}\right) \leqslant 10^{-6}.$$

Notez qu'il ne s'annule que si $|\min\left(x_i, F_i(x)\right)| \simeq 0$ pour $i \in \underline{m}$ et si $|F_i(x)| \simeq 0$ pour $m < i \leqslant n$, c'est-à-dire si $H(x) \simeq 0$. Un autre critère de terminaison fut aussi défini pour marquer que l'algorithme a échoué :

$$\Delta_k \leqslant 10^{-10} \quad \text{ou} \quad k \geqslant 200.$$

Les paramètres de la région de confiance ont pris les valeurs suivantes : $\Delta_{\min} = 1$, $\Delta_0 = 100$, $\eta_1 = 10^{-4}$, $\eta_2 = 0,75$, $\gamma_1 = \frac{1}{2}$, $\gamma_2 = 2$ et

$$\Delta_{k+1} = \begin{cases} \gamma_1 \Delta_k & \text{si } \rho_k \leqslant \eta_1, \\ \max\left(\Delta_{\min}, \Delta_k\right) & \text{si } \eta_1 < \rho_k < \eta_2, \\ \max\left(\Delta_{\min}, \gamma_2 \Delta_k\right) & \text{si } \rho_k \geqslant \eta_2. \end{cases}$$

Les paramètres pour la décroissance non monotone furent fixés à $m = 4$, $\lambda = 0,01$ et

$$\lambda_{k,r} = \begin{cases} \lambda & \text{si } r \neq r_k\,; \\ 1 - (m_k - 1)\lambda & \text{si } r = r_k\,; \end{cases}$$

où $r_k = \operatorname{argmax}_{0 \leqslant r < m_k} h(x_{j_i - r})$. Les paramètres de préconditionnement furent quant à eux pris comme $\gamma = 1 = \kappa_D$.

Un point de départ est fourni par MATLAB pour chaque problème de MCPLIB. Toutefois, il s'est avéré que pour les versions avec contraintes il est plus avantageux de prendre un point à l'intérieur du domaine admissible : le point de départ a donc été projeté préalablement sur $[l + 0,1; u - 0,1]$ pour ces versions. Selon M. Ulbrich, le fait de partir d'un point intérieur permet à l'algorithme de déterminer plus efficacement quelles sont les contraintes actives. Il insiste toutefois sur le fait que cette projection du point de départ sur $[l+0,1; u-0,1]$ n'apporte pas d'amélioration significative.

Les résultats complets sont repris dans la table 3.1

Pour faciliter leur analyse, nous avons construit les profils de performance correspondant aux deus premières variables mesurées : le nombre d'itérations réussies (figure 3.9) et le nombre total d'itérations (figure 3.10). Concernant le nombre de sous-problèmes complètement résolus — c'est-à-dire pour lequel le pas de type Newton projeté s_k^P n'a pas été accepté — nous ne présentons pas le profil de performance correspondant, car il serait faussé par la présence de nombreux zéros, c'est-à-dire de performances qui sont infinies !

Nous remarquons ainsi que les versions avec contraintes sont plus efficaces et plus robustes que celles sans contraintes ; nous pouvons aussi observer que le choix $q = 0,3$ est préférable

	Itér. réussies				Itér.				Quad. prob.			
Algorithme	1	2	3	4	1	2	3	4	1	2	3	4
badfree	4	5	4	5	4	5	4	5	0	0	0	0
bertsekas	19	19	20	20	43	48	40	39	27	31	21	23
billups	–	–	–	–	–	–	–	–	–	–	–	–
colvdual	9	12	21	12	9	12	30	14	0	0	9	0
colvnlp	8	12	11	12	8	12	12	14	0	0	0	0
cycle	5	5	7	7	5	5	7	7	0	0	0	0
degen	4	5	4	5	4	5	4	5	0	0	0	0
duopoly	11	–	12	–	11	–	12	–	9	–	10	–
ehl_k40	11	11	24	–	11	11	45	–	0	0	5	–
ehl_k60	11	13	32	31	11	13	57	65	2	0	8	12
ehl_k80	12	12	25	–	12	12	46	–	1	0	4	–
ehl_kost	12	13	–	–	31	13	–	–	22	0	–	–
freebert	8	10	12	22	8	17	23	83	0	0	0	66
games	12	–	–	20	22	–	–	40	13	–	–	21
hanskoop	9	10	11	10	24	19	31	19	6	2	9	2
hydroc06	12	12	7	7	22	19	7	7	8	5	1	1
hydroc20	12	12	11	11	13	15	12	12	1	2	0	0
jel	6	6	7	7	6	6	7	7	0	0	0	0
josephy	5	5	6	7	5	11	6	7	1	0	1	0
kojshin	6	6	7	6	12	12	7	11	0	0	0	0
mathinum	7	7	8	4	14	14	18	4	0	0	1	0
mathisum	5	6	5	5	5	6	12	12	0	0	0	0
methan08	4	4	4	4	4	4	4	4	0	0	0	0
nash	8	8	8	8	8	8	8	8	0	0	0	0
ne-hard	–	–	–	–	–	–	–	–	–	–	–	–
pgvon105	–	–	–	–	–	–	–	–	–	–	–	–
pgvon106	–	–	–	–	–	–	–	–	–	–	–	–
powell	7	9	7	9	7	16	7	16	0	0	0	0
powell_mcp	6	6	6	6	6	6	6	6	0	0	0	0
qp	4	5	3	5	4	5	3	5	0	0	0	0
scarfanum	9	12	9	11	9	12	15	19	2	0	1	1
scarfasum	8	9	7	7	8	9	18	18	1	0	0	0
scarfbnum	12	15	18	20	12	21	18	31	1	3	1	5
scarfbsum	11	15	14	14	11	22	14	24	4	8	4	7
shubik	16	25	25	32	31	62	69	82	23	51	59	70
simple-ex	–	–	–	–	–	–	–	–	–	–	–	–
simple-red	11	11	10	10	11	11	10	10	0	0	0	0
sppe	8	9	8	8	8	9	8	8	0	0	0	0
tinloi	9	7	8	7	9	7	27	7	0	0	3	0
tobin	8	10	11	14	11	16	15	17	0	0	3	0
Moyenne	8,1	9,4	10,0	10,5	10,7	13,8	16,1	18,3	2,5	3,3	3,9	6,0

TAB. 3.1. — *Résultats numériques de l'algorithme de M. Ulbrich (la moyenne est prise sur les seuls problèmes résolus par les quatre algorithmes).*

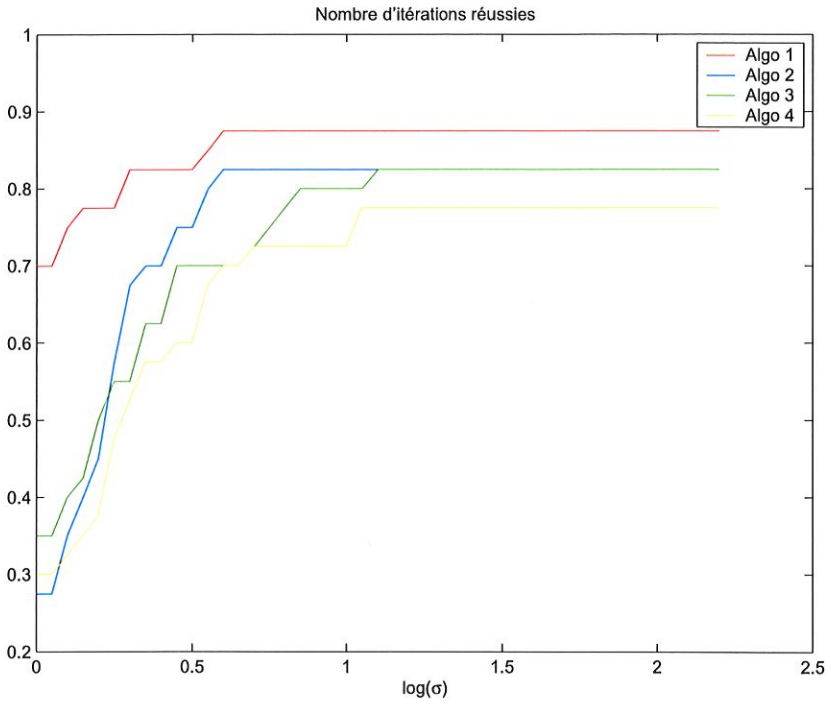


FIG. 3.9. — *Profil de performance sur base du nombre d'itérations réussies.*

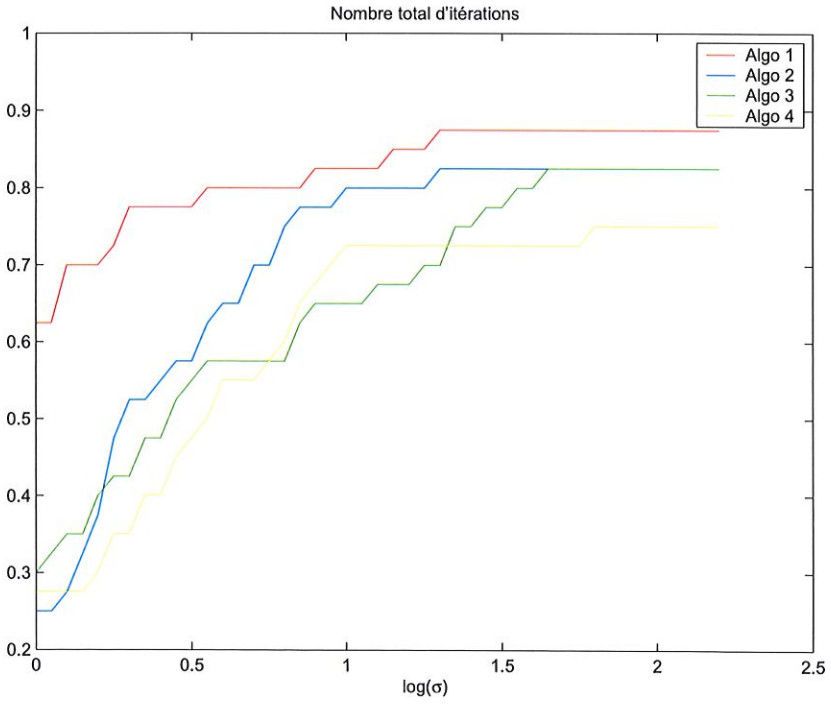


FIG. 3.10. — *Profil de performance sur base du nombre total d'itérations.*

$q = 0,05$ tant sur le plan de l'efficacité que de la robustesse

Avec une version monotone avec contrainte et $q = 0,3$, l'algorithme échoue pour 9 problèmes au lieu de 5 autrement. Le choix de la non-monotonicité permet donc ici de résoudre davantage de problèmes.

Chapitre 4.

Les méthodes multi-niveaux

Un nombre important de problèmes d'optimisation sont formulés dans des espaces de dimension infinie ou de dimension élevée. Nous pensons ici aux problèmes de contrôle des systèmes dynamiques typiquement décrits par des équations différentielles ordinaires ou aux dérivées partielles, mais aussi aux nombreuses problématiques de l'assimilation de données dans les domaines de la météorologie, de l'exploitation de l'énergie nucléaire ou de l'hydrologie.

Cette grande dimension nous pousse à les discrétiser pour être capable de les résoudre plus facilement. Deux méthodes sont en pratique souvent utilisées : les différences finies et les éléments finis. Mais, la réduction de dimension peut intervenir via d'autres moyens comme par exemple, une simplification du modèle physique qui se trouve derrière les équations.

Mais de la sorte, différents niveaux sont en général possible et nous souhaiterions tirer profit de cette diversité afin d'obtenir des algorithmes performants. Nous ne voulons pas choisir un seul niveau auquel le problème serait traité, mais plusieurs, en passant intelligemment de l'un à l'autre afin de tirer profit des avantages que chacun peut nous offrir.

Les problèmes rencontrés s'expriment généralement sous la forme d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ à annuler ou à optimiser. Les divers niveaux de discrétisation sont des fonctions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ ($i \in \underline{r}$) avec typiquement $n_i \leqslant n_{i+1} \leqslant n$, $\forall\ i \in \underline{r-1}$; elles représentent la fonction f aux différents niveaux, par exemple sur des grilles de discrétisation de plus en plus fines (on parle alors d'algorithmes multi-grilles).

