



HAL
open science

Definition et Applications des Extensions des Fonctions Reelles aux Intervalles Généralisés; reformulation de la theorie des intervalles modaux.

Alexandre Goldsztejn

► **To cite this version:**

Alexandre Goldsztejn. Definition et Applications des Extensions des Fonctions Reelles aux Intervalles Généralisés; reformulation de la theorie des intervalles modaux.. Modélisation et simulation. Université Nice Sophia Antipolis, 2005. Français. NNT : . tel-00011922

HAL Id: tel-00011922

<https://theses.hal.science/tel-00011922v1>

Submitted on 10 Mar 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE NICE-SOPHIA ANTIPOLIS - UFR Sciences
Ecole Doctorale STIC

THESE

pour obtenir le titre de
Docteur en Sciences
de l'UNIVERSITE de Nice-Sophia Antipolis

Discipline : Informatique

présentée et soutenue par **Alexandre GOLDSZTEJN**

Définition et Applications des Extensions des Fonctions Réelles aux Intervalles Généralisés

Révision de la Théorie des Intervalles Modaux et Nouveaux Résultats

Thèse dirigée par **Michel RUEHER**

Soutenue le 10 novembre 2005

Rapporteurs :

M. Gardenyes, Ernest	Professeur
M. Jaulin, Luc	Professeur
M. Shary, Sergey P.	Professeur

Jury :

M. Formenti, Enrico	Professeur	Président
M. Gardenyes, Ernest	Professeur	
M. Jaulin, Luc	Professeur	
M. Rueher, Michel	Professeur	Directeur de thèse
M. Markov, Svetoslav	Professeur	
M. Taillibert, Patrick	Ingénieur	Co-Directeur de thèse

Remerciements

Je souhaite remercier chaleureusement Patrick Taillibert qui a proposé et encadré cette thèse. L'intérêt scientifique et la réussite de cette thèse reposent tout autant sur le travail qui a été effectué que sur l'originalité et l'intérêt du sujet qui a été proposé. Nos longues discussions ont permis à cette thèse de naître et d'aboutir aujourd'hui. Grâce à Patrick, j'ai découvert le métier d'enseignant-chercheur qui me passionne aujourd'hui. Je remercie aussi Thales Airborne Systems de m'avoir accueilli et financé pendant trois années.

Je remercie mon directeur de thèse Michel Rueher pour son soutien et pour avoir cru en moi malgré les moments de doutes que nous avons traversés.

Je tiens à remercier particulièrement les rapporteurs et les membres du jury pour l'intérêt qu'ils ont eu à l'égard de mon travail.

Je voudrais remercier les chercheurs (Bernard Botella, David Daney, ...), stagiaires et personnels administratifs (Élisabeth Brissot, Patricia Lachaume, ...) que j'ai côtoyés pour leur aide et les échanges que nous avons eus.

Je remercie aussi mes parents qui m'ont toujours accompagné dans les choix que j'ai fait et qui m'ont permis de trouver ma voie.

Finalement, cette thèse n'aurait certainement pas été possible sans mon amie Sabine qui m'a aidé à surmonter de nombreuses épreuves.

Résumé

La théorie des intervalles permet de construire des sur-ensembles du domaine de variation d'une fonction réelle. Ainsi, de manière très naturelle, elle permet de construire une approximation extérieure de l'ensemble des solutions d'un système d'équations. Couplée aux théorèmes usuels d'existence (par exemple les théorèmes de Brouwer ou de Miranda) la théorie des intervalles permet aussi de prouver rigoureusement l'existence de solutions pour un système d'équations.

La théorie des intervalles modaux propose des interprétations plus riches que la théorie de intervalles classiques. En particulier, l'interprétation des extensions aux intervalles modaux permet de prouver directement l'existence de solution d'un système d'équations (sans faire intervenir explicitement les théorèmes d'existence). Malgré les récents développements qui ont montré le potentiel applicatif de la théorie des intervalles modaux, l'utilisation de cette théorie reste fort limitée. Cela peut s'expliquer de la manière suivante :

- La théorie des intervalles modaux a une construction originale mais compliquée qui est assez éloignée de la construction de la théorie des intervalles classiques. Cela rend par exemple difficile l'ajout de nouveaux concepts.
- Aucun préconditionnement compatible avec les interprétations offertes par la théorie des intervalles modaux n'a été proposé.
- Aucun protocole de linéarisation compatible avec les interprétations offertes par la théorie des intervalles modaux n'a été proposé.

Dans le cadre de cette thèse, ces trois points sont développés. D'une part, une nouvelle formulation des principaux résultats de la théorie des intervalles modaux est proposée. Cette nouvelle formulation est faite dans le cadre des intervalles généralisés (intervalles dont les bornes ne sont pas contraintes à être ordonnées) et reprend la construction de la théorie des intervalles classiques. D'autre part, un protocole de préconditionnement et un protocole de linéarisation compatibles avec les interprétations des nouvelles extensions aux intervalles généralisés sont proposés. Le protocole de linéarisation proposé aura la forme d'une nouvelle extension de la valeur moyenne aux intervalles généralisés.

Ces développements théoriques aboutissent à deux applications : d'une part, la nouvelle extension de la valeur moyenne aux intervalles généralisés est utilisée pour construire une approximation intérieure du domaine de variation d'une fonction à valeurs vectorielles. Ce problème est aujourd'hui mal traité par la théorie des intervalles classiques. D'autre part, un opérateur généralisé de Hansen-Sengupta dédié à l'approximation extérieure des "AE-solution sets" est proposé. Il est beaucoup plus simple et moins coûteux en temps de calcul que les autres techniques permettant de résoudre ce type de problèmes. Une comparaison de la puissance de résolution de ces différentes techniques nécessitera d'intégrer l'opérateur généralisé de Hansen-Sengupta au sein d'un algorithme de bisection.

Abstract

The intervals theory allows constructing supersets of the range of real functions. Therefore, in a very natural way it allows constructing some outer approximation of the solution set of systems of real equations. When it is used in conjunction to some usual existence theorems (e.g. Brouwer or Miranda theorems), the intervals theory also allows to rigorously prove the existence of solutions to such systems of equations.

The modal intervals theory proposed some richer interpretations. In particular, the construction of both subsets and supersets of the range of real functions are in the scope of extensions to modal intervals. As a consequence, the extensions of real functions to modal intervals have the intrinsic power of proving the existence of solutions to systems of equations. In spite of some recent developments that have shown the promising potential applications of these richer interpretations, the modal intervals theory remains unused by most of the interval community. This may be explained by the following arguments :

- The modal intervals theory has an original and complicated construction. It is not similar to the construction of the classical intervals theory. This makes for example difficult the addition of new developments in the framework of modal intervals.
- No preconditioning has yet been proposed that would be compatible with the richer interpretations provided by the modal intervals theory.
- No linearization process has yet been proposed that would be compatible with the richer interpretations provided by the modal intervals theory.

These three points are developed in the present thesis. On one hand, a new formulation of the modal intervals theory is proposed. This new formulation uses only generalized intervals (intervals whose bounds are not constrained to be ordered) and follows the construction of the classical intervals theory. On the other hand, some new preconditioning and linearization processes are proposed which are compatible with the richer interpretations provided by the modal interval theory. The new linearization process which is proposed will have the form of a new mean-value extension to generalized intervals.

These theoretical developments lead to two applications : on one hand, the new mean-value extension to generalized intervals is used to construct an inner approximation of the range of a vector-valued function. This problem is not well solved today using the classical intervals theory. On the other hand, a generalized Hansen-Sengupta operator is proposed. It is dedicated to the outer approximation of non-linear AE-solution sets. It is much simpler and less expensive in computations than the other techniques that can solve the same problems. A comparison of these different techniques remains to be conducted, and will need the integration of the Hansen-Sengupta operator within some bisection algorithm.

Glossaire des notations

\mathbb{R}	ensemble des nombres réels
\mathbb{IR}	ensemble des intervalles réels fermés bornés et non-vides
\mathbb{KR}	ensemble des intervalles généralisés
$\overline{\mathbb{IR}}$	ensemble des intervalles impropres ($\mathbb{KR} = \mathbb{IR} \cup \overline{\mathbb{IR}}$)
\mathbb{F}	ensemble des nombres flottants
\mathbb{IF}	ensemble des intervalles réels fermés bornés et non-vides à bornes flottantes
\mathbb{KF}	ensemble des intervalles généralisés à bornes flottantes
$E \cup F$	union d'ensembles
$E \cap F$	intersection d'ensembles
$E \setminus F$	différence d'ensembles
$E \subseteq F$	inclusion entre ensembles
$\inf_{\leq} E$	plus grand minorant de E , où (E, \leq) est un ensemble ordonné
$\sup_{\leq} E$	plus petit majorant de E , où (E, \leq) est un ensemble ordonné
x	nombre réel ou vecteur réel
\mathbf{x}	intervalle (généralisé) ou vecteur d'intervalles (généralisés)
$\inf \mathbf{x}$	borne inférieure de \mathbf{x}
$\sup \mathbf{x}$	borne supérieure de \mathbf{x}
$\mathbf{x} \wedge \mathbf{y}$	intersection d'intervalles ¹ (généralisés), i.e. $\inf_{\subseteq} \{\mathbf{x}, \mathbf{y}\}$
$\mathbf{x} \vee \mathbf{y}$	union d'intervalles ² (généralisés), i.e. $\sup_{\subseteq} \{\mathbf{x}, \mathbf{y}\}$
$\text{mid } \mathbf{x}$	milieu de \mathbf{x}
$\text{rad } \mathbf{x}$	rayon de \mathbf{x}
$\text{wid } \mathbf{x}$	largeur de \mathbf{x}
$\text{dual } \mathbf{x}$	dual de l'intervalle généralisé \mathbf{x}
$\text{pro } \mathbf{x}$	projection propre de l'intervalle généralisé \mathbf{x}
$\text{imp } \mathbf{x}$	projection impropre de l'intervalle généralisé \mathbf{x}
$\diamond \mathbf{x}$	arrondi extérieur de \mathbf{x}
$\square E$	enveloppe d'intervalle ³ de $E \subseteq \mathbb{R}^n$, i.e. $\inf_{\subseteq} \{\mathbf{x} \in \overline{\mathbb{IR}}^n \mid E \subseteq \mathbf{x}\}$
A	matrice réelle
\mathbf{A}	matrice d'intervalle
$\text{diag}(x)$	matrice diagonale dont les entrées sont les x_i
$\langle \mathbf{A} \rangle$	matrice de comparaison de \mathbf{A}
f	fonction réelle ou vecteur de fonctions réelles
f'	dérivée, gradient ou jacobienne de f , i.e. $f'_{ij}(x) = \frac{\partial f_i}{\partial x_j}(x)$
$\text{range}(f, E)$	image de E par f
\mathbf{g}	fonction d'intervalle
\mathbf{g}^\diamond	fonction d'intervalle arrondie

¹"Interval hull" en anglais.