Nous supposons ici que plus le problème est représenté à un niveau grossier (i petit), plus il est facile à résoudre. Afin de pouvoir passer d'un niveau à un autre, deux opérateurs sont nécessaires : un opérateur de restriction R_i pour descendre d'un niveau fin vers un niveau grossier (de $\mathbb{R}^i$ à $\mathbb{R}^{i-1}$) ainsi qu'un opérateur de prolongation P_i pour remonter d'un niveau grossier à un niveau plus fin (de $\mathbb{R}^{i-1}$ à $\mathbb{R}^i$). En règle générale, ces deux familles d'opérateurs sont linéaires, de rang plein et satisfont la condition de Galerkin :

$$\sigma_i P_i \quad = \quad R_i^T$$

pour un certain $\sigma_i > 0$ ($i \in \underline{r}$). Les méthodes de résolution que nous pouvons raisonnablement

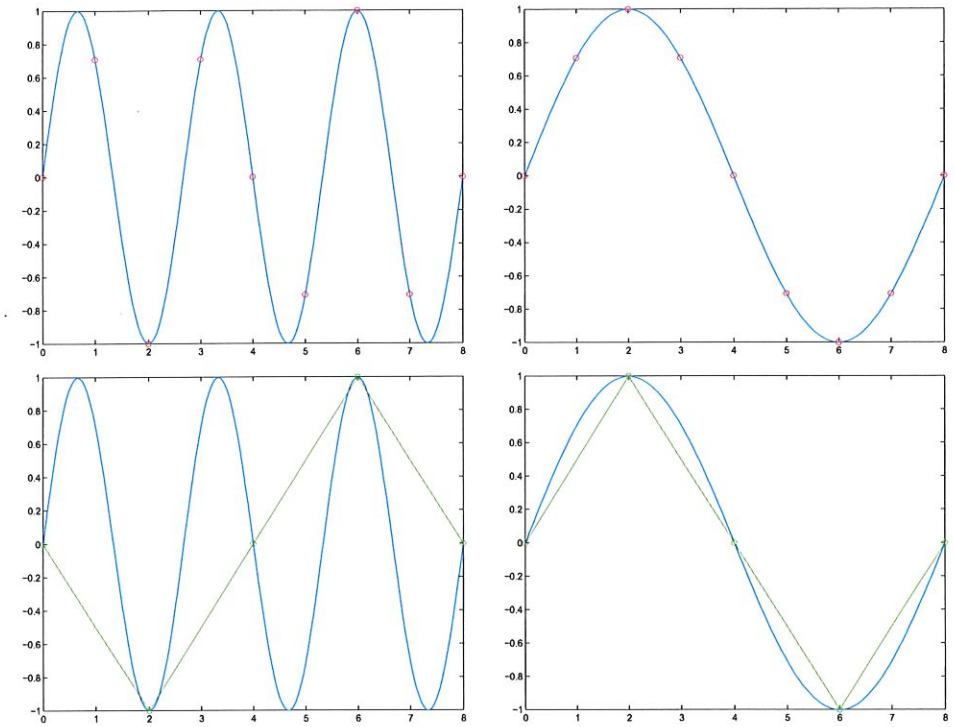


FIG. 4.1. — *Comparaison de modes à haute et basse fréquence sur des grilles fine (en haut) ou grossière (en bas).*

envisager aux différents niveaux sont des méthodes itératives. En effet, elles permettent d'avoir rapidement une ébauche de la solution, tandis que les méthodes directes sont obligées de résoudre complètement le problème avant de fournir une solution. Ces dernières ne pourront donc être utiles qu'aux niveaux les plus grossiers où le coût de résolution est relativement faible. Le choix de méthodes itératives est aussi justifié par notre volonté de tirer parti des informations aux différents niveaux de manière judicieuse : ainsi, il est possible de ne réaliser que quelques itérations à un niveau, puis de passer à un autre qui se révèle plus intéressant pour avancer dans la résolution.

Il convient d'ailleurs de remarquer une différence importante entre les niveaux fins et les niveaux grossiers, à savoir qu'il est impossible de représenter des hautes fréquences à un niveau grossier, mais que par contre, les basses fréquences y apparaissent comme ayant une fréquence plus élevée (en ce sens que nombre de points entre deux crêtes diminue). Cette observation va s'avérer fort utile dans un cadre multi-niveaux puisqu'en général, les méthodes de résolution classiques (Jacobi, Gauss-Seidel, etc.) éliminent rapidement des composantes hautes fréquences de l'erreur commise sur la solution, mais peinent à faire de même avec ses composantes basses fréquences, ce qui les empêche de converger rapidement.

Le motif de base des algorithmes multi-niveaux comprend donc deux niveaux. Ils partent en général d'un point de départ au niveau fin, soit $\hat{x}_0$; là, quelques itérations sont effectuées

pour éliminer les modes hautes fréquences de l'erreur. Ensuite, pour améliorer la résolution qui devient lente au niveau fin, on prend la restriction de l'itérée courante au niveau grossier $\check{x}_i = R\hat{x}_i$. À ce niveau, on résout — partiellement ou totalement — le problème grossier, ce qui conduit à une solution $\check{x}_j$ qui est prolongée au niveau fin $\hat{x}_j = P\check{x}_j$. Enfin, quelques itérations de lissage peuvent encore être réalisées pour améliorer la solution après la réapparition de modes hautes fréquences au niveau fin lors de la remontée. Avec les conventions L : lissage, R : restriction, P : prolongation et S : résolution complète, ce schéma se représente graphiquement comme ceci :

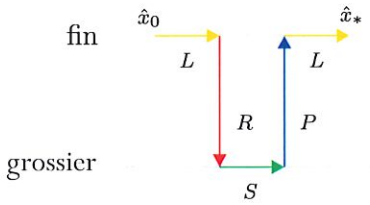


FIG. 4.2. — *Motif de base des techniques multi-niveaux.*

La solution obtenue n'est alors probablement pas satisfaisante. C'est pourquoi nous allons assembler ces motifs de base horizontalement et/ou verticalement.

Dans le premier cas, nous choisissons comme point de départ, la solution du motif précédent comme illustré ci-dessous :

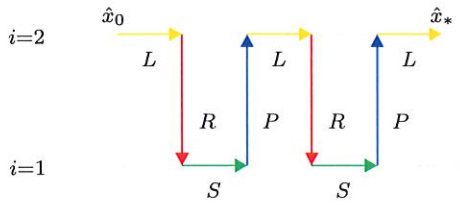


FIG. 4.3. — *Assemblage horizontal du motif de base*

Ainsi, après être remonté la première fois au niveau fin, on applique quelques étapes de lissage (de l'erreur), puis on descend à nouveau au niveau grossier comme précédemment. Là, on résout le problème grossier dont on prolonge la solution au niveau fin et ainsi de suite.

Dans le cas d'une combinaison verticale, nous utilisons un niveau supplémentaire : en effet, la résolution à un niveau grossier peut encore nous sembler trop difficile à mener jusqu'au bout. L'idée est alors d'appliquer le procédé de résolution multi-niveaux au niveau grossier comme s'il s'agissait du niveau le plus fin. Dans cette perspective, au lieu de remonter au niveau supérieur après quelques itérations, nous descendons au niveau inférieur. Là, le problème est résolu — partiellement ou totalement — puis nous remontons au niveau intermédiaire où sont effectuées quelques itérations de lissage, avant de retourner au niveau fin, comme illustré dans la figure 4.4 :

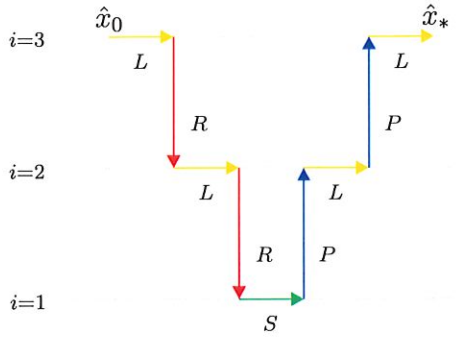


FIG. 4.4. — *Assemblage vertical du motif de base*

De la sorte, de très nombreuses combinaisons du motif de base sont possibles. Les plus connues sont le V-cycle, qui est un enchaînement exclusivement vertical du motif de base :

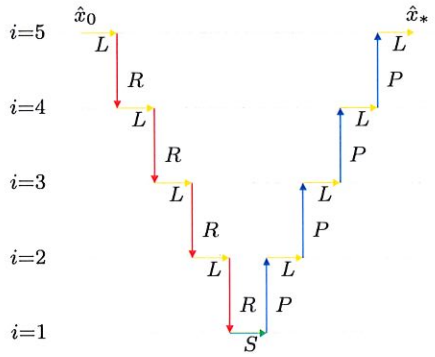


FIG. 4.5. — *Schéma multi-niveaux en V-cycle*

le W-cycle, sorte de V-cycle doublé à chaque niveau :

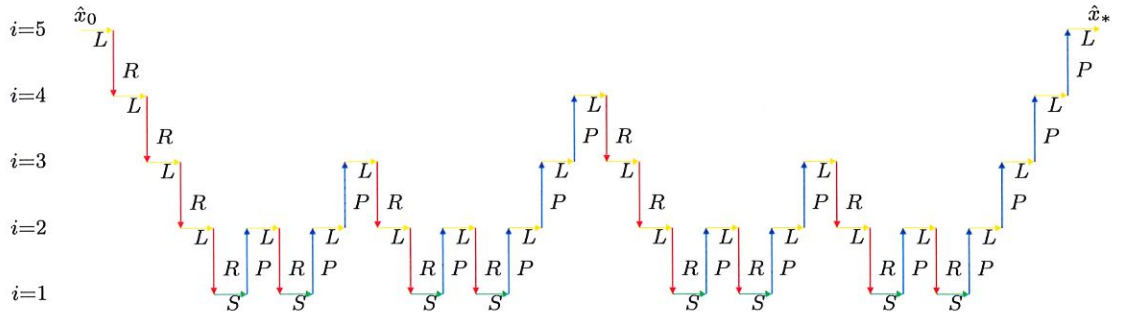


FIG. 4.6. — *Schéma multi-niveaux en W-cycle*

et le schéma multi-niveaux complet, qui part du niveau grossier et remonte progressivement en effectuant un V-cycle à partir de chaque niveau de cette remontée :

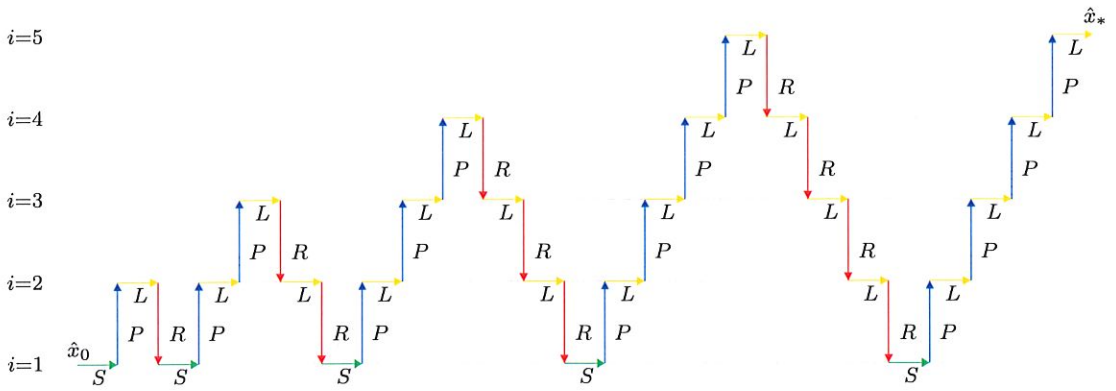


FIG. 4.7. — *Schéma multi-niveaux complet*

Au delà de ces schémas prédéfinis, il est possible de construire un algorithme multi-niveaux qui décide lui-même s'il reste à un niveau donné pour continuer la résolution ou s'il passe au niveau inférieur ou supérieur (voir [15]). Dès lors, l'algorithme travaille au niveau auquel il est le plus productif, c'est-à-dire où il réduit rapidement des composantes de l'erreur sur la solution.

Chapitre 5.

Aspects multi-niveaux des problèmes de complémentarité

5.1. Introduction

Pour créer une méthodes multi-niveaux applicable à des problèmes de complémentarité, il faut pouvoir les définir aux différents niveaux de discrétisation. Ceux-ci peuvent être inclus dans le problème lui-même, par exemple s'il s'agit de la discrétisation d'un problème de dimension infinie qui est tout naturellement réalisable à n'importe quel niveau.

Mais dans bien d'autres cas, nous ne disposons que du problème au niveau le plus fin. Deux possibilités sont dès lors envisageables :

- soit transformer le problème de complémentarité en un système d'équations $\Phi(x) = 0$ au moyen d'une fonction de complémentarité ϕ, puis à partir de là, déterminer la version du problème $\Phi(x) = 0$ aux différents niveaux ;
- soit créer une version du problème de complémentarité à chaque niveaux et seulement ensuite, reformuler chacune de ces versions au moyen d'une fonction de complémentarité.

La première approche est plus classique en ce sens qu'elle n'utilise pas le fait que le système d'équations à résoudre par une méthode multi-niveaux provient d'un problème de complémentarité. La seconde démarche est plus délicate à mettre en œuvre car il faut réussir à trouver une version cohérente du problème de complémentarité à chaque niveau.

Une autre possibilité aurait encore été d'appliquer une opération de linéarisation à l'un des stades de la transformation du problème (avant ou après la reformulation ou la discrétisation). Nous ne nous y attarderons pas ici car cela implique d'avoir un point autour duquel linéariser — typiquement l'itérée courante — et donc de changer de problème à chaque itération.

Dans ce chapitre, nous essayerons de donner une intuition des conséquences de ces choix, au travers de plusieurs exemples. En général, nous nous limiterons à deux niveaux de discrétisation pour ne pas surcharger le texte d'indice et étant donné que le multi-niveaux n'est qu'une généralisation du bi-niveaux. Notre intérêt se portera en outre sur les versions du problème

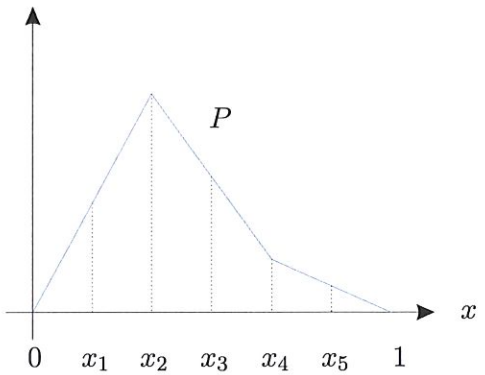


FIG. 5.1. — *Prolongation de $\mathbb{R}^2$ dans $\mathbb{R}^5$ en considérant une annulation sur les bords.*

au niveau grossier car c'est seulement à ce moment qu'une différence apparaît entre les deux approches. Enfin, pour permettre d'illustrer graphiquement les exemples développés par la suite, nous travaillerons avec les niveaux $\mathbb{R}^2$ et $\mathbb{R}^5$, ainsi qu'avec éventuellement un ensemble de fonctions suffisamment lisses pour la dimension infinie.

5.2. Formalisation des deux démarches

5.2.1. Opérateurs de prolongation et de restriction

Comme nous le verrons par la suite, le choix des opérateurs de prolongation et de restriction peut avoir de l'importance sur les résultats qu'il est possible d'obtenir à un niveau donné.

Un opérateur de prolongation habituel entre les espaces $\mathbb{R}^2$ et $\mathbb{R}^5$ est donnée par interpolation linéaire :

$$P : \mathbb{R}^2 \to \mathbb{R}^5 : y \mapsto \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 1 & 1 \\ 0 & 2 \\ 0 & 1 \end{pmatrix} y.$$

Dans la figure 5.1, cela correspondant au passage des valeurs en x_2, x_4 à celles en x_1, x_2, x_3, x_4, x_5 étant donné que la fonction modélisée par ces points est nulle aux bords (en 0 et en 1). L'opérateur de restriction correspondant selon la relation de Galerkin est alors :

$$R : \mathbb{R}^5 \to \mathbb{R}^2 : x \mapsto \frac{1}{4}\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 \end{pmatrix} x$$

vérifiant $R^T = \frac{1}{2}P$. Cet opérateur de restriction R signifie que la valeur de la fonction en un point du niveau grossier est la moyenne entre celle au même point au niveau fin et sa valeur approximée linéairement à partir de ses deux voisins ; par exemple, la valeur en x_2 au niveau grossier est égale à la moyenne entre u_2 et $\frac{u_1+u_3}{2}$ avec u_i, la valeur de la fonction au niveau fin en x_i. Nous avons représenté cela dans la figure 5.2.

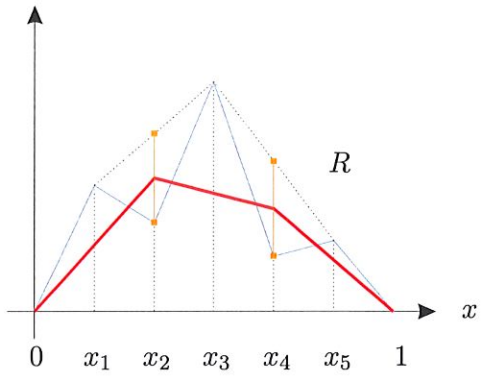


FIG. 5.2. — *Restriction de $\mathbb{R}^5$ dans $\mathbb{R}^2$ en considérant une annulation sur les bords.*

De façon générale, pour passer de $\mathbb{R}^n$ à $\mathbb{R}^{2n+1}$ l'opérateur de prolongation par interpolation avec valeurs nulles sur les bords est :

$$P = \frac{1}{2}\begin{pmatrix} 1 & 2 & 1 & & & & & & \\ & & 1 & 2 & 1 & & & & \\ & & & & 1 & 2 & 1 & & \\ & & & & & & \ddots & & \\ & & & & & & 1 & 2 & 1 \end{pmatrix}^T . \tag{5.2.1}$$

Notez que si les conditions-frontière sont plutôt une « annulation de la dérivée » sur les bords, un choix plus judicieux est de prendre ce même opérateur de prolongation mais en remplaçant les deux valeurs « 1 » aux extrémités de P par des « 2 » :

$$P = \frac{1}{2}\begin{pmatrix} 2 & 2 & 1 & & & & & & \\ & & 1 & 2 & 1 & & & & \\ & & & & 1 & 2 & 1 & & \\ & & & & & & \ddots & & \\ & & & & & & 1 & 2 & 1 \end{pmatrix}^T . \tag{5.2.2}$$

Graphiquement, cela donne cette fois l'interpolation de la figure 5.3. La restriction correspondante est toujours donnée par la relation de Galerkin $R^T = \frac{1}{2}P$.

5.2.2. Explicitation des deux démarches

Considérons le problème de complémentarité

$$\langle f(x), g(x) \rangle , \ f(x) \geqslant 0, \ g(x) \geqslant 0, \tag{5.2.3}$$

avec $x \in \mathbb{R}^n$ et $f, g : \mathbb{R}^n \to \mathbb{R}^n$ suffisamment lisses.

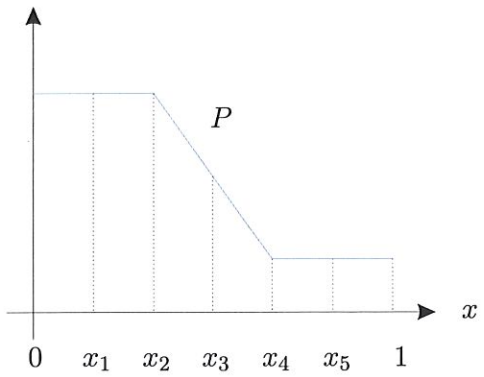


FIG. 5.3. — *Prolongation de $\mathbb{R}^2$ dans $\mathbb{R}^5$ en considérant une annulation de la dérivée sur les bords.*

Première approche. La première démarche envisagée consiste à transformer ce problème grâce à une fonction de complémentarité ϕ :

$$\phi(f_i(x), g_i(x)) = 0, \ \forall\, i \in \underline{n} \quad \Leftrightarrow \quad \Phi(x) = 0$$

avec $x \in \mathbb{R}^n$. Ensuite, nous discrétisons cette reformulation pour obtenir des versions du problème $\Phi(x) = 0$ dans les espaces moins fins. Le niveau grossier est alors donné par :

$$R\phi(f_i(Py), g_i(Py)) = 0, \ \forall\, i \in \underline{n} \quad \Leftrightarrow \quad R\Phi(Py) = 0.$$

Par facilité, nous parlerons ainsi du problème « $R\Phi$ ».

Dans la perspective de l'utilisation d'une méthode de région de confiance à un problème de minimisation, il serait aussi parfaitement concevable de considérer une version multi-niveaux du problème $\min \frac{1}{2} \|\Phi(x)\|^2$.

Seconde approche. Avec cette seconde possibilité, nous devons définir le problème de complémentarité aux différents niveaux, c'est-à-dire trouver des fonctions $\check{f}, \check{g} : \mathbb{R}^{\check{n}} \to \mathbb{R}^{\check{n}}$ (avec $\check{n} < n$, la dimension de l'espace grossier) pour définir le problème de complémentarité grossier

$$\left\langle \check{f}(y), \check{g}(y) \right\rangle, \ \check{f}(y) \geqslant 0, \ \check{g}(y) \geqslant 0,$$

où $y \in \mathbb{R}^{\check{n}}$. Une idée pour construire ces fonctions $\check{f}$ et $\check{g}$ est de les choisir égales à $Rf(P\cdot)$ et $Rg(P\cdot)$, c'est-à-dire que nous prenons la valeur des fonctions f et g en l'argument prolongé Py, pour ensuite la restreindre dans l'espace grossier. Après, nous appliquons une fonction de complémentarité à chacun des problèmes ainsi obtenus. Le problème au niveau grossier se présente alors sous la forme :

$$\phi([Rf(Py)]_i, [Rg(Py)]_i) = 0, \ \forall\, i \in \underline{n}.$$

Cette fois, nous dirons qu'il s'agit du problème « ΦR ».

Nous allons maintenant montrer que ce choix est cohérent avec l'optique multi-niveaux classique. Considérons le problème de minimisation d'un modèle quadratique

$$q : \mathbb{R}^n \to \mathbb{R} : x \mapsto g^T x + x^T H x.$$

Nous prenons les prolongation P et restriction R habituelles pour voyager entre $\mathbb{R}^n$ et $\mathbb{R}^{\check{n}}$ avec $n = 2\check{n} + 1$, à savoir celles données par (5.2.1). Classiquement [15], l'information qui est passée au niveau inférieur est la version grossière du gradient g et du hessien H, c'est-à-dire Rg et RHP ; la version grossière du modèle est alors construite comme

$$\check{q} : \mathbb{R}^{\check{n}} \to \mathbb{R} : y \mapsto (Rg)^T y + y^T (RHP) y = \sigma g^T (Py) + \sigma (Py)^T H (Py) = \sigma q(Py),$$

c'est-à-dire, à un facteur positif près, la fonction q prise en la prolongation de l'argument y.

Si nous prenons cette fois un modèle quadratique q à valeur dans $\mathbb{R}^n$, autrement dit n fonctions

$$q_i : \mathbb{R}^n \to \mathbb{R} : x \mapsto g_i^T x + x^T H_i x \qquad (i \in \underline{n}),$$

il est possible de transmettre la même information Rg_i et RH_iP, ce qui conduit à n fonctions $\tilde{q}_i$ ($i \in \underline{n}$) allant de $\mathbb{R}^{\check{n}}$ dans $\mathbb{R}$. Hors, nous souhaitons ici trouver une version de q allant de $\mathbb{R}^{\check{n}}$ dans $\mathbb{R}^{\check{n}}$; il faut par conséquent réduire la dimension de l'ensemble d'arrivée, ou en d'autres termes, agréger l'information transmise pour qu'elle ne donne plus naissance qu'à $\check{n}$ fonctions à valeurs dans $\mathbb{R}$ au final. Pour ce faire, nous « appliquons » l'opérateur de restriction R aux vecteurs $(Rg_1, \ldots, Rg_n)$ et $(RH_1P, \ldots, RH_nP)$ en considérant leurs composantes comme de simples éléments et non comme des vecteurs ou des matrices ; nous obtenons ainsi

$$\tfrac{1}{4}\left(R(g_1 + 2g_2 + g_3), R(g_3 + 2g_4 + g_5), \ldots, R(g_{n-2} + 2g_{n-1} + g_n)\right)$$

et

$$\tfrac{1}{4}\left(R(H_1 + 2H_2 + H_3)P, R(H_3 + 2H_4 + H_5)P, \ldots, R(H_{n-2} + 2H_{n-1} + H_n)P\right).$$

Dès lors, nous avons trouvé $\check{n}$ valeurs pour les gradients et hessiens des fonctions $\check{q}_i$ ($i \in \underline{\check{n}}$) :

$$\begin{aligned} \check{q}_i : \mathbb{R}^{\check{n}} \to \mathbb{R} : y \mapsto \check{q}_i(y) &= \left(R\tfrac{g_{2i-1}+2g_{2i}+g_{2i+1}}{4}\right)^T y + y^T \left(R\tfrac{H_{2i-1}+2H_{2i}+H_{2i+1}}{4} P\right) y \\ &= \sigma \left[\left(\tfrac{g_{2i-1}+2g_{2i}+g_{2i+1}}{4}\right)^T (Py) + (Py)^T \left(\tfrac{H_{2i-1}+2H_{2i}+H_{2i+1}}{4}\right) (Py)\right] \end{aligned}$$

Par ailleurs, si nous développons $Rq(Py)$ avec $q : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto (q_i)_{i \in \underline{n}}$, nous trouvons les mêmes résultats (toujours au facteur σ près) :

$$Rq(Py) = \tfrac{1}{4}\left(q_1(Py) + 2q_2(Py) + q_3(Py), \ldots, q_{n-2}(Py) + 2q_{n-1}(Py) + q_n(Py)\right)$$

dont chacune des $\check{n}$ composante peut se développer comme ceci :

$$\begin{aligned} & \tfrac{1}{4}\left[q_{2i-1}(x) + 2q_{2i}(x) + q_{2i+1}(x)\right] \\ = \; & \tfrac{1}{4}\left[g_{2i-1}^T x + x^T H_{2i-1} x + 2(g_{2i}^T x + x^T H_{2i} x) + g_{2i+1}^T x + x^T H_{2i+1} x\right] \\ = \; & \left(\tfrac{g_{2i-1}+2g_{2i}+g_{2i+1}}{4}\right)^T (Py) + (Py)^T \left(\tfrac{H_{2i-1}+2H_{2i}+H_{2i+1}}{4}\right)(Py) \end{aligned}$$

où nous avons posé $x = Py$ pour alléger les écritures.