²"Meet" en anglais

³"Join" en anglais

Table des matières

I	Introduction	9
II	Les intervalles et leurs différentes généralisations	15
1	Rappels sur les nombres réels	17
1.1	Les nombres réels	17
1.2	Vecteurs et matrices réels	17
1.3	Les fonctions réelles	18
1.4	Représentation finie des nombres réels	20
2	La théorie des intervalles	21
2.1	Les intervalles	22
2.2	Vecteurs et matrices intervalles	25
2.3	Fonctions d'intervalles	28
2.4	Extensions des fonctions continues	28
2.5	Extensions optimales des fonctions continues $f : \mathbb{R}^n \longrightarrow \mathbb{R}$	32
2.6	L'extension naturelle	34
2.7	L'extension de la valeur moyenne	35
2.8	Extensions des fonctions continues à valeurs vectorielles	36
2.9	Autres extensions	36
2.10	Opérations sur les vecteurs et matrices d'intervalles	36
3	Intervalles généralisés	37
3.1	Les intervalles généralisés	37
3.2	Vecteurs et matrices d'intervalles généralisés	40
3.3	Fonctions d'intervalles généralisés	40
3.4	Extensions des fonctions continues aux intervalles généralisés	40
3.5	Développements supplémentaires	44

3.6	Liens entre les intervalles généralisés et les intervalles classiques	45
4	Intervalles modaux	49
4.1	Les intervalles modaux	49
4.2	Extensions des fonctions continues aux intervalles modaux	53
4.3	Lien entre les intervalles modaux et intervalles généralisés	61
4.4	Conclusion	62
5	Systèmes d'intervalles	63
5.1	"United Solution set"	63
5.2	"AE-Solution set"	67
III	Nouveaux développements pour les "AE-solution sets linéaires"	79
6	Introduction	81
7	Préconditionnement à droite des "AE-solution sets" linéaires	85
7.1	Existence et unicité de la solution de l'équation (dual \mathbf{A}) $\mathbf{x} = \mathbf{b}$	85
7.2	Préconditionnement à droite des "AE-solution sets" linéaires	88
7.3	Sur la qualité du preconditionnement à droite	94
7.4	Conclusion	104
8	Amélioration de l'opérateur généralisé de Gauss-Seidel	105
8.1	L'opérateur généralisé de Gauss-Seidel (cas $n = 1$)	105
8.2	L'opérateur généralisé de Gauss-Seidel (cas général)	107
8.3	Conclusion	108
IV	AE-extensions des fonctions réelles continues	109
9	Introduction	111
10	La nouvelle formulation de la théorie des intervalles modaux	115
10.1	Introduction	115
10.2	Description du problème à résoudre	116
10.3	Intervalles généralisés et quantificateurs	117
10.4	L'arithmétique de Kaucher	117
10.5	Une interprétation simplifiée de l'évaluation d'intervalles généralisés	118

10.6	Une extension de la valeur moyenne aux intervalles généralisés	119
10.7	Cas des fonctions à valeurs vectorielles	120
11	AE-extensions des relations réelles	123
11.1	L'inclusion entre intervalles généralisés	123
11.2	AE-extensions des relations réelles	125
12	AE-extensions des fonctions continues $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$	131
12.1	Définition des AE-extensions	131
12.2	Interprétation des AE-extensions	134
12.3	Minimalité des AE-extensions	135
12.4	Ordre de convergence d'une AE-extension	140
12.5	Arrondi d'une AE-extension	142
12.6	Conclusion	144
13	AE-extensions minimales des fonctions $f : \mathbb{R}^n \longrightarrow \mathbb{R}$	145
13.1	L'AE-extension minimale f^*	145
13.2	Commutation avec l'opération dual	148
14	AE-extensions des fonctions élémentaires	151
14.1	Fonctions à une variable	151
14.2	Fonctions à deux variables	153
14.3	Démonstrations	155
15	L'AE-extension naturelle	159
15.1	Évaluation d'intervalles	159
15.2	L'AE-extension naturelle (cas particulier $m = 1$)	162
15.3	L'AE-extension naturelle (cas général)	166
15.4	Qualité de l'AE-extension naturelle	169
16	L'AE-extension de la valeur moyenne	171
16.1	Généralisation du théorème de la valeur moyenne	171
16.2	L'AE-extension de la valeur moyenne (cas particulier $m = 1$)	173
16.3	L'AE-extension de la valeur moyenne (cas général)	177
16.4	Qualité de l'AE-extension de la valeur moyenne	181

V Applications	185
17 Approximation intérieure de l'image d'une fonction à valeurs vectorielles	187
18 L'opérateur généralisé de Hansen-Sengupta	191
18.1 Définition de l'opérateur généralisé de Hansen-Sengupta	191
18.2 Exemple	193
VI Conclusion	197
A Un corollaire du théorème de Brouwer	201
Bibliographie	203

Table des figures

2.1	Vecteur d'intervalle : deux définitions équivalentes.	25
2.2	Un disque fermé possède une infinité d'approximations intérieures non-extensibles.	27
2.3	Effet d'enveloppe.	30
5.1	"United solution set", "tolerable solution set" et "controlable solution set" de l'exemple II.40.	70
5.2	Autres "AE-solution sets" de l'exemple II.40.	71
6.1	Un "AE-solution set" avec ses approximations intérieures et extérieures sous forme de parallélogrammes.	83
7.1	Le "united solution set" de l'Exemple III.4 avec ses différentes approximations.	102
11.1	Graphe de la relation $x_1^2 + x_2^2 \leq 9$	126
14.1	Une fonction simple qui ne peut pas être considérée comme fonction élémentaire.	153
16.1	Graphe de la fonction $f(x) = x^2$ avec le cône des dérivées $f(\tilde{\mathbf{x}}) + f'(\mathbf{x})(x - \tilde{\mathbf{x}})$.176	
17.1	Approximation intérieure de l'image d'un intervalle par une fonction à valeur vectorielle.	189
18.1	Le robot parallèle à deux bras.	194

Première partie

Introduction

Les intervalles

Les intervalles classiques

La théorie des intervalles (Moore[74], Alefeld[31], Neumaier[8], Jaulin[55]) a été initialement développée à partir des années 60 pour prendre en compte de façon rigoureuse différents types d'incertitudes, eg. erreurs d'arrondi dues aux calculs en précision finie, incertitudes de mesures, erreurs de linéarisation. De nombreux développements ont été faits depuis cette époque : la théorie des intervalles propose aujourd'hui des outils originaux et utiles indépendamment de la prise en compte des erreurs d'arrondi (cette dernière restant malgré tout un avantage de l'utilisation des intervalles), voir e.g. Kearfott[78].

En particulier, les intervalles classiques sont intensivement utilisés pour approximer l'ensemble des solutions d'un système d'équations. Ainsi, étant donné un système d'équations $f(x) = 0$, les intervalles classiques permettent de prouver que ce système d'équations n'a aucune solution dans une région donnée, et ainsi de réduire le domaine des solutions potentielles de cette équation. Dans le cas des systèmes d'équations bien déterminés, i.e. autant d'équations que d'inconnues et sans équation redondante, les intervalles classiques peuvent aussi être utilisés pour valider rigoureusement les hypothèses de certains théorèmes d'existence (comme le théorème de Brouwer ou le théorème de Miranda), voir e.g. Istratescu[101] ou Neumaier[8]. L'utilisation de la théorie des intervalles classiques permet ainsi de prouver l'existence de solutions pour des systèmes d'équations.

Les intervalles généralisés et les intervalles modaux

Les intervalles généralisés, i.e. intervalles dont les bornes ne sont pas contraintes à être ordonnées (Ortolf[42], Kaucher[20, 21]), et les intervalles modaux, i.e. couples (intervalle, quantificateur) (Gardenyes[16]), permettent de généraliser certains algorithmes développés dans le cadre de la théorie des intervalles classiques aux systèmes d'équations paramétrés en prenant en compte des quantifications différentes pour les paramètres. Considérons l'ensemble suivant :

$$\{x \in \mathbb{R}^n \mid (\forall a \in \mathbf{a})(\exists b \in \mathbf{b})(f(a, b, x) = 0)\}, \quad (1)$$

où \mathbf{a} et \mathbf{b} sont des vecteurs d'intervalles. Si des techniques reposant sur les intervalles classiques ont été proposées pour approximer ce type d'ensembles, leur champ d'application reste aujourd'hui limité. Les intervalles généralisés et les intervalles modaux sont des cadres théoriques prometteurs pour l'approximation de tels ensembles. Dans le cas des systèmes linéaires, les intervalles généralisés ont permis de construire des généralisations d'algorithmes classiques (Gauss-Seidel, Krawczyk) qui prennent en compte les quantifications des paramètres (Shary[95, 96], Markov[90]). Dans le cadre des systèmes non-linéaires, les intervalles modaux proposent un cadre prometteur pour la prise en compte de ces quantifications (Vehí[49], Sainz[60, 61], Herrero[69, 70]). Bien que la théorie des intervalles modaux soit jugée potentiellement prometteuse, elle n'est guère utilisée à ce jour : cela s'explique certainement par la construction originale mais compliquée de la théorie des intervalles modaux.

Contributions

Cette thèse propose plusieurs contributions dans le contexte des systèmes quantifiés de type AE, i.e. dans lesquels les quantificateurs universels précèdent les quantificateurs

existentiels :

Systèmes linéaires

Un nouveau préconditionnement compatible avec l'approximation intérieure des ensembles de solutions quantifiés de type AE est proposé. Ce procédé de préconditionnement présente le double avantage de simplifier la détermination d'une approximation intérieure (et dans certains cas où l'approche sans préconditionnement échoue, de rendre possible cette approximation intérieure) et d'accélérer la convergence des algorithmes utilisés.

Une amélioration de l'opérateur généralisé de Gauss-Seidel est proposée. L'opérateur généralisé de Gauss-Seidel est aujourd'hui restreint aux matrices n'ayant pas d'intervalles contenant zéro sur la diagonale. Cette restriction sera levée en généralisant la méthode utilisée dans le cadre de l'opérateur classique de Gauss-Seidel.

Reformulation de la théorie des intervalles modaux

La principale contribution de cette thèse est une nouvelle formulation de la théorie des intervalles modaux. Après avoir expliciter les points forts et les points faibles de la théorie des intervalles modaux, la nouvelle formulation de la théorie permettra de conserver uniquement les premiers. Seront conservés en particulier :

- Les interprétations enrichies proposées par la théorie des intervalles modaux.
- Le procédé de composition des extensions, correspondant à une adaptation pour les interprétation enrichies proposée par la théorie de l'extension naturelle des intervalles classiques.
- Le procédé d'arrondi des extensions aux intervalles modaux.