Finalement, avec $\check{q} = \sigma R q(Py)$, nous constatons que notre manière de définir le problème grossier est bien cohérente avec la façon d'agir habituellement en optimisation multi-niveaux. À vrai dire, il en est de même pour la première approche où nous construisons la version grossière du problème $\Phi(x) = 0$ en « pré- et postmultipliant » Φ par R et P respectivement.

Influence de la fonction ϕ. Il faut par ailleurs remarquer que les solutions de ce second problème sont indépendantes du choix de la fonction de complémentarité ϕ. En effet, l'annulation de cette fonction revient à dire que

$$[Rf(Py)]_i \cdot [Rg(Py)]_i = 0, \quad [Rf(Py)]_i \geqslant 0 \quad \text{et} \quad [Rg(Py)]_i \geqslant 0, \; \forall\, i \in \underline{n},$$

condition qui est identique quelle que soit la fonction ϕ considérée.

Au contraire, les solutions au niveau grossier peuvent être fortement influencées dans la première approche par le choix de la fonction de complémentarité comme nous l'observerons dans les exemples qui suivent.

5.3. Quelques exemples

Premier exemple : cas linéaire. Considérons le problème (5.2.3) avec les fonctions linéaires $f, g : \mathbb{R}^5 \to \mathbb{R}^5$ correspondant aux matrices

$$A = \begin{pmatrix} 1 & 0 & 6 & 4 & 7 \\ 0 & 2 & 7 & 4 & 5 \\ 6 & 6 & 7 & 2 & 6 \\ 6 & 1 & 1 & 7 & 1 \\ 0 & 4 & 5 & 7 & 5 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 7 & 2 & 0 & 2 & 3 \\ 9 & 8 & 1 & 2 & 4 \\ 3 & 9 & 6 & 10 & 4 \\ 3 & 2 & 2 & 4 & 6 \\ 9 & 2 & 8 & 3 & 1 \end{pmatrix}$$

Selon l'approche $R\Phi$, nous trouvons un problème grossier en dimension 2 qui est donnée par $R\Phi(Py) = 0$ avec $y = (y_1, y_2) \in \mathbb{R}^2$, i.e.

$$\begin{pmatrix} \frac{1}{4}\phi\left(\frac{7}{2}y_1+\frac{21}{2}y_2, \frac{11}{2}y_1+\frac{7}{2}y_2\right)+\frac{1}{2}\phi\left(\frac{11}{2}y_1+10y_2, 13y_1+\frac{9}{2}y_2\right)+\frac{1}{4}\phi\left(\frac{25}{2}y_1+\frac{17}{2}y_2, \frac{27}{2}y_1+15y_2\right) \\ \frac{1}{4}\phi\left(\frac{25}{2}y_1+\frac{17}{2}y_2, \frac{27}{2}y_1+15y_2\right)+\frac{1}{2}\phi\left(\frac{9}{2}y_1+8y_2, \frac{9}{2}y_1+8y_2\right)+\frac{1}{4}\phi\left(\frac{13}{2}y_1+12y_2, \frac{21}{2}y_1+\frac{15}{2}y_2\right) \end{pmatrix} = 0. \quad (5.3.1)$$

En revanche, suivant l'autre possibilité ΦR, les fonctions linéaires $\check{f}, \check{g} : \mathbb{R}^2 \to \mathbb{R}^2$ correspondent

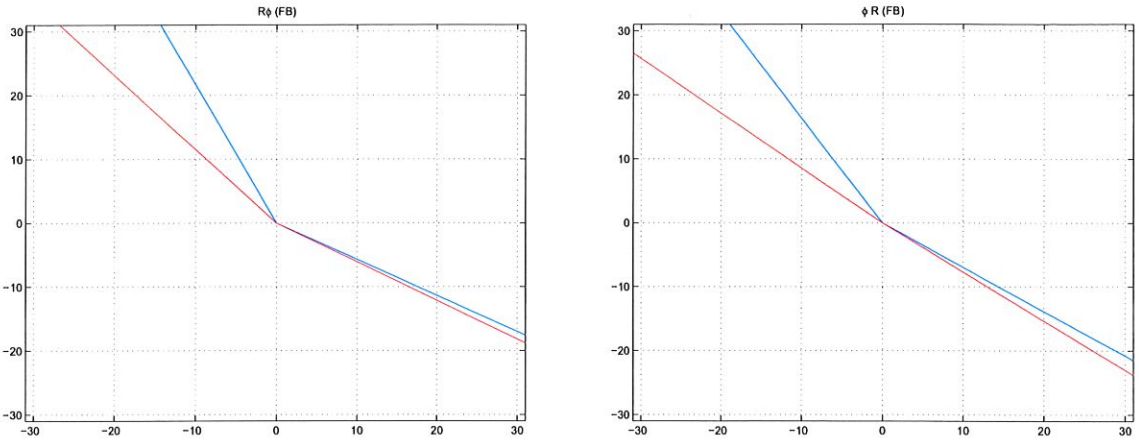


FIG. 5.4. — *Illustration des problèmes* (5.3.1) *(à gauche) et* (5.3.2) *(à droite) avec la fonction de Fischer-Burmeister.*

aux matrices

$$\check{A} = RAP = \begin{pmatrix} \frac{27}{4} & \frac{39}{4} \\ 7 & \frac{73}{8} \end{pmatrix} \quad \text{et} \quad \check{B} = RBP = \begin{pmatrix} \frac{45}{4} & \frac{55}{8} \\ \frac{33}{4} & \frac{77}{8} \end{pmatrix}.$$

Après la transformation de complémentarité, nous trouvons alors comme problème grossier :

$$\begin{pmatrix} \phi\left(\frac{27}{4}y_1 + \frac{39}{4}y_2, \frac{45}{4}y_1 + \frac{55}{8}y_2\right) \\ \phi\left(7y_1 + \frac{73}{8}y_2, \frac{33}{4}y_1 + \frac{77}{8}y_2\right) \end{pmatrix} = 0. \tag{5.3.2}$$

La plus grande simplicité d'écriture de ce problème provient du fait que les fonctions $\check{f} = RfP$, $\check{g} = RgP$ sont linéaires étant donné que les fonctions f, g le sont au départ.

Pour donner une idée du comportement graphique des problèmes ainsi obtenus, nous allons chaque fois illustrer notre propos par les courbe de niveau 0 des deux composantes de ces petits vecteurs de $\mathbb{R}^2$ à annuler (la première composante en bleu, la seconde en rouge). Les solutions du problème sont alors les points d'intersection de la courbe bleue (première composante nulle) et de la rouge (seconde composante nulle).

Pour l'exemple précédant, la figure 5.4 nous confirme que la solution du problème en dimension 2 est $(0,0)$, ce qui est logique puisqu'il s'agit d'annuler $\langle \check{f}(y), \check{g}(y) \rangle = y^T(A^TB)y$. Nous constatons que même si les deux graphes ne sont pas parfaitement identiques, ils présentent cependant la même allure et le même point d'intersection. Mais tel n'est pas toujours le cas.

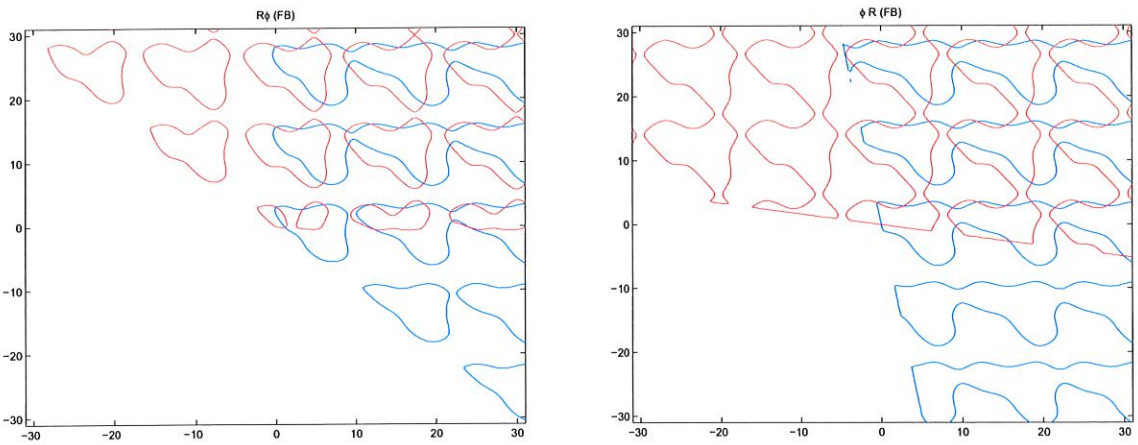


FIG. 5.5. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g définis par* (5.3.3) *et* (5.3.4) *en utilisant la fonction de Fischer-Burmeister.*

Deuxième exemple : cas trigonométrique. Prenons par exemple

$$f : \mathbb{R}^5 \to \mathbb{R}^5 : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \mapsto \begin{pmatrix} \cos x_2 \\ \sin x_3 - \cos x_1 \\ \sin x_4 + 3\cos x_5 \\ \cos x_3 - \sin x_2 \\ 2\sin x_1 \end{pmatrix} \tag{5.3.3}$$

$$\text{et} \qquad g : \mathbb{R}^5 \to \mathbb{R}^5 : (x_1, x_2, x_3, x_4, x_5) \mapsto (x_1, x_2, x_3, x_4, x_5). \tag{5.3.4}$$

Les graphiques des problèmes $R\Phi$ et ΦR correspondant à ces deux fonctions f et g sont présentés à la figure 5.5 dans le cas de la fonction ϕ de Fischer-Burmeister. Nous donnons aussi ceux pour la fonction de Chen, Chen et Kanzow pour les valeurs 0,2 ; 0,5 et 0,8 du paramètre q (figures 5.6, 5.7 et 5.8). Comme les solutions du problème ΦR sont indépendantes de cette fonction ϕ et étant donné que les courbes trouvées sont rigoureusement identiques, nous ne les dessinerons plus à l'avenir que pour une seule fonction de complémentarité.

Nous observons des motifs complexes variant de figure en figure ; il n'y a par ailleurs pas forcément de correspondance entre les solutions découvertes sur l'une ou l'autre illustration, tout au plus une certaine proximité.