Les points suivants seront abandonnés dans la nouvelle formulation de la théorie :

- L'association d'un quantificateur à un intervalle classique est abandonnée car elle n'est pas compatible avec une généralisation de la théorie des intervalles classiques. Ainsi, la nouvelle formulation de la théorie des intervalles modaux est développée dans le cadre des intervalles généralisés.
- La théorie des intervalles modaux propose deux types d'extensions f^* et f^{**} . Ces deux types d'extensions sont en fait redondants. Seul les extensions f^* seront conservées, ce qui demandera d'adapter les notions de la théorie des intervalles modaux qui mettent en jeux ces deux types d'extensions.
- La définition de l'optimalité d'une extension aux intervalles modaux présente plusieurs défauts : d'une part, elle n'est pas intuitive. D'autre part, elle ne reflète pas réellement la qualité de l'extension. Cette définition sera remplacée par une définition plus intuitive et plus représentative de la qualité des extensions.

Le point de départ de la nouvelle formulation est une généralisation de la définition de l'extension d'une fonction aux intervalles classiques. Elle s'applique aux intervalles généralisés et permet une plus grande liberté dans la quantification des variables. Ensuite, la nouvelle formulation de la théorie des intervalles modaux sera construite sur le modèle de la théorie des intervalles classiques. Les principaux résultats de la théorie des intervalles modaux sont retrouvés dans la nouvelle formulation de la théorie, eg.

- extensions des fonctions élémentaires, avec la coïncidence des extensions avec l'arithmétique de Kaucher (arithmétique des intervalles généralisés définie dans Ortolfo[42]) pour les fonctions $+$, $-$, \times et \div ,

- adaptation de l’extension naturelle classique pour son utilisation avec les interprétations enrichies proposées par la théorie
- optimalité de l’extension naturelle pour les fonctions linéaires.

La nouvelle formulation de la théorie des intervalles modaux permettra l’introduction de nouveaux concepts : d’une part, l’ordre de convergence des extensions aux intervalles généralisés sera défini en généralisant la notion classique correspondante. Cette notion permettra de quantifier asymptotiquement la qualité d’une extension lorsqu’elle est utilisée sur de petits intervalles. Comme dans le cadre des intervalles classiques, il sera montré que l’extension naturelle généralisée a un ordre de convergence linéaire. D’autre part, une nouvelle extension de la valeur moyenne aux intervalles généralisés sera proposée. Celle-ci représente un nouveau procédé de linéarisation qui est compatible avec les interprétations enrichies de la théorie des intervalles modaux. Son ordre de convergence quadratique sera prouvé dans certains cas particuliers.

Applications

Sur la base de ces contributions, deux applications sont proposées : d’une part, la nouvelle extension de la valeur moyenne est utilisée pour construire un sous-ensemble de l’image d’un intervalle par une fonction à valeurs vectorielles, i.e. $f = (f_1, \dots, f_n)$. Ce problème est aujourd’hui mal résolu par la théorie des intervalles classiques. D’autre part, un opérateur généralisé de Hansen-Sengupta dédié à l’approximation extérieure des ensembles de solution du type (1) est proposé. Le potentiel applicatif prometteur de cet opérateur généralisé de Hansen-Sengupta est illustré sur un exemple où il est comparé aux autres méthodes capables de résoudre le problème (élimination de quantificateurs, voir e.g. Collins[15], et l’algorithme de "continuous first-order constraint satisfaction", voir Ratschan[83]) : tout en étant beaucoup plus simple, l’opérateur généralisé de Hansen-Sengupta est plus rapide que les autres méthodes pour trouver la solution de cet exemple.

Plan de la thèse

La thèse est composée de quatre parties. La première partie décrit la théorie des intervalles (Chapitre 2), les intervalles généralisés (Chapitre 3) et la théorie originale des intervalles modaux (Chapitre 4). Les intervalles généralisés sont le cadre théorique dans lequel sera faite la reformulation de la théorie originale des intervalles modaux. La description des trois ensembles d’intervalles est faite en suivant un plan commun, ce qui permettra une comparaison de ces structures et de leurs propriétés. La deuxième partie présente des nouveaux développements dans le cadre des systèmes linéaires. La reformulation de la théorie des intervalles modaux ainsi que les nouveaux développements qu’elle permet sont présentés dans la troisième partie. La quatrième partie est consacrée aux applications : particulier, un opérateur généralisé de Hansen-Sengupta dédié à l’approximation extérieure des AE-solution sets non-linéaires y est en proposé.

Deuxième partie

Les intervalles et leurs différentes généralisations

Chapitre 1

Rappels sur les nombres réels

1.1 Les nombres réels

Le corps des réels $(\mathbb{R}, +, \times)$ est l'aboutissement de siècles de tâtonnements. La première construction formelle fut donnée par Dedekind à la fin du XIX^{ème} siècle : les coupures de Dedekind permettent de construire les nombres réels à partir des nombres rationnels. Elles sont encore aujourd'hui une des méthodes les plus utilisées pour formaliser la construction des nombres réels (voir Birkhoff[32] pour les coupures de Dedekind et Spivak[58] ou Dixmier[47] pour une construction alternative au moyen de suites de Cauchy de nombres rationnels). Les réels sont naturellement séparés en nombres positifs et nombres négatifs, ce qui permet d'introduire un ordre total sur \mathbb{R} par $x \leq y \iff 0 \leq y - x$. Une distance est définie sur \mathbb{R} par $\text{dist}(x, y) = |x - y|$.

Le principal avantage de \mathbb{R} sur \mathbb{Q} est d'être complet, i.e. toute partie majorée de \mathbb{R} possède une borne supérieure dans \mathbb{R} . Ou, de manière équivalente, toute suite de Cauchy de \mathbb{R} converge dans \mathbb{R} . Cette complétude permet d'obtenir les théorèmes fondamentaux de l'analyse réelle, par exemple le théorème des valeurs intermédiaires ou le théorème de la valeur moyenne (voir section 1.3).

1.2 Vecteurs et matrices réels

Vecteurs réels

L'espace vectoriel \mathbb{R}^n est constitué des n-uples $(x_i)_{i \in [1..n]}$ de réels.

Notation (Kearfott[25]). *Il sera utile d'utiliser des indices quelconques pour la numérotation des composantes d'un vecteur. Si $E = \{e_1, e_2, \dots, e_n\}$ est un ensemble d'indices totalement ordonné (l'ordre lexicographique sera utilisé) alors le vecteur dont les composantes sont x_{e_i} pour $i \in [1..n]$ sera noté x_E . Lorsque cela ne prête pas à confusion, le vecteur $(x_i)_{i \in [1..n]} = x_{[1..n]}$ sera noté x .*

Les opérations suivantes sont définies : soient $x, y \in \mathbb{R}^n$ et $a \in \mathbb{R}$

- $x + y = (x_i + y_i)_{i \in [1..n]}$;
- $ax = (ax_i)_{i \in [1..n]}$;
- $|x| = (|x_i|)_{i \in [1..n]}$.

Les nombres réels sont identifiés aux éléments de \mathbb{R}^1 . Un ordre partiel est défini sur \mathbb{R}^n de la manière suivante : $x \leq y \iff (\forall i \in [1..n])(x_i \leq y_i)$ (l'ordre partiel est total si

et seulement si $n = 1$). La norme $\|x\|$ est définie par $\|x\| = \max_{i \in [1..n]} |x_i|$ et la distance correspondante par $\text{dist}(x, y) = \|x - y\|$.

Remarque. Ces normes et distances sont habituellement notées $\|\cdot\|_\infty$ et dist_∞ . Leur utilisation intensive et presque exclusive dans cette thèse justifie ces notations simplifiées.

Matrices réelles

Une matrice réelle est un tableau de nombres réels : $M \in \mathbb{R}^{n \times m}$ est définie par $n \times m$ nombres réels a_{ij} où $i \in [1..n]$ et $j \in [1..m]$ et est notée

$$M = (a_{ij})_{\substack{i \in [1..n] \\ j \in [1..m]}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

Lorsque cela ne prête pas à confusion, l'écriture (a_{ij}) sera utilisée sans la précision du domaine des indices i et j . Les opérations suivantes sont définies : soient $A \in \mathbb{R}^{n_A \times m_A}$, $B \in \mathbb{R}^{n_B \times m_B}$, $C \in \mathbb{R}^{n_C \times m_C}$, $x \in \mathbb{R}^{m_A}$, $y \in \mathbb{R}^{n_A}$ et $a \in \mathbb{R}$

- si $n_A = n_B$ et $m_A = m_B$ alors $A + B = C$ avec $n_C = n_A$, $m_C = m_A$ et $C_{ij} = A_{ij} + B_{ij}$;
- si $m_A = n_B$ alors $A \times B = C$ avec $n_C = n_A$, $m_C = m_B$ et $C_{ij} = \sum_{k \in [1..m_A]} A_{ik} B_{kj}$;
- $aA = C$ avec $n_C = n_A$, $m_C = m_A$ et $C_{ij} = aA_{ij}$;
- $|A| = C$ avec $n_C = m_A$, $m_C = n_A$ et $C_{ij} = |A_{ij}|$;
- $A^T = C$ avec $n_C = m_A$, $m_C = n_A$ et $C_{ij} = A_{ji}$;
- $Ax = y$ avec $y_i = \sum_{j \in [1..m_A]} A_{ij} x_j$.

L'ordre des réels conduit à un ordre partiel entre matrices défini par

$$A \leq B \iff (\forall i \in [1..n])(\forall j \in [1..m])(A_{ij} \leq B_{ij}).$$

Les vecteurs réels de \mathbb{R}^n sont identifiés aux matrices de $\mathbb{R}^{n \times 1}$ (et donc les nombres réels aux matrices de $\mathbb{R}^{1 \times 1}$). Ainsi, un vecteur peut-être noté $(x_1, \dots, x_n)^T$. Une matrice carrée, i.e. $n = m$, M est dite régulière si l'application $x \mapsto Mx$ est injective. La matrice M est régulière si et seulement si $Mx = 0 \iff x = 0$ et dans ce cas elle admet une unique inverse, notée M^{-1} , qui vérifie $MM^{-1} = M^{-1}M = I$. La norme $\|M\|$ est définie de la manière suivante :

$$\|M\| = \max_{i \in [1..n]} \sum_{j \in [1..m]} |M_{ij}|$$

Remarque. Cette norme est habituellement notée $\|M\|_\infty$.

Cette définition permet d'obtenir la relation utile $\|Mx\| \leq \|M\| \|x\|$.

1.3 Les fonctions réelles

Une fonction de $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ associe un vecteur de \mathbb{R}^m à chaque vecteur de \mathbb{R}^n .

Remarque. Les fonctions peuvent être définies pour un sous-ensemble de \mathbb{R}^n seulement. Par exemple, \sqrt{x} est définie pour les nombres positifs seulement. Un exposé simplifié est fait en considérant uniquement les fonctions définies dans \mathbb{R}^n . L'utilisation de fonctions définies sur un sous-ensemble de \mathbb{R}^n demandera une adaptation des énoncés présentés.

Une fonction de $\mathbb{R}^n \rightarrow \mathbb{R}^m$ est définie de manière équivalente par m fonctions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$. Il existe de nombreuses définitions différentes (et non équivalentes) de la continuité. Celle qui sera utile pour les développements faits dans cette thèse est la suivante :

Définition II.1. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. f est localement Lipschitz continue si pour tout $a, b \in \mathbb{R}^n$ tels que $a \leq b$, il existe $\lambda > 0$ tel que pour tout $x, y \in \mathbb{R}^n$ tels que $a \leq x \leq b$ et $a \leq y \leq b$,

$$\text{dist}(f(x), f(y)) \leq \lambda \text{dist}(x, y)$$

Ainsi, une fonction localement Lipschitz continue est Lipschitz continue lorsqu'elle est restreinte à n'importe quel ensemble $\{x \in \mathbb{R}^n \mid a \leq x \leq b\}$, i.e. à n'importe quel intervalle de \mathbb{R}^n (voir section 2.2 pour la définition des intervalles de \mathbb{R}^n). Évidemment, une fonction Lipschitz continue est localement Lipschitz continue. Une fonction localement Lipschitz continue est continue au sens habituel, i.e. toute suite convergente a pour image une suite convergente dont la limite est l'image de la limite de la suite originale. La réciproque est fautive, par exemple $f(x) = \sqrt{x}$ est continue dans $[0, 1]$ mais n'est pas Lipschitz continue dans cet intervalle (car $f'(x) = \frac{1}{2\sqrt{x}}$ tend vers l'infini lorsque x tend vers 0). Remarquons que \sqrt{x} est Lipschitz continue dans $(0, 1] = \{x \in \mathbb{R} \mid 0 < x \leq 1\}$.

Les fonctions les plus simples sont les fonctions linéaires, i.e. vérifiant $f(x + y) = f(x) + f(y)$ et $f(ax) = af(x)$. Elles correspondent au produit par une matrice : f est linéaire si et seulement si $f(x) = Mx$ pour une matrice $M \in \mathbb{R}^{m \times n}$ quelconque (voir par exemple Dixmier[47]). Ainsi, une fonction linéaire est Lipschitz continue car

$$\text{dist}(f(x), f(y)) = \|Mx - My\| \leq \|M\| \|x - y\| = \|M\| \text{dist}(x, y).$$

La dérivabilité et la dérivée d'une fonction sont maintenant définies (voir Dixmier[48] pour une présentation plus formelle). Tout d'abord, la tangente d'une fonction f en \tilde{x} , si elle existe, est l'unique fonction affine $Tf(x) = f(\tilde{x}) + M(x - \tilde{x})$ qui vérifie $f(x) \approx Tf(x)$ au voisinage de \tilde{x} . La fonction f est dérivable en x si elle admet une tangente en x . La dérivée $f'(x)$ de f en x est alors la matrice M associée à la tangente. Dans ce cas, les dérivées partielles de f sont définies en x et $f'(x)$ peut-être écrite en les utilisant :

$$f'(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

Remarque. La réciproque est fautive en général : il ne suffit pas en général que les dérivées partielles soient définies et continues pour que la fonction soit dérivable (voir Gelbaum[82] pour des contre-exemples).

Une fonction est dite continûment dérivable si elle est dérivable et si ses dérivées sont continues. Dans le cas des fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$, la dérivée est le coefficient directeur de la tangente (les matrices 1×1 sont identifiées aux nombres réels) et dans le cas des fonctions $f : \mathbb{R}^n \rightarrow \mathbb{R}$, la dérivée est le gradient de f , i.e.

$$f'(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) \in \mathbb{R}^{1 \times n}$$

Cette section est terminée par trois théorèmes fondamentaux qui seront utiles pour les développements à venir.

Théorème des valeurs intermédiaires. Soit $f : [a, b] \rightarrow \mathbb{R}$ continue telle que $f(a) \leq 0$ et $f(b) \geq 0$. Alors, il existe $x \in [a, b]$ tel que $f(x) = 0$.

Le théorème des valeurs intermédiaires peut-être formulé de manière équivalente en énonçant que toute fonction continue $f : [a, b] \rightarrow [a, b]$ admet un point fixe, i.e. il existe $x \in [a, b]$ tel que $f(x) = x$.

Théorème de la valeur moyenne. Soit $f : [a, b] \rightarrow \mathbb{R}$ continûment dérivable. Alors, il existe $c \in [a, b]$ tel que $f(b) = f(a) + c(b - a)$.

Le théorème de Brouwer est une généralisation importante du théorème des valeurs intermédiaires.

Théorème de Brouwer. Soient $D \subseteq \mathbb{R}^n$ non-vide, compact⁴ et convexe et $f : D \rightarrow D$ continue. Alors, il existe $x \in D$ tel que $f(x) = x$.

1.4 Représentation finie des nombres réels

Lorsqu'un ordinateur est utilisé pour effectuer des calculs, seule une précision finie peut être utilisée. En particulier, seul un nombre fini de réels peuvent être représentés. Ainsi, en suivant les notations proposées dans Neumaier[8], notons $\mathbb{F} \subseteq \mathbb{R}$ l'ensemble fini des nombres représentables (une description précise du standard IEEE754 peut-être trouvée dans Goldberg[14]). Deux arrondis optimaux, qui associent à un nombre réel x un nombre représentable le plus proche possible de x , peuvent alors être définis :

$$\begin{aligned}\nabla x &:= \max \{x_m \in \mathbb{F} \mid x_m \leq x\} \\ \Delta x &:= \min \{x_m \in \mathbb{F} \mid x \leq x_m\}\end{aligned}$$

Exemple II.1. Le nombre π n'est pas représentable en virgule flottante. Nous avons $\nabla \pi = 3.1$ et $\Delta \pi = 3.2$.

Remarque. Les cas extrêmes où $\{x_m \in \mathbb{F} \mid x_m \leq x\} = \emptyset$ ou $\{x_m \in \mathbb{F} \mid x \leq x_m\} = \emptyset$ ne sont pas explicitement traités dans cette thèse. Il est aisé d'introduire un élément ∞ qui permet de traiter ces dépassements de capacité (voir par exemple Hickey[43]).

⁴Par le théorème de Heine-Borel, $D \subseteq \mathbb{R}^n$ est compact si et seulement si il est fermé borné.

Chapitre 2

La théorie des intervalles

La théorie moderne des intervalles a été initiée dans les années 60 par Warmus[59], Sunaga[97] et Moore[74]. L'article Hayes[11] donne une bonne introduction technique et historique de la théorie des intervalles. Cette théorie permet en particulier d'effectuer des calculs rigoureux au moyen d'une implémentation en précision finie des nombres et opérations utilisées par les ordinateurs. Les deux exemples suivants montrent à quel point l'utilisation d'une précision finie peut être dommageable. Le premier est une version adaptée au format IEEE 754 de l'exemple proposé dans Rump[91].

Exemple II.2 (Loh[22]). *Soit la fonction*

$$f(x, y) = (333.75 - x^2)y^6 + x^2(11x^2y^2 - 121y^4 - 2) + 5.5y^8 + \frac{x}{2y}.$$

Son évaluation en utilisant une implémentation compatible avec le standard IEEE 754 pour $x = 77617$ et $y = 33096$ donne les résultats suivants :

- $f(x, y) = 1.172604$ avec 32 bits ;
- $f(x, y) = 1.1726039400531786$ avec 64 bits.
- $f(x, y) = 1.1726039400531786318588349045201838$ avec 128 bits.

L'augmentation apparente de la précision du calcul est trompeuse : la véritable valeur de $f(x, y)$ est contenue dans l'intervalle $[-0.8274, -0.8273]$, qui est très éloigné des calculs approchés obtenus par une machine respectant le standard IEEE 754.

Exemple II.3 (Muller[51][52], Ménissier[99]). *Soit la suite définie par la relation de récurrence suivante :*

$$a^{(0)} = \frac{11}{2} ; \quad a^{(1)} = \frac{61}{11} ; \quad a^{(n+1)} = 111 - \frac{1130 - 3000/a^{(n-1)}}{a^{(n)}}$$

En utilisant une implémentation compatible avec le standard IEEE 754, cette suite converge vers 100 en utilisant la simple ou la double précision. Néanmoins, la limite réelle de la suite est 6.

En plus de traiter rigoureusement les erreurs d'arrondi, les intervalles sont utiles pour modéliser d'autres sources d'incertitudes, comme par exemple les incertitudes physiques ou les incertitudes de mesure sur les paramètres d'un système ou les incertitudes inhérentes à certains algorithmes (voir Kearfott[78] pour de nombreuses applications de la théorie des intervalles).

2.1 Les intervalles

2.1.1 Définition

Un intervalle de \mathbb{R} est défini comme l'ensemble des réels compris entre deux réels donnés : étant donnés $a \in \mathbb{R}$ et $b \in \mathbb{R}$ tels que $a \leq b$, l'intervalle

$$\{ x \in \mathbb{R} \mid a \leq x \leq b \}$$

est noté $[a, b]$. En suivant les notations proposées dans Kearfott[25], les intervalles sont notés en lettres grasses et l'ensemble des intervalles est noté \mathbb{IR} (\mathbb{IR} est donc l'ensemble des intervalles fermés bornés et non-vides). La borne inférieure d'un intervalle $\mathbf{x} = [a, b]$ est noté $\inf \mathbf{x} = a$ et sa borne supérieure $\sup \mathbf{x} = b$. Il pourra être utile de noter les bornes d'un intervalle \mathbf{x} par $\underline{\mathbf{x}}$ et $\overline{\mathbf{x}}$. Cette notation n'étant pas standard, elle sera explicitement définie par $\mathbf{x} = [\underline{\mathbf{x}}, \overline{\mathbf{x}}]$ à chaque utilisation.

Remarque. *Il existe des extensions des intervalles qui prennent en compte l'intervalle vide et les intervalles non bornés (voir par exemple Hickey[43]).*

Ainsi, nous utiliserons le terme "intervalle" pour les intervalles non-vides, fermés et bornés. L'intérieur d'un intervalle $\mathbf{x} = [a, b]$ est noté $\text{int } \mathbf{x} = \{x \in \mathbb{R} \mid a < x < b\}$. Un intervalle dégénéré, i.e. $\mathbf{x} = [x, x]$, est identifié au nombre réel x et est aussi noté x . Les intervalles héritent naturellement de l'inclusion entre les ensembles de nombres réels :

$$\mathbf{x} \subseteq \mathbf{y} \iff (\forall t \in \mathbb{R})(t \in \mathbf{x} \implies t \in \mathbf{y})$$

Les intervalles étant des sous-ensembles de \mathbb{R} très simples, l'inclusion peut-être caractérisée par une formule elle aussi très simple mettant en jeu les bornes des intervalles :

$$\mathbf{x} \subseteq \mathbf{y} \iff (\inf \mathbf{y} \leq \inf \mathbf{x} \wedge \sup \mathbf{x} \leq \sup \mathbf{y})$$

Comme l'illustre l'exemple suivant, l'intersection et la réunion ensemblistes ne sont pas des opérations internes à \mathbb{IR} .