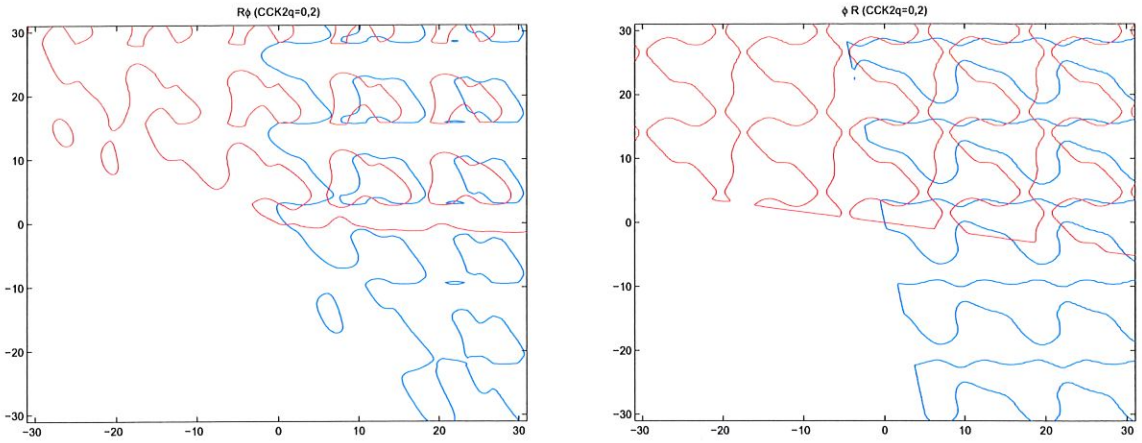


FIG. 5.6. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g définis par* (5.3.3) *et* (5.3.4) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,2$.*

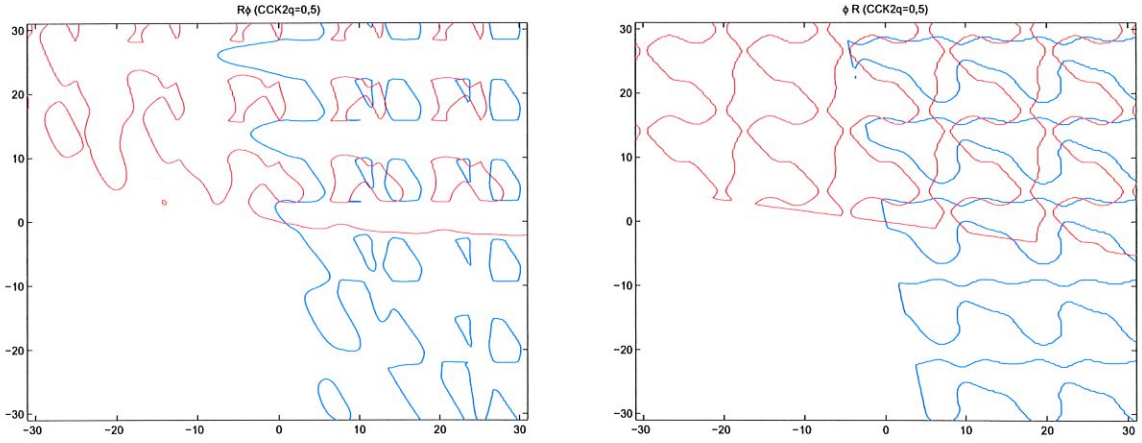


FIG. 5.7. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g définis par* (5.3.3) *et* (5.3.4) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,5$.*

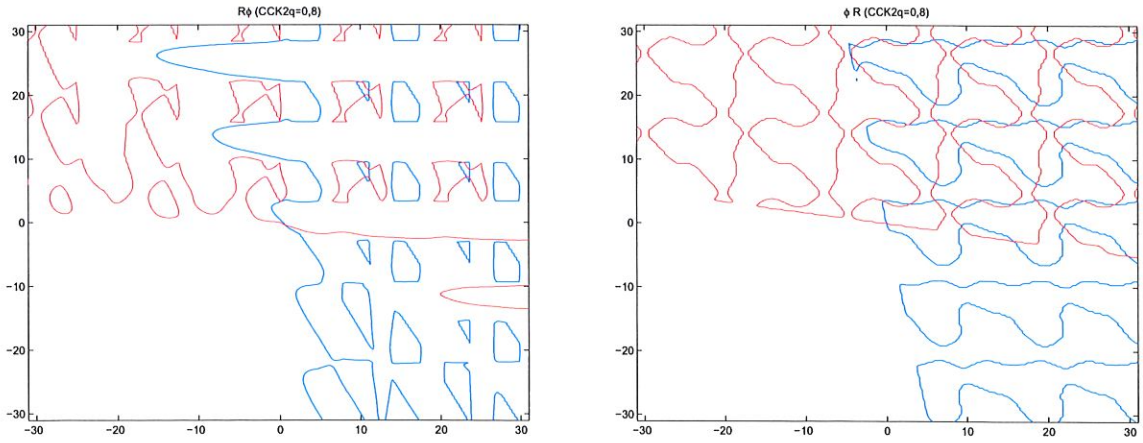


FIG. 5.8. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g définis par* (5.3.3) *et* (5.3.4) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,8$.*

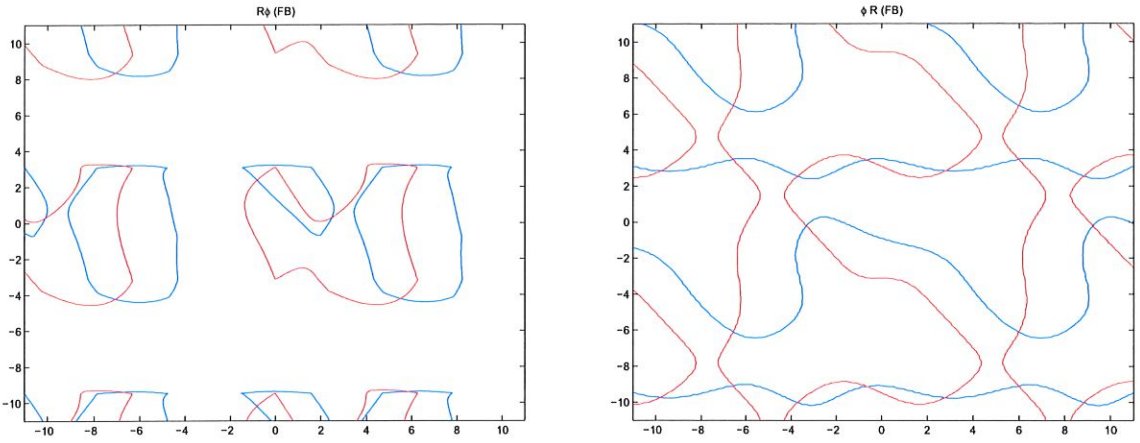


FIG. 5.9. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g égaux définis par* (5.3.3) *en utilisant la fonction de Fischer-Burmeister.*

Troisième exemple : cas trigonométrique (*bis*). Nous avons ici choisi de prendre $g := f$ (avec la même définition pour f que dans l'exemple précédent) pour voir ce qui se passait alors. Remarquez que le problème de complémentarité se réduit alors à

$$\|f(x)\|^2 = 0, f(x) \geqslant 0 \quad \Leftrightarrow \quad f(x) = 0.$$

Il est facile de montrer que contrairement au précédent, cet exemple n'a pas de solution : considérons les quatre composantes de f faisant intervenir les variables x_1, x_2 et x_3 ; pour les annuler toutes, il faudrait avoir $\cos^2(x_1) = 1 = \sin^2(x_2)$, d'où $\cos^2(x_3) + \sin^2(x_3) = \sin^2(x_2) + \cos^2(x_1) = 2$, ce qui est absurde !

Il faut cependant noter que de nombreuses solutions sont pourtant présentes au niveau grossier à cause, selon nous, du brassage des composantes qu'effectuent les opérateurs R et P.

Les figures 5.9, 5.10 et 5.11 illustrent les problèmes $R\Phi$ et ΦR pour cet exemple. Nous y remarquons directement le comportement périodique attendu vu la forme des fonctions f et g.

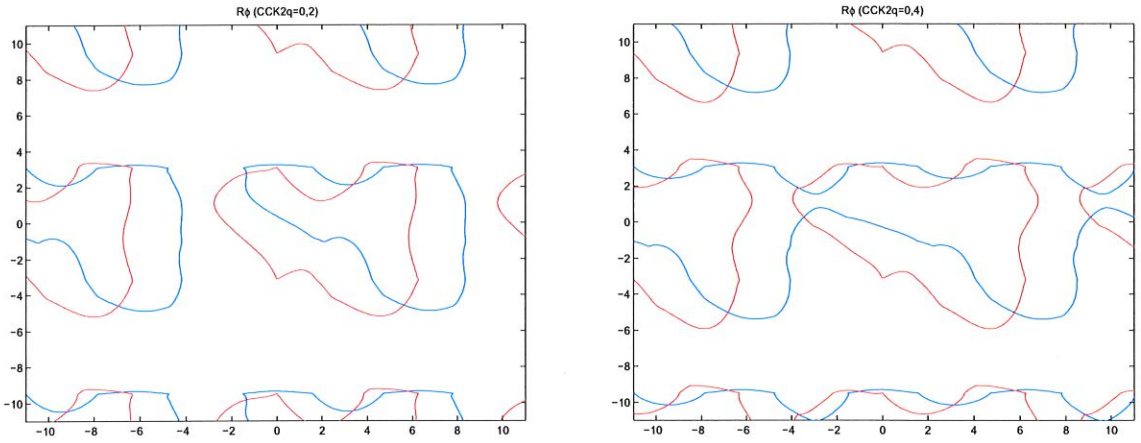


FIG. 5.10. — *Illustration du problèmes RΦ pour des f et g égaux définis par* (5.3.3) *en utilisant la fonction de Chen-Chen-Kanzow avec* $q = 0,2$ *(à gauche) et* $q = 0,4$ *(à droite).*

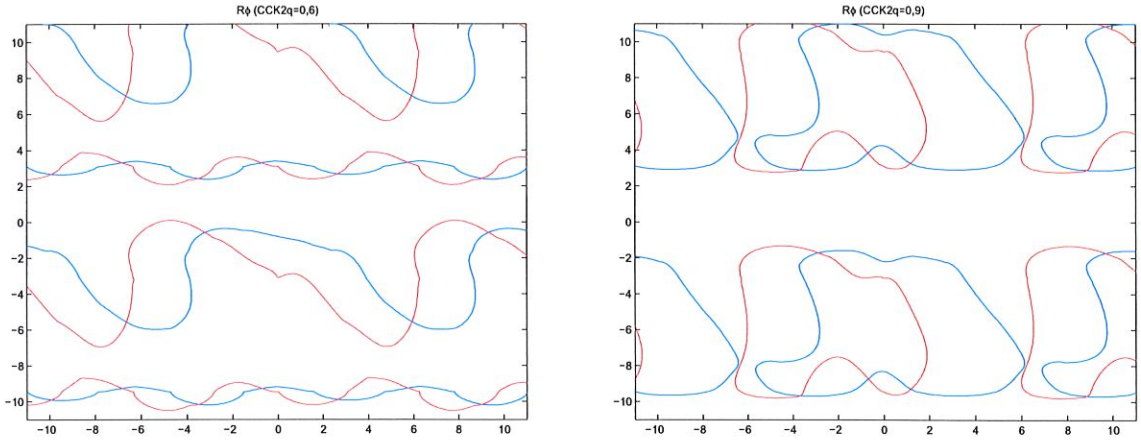


FIG. 5.11. — *Illustration du problème RΦ pour des f et g égaux définis par* (5.3.3) *en utilisant la fonction de Chen-Chen-Kanzow avec* $q = 0,6$ *(à gauche) et* $q = 0,9$ *(à droite).*

Quatrième exemple : cas quadratique. Nous allons considérer ici la discrétisation d'un problème de dimension infinie. Soient $\mathcal{H} = L_2[0,1]$, l'espace de Hilbert des fonctions de carré intégrable sur l'intervalle $[0,1]$, ainsi que les deux opérateurs $F, G : \mathcal{H} \mapsto \mathcal{H}$ définis par

$$\begin{aligned} & [F(u)](x) = u^2(x) - 5, \quad \forall\, u \in \mathcal{H},\ \forall\, x \in [0,1] \\ \text{et} \quad & [G(u)](x) = 3 - u(x), \quad \forall\, u \in \mathcal{H},\ \forall\, x \in [0,1]. \end{aligned}$$

Le problème de complémentarité en dimension infinie (en la variable $u \in \mathcal{H}$) s'écrit :

$$\langle F(u), G(u) \rangle = 0, \quad F(u) \geqslant 0, \quad G(u) \geqslant 0,$$

où le produit scalaire est donné par $\langle u, v \rangle = \int_0^1 u(x)v(x)\,dx$ $(u, v \in \mathcal{H})$ et la positivité par $u \geqslant 0 \Leftrightarrow u(x) = 0,\ \forall\, x \in [0,1]$. Notez que l'annulation de l'intégrale signifie ici que

$$[F(u)](x) \cdot [G(u)](x) = 0$$

pour presque tout $x \in [0,1]$ et même pour tous, si l'espace $\mathcal{H}$ est restreint aux fonctions continues, ce que nous supposerons réalisé par la suite ; ainsi dès à présent, $\mathcal{H} = C[0,1]$.