Exemple II.4. *Considérons $\mathbf{x} = [-1, 1]$ et $\mathbf{y} = [2, 3]$. Alors d'une part,*

- $\mathbf{x} \cap \mathbf{y} = \emptyset$ *qui n'est pas un intervalle par définition de \mathbb{IR} .*
- $\mathbf{x} \cup \mathbf{y} = \{t \in \mathbb{R} \mid t \in \mathbf{x} \vee t \in \mathbf{y}\}$ *qui n'est pas un intervalle car il n'est pas connexe.*

Néanmoins, les trois conditions suivantes sont équivalentes quelques soient $\mathbf{x}, \mathbf{y} \in \mathbb{IR}$:

1. $\mathbf{x} \cap \mathbf{y} \neq \emptyset$.
2. $\mathbf{x} \cap \mathbf{y}$ est un intervalle.
3. $\mathbf{x} \cup \mathbf{y}$ est un intervalle.

Les intervalles héritent aussi de l'ordre \leq des nombres réels :

$$\mathbf{x} \leq \mathbf{y} \iff (\forall x \in \mathbf{x})(\forall y \in \mathbf{y})(x \leq y)$$

Il peut être caractérisé grâce aux bornes des intervalles mis en jeu :

$$\mathbf{x} \leq \mathbf{y} \iff \sup \mathbf{x} \leq \inf \mathbf{y}$$

Les opérations suivantes donnent des informations importantes qui caractérisent un intervalle : étant donné un intervalle $\mathbf{x} = [\underline{\mathbf{x}}, \bar{\mathbf{x}}]$,

$$\begin{aligned}\text{mid } \mathbf{x} &= 1/2 (\underline{\mathbf{x}} + \bar{\mathbf{x}}) \\ \text{rad } \mathbf{x} &= 1/2 (\bar{\mathbf{x}} - \underline{\mathbf{x}}) \\ |\mathbf{x}| &= \max\{|x| \mid x \in \mathbf{x}\} = \sup\{|\underline{\mathbf{x}}|, |\bar{\mathbf{x}}|\} \\ \langle \mathbf{x} \rangle &= \min\{|x| \mid x \in \mathbf{x}\}\end{aligned}$$

On a alors $\langle \mathbf{x} \rangle = \min\{|\underline{\mathbf{x}}|, |\bar{\mathbf{x}}|\}$ si $0 \notin \mathbf{x}$ ou $\langle \mathbf{x} \rangle = 0$ sinon. De plus, la largeur d'un intervalle est définie comme deux fois son rayon : $\text{wid } \mathbf{x} = 2 \text{rad } \mathbf{x}$.

2.1.2 Intersection et réunion d'intervalles

L'intersection et la réunion ensemblistes ne sont pas des opérations internes à \mathbb{IR} . L'inclusion étant un ordre partiel pour intervalles, il est naturel de définir de nouvelles opérations d'intersection et de réunion d'intervalles en considérant les bornes inférieures et supérieures de cet ordre partiel : étant donné un ensemble d'intervalles $E \subseteq \mathbb{IR}$,

- l'intersection d'intervalles⁵ $\bigwedge E$ est la borne inférieure de E dans (\mathbb{IR}, \subseteq) , c'est à dire le plus grand intervalle qui est inclus dans tous les intervalles de E .
- la réunion d'intervalles⁶ $\bigvee E$ est la borne supérieure de E dans (\mathbb{IR}, \subseteq) , c'est à dire le plus petit intervalle qui contient tous les intervalles de E .

Notation. Lorsque $E = \{\mathbf{x}, \mathbf{y}\}$ ne contient que deux éléments, l'intersection et la réunion des intervalles de E sont aussi notées $\mathbf{x} \bigwedge \mathbf{y}$ et $\mathbf{x} \bigvee \mathbf{y}$.

Dans l'exemple suivant, l'intersection et la réunion d'intervalles coïncident avec l'intersection et la réunion ensemblistes.

Exemple II.5. Soit $E = \{[-1, 1], [0, 2]\}$. Les intervalles qui sont inclus à la fois dans $[-1, 1]$ et dans $[0, 2]$ sont les intervalles qui sont inclus dans $[0, 1]$. Le plus grand d'entre eux est $[0, 1] = [-1, 1] \bigwedge [0, 2]$. Les intervalles qui contiennent à la fois $[-1, 1]$ et $[0, 2]$ sont les intervalles qui contiennent $[-1, 2]$. Le plus petit d'entre eux est $[-1, 2] = [-1, 1] \bigvee [0, 2]$. Nous avons donc $[-1, 1] \bigwedge [0, 2] = [-1, 1] \cap [0, 2]$ et $[-1, 1] \bigvee [0, 2] = [-1, 1] \cup [0, 2]$.

Dans l'exemple suivant, l'intersection et la réunion d'intervalles ne coïncident pas avec l'intersection et la réunion ensemblistes.

Exemple II.6. Soit $E = \{[-1, 1], [2, 3]\}$. L'ensemble des intervalles inclus à la fois dans $[-1, 1]$ et dans $[2, 3]$ est vide et n'a donc pas de plus grand élément. L'intersection d'intervalles $[-1, 1] \bigwedge [2, 3]$ n'est donc pas définie alors que $[-1, 1] \cap [2, 3] = \emptyset$. Les intervalles qui contiennent à la fois $[-1, 1]$ et $[2, 3]$ sont les intervalles qui contiennent $[-1, 3]$. Le plus petit d'entre eux est $[-1, 3] = [-1, 1] \bigvee [2, 3]$. Ce dernier est différent de $[-1, 1] \cup [2, 3]$.

Le fait que l'intersection d'intervalle $\mathbf{x} \bigwedge \mathbf{y}$ ne soit pas définie pour des intervalles disjoints est un défaut théorique mais n'a pas de réelle influence pratique : en effet, si l'union d'intervalles $\mathbf{x} \bigvee \mathbf{y}$ est beaucoup utilisée, l'intersection d'ensemble $\mathbf{x} \cap \mathbf{y}$ est généralement utilisée à la place de l'intersection d'intervalles $\mathbf{x} \bigwedge \mathbf{y}$. De plus, l'utilisation en pratique d'un intervalle vide ne pose pas de difficulté (Walster[35, 36]).

⁵"Meet" en anglais.

⁶"Join" en anglais.

Formellement, $\bigwedge E$ et $\bigvee E$ sont donc définies de la manière suivante : lorsqu'ils existent, ce sont les intervalles qui vérifient respectivement

$$\begin{aligned} (\forall \mathbf{z} \in \mathbb{IR}) \quad ((\forall \mathbf{y} \in E, \mathbf{z} \subseteq \mathbf{y}) &\iff \mathbf{z} \subseteq (\bigwedge E)) \\ (\forall \mathbf{z} \in \mathbb{IR}) \quad ((\forall \mathbf{y} \in E, \mathbf{y} \subseteq \mathbf{z}) &\iff (\bigvee E) \subseteq \mathbf{z}) \end{aligned}$$

On peut vérifier que l'intersection d'intervalles $\bigwedge E$ est définie si et seulement si E est minoré, i.e. $(\exists \mathbf{x} \in \mathbb{IR})(\forall \mathbf{y} \in E)(\mathbf{x} \subseteq \mathbf{y})$; la réunion d'intervalles $\bigvee E$ est définie si et seulement si E est majoré, i.e. $(\exists \mathbf{x} \in \mathbb{IR})(\forall \mathbf{y} \in E)(\mathbf{y} \subseteq \mathbf{x})$. Lorsqu'ils sont définis, ces intervalles sont uniques. Lorsque l'intersection d'intervalles est définie, elle coïncide avec l'intersection ensembliste :

$$\mathbf{x} \cap \mathbf{y} \neq \emptyset \implies \mathbf{x} \cap \mathbf{y} = \mathbf{x} \bigwedge \mathbf{y}$$

Par contre, la réunion d'intervalles ne coïncide pas en général avec la réunion ensembliste (voir exemple II.4). Nous avons néanmoins la propriété suivante :

$$\mathbf{x} \cap \mathbf{y} \neq \emptyset \implies \mathbf{x} \cup \mathbf{y} = \mathbf{x} \bigvee \mathbf{y}$$

Ces opérations peuvent être caractérisées simplement grâce aux bornes des intervalles mis en jeu. Lorsqu'elles sont définies, elles peuvent être déterminées grâce aux formules suivantes :

$$\begin{aligned} \bigwedge E &= [\sup\{\inf \mathbf{x} | \mathbf{x} \in E\} , \inf\{\sup \mathbf{x} | \mathbf{x} \in E\}] \\ \bigvee E &= [\inf\{\inf \mathbf{x} | \mathbf{x} \in E\} , \sup\{\sup \mathbf{x} | \mathbf{x} \in E\}] \end{aligned}$$

Dans le cas de où E contient deux intervalles $\mathbf{x} = [\underline{\mathbf{x}}, \bar{\mathbf{x}}]$ et $\mathbf{y} = [\underline{\mathbf{y}}, \bar{\mathbf{y}}]$, nous obtenons les expressions suivantes :

$$\begin{aligned} \mathbf{x} \bigwedge \mathbf{y} &= [\sup\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\} , \inf\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}] \\ \mathbf{x} \bigvee \mathbf{y} &= [\inf\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\} , \sup\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}] \end{aligned}$$

En conclusion, si l'inclusion est un ordre partiel pour les intervalles, (\mathbb{IR}, \subseteq) n'est pas un treillis car il existe des paires d'intervalles de \mathbb{IR} qui n'ont pas de borne inférieure dans \mathbb{IR} (par exemple $\{-1, 1\}, [2, 3\}$).

2.1.3 Distances entre intervalles

Les intervalles étant des sous-ensembles compacts de \mathbb{R} , la distance de \mathbb{R} peut être utilisée pour construire une distance sur \mathbb{IR} en utilisant la distance de Hausdorff (Preparata[30]). Cela conduit à la définition suivante :

$$\text{dist}_H(\mathbf{x}, \mathbf{y}) = \max\left\{ \max_{y \in \mathbf{y}} \min_{x \in \mathbf{x}} \text{dist}(x, y) , \max_{x \in \mathbf{x}} \min_{y \in \mathbf{y}} \text{dist}(x, y) \right\}$$

A cause de la forme simple des ensembles mis en jeu, cette distance peut-être facilement exprimée en utilisant les bornes des intervalles mis en jeu :

$$\text{dist}_H(\mathbf{x}, \mathbf{y}) = \max\{ |\inf \mathbf{x} - \inf \mathbf{y}|, |\sup \mathbf{x} - \sup \mathbf{y}| \}$$

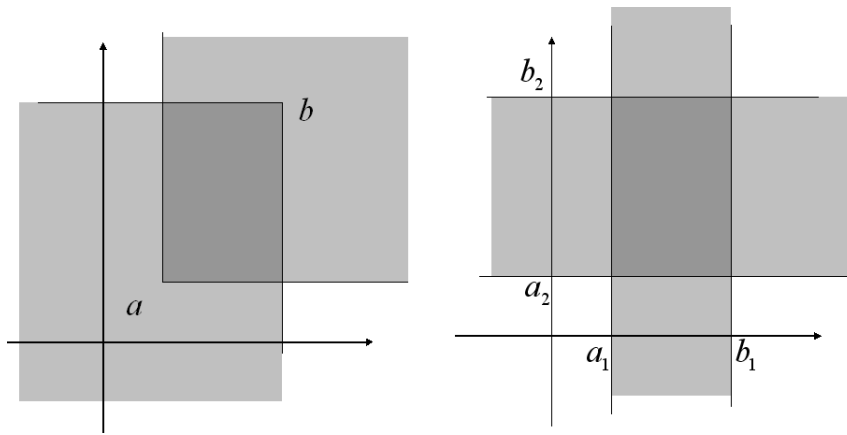


FIG. 2.1 – Vecteur d'intervalle : deux définitions équivalentes.

Une autre façon de construire des distances entre intervalles consiste à considérer un intervalle comme un couple de réels. De manière générale, toute distance dist' de \mathbb{R}^2 peut-être utilisée pour construire une distance entre intervalles (qui est notée avec le même symbole) :

$$\text{dist}'(\mathbf{x}, \mathbf{y}) = \text{dist}'((\inf \mathbf{x}, \sup \mathbf{x})^T, (\inf \mathbf{y}, \sup \mathbf{y})^T)$$

Ainsi, en utilisant la distance dist de \mathbb{R}^2 définie dans la chapitre 1, la distance entre intervalles $\text{dist}(\mathbf{x}, \mathbf{y})$ est définie par $\max\{|\inf \mathbf{x} - \inf \mathbf{y}|, |\sup \mathbf{x} - \sup \mathbf{y}|\}$. Cette expression coïncide avec la distance de Hausdorff et est la distance habituellement utilisée pour les intervalles.

2.2 Vecteurs et matrices intervalles

2.2.1 Vecteurs d'intervalles

Un vecteur d'intervalles peut être défini de trois manières équivalentes :

- comme un vecteur dont les composantes sont des intervalles ;
- comme le produit cartésien de n intervalles ;
- comme un intervalle de \mathbb{R}^n .

Tout d'abord, les deux dernières définitions sont équivalentes. En effet, étant donnés $a, b, x \in \mathbb{R}^n$

$$a \leq x \leq b \iff (\forall i \in [1..n], a_i \leq x_i \leq b_i)$$

Cette équivalence est illustrée par la figure 2.1. Enfin, le produit cartésien de n intervalles étant décrit par n intervalles, il est identifié à un vecteur dont les composantes sont ces intervalles. Un tel vecteur d'intervalle est aussi appelé une boîte de \mathbb{R}^n .

Exemple II.7. Soit l'intervalle de \mathbb{R}^2

$$\mathbf{x} = [(0, 0)^T, (1, 2)^T] = \{x \in \mathbb{R}^2 \mid (0, 0)^T \leq x \leq (1, 2)^T\}.$$

Nous avons aussi l'égalité suivante :

$$\mathbf{x} = \{x \in \mathbb{R}^2 \mid 0 \leq x_1 \leq 1 \wedge 0 \leq x_2 \leq 2\}.$$

Cet intervalle s'écrit aussi comme un produit cartésien d'intervalles, i.e. $\mathbf{x} = [0, 1] \times [1, 2]$, et comme un vecteur dont les composantes sont des intervalles, i.e. $\mathbf{x} = ([0, 1], [1, 2])^T$.

Remarque. La notation des vecteurs d'intervalles sous la forme de produit cartésien doit être utilisée avec prudence car elle entre en conflit avec la notation de la multiplication de l'arithmétique des intervalles, cf. Section 2.5.2 page 32.

L'ensemble des vecteurs d'intervalles de dimension n est noté \mathbb{IR}^n . Les vecteurs d'intervalles étant des parties de \mathbb{R}^n , y sont définis l'appartenance, l'inclusion, l'intersection et la réunion. Les opérations mid, rad, wid et $|\cdot|$ sont définies en appliquant leurs versions définies dans \mathbb{IR} composante par composante. Comme dans le cas des intervalles, l'intersection et la réunion ne sont pas en général des opérations internes à \mathbb{IR}^n . Ainsi, l'intersection et la réunion de vecteurs d'intervalles sont définies comme dans le cas des intervalles en utilisant les bornes inférieures et supérieures de l'inclusion. Leur expression peut-être donnée en utilisant l'intersection et la réunion d'intervalles :

- $\mathbf{x} \wedge \mathbf{y} = (\mathbf{x}_1 \wedge \mathbf{y}_1, \dots, \mathbf{x}_n \wedge \mathbf{y}_n)^T$ est définie si et seulement si $\mathbf{x} \cap \mathbf{y} \neq \emptyset$;
- $\mathbf{x} \vee \mathbf{y} = (\mathbf{x}_1 \vee \mathbf{y}_1, \dots, \mathbf{x}_n \vee \mathbf{y}_n)^T$.

Il est facile de vérifier que si $\mathbf{x} \wedge \mathbf{y}$ est définie alors $\mathbf{x} \wedge \mathbf{y} = \mathbf{x} \cap \mathbf{y}$. De plus $\mathbf{x} \vee \mathbf{y} = \mathbf{x} \cup \mathbf{y}$ si et seulement si $\mathbf{x} \subseteq \mathbf{y}$ ou $\mathbf{y} \subseteq \mathbf{x}$. La distance entre vecteurs d'intervalles est définie de la manière suivante :

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \max_{k \in [1..n]} \text{dist}(\mathbf{x}_k, \mathbf{y}_k).$$

Cette distance coïncide avec la distance de Hausdorff entre les ensembles de réels correspondant aux vecteurs d'intervalles.

2.2.2 Matrices d'intervalles

Comme dans le cas des vecteurs d'intervalles, une matrice d'intervalles \mathbf{M} peut être définie de deux manières équivalentes :

- étant données deux matrices réelles $\underline{M} \in \mathbb{R}^{n \times m}$ et $\overline{M} \in \mathbb{R}^{n \times m}$,

$$\mathbf{M} = [\underline{M}, \overline{M}] = \{M \in \mathbb{R}^{n \times m} \mid \underline{M} \leq M \leq \overline{M}\},$$

i.e. \mathbf{M} est un intervalle de matrices ;

- étant donnés $n \times m$ intervalles \mathbf{m}_{ij} , $\mathbf{M} = (\mathbf{m}_{ij})$, i.e. \mathbf{M} est une matrice d'intervalles.

Les matrices carrées particulières suivantes seront utiles :

- \mathbf{M} est régulière si $\forall M \in \mathbf{M}$, M est régulière ;
- \mathbf{M} est à diagonale strictement dominante si pour tout $i \in [1..n]$, $\langle \mathbf{M}_{ii} \rangle \geq \sum_{j \neq i} |\langle \mathbf{M}_{ij} \rangle|$;
- \mathbf{M} est une M-matrices si $\mathbf{M}_{ij} \leq 0$ pour $i \neq j$ et si il existe un vecteur $u \in \mathbb{R}^n$ tel que $\mathbf{M}u > 0$.

La matrice de comparaison⁷ associée à une matrice \mathbf{M} est notée $\langle \mathbf{M} \rangle$ et définie par $\langle \mathbf{M} \rangle_{ii} = |\mathbf{M}_{ii}|$ et $\langle \mathbf{M} \rangle_{ij} = -\langle \mathbf{M} \rangle$ pour $i \neq j$. Alors, les H-matrices sont définies de la manière suivante :

- \mathbf{M} est une H-matrice si sa matrice de comparaison est une M-matrice.

Lorsque la matrice mise en jeux est dégénérée, les définitions précédentes coïncident avec les définitions correspondantes pour les matrices réelles. La définition suivante explicite une sous-classe des matrices d'intervalles régulières qui sera utile :

⁷"Comparison matrix" en anglais.

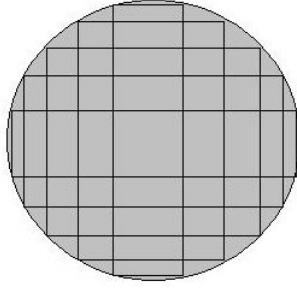


FIG. 2.2 – Un disque fermé possède une infinité d'approximations intérieures non-extensibles.

Définition II.2 (Neumaier[8]). *La matrice d'intervalles $\mathbf{A} \in \mathbb{IR}^{n \times n}$ est fortement régulière si et seulement si $(\text{mid } \mathbf{A})$ est régulière et si la matrice d'intervalles $(\text{mid } \mathbf{A})^{-1} \mathbf{A}$ est régulière, i.e. ne contient que des matrices réelles régulières.*

De nombreuses propriétés de ces matrices d'intervalles sont décrites dans Neumaier[8].

2.2.3 Approximations d'ensembles quelconques

Les intervalles peuvent être utilisés pour approximer des sous-ensembles de \mathbb{R}^n quelconques. L'idéal pour approximer un sous ensemble A de \mathbb{R}^n serait de disposer de deux intervalles $\mathbf{x} \in \mathbb{IR}^n$ et $\mathbf{y} \in \mathbb{IR}^n$ qui encadrent A , i.e. $\mathbf{x} \subseteq A \subseteq \mathbf{y}$. Ces intervalles \mathbf{x} et \mathbf{y} sont alors appelés respectivement une approximation intérieure et une approximation extérieure de A . Une approximation intérieure de A est dite non-extensible si

$$(\forall \mathbf{z} \in \mathbb{IR}^n) (\mathbf{x} \subseteq \mathbf{z} \subseteq A \implies \mathbf{x} = \mathbf{z}).$$

Une approximation extérieure de A est dite non-rétractable si

$$(\forall \mathbf{z} \in \mathbb{IR}^n) (A \subseteq \mathbf{z} \subseteq \mathbf{x} \implies \mathbf{x} = \mathbf{z}).$$

L'approximation intérieure d'un sous-ensemble de \mathbb{R}^n pose de nombreux problèmes, parmi lesquels l'existence possible d'une infinité d'approximations intérieures non-extensibles différentes, voir Figure 2.2, ou l'absence d'approximation intérieure non-extensibles. Par contre, l'approximation extérieure d'un sous-ensemble de \mathbb{R}^n possède de meilleures propriétés : tout sous-ensemble non vide de \mathbb{R}^n qui est borné possède une unique approximation extérieure non-rétractable, qui est notée $\square A$ ⁸.