Les fonctions solutions de ce problème sont alors évidentes par continuité :

$$u(x) \equiv 3,\ u(x) \equiv \sqrt{5} \text{ ou } u(x) \equiv -\sqrt{5}.$$

Nous considérons à présent la discrétisation de ce problème sur $\mathbb{R}^5$ comme dans la figure 5.1. Les fonctions f et g correspondant à ces opérateurs sur $\mathcal{H}$ sont définies par :

$$\begin{aligned} & f : \mathbb{R}^5 \to \mathbb{R}^5 : (x_1, x_2, x_3, x_4, x_5) \mapsto (x_1^2 - 5, x_2^2 - 5, x_3^2 - 5, x_4^5 - 5, x_5^2 - 5) \quad (5.3.5) \\ \text{et} \quad & g : \mathbb{R}^5 \to \mathbb{R}^5 : (x_1, x_2, x_3, x_4, x_5) \mapsto (3 - x_1, 3 - x_2, 3 - x_3, 3 - x_4, 3 - x_5). \quad (5.3.6) \end{aligned}$$

Le problème de complémentarité qui en découle s'exprime comme :

$$(x_i^2 - 5)(3 - x_i) = 0,\ x_i^2 - 5 \geqslant 0,\ 3 - x_i \geqslant 0, \quad \forall\, i \in \underline{5}$$

et ses solutions sont données par $x_i = 3$, $\sqrt{5}$ ou $-\sqrt{5}$ $(i \in \underline{5})$. Nous constatons donc une multiplication des solutions par rapport au problème initial, puisque toute les combinaisons sont possibles.

La discrétisation des fonctions f et g au niveau inférieur va être effectuée via le procédé que nous avons décrit précédemment : $\check{f} = Rf(P\cdot)$ et $\check{g} = Rg(P\cdot)$ avec P et R toujours définis

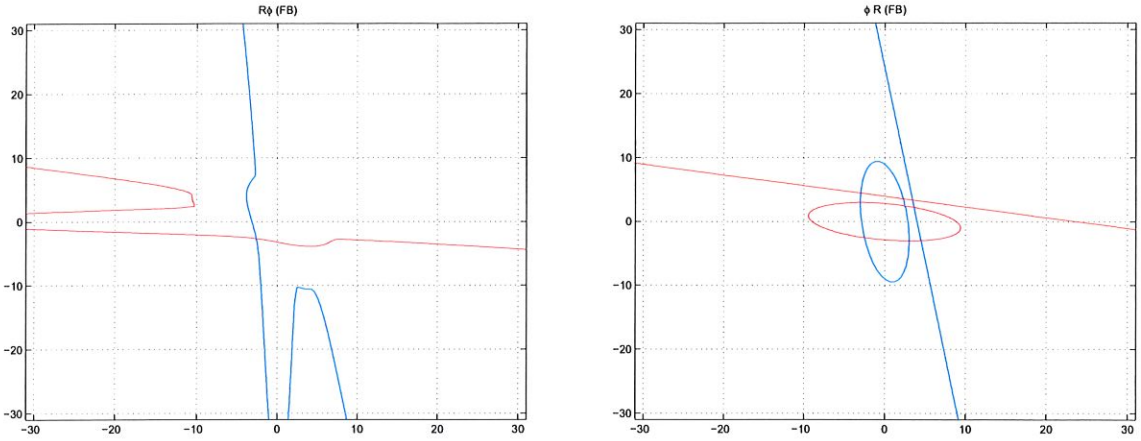


FIG. 5.12. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g définis par* (5.3.5) *et* (5.3.6) *en utilisant la fonction de Fischer-Burmeister.*

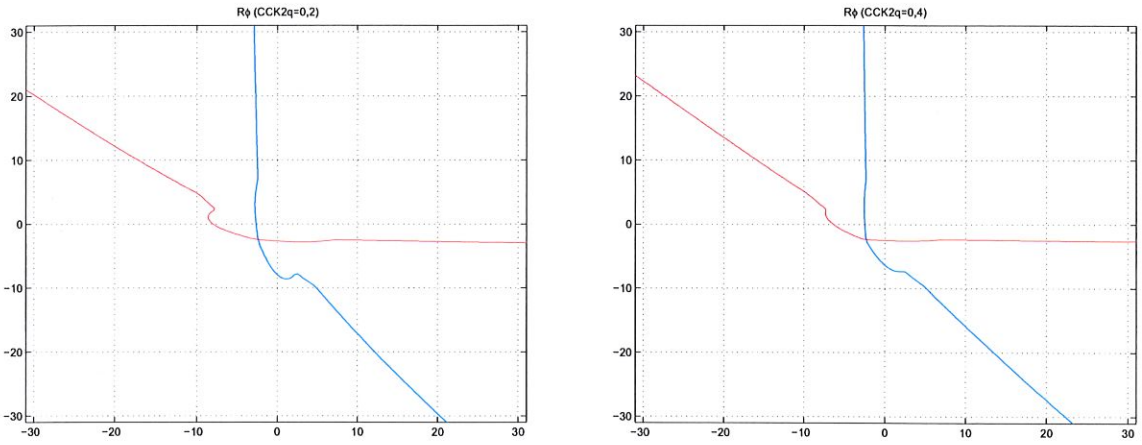


FIG. 5.13. — *Illustration du problèmes $R\Phi$ pour des f et g définis par* (5.3.5) *et* (5.3.6) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,2$ (à gauche) et $q = 0,4$ (à droite).*

par (5.2.1). Nous obtenons ainsi

$$\check{f}(y_1, y_2) = \frac{1}{16}(10y_1^2 + 2y_1y_2 + y_2^2 - 80, y_1^2 + 2y_1y_2 + 10y_2^2 - 80) \tag{5.3.7}$$

$$\text{et} \quad \check{g}(y_1, y_2) = \frac{1}{8}(24 - 6y_1 - y_2, 24 - y_1 - 6y_2). \tag{5.3.8}$$

Nous ne retrouvons pas dans ces deux fonctions, la même forme que celle que nous aurions obtenue à partir du problème initial de dimension infinie ; il reste néanmoins une bonne partie des $y_i^2 - 5$ dans les composantes de $\check{f}$ et des $3 - y_i$ dans celles de $\check{g}$. Nous observons ainsi que les opérateurs R et P mélangent légèrement les composantes en incorporant chaque fois un peu d'information provenant les composantes voisines au niveau fin. Cela aura notamment pour effet de créer un effet de lissage (lorsque P est comme ici, une fonction d'interpolation).

Les graphiques correspondants à ce problème quadratique sont présentés aux figures 5.12,

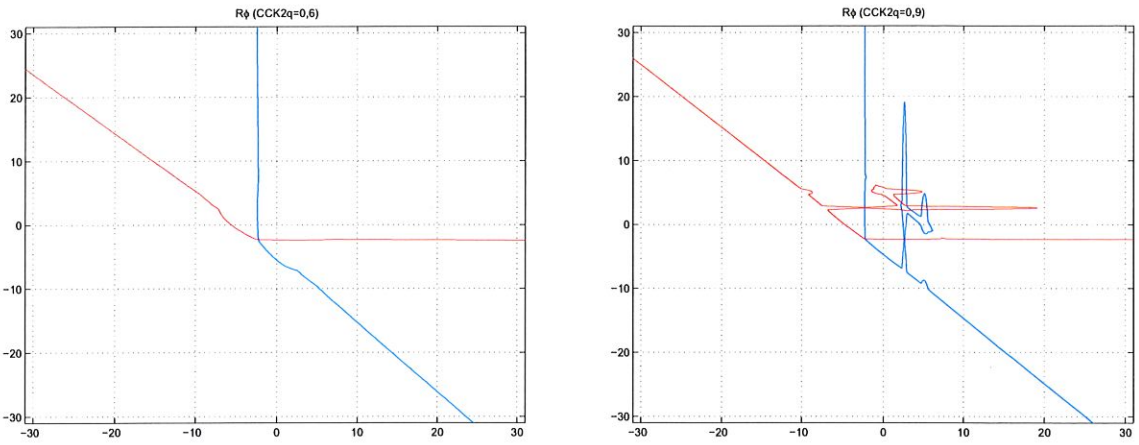


FIG. 5.14. — *Illustration du problème $R\Phi$ pour des f et g définis par* (5.3.5) *et* (5.3.6) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,6$ (à gauche) et $q = 0,9$ (à droite).*

5.13 et 5.14. Nous remarquons que pour le problème $R\Phi$, la seule solution trouvée est proche de $(-\sqrt{5}, -\sqrt{5})$ mais que les deux autres solutions théoriques du problème sont oubliées, sauf pour des valeurs forts élevées ($q > 0,8$) du paramètre de la fonction de Chen-Chen-Kanzow ; elles approximent alors $(\sqrt{5}, \sqrt{5})$ et $(3, 3)$.

Du côté du problème ΦR, les trois solutions situées sur la première bissectrice du plan sont approximativement retrouvées, mais six autres solutions apparaissent. Elles correspondent en réalité au phénomène de multiplication que nous évoquions précédemment du fait de la perte de la continuité en discrétisant. La forme des courbes de niveau (deux ellipses et deux droites) correspond exactement à l'annulation des composantes de $\check{f}$ et $\check{g}$, étant donné qu'ici, les contraintes de positivité sont satisfaites dès que le produit scalaire est nul.

L'erreur sur les solutions qui existe dans ces deux problèmes est due à l'impossibilité de trouver un vecteur $(x, x, x, x, x) \in \mathbb{R}^5$ dans Im P, alors que toutes les véritables solutions se présentent sous cette forme. Pour parvenir aux solutions correctes, il faut utiliser la seconde prolongation dont nous avons parlé à la sous-section précédente. Avec celle-ci, l'image de (x, x) par P est (x, x, x, x, x), c'est-à-dire de la forme de la solution.

Avec ce nouvel opérateur de prolongation (5.2.2), nous obtenons les illustrations présentées aux figures 5.15, 5.16 et 5.17. Nous remarquons qu'il n'y pas énormément de différence entre les deux familles de graphiques, mais que néanmoins, la modification que nous avons réalisé, a porté ses fruits : les trois solutions $(-\sqrt{5}, -\sqrt{5})$, $(\sqrt{5}, \sqrt{5})$ et $(3, 3)$ sont cette fois trouvées de façon précise.

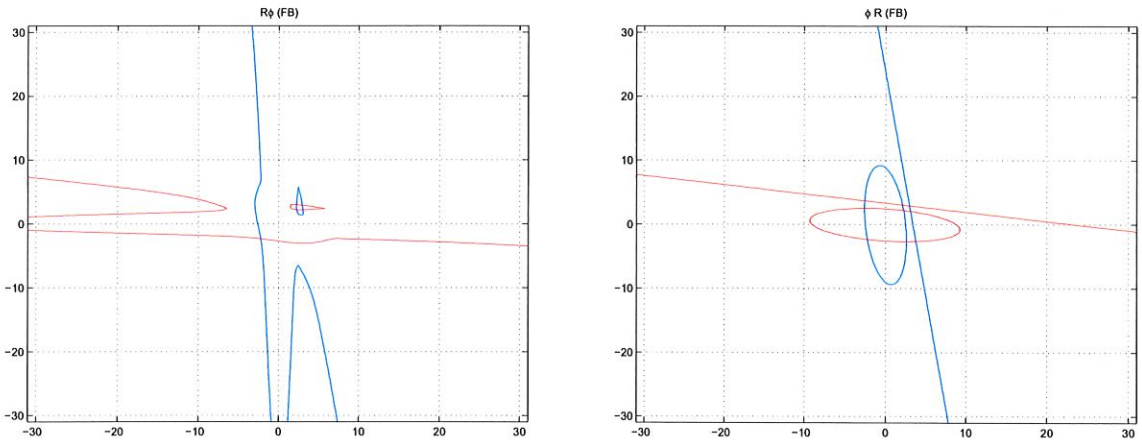


FIG. 5.15. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g définis par* (5.3.5) *et* (5.3.6) *en utilisant la fonction de Fischer-Burmeister et la prolongation modifiée* (5.2.2).