Exemple II.8. *Considérons $E = \{0, 1\}$, $E' = \{(x, y)^T \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$ et $E'' = \{(x, y)^T \in \mathbb{R}^2 | x^2 + y^2 < 1\}$. Alors,*

- E a deux approximations intérieures non-extensibles : $[0, 0]$ et $[1, 1]$. L'unique approximation extérieure non-rétractable de E est $\square E = [0, 1]$.
- E' a une infinité d'approximations intérieures non-extensibles : chaque boîte

$$([-\cos t, \cos t], [-\sin t, \sin t])^T, \quad t \in [0, \pi/2]$$

est une approximation intérieure non extensible (voir Figure 2.2). L'unique approximation extérieure non-rétractable de E est $\square E' = [-1, 1] \times [-1, 1]$.

⁸"Interval hull" dans la littérature anglaise.

- E'' n'a aucune approximation intérieure non-extensible. En effet toute approximation intérieure de E'' peut être étendue car E'' est ouvert. L'unique approximation extérieure non-rétractable de E'' est $\square E'' = [-1, 1] \times [-1, 1]$.

2.3 Fonctions d'intervalles

Une fonction d'intervalles est une fonction $\mathbf{f} : \mathbb{IR}^n \longrightarrow \mathbb{IR}^m$ qui associe à un n -uple d'intervalles un m -uple d'intervalles. Elle peuvent être définies de deux manières : d'une part en utilisant des opérations sur les intervalles, eg. $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{x} \vee \mathbf{y}$. D'autre part, une fonction d'intervalles peut être définie par la donnée de deux fonctions réelles, eg. $\mathbf{f}(\mathbf{x}, \mathbf{y}) = [\inf \mathbf{x} - \sup \mathbf{y}, \sup \mathbf{x} - \inf \mathbf{y}]$.

2.3.1 Continuité

La distance entre intervalles permet d'étudier la continuité des fonctions d'intervalles. Parmi les nombreuses définitions de continuité, celle qui aura le plus d'utilité pour les développements à venir est la continuité au sens de Lipschitz.

Définition II.3. Soit $\mathbf{g} : \mathbb{IR}^n \longrightarrow \mathbb{IR}^m$ une fonction d'intervalles. \mathbf{g} est localement Lipschitz continue si et seulement si quelque soit $\mathbf{z} \in \mathbb{IR}^n$, il existe une constante γ telle que quelque soient $\mathbf{x} \subseteq \mathbf{z}$ et $\mathbf{y} \subseteq \mathbf{z}$

$$\text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y})) \leq \gamma \text{dist}(\mathbf{x}, \mathbf{y})$$

Un résultat classique prouve que la continuité locale de Lipschitz implique la continuité usuelle (l'image de toute suite convergente est convergente et converge vers l'image de la suite originale). Lorsque $m > 1$, les deux conditions suivantes sont équivalentes :

- $\mathbf{g} = (\mathbf{g}_1, \dots, \mathbf{g}_m)^T$ est localement Lipschitz continue ;
- pour tout $i \in [1..m]$, \mathbf{g}_i est localement Lipschitz continue.

Finalement, la composition de fonctions localement Lipschitz continues est localement Lipschitz continue.

2.3.2 Croissance par rapport à l'inclusion

La croissance par rapport à l'inclusion des fonctions d'intervalles est définie de la manière suivante :

Définition II.4. Soit $\mathbf{g} : \mathbb{IR}^n \longrightarrow \mathbb{IR}^m$ une fonction d'intervalles. \mathbf{g} est croissante par rapport à l'inclusion si et seulement si quelque soient $\mathbf{x}, \mathbf{y} \in \mathbb{IR}^n$,

$$\mathbf{x} \subseteq \mathbf{y} \implies \mathbf{g}(\mathbf{x}) \subseteq \mathbf{g}(\mathbf{y})$$

La croissance par rapport à l'inclusion jouera un rôle important dans les développements à venir.

2.4 Extensions des fonctions continues

Les extensions aux intervalles⁹ des fonctions réelles continues sont les objets principaux de la théorie des intervalles. La démarche suivie pour leur construction est la suivante :

⁹"Interval extensions" en anglais.

1. définir la condition pour qu'une fonction d'intervalle soit une extension d'une fonction réelle donnée (section 2.4) ;
2. calculer formellement les extensions des fonctions élémentaires (section 2.5) ;
3. construire des extensions pour les fonctions générales à partir des extensions élémentaires calculée précédemment (sections 2.6 pour l'extension naturelle et 2.7 pour l'extension de la valeur moyenne).

2.4.1 Définition

Les extensions aux intervalles d'une fonction réelle continue f sont introduites pour construire des approximations extérieures de l'image de f .

Définition II.5 (Neumaier[8]). Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \longrightarrow \mathbb{IR}^m$ une fonction d'intervalles. \mathbf{g} est une extension aux intervalles de f si et seulement si les deux conditions suivantes sont vérifiées :

1. $(\forall x \in \mathbb{R}^n)(\mathbf{g}(x) = f(x))$
2. $(\forall \mathbf{x} \in \mathbb{IR}^n)(\mathbf{g}(\mathbf{x}) \supseteq \text{range}(f, \mathbf{x}))$

Remarque. La condition 1 ne pourra pas toujours être vérifiée (en particulier lorsque les extensions sont arrondies). En suivant la dénomination introduite dans Neumaier[8], une fonction d'intervalles ne vérifiant que la condition 2 sera appelée une extension faible aux intervalles¹⁰ de f . Les extensions faibles aux intervalles sont parfois appelées fonction d'inclusion¹¹, par exemple dans Jaulin[55], ou extension aux intervalles, par exemple dans Kearfott[80]. La condition 1 est importante pour l'étude de la qualité des extensions, i.e. de leur ordre de convergence.

La croissance par rapport à l'inclusion permet de caractériser facilement certaines extensions. Le théorème suivant est souvent nommé le théorème fondamental de l'analyse par intervalles :

Théorème II.1 (Moore[74]). Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \longrightarrow \mathbb{IR}^m$ une fonction d'intervalles croissante par rapport à l'inclusion qui vérifie

$$(\forall x \in \mathbb{R}^n)(\mathbf{g}(x) = f(x))$$

Alors \mathbf{g} est une extension de f .

Démonstration (Moore[74]). Considérons un intervalle quelconque $\mathbf{x} \in \mathbb{IR}^n$. Alors, quelque soit $x \in \mathbf{x}$, nous avons $x \subseteq \mathbf{x}$ et la croissance de \mathbf{g} implique $\mathbf{g}(x) \subseteq \mathbf{g}(\mathbf{x})$. C'est à dire $f(x) \in \mathbf{g}(\mathbf{x})$. \square

2.4.2 Optimalité

L'inclusion entre intervalles permet de comparer différentes extensions d'une fonction : l'extension \mathbf{g} est plus précise que l'extension \mathbf{h} en \mathbf{x} si $\mathbf{g}(\mathbf{x}) \subseteq \mathbf{h}(\mathbf{x})$. Cela conduit naturellement à la définition suivante d'une extension optimale.

¹⁰"Weak interval extension" dans la littérature anglaise.

¹¹"inclusion function" dans la littérature anglaise.

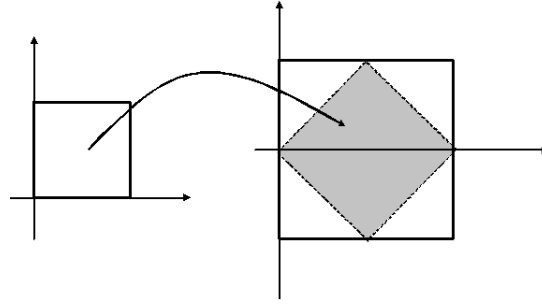


FIG. 2.3 – Effet d'enveloppe.

Définition II.6. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \rightarrow \mathbb{IR}^m$ une extension de f . L'extension \mathbf{g} est optimale en $\mathbf{x} \in \mathbb{IR}^n$ si et seulement si

$$(\forall \mathbf{z} \in \mathbb{IR}^m) (\text{range}(f, \mathbf{x}) \subseteq \mathbf{z} \subseteq \mathbf{g}(\mathbf{x}) \implies \mathbf{z} = \mathbf{g}(\mathbf{x})).$$

De plus, \mathbf{g} est optimale si et seulement si elle est optimale pour tout $\mathbf{x} \in \mathbb{IR}^n$.

C'est à dire, \mathbf{g} est optimale en \mathbf{x} si $\mathbf{g}(\mathbf{x})$ est une approximation extérieure optimale de $\text{range}(f, \mathbf{x})$. Une fonction f possède donc une *unique* extension optimale qui sera notée $f^{\mathbb{IR}}$:

$$f^{\mathbb{IR}}(\mathbf{x}) = \square \text{range}(f, \mathbf{x})$$

L'extension optimale peut être exprimée grâce à la réunion d'intervalles :

$$f^{\mathbb{IR}}(\mathbf{x}) = \bigvee_{x \in \mathbf{x}} f(x) = [\inf_{x \in \mathbf{x}} f(x), \sup_{x \in \mathbf{x}} f(x)] \quad (2.1)$$

L'expression de l'extension optimale d'une fonction $f = (f_i)_{i \in [1..m]} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ peut être obtenue grâce aux expressions des extensions optimales de ses composantes f_i :

$$f^{\mathbb{IR}}(\mathbf{x}) = (f_1^{\mathbb{IR}}(\mathbf{x}), \dots, f_m^{\mathbb{IR}}(\mathbf{x}))^T \quad (2.2)$$

En effet quelque soit $A \subseteq \mathbb{R}^n$, et donc en particulier pour $A = \text{range}(f, \mathbf{x})$, l'approximation optimale de A est le produit cartésien des approximations optimales des projection de A sur les différent axes, i.e. $\square A = (\square A_1) \times \dots \times (\square A_n)$ où A_i est la projection de A sur l'axe x_i . L'optimalité est un gage de qualité pour les fonctions $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Cependant, une extension optimale pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ peut donner une mauvaise description de $\text{range}(f, \mathbf{x})$. Ce problème est appelé l'effet d'enveloppe¹² (Neumaier[9]).

Exemple II.9. Soit $f(x) = (x_1 + x_2, x_1 - x_2)^T$. Alors, comme illustré par la figure 2.3, $f^{\mathbb{IR}}([0, 1] \times [0, 1]) = ([0, 2], [-1, 1])^T$. Ce dernier intervalle est l'approximation extérieure optimale de

$$\text{range}(f, [0, 1] \times [0, 1])$$

mais n'est malgré tout qu'une approximation grossière.