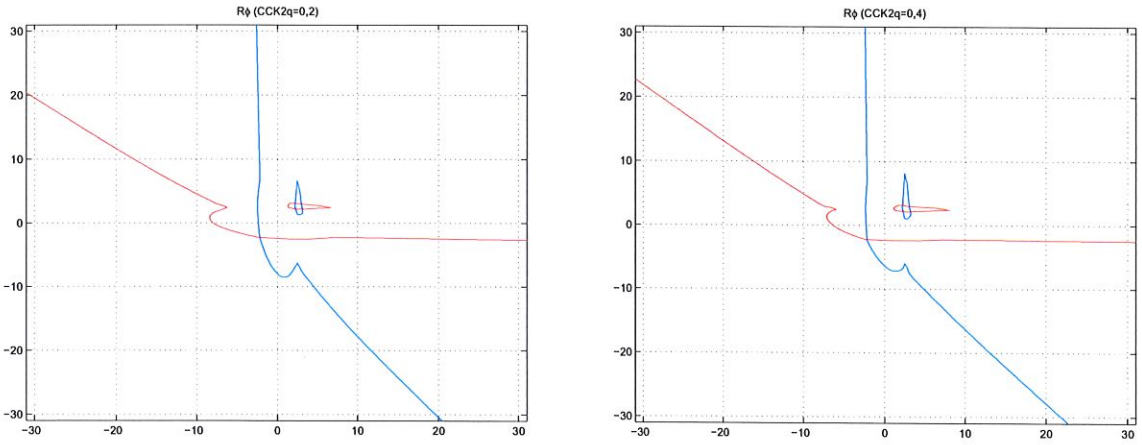


FIG. 5.16. — *Illustration du problèmes $R\Phi$ pour des f et g définis par* (5.3.5) *et* (5.3.6) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,2$ (à gauche) et $q = 0,4$ (à droite) et la prolongation modifiée* (5.2.2).

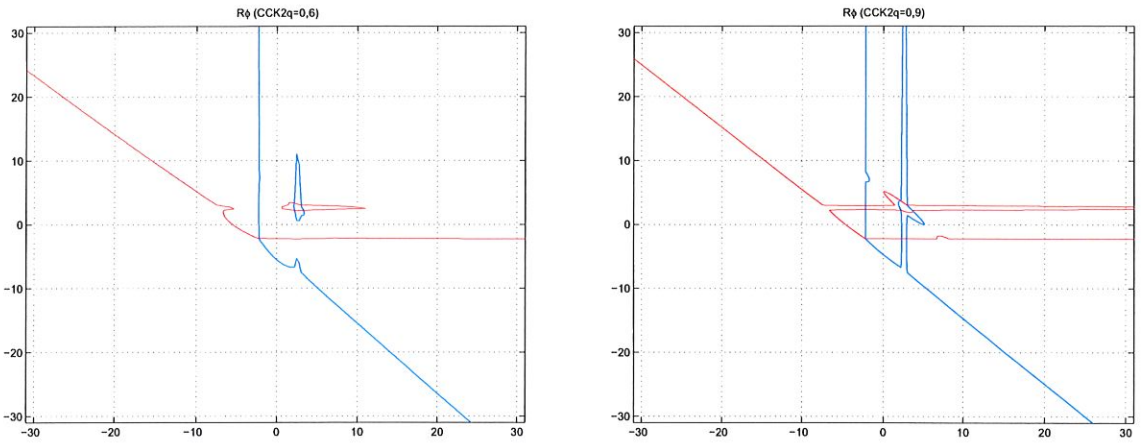


FIG. 5.17. — *Illustration du problème RΦ pour des f et g définis par* (5.3.5) *et* (5.3.6) *en utilisant la fonction de Chen-Chen-Kanzow avec* $q = 0,6$ *(à gauche) et* $q = 0,9$ *(à droite) et la prolongation modifiée* (5.2.2).

Cinquième exemple : cas quadratique (*bis*). Nous reprenons la même fonction pour f que dans le cas précédent et pour g, nous prenons *une* discrétisation de l'opérateur de dérivation :

$$g : \mathbb{R}^5 \to \mathbb{R}^5 : (x_1, x_2, x_3, x_4, x_5) \mapsto (x_2 - x_1, x_2 - x_1, x_3 - x_2, x_4 - x_3, x_5 - x_4). \quad (5.3.9)$$

Notez au passage que si le passage entre des espaces de dimension finie peut être effectué de diverses façons, la liberté pour la discrétisation en dimension finie d'un opérateur de dimension infinie est incroyablement plus large.

La contrainte de positivité de g va imposer que les variables soient en ordre croissant, alors que celle sur f va les empêcher d'être dans l'intervalle $]-\sqrt{5}, \sqrt{5}[$. Au final, les solutions sont données par (a, a, a, a, a) ; $(b, b, b, b, \pm\sqrt{5})$; $(b, b, b, \pm\sqrt{5}, \pm\sqrt{5})$; $(b, b, \pm\sqrt{5}, \pm\sqrt{5}, \pm\sqrt{5})$; $(b, \pm\sqrt{5}, \pm\sqrt{5}, \pm\sqrt{5}, \pm\sqrt{5})$ avec $a \in \mathbb{R}\backslash]-\sqrt{5}, \sqrt{5}[$ et $b \in]-\infty, \sqrt{5}]$.

La version de grossière de g est donnée par

$$\check{g}(y_1, y_2) = \frac{1}{8}(2y_1 + y_2, 2y_2 - 3y_1). \quad (5.3.10)$$

Les illustrations de cet exemple se trouvent aux figures 5.18, 5.19 et 5.20.

Concernant le problème $R\Phi$, une seule solution proche de $(\sqrt{5}, \sqrt{5})$ est trouvée ; toutefois d'autres apparaissent (et ce, plus rapidement que dans l'exemple précédent), lorsque la valeur du paramètre q de la fonction de Chen-Chen-Kanzow augmente.

Quant au problème ΦR, nous constatons cette fois que les contraintes de positivité ne sont plus nécessairement vérifiées dès lors que le produit scalaire s'annule ; la figure comporte ainsi deux droites amputées chacune d'un segment, de même que seulement deux demi-ellipses. La seule solution trouvée est proche de $(\sqrt{5}, \sqrt{5})$ pour les mêmes raisons que précédemment.

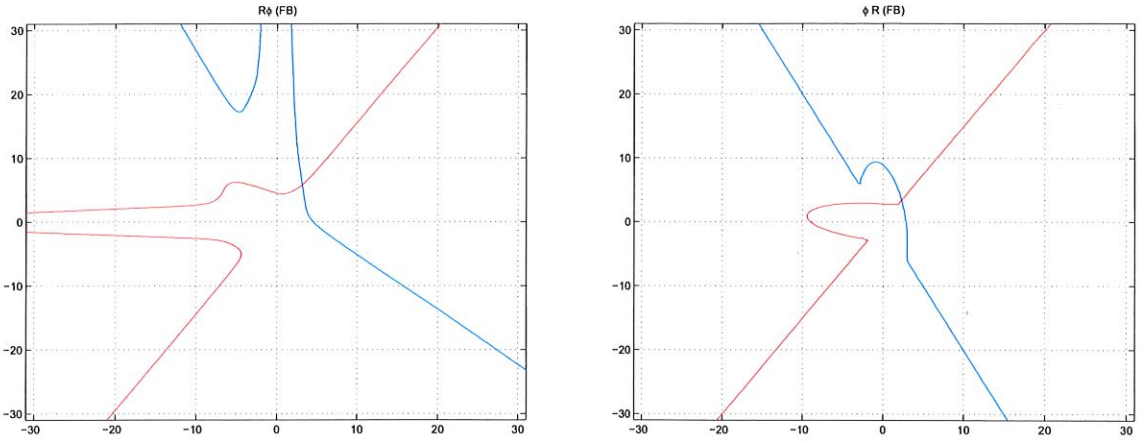


FIG. 5.18. — *Illustration des problèmes $R\Phi$ (à gauche) et ΦR (à droite) pour des f et g définis par* (5.3.5) *et* (5.3.9) *en utilisant la fonction de Fischer-Burmeister.*

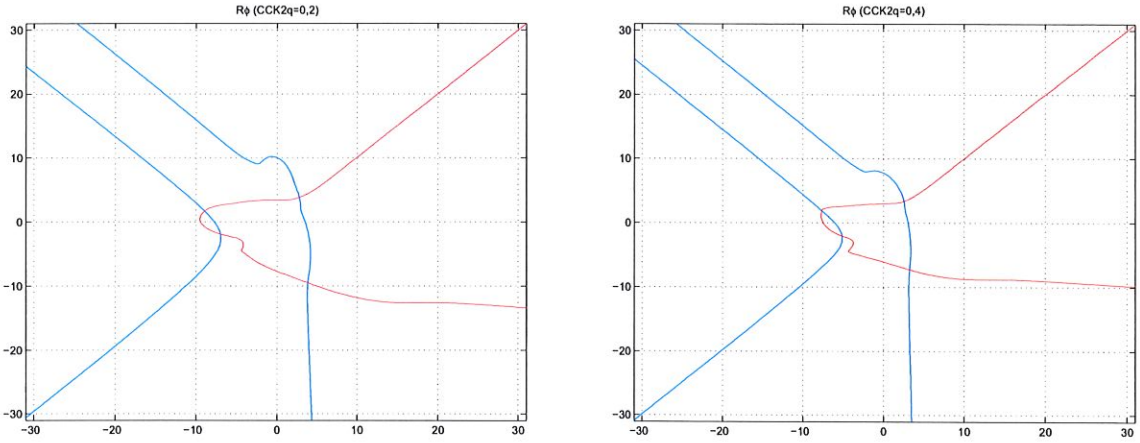


FIG. 5.19. — *Illustration du problèmes $R\Phi$ pour des f et g définis par* (5.3.5) *et* (5.3.9) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,2$ (à gauche) et $q = 0,4$ (à droite).*

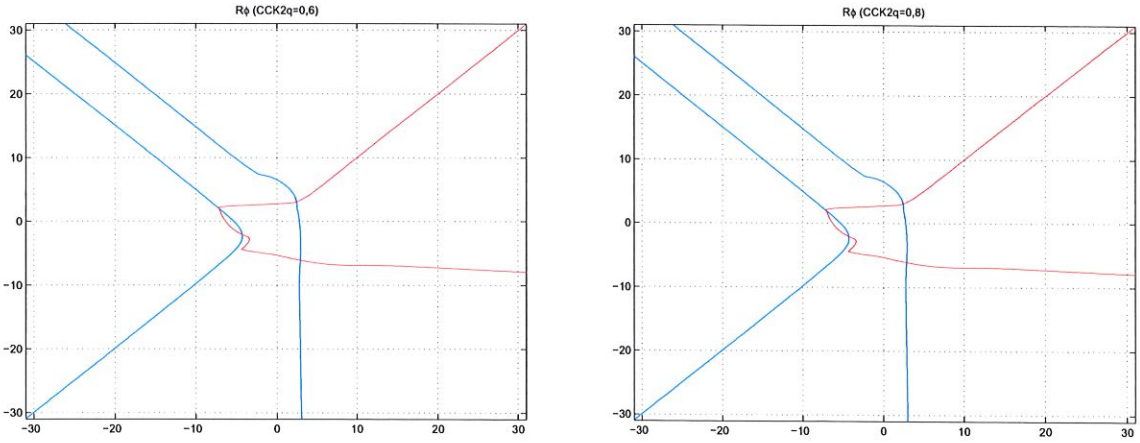


FIG. 5.20. — *Illustration du problème $R\Phi$ pour des f et g définis par* (5.3.5) *et* (5.3.9) *en utilisant la fonction de Chen-Chen-Kanzow avec $q = 0,6$ (à gauche) et $q = 0,8$ (à droite).*

5.4. Quelques réflexions

Nous avons observé au travers de ces exemples que les deux approches $R\Phi$ et ΦR peuvent être fort différentes ; il n'est donc pas inutile de s'interroger pour savoir laquelle est la plus appropriée et la plus cohérente pour résoudre des problèmes de complémentarité.

Tout d'abord, il faut noter que l'indépendance des solutions du problème ΦR par rapport à la fonction de complémentarité ϕ ne signifie pas que ces problèmes sont identiques ; ils pourraient être plus faciles à résoudre pour l'une ou l'autre de ces fonctions, tout comme c'est le cas pour la résolution de systèmes linéaires possédant la même solution.

Par ailleurs, il semblerait que davantage de solutions soient en général présentes dans le problème ΦR que dans celui $R\Phi$. La question de savoir s'il est préférable d'avoir beaucoup ou peu de solutions dépend surtout de l'apparition ou pas de solutions parasites, qui ne représentent rien d'intéressant pour la résolution du problème au niveau fin. Sous cet angle, la première démarche apparaît comme étant meilleure en général.

Dans le cas linéaire où une certaine littérature est déjà disponible, les problèmes de complémentarité ne sont en général pas reformulés par une fonction de complémentarité et la question de l'ordre des opérations ne se pose pas : les différentes versions du problèmes sont établies, puis des algorithmes appropriés (par exemple, celui de Lemke) sont utilisés lorsqu'une résolution à un niveau donné est nécessaire (voir notamment [22]). Cet démarche rejoint évidemment notre seconde proposition.

Enfin, au niveau des opérateurs de prolongation et de restriction, nous avons pu constater qu'ils jouaient un rôle important quant aux solutions qui peuvent être trouvées à un niveau donné, puisqu'ils préformatent en quelque sorte le type de solution qui est accessible. Ces opérateurs agissent notamment en mélangeant légèrement les composantes qui sont voisines, induisant par là un effet de lissage, au sens où elle tend à plutôt faire apparaître des solutions qui sont « plates » plutôt qu'oscillantes. Toutefois, cet effet dépend de la prolongation et de la restriction qui sont choisies ; le lissage est ici une conséquence du caractère interpolant de la prolongation P.

Conclusion

Dans ce mémoire, nous avons découvert ce que sont les problèmes de complémentarité et les différentes variantes qui en existent. Plusieurs exemples ont permis de constater la diversité et le nombre des applications pour lesquelles ils peuvent être utilisés, justifiant d'autant plus leur étude.

Nous nous sommes attachés aux versions non linéaires de cette classe de problèmes et sommes passé par leur reformulation au moyen de fonctions de complémentarité pour pouvoir les résoudre. Au lieu d'une relation de complémentarité, nous trouvions alors simplement un système d'équations à résoudre. Cependant, celles-ci n'étaient pas forcément différentiables, ce qui nous a amené à étudier auparavant quelques notions d'analyse non lisse.

Ensuite, nous avons entrepris d'exposer l'algorithme proposé par M. Ulbrich dans [29]. Tout d'abord, nous avons transformé le système d'équations en un problème de minimisation d'une fonction mérite, différentiable sous certaines hypothèses. Il était alors possible d'appliquer une méthode de type Newton au problème. Mais puisque nous envisagions l'éventualité que le problème présente des contraintes, il fallait maintenir la réalisabilité des itérées successives ; pour ce faire, le pas de type Newton fut chaque fois projeté sur le domaine admissible défini par ces contraintes. La convergence locale fut alors établie pour cet algorithme de type Newton avec projection.

Il se posait aussi la question de la convergence globale, c'est-à-dire même si le point de départ n'est pas à proximité de la solution. Nous savions d'avance que les méthodes de type Newton ne sont pas directement globalement convergente, c'est pourquoi nous l'avons intégrée dans une technique supplémentaire comme le fait M. Ulbrich dans l'article précité. Il s'agit ici d'une méthode de région de confiance non monotone. Grâce à elle, la convergence globale de l'algorithme ainsi obtenu est prouvée. Par rapport à l'approche classique des régions de confiance, seule une décroissance non monotone est ici imposée au moyen d'une moyenne pondérée sur les dernières itérations. Davantage de pas sont ainsi acceptés et il est constant dans la littérature que cela conduit presque toujours à une amélioration de la vitesse de convergence des algorithmes qui usent de ce procédé. Nous avons aussi indiqué dans le chapitre 3, comment choisir le pas à chaque itération. Au lieu de projeter le pas de type Newton simplement sur le domaine défini par les contraintes du problème, le pas est projeté sur son intersection avec la région de confiance. Pour la convergence, cela ne change rien en fin de compte, étant donné

que le pas de type Newton finit par toujours être dans la région de confiance lorsque que les itérées sont suffisamment proches d'une solution. Toutefois, nous avions imposé une condition de décroissance suffisante pour accepter le pas de type Newton projeté ; nous ne voulions en effet pas qu'une itération puisse se solder par des progrès quasiment inexistants. Il fallait dès lors pallier la possibilité que le pas de type Newton projeté ne satisfasse pas cette condition, soit en garantissant que tel ne serait pas le cas, soit en essayant de faciliter la vérification de cette condition en prenant une mesure de criticalité appropriée.

Même si la méthode de résolution proposée par M. Ulbrich est performante, nous nous sommes interrogés pour savoir s'il n'y aurait pas moyen de l'améliorer par des techniques multi-niveaux, comme cela a déjà été réalisé dans d'autres branches de l'analyse numérique.

Une brève introduction aux méthodes multi-niveaux fut ainsi inclue à ce mémoire pour en expliquer le cadre général. Pour leur application aux problèmes de complémentarité, nous fûmes confrontés à plusieurs difficultés. D'une part, il fallait choisir la représentation du problème de complémentarité aux niveaux grossiers et donc la façon d'utiliser les opérateurs de restriction et de prolongation ; nous nous aperçûmes aussi que leur choix pouvait jouer un rôle important dans le fait que l'algorithme trouve une solution correcte ou pas. D'autre part, une alternative importante se présenta à nous : était-il plus intéressant de reformuler le problème de complémentarité en un système d'équations puis de leur appliquer une méthode multi-niveaux, ou valait-il mieux commencer par discrétiser le problème de complémentarité puis appliquer une fonction de complémentarité pour transformer chaque niveau du problème en systèmes d'équations.

À travers différents exemples, nous avons observé que les deux approches étaient distinctes et ne menaient pas forcément aux mêmes solutions. L'importance du choix des opérateurs de prolongation P et de restriction R fut soulignée en montrant qu'ils permettaient de faire apparaître ou disparaître des solutions. La façon de discrétiser le problème de complémentarité par pré- et postapplication des opérateurs P et R s'est avérée être cohérente avec la manière de transférer l'information au niveau inférieur qui est habituellement utilisée en optimisation multi-niveaux.

Au final, nous n'avons pas pu trancher définitivement en faveur de l'une ou l'autre approche. Cette question mériterait ainsi une étude plus approfondie et notamment via l'implémentation d'un algorithme dans les deux cas. Il serait alors possible de comparer non seulement les problèmes de manière intrinsèque, mais aussi de voir en pratique laquelle de ces démarches est la plus rapide.

Bibliographie

[1] M. Anitescu et F. A. Potra, « Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems », Reports on Computational Mathematics 93, Department of Mathematics, University of Iowa, Iowa City, USA, 1996.

[2] M. Anitescu et F. A. Potra, « A time-stepping method for stiff multibody dynamics with contact and friction », *International Journal on Numerical Methods in Engineering*, vol. 55, pp. 753–784, 2002.

[3] B. Chen, X. Chen et C. Kanzow, « A penalized Fischer-Burmeister NCP-function : theoretical investigation and numerical results », Tech. Rep. A-126, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, 1997.

[4] F. H. Clarke, *Optimization and Nonsmooth Analysis.* Canadian Mathematical Society Series in Mathematics, Chichester, England, J. Wiley and Sons, 1983. Reprinted as *Classics in Applied Mathematics 5*, SIAM, Philadelphia, USA, 1990.

[5] A. R. Conn, N. I. M. Gould et Ph. L. Toint, « Global convergence of a class of trust region algorithms for optimization with simple bounds », *SIAM Journal on Numerical Analysis*, vol. 25, no. 182, pp. 433–460, 1988.

[6] A. R. Conn, N. I. M. Gould et Ph. L. Toint, *Trust-Region Methods.* No. 01 in MPS-SIAM Series on Optimization, Philadelphia, USA, SIAM, 2000.

[7] R. W. Cottle, J.-S. Pang et R. E. Stone, *The Linear Complementarity Problem.* London, Academic Press, 1992.

[8] S. P. Dirkse et M. C. Ferris, « MCPLIB : A collection of non-linear mixed complementarity problems », *Optimization Methods and Software*, vol. 5, pp. 319–345, 1995.

[9] S. P. Dirkse et M. C. Ferris, « The PATH solver : a non-monotone stabilization scheme for mixed complementarity problems », *Optimization Methods and Software*, vol. 5, no. 2, pp. 123–156, 1995.

[10] B. K. Driver, *Analysis Tools with Applications.* Heidelberg, Berlin, New York, Springer Verlag, 2003.

[11] F. Facchinei et C. Kanzow, « A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems », *Mathematical Programming*, vol. 76, pp. 493–512, 1997.

[12] F. Facchinei et J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vol. 1. Heidelberg, Berlin, New York, Springer Verlag, 2003.

[13] F. Facchinei et J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vol. 2. Heidelberg, Berlin, New York, Springer Verlag, 2003.

[14] A. Fischer, « Solution of monotone complementarity problems with locally Lipschitzian functions », *Mathematical Programming*, vol. 76, pp. 513–532, 1997.

[15] S. GRATTON, A. SARTENAER et PH. L. TOINT, « Recursive trust-region methods for multiscale nonlinear optimization », Tech. Rep. 04/06, Department of Mathematics, University of Namur, Namur, Belgium, 2004.

[16] C. KANSOW et H. KLEINMICHEL, « A new class of semismooth newton-type methods for nonlinear complementarity problems », *Computational Optimization and Applications*, vol. 11, pp. 227–251, 1999.

[17] K. KEDLAYA, « $a < b$ (a is less than b) ». Based on notes for the Math Olympiad Program (MOP), 1999. (http ://www.artofproblemsolving.com/Resources/Papers/-KedlayaInequalities.pdf).

[18] Z. Q. LUO et P. TSENG, « A new class of merit functions for the nonlinear complemetarity problem », in *Complementarity and variational problems : State of the Art* (M. C. FERRIS et J.-S. PANG, édits.), (Philadelphia, USA), pp. 204–225, SIAM, 1997.

[19] O. L. MANGASARIAN, « Equivalence of the complementarity problem to a system of nonlinear equations », *SIAM Journal on Applied Mathematics*, vol. 31, pp. 89–92, 1976.

[20] J. MARTÍNEZ et L. QI, « Inexact Newton methods for solving nonsmooth equations », *Journal of Computational and Applied Mathematics*, vol. 60, pp. 127–145, 1995.

[21] C. MÉZÁROS, « The BPMPD Interior Point Solver for Convex Quadratic Problems », Working paper WP 98-8, Computer and Automation Research Institute, Hungarian Academy of Sciences, 1998.

[22] C. W. OOSTERLEE, « On multigrid for linear complementarity problems with application to american-style options », *Electronic Transactions on Numerical Analysis*, vol. 15, pp. 165–185, 2003.

[23] J.-S. PANG et L. QI, « Nonsmooth equations : Motivation and algorithms », *SIAM Journal on Optimization*, vol. 3, pp. 443–465, 1993.

[24] M. J. D. POWELL, « A hybrid method for nonlinear equations », in *Numerical Methods for Nonlinear Algebraic Equations* (P. RABINOWITZ, édit.), (London), pp. 87–114, Gordon and Breach, 1970.

[25] M. J. D. POWELL, « A new algorithm for unconstrained optimization », in *Nonlinear Programming* (J. B. ROSEN, O. L. MANGASARIAN et K. RITTER, édits.), (London), pp. 31–65, Academic Press, 1970.

[26] M. J. D. POWELL, « Convergence properties of a class of minimization algorithms », in *Nonlinear Programming, 2* (O. L. MANGASARIAN, R. R. MEYER et S. M. ROBINSON, édits.), (London), pp. 1–27, Academic Press, 1975.

[27] L. QI et J. SUN, « A nonsmooth version of Newton's method », *Mathematical Programming*, vol. 58, pp. 353–367, 1993.

[28] S. THOMAS, *Sequential estimation techniques for quasi-Newton algorithms.* PhD thesis, Cornell University, Ithaca, New York, USA, 1975.

[29] M. ULBRICH, « Nonmonotone Trust-Region Methods for Bound-Constrained Semismooth Equations with Applications to Nonlinear Mixed Complementarity Problems », *SIAM Journal on Optimization*, vol. 11, pp. 889–917, 2001.

[30] M. ULBRICH, « Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces ». Habilitationsschrift, 2002.

[31] Y. YUAN, « An example of non-convergence of trust region algorithms », in *Advances in Nonlinear Programming* (Y. YUAN, édit.), (Dordrecht, The Netherlands), pp. 205–218, Kluwer Academic Publishers, 1998.