Les extensions aux intervalles ne peuvent pas surmonter directement ce problème. Seules des techniques spécifiques aux applications, recourant généralement à un changement de base bien choisi, peuvent diminuer l'influence de l'effet d'enveloppe (Nickel[64], Neumaier[9] ou Kuhn[53]).

¹²"Wrapping effect" dans la littérature anglaise.

2.4.3 Ordre de convergence

Les extensions qui vont être construites par la suite présentent généralement un mauvais comportement lorsqu'elles sont utilisées sur de larges intervalles : elles permettent bien de construire une approximation extérieure de l'image de larges intervalles par la fonction qu'elles étendent, mais cette approximation est souvent de mauvaise qualité. Ainsi, la qualité d'une extension est plutôt mesurée par rapport à son comportement sur de petits intervalles. Cette mesure est formalisée par le concept d'ordre de convergence qui est généralement défini de la manière suivante :

Définition II.7 (Moore[74]). Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \rightarrow \mathbb{IR}^m$ une extension de f . L'extension \mathbf{g} a un ordre de convergence $\alpha > 0$ si et seulement si quelque soit $\mathbf{x}^{\text{ref}} \in \mathbb{IR}^n$, il existe $\gamma > 0$ tel que quelque soit $\mathbf{x} \subseteq \mathbf{x}^{\text{ref}}$,

$$\|\text{wid } \mathbf{g}(\mathbf{x}) - \text{wid } f^{\mathbb{IR}}(\mathbf{x})\| \leq \gamma \|\text{wid } \mathbf{x}\|^\alpha$$

Les ordres de convergence généralement considérés sont entiers. Pour $\alpha = 1$, l'expression "ordre de convergence linéaire" est généralement utilisée et pour $\alpha = 2$, "ordre de convergence quadratique". Si une extension a un ordre de convergence α , elle a aussi un ordre de convergence α' pour tout $\alpha' < \alpha$. L'ordre de convergence d'une extension est lié à la distance entre cette extension et l'extension optimale : informellement, l'inclusion $f^{\mathbb{IR}}(\mathbf{x}) \subseteq \mathbf{g}(\mathbf{x})$ est toujours vérifiée ; donc si les deux intervalles $f^{\mathbb{IR}}(\mathbf{x})$ et $\mathbf{g}(\mathbf{x})$ ont presque la même largeur, alors, ils sont très proches l'un de l'autre. La proposition suivante formalise cette idée.

Proposition II.1. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \rightarrow \mathbb{IR}^m$ une extension de f . L'extension \mathbf{g} a un ordre de convergence $\alpha > 0$ si et seulement si quelque soit $\mathbf{x}^{\text{ref}} \in \mathbb{IR}^n$, il existe $\gamma > 0$ tel que quelque soit $\mathbf{x} \subseteq \mathbf{x}^{\text{ref}}$,

$$\text{dist}(\mathbf{g}(\mathbf{x}), f^{\mathbb{IR}}(\mathbf{x})) \leq \gamma \|\text{wid } \mathbf{x}\|^\alpha$$

Cette section est terminée par le résultat suivant, qui justifie l'introduction de la continuité au sens de Lipschitz.

Proposition II.2. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \rightarrow \mathbb{IR}^m$ une extension de f . Si \mathbf{g} est localement Lipschitz continue alors elle a un ordre de convergence linéaire.

2.4.4 Arrondi extérieur des extensions aux intervalles

Lorsque des ordinateurs sont utilisés pour effectuer des calculs, la représentation finie des nombres doit être prise en compte. Dans le contexte des extensions aux intervalles, l'arrondi extérieur est compatible avec la sémantique des extensions. L'arrondi extérieur d'un intervalle $\mathbf{x} \in \mathbb{IR}^n$ est noté $\diamond \mathbf{x} \in \mathbb{IF}^n$, où \mathbb{IF} est l'ensemble des intervalles dont les bornes sont dans \mathbb{F} , et vérifie $\mathbf{x} \subseteq \diamond \mathbf{x}$. L'arrondi d'une extension aux intervalles est effectué en deux étapes :

1. arrondi extérieur des arguments ;
2. arrondi extérieur de l'extension.

Ces deux arrondis sont utilisés de la manière suivante : soit \mathbf{g} une extension de f , nous devons être capable de calculer une extension arrondi \mathbf{g}^\diamond à partir de \mathbf{g} , i.e. vérifiant $\mathbf{g}(\mathbf{x}) \subseteq \mathbf{g}^\diamond(\mathbf{x})$ quelque soit \mathbf{x} . L'extension \mathbf{g} est souvent exprimée comme une composition de fonctions d'intervalles et \mathbf{g}^\diamond est alors obtenue en composant les même fonctions et en utilisant un arrondi extérieur. Alors la fonction d'intervalles $\mathbf{h} : \mathbb{IR}^n \longrightarrow \mathbb{IF}^m$ définie par $\mathbf{h}(\mathbf{x}) = \diamond \mathbf{g}(\diamond \mathbf{x})$ est une extension faible de f . L'exemple suivant illustre ce protocole d'arrondi.

Exemple II.10. *Supposons que l'on veuille calculer une approximation extérieure de*

$$\text{range}(\ln, \mathbf{x}) \quad \text{avec } \mathbf{x} = [e, e^2]$$

en utilisant une précision de deux décimales. Premièrement, \mathbf{x} est arrondi de la manière suivante : $\diamond \mathbf{x} = [2.71, 7.39]$. Ensuite, l'extension arrondie de \ln donne

$$\ln(\diamond \mathbf{x}) = \diamond[0.99694863\dots, 2.0001277\dots]$$

et

$$\ln^\diamond(\diamond \mathbf{x}) = [0.99, 2.01]$$

qui est bien un arrondi extérieur de $\text{range}(\ln, \mathbf{x}) = [1, 2]$.

2.5 Extensions optimales des fonctions continues $f : \mathbb{R}^n \longrightarrow \mathbb{R}$

Dans cette sous-section, f est une fonction continue à valeurs réelles, i.e. $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

2.5.1 Expression générale

L'expression générale (2.1) reste valable mais peut être précisée. Il est aisé de démontrer grâce au théorème des valeurs intermédiaires que l'extension optimale des fonctions à valeurs réelles coïncide avec l'image exacte de l'intervalle :

$$f^{\mathbb{IR}}(\mathbf{x}) = [\min_{x \in \mathbf{x}} f(x), \max_{x \in \mathbf{x}} f(x)] = \text{range}(f, \mathbf{x}).$$

La démarche suivie pour construire des extensions aux intervalles est de commencer par calculer les extensions optimales des fonctions élémentaires :

- fonctions à deux variables : $\Omega = \{ x + y, x - y, x \times y, x \div y \}$
- fonctions à une variable : $\Phi = \{ \exp x, \ln x, \sin x, \cos x, \tan, \arccos x, \arcsin x, \arctan x, \text{abs } x, x^n, \sqrt[n]{x} \text{ pour } n \in \mathbb{N}^* \}$

Les domaines de définition habituels sont associés à ces fonctions. Les extensions optimales des fonctions de Ω sont calculées formellement dans la prochaine sous-section et conduisent à l'arithmétique des intervalles.

2.5.2 Arithmétique des intervalles

Les opérations $+$, $-$, \times et \div réelles sont étendues aux intervalles en utilisant leur extension optimale.

Définition II.8. Soient $\circ \in \{+, -, \times, /\}$ et $\mathbf{x}, \mathbf{y} \in \mathbb{IR}$ (l'opération $/$ est définie pour \mathbf{y} vérifiant $0 \notin \mathbf{y}$). Alors,

$$\mathbf{x} \circ \mathbf{y} = \{x \circ y \in \mathbb{R} \mid x \in \mathbf{x} \wedge y \in \mathbf{y}\}.$$

Grâce aux propriétés de monotonie des fonctions considérées, leur extension peut-être exprimée formellement en fonction des arguments : quelque soient $\mathbf{x} = [\underline{\mathbf{x}}, \bar{\mathbf{x}}] \in \mathbb{IR}$ et $\mathbf{y} = [\underline{\mathbf{y}}, \bar{\mathbf{y}}] \in \mathbb{IR}$, avec la restriction $0 \notin \mathbf{y}$ pour l'opération $/$,

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= \square\{ \underline{\mathbf{x}} + \underline{\mathbf{y}}, \underline{\mathbf{x}} + \bar{\mathbf{y}}, \bar{\mathbf{x}} + \underline{\mathbf{y}}, \bar{\mathbf{x}} + \bar{\mathbf{y}} \} = [\underline{\mathbf{x}} + \underline{\mathbf{y}}, \bar{\mathbf{x}} + \bar{\mathbf{y}}] \\ \mathbf{x} - \mathbf{y} &= \square\{ \underline{\mathbf{x}} - \underline{\mathbf{y}}, \underline{\mathbf{x}} - \bar{\mathbf{y}}, \bar{\mathbf{x}} - \underline{\mathbf{y}}, \bar{\mathbf{x}} - \bar{\mathbf{y}} \} = [\underline{\mathbf{x}} - \bar{\mathbf{y}}, \bar{\mathbf{x}} - \underline{\mathbf{y}}] \\ \mathbf{x} \times \mathbf{y} &= \square\{ \underline{\mathbf{x}} \underline{\mathbf{y}}, \underline{\mathbf{x}} \bar{\mathbf{y}}, \bar{\mathbf{x}} \underline{\mathbf{y}}, \bar{\mathbf{x}} \bar{\mathbf{y}} \} \\ \mathbf{x} / \mathbf{y} &= \square\{ \underline{\mathbf{x}} / \underline{\mathbf{y}}, \underline{\mathbf{x}} / \bar{\mathbf{y}}, \bar{\mathbf{x}} / \underline{\mathbf{y}}, \bar{\mathbf{x}} / \bar{\mathbf{y}} \}. \end{aligned}$$

Notation. Le produit $\mathbf{x} \times \mathbf{y}$ pourra être noté $\mathbf{x} \mathbf{y}$.

Ces opérations entre intervalles sont croissantes par rapport à l'inclusion et localement Lipschitz continues. De nombreuses propriétés de ces opérations sont décrites dans Neumaier[8]. En particulier,

$$\begin{array}{lll} \text{associativité :} & \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z} & \mathbf{x} (\mathbf{y} \mathbf{z}) = (\mathbf{x} \mathbf{y}) \mathbf{z} \\ \text{commutativité :} & \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} & \mathbf{x} \mathbf{y} = \mathbf{y} \mathbf{x} \\ \text{élément neutre :} & \mathbf{x} + 0 = \mathbf{x} & \mathbf{x} \times 1 = \mathbf{x} \\ & \mathbf{x} - \mathbf{y} = \mathbf{x} + (-\mathbf{y}) & \mathbf{x} / \mathbf{y} = \mathbf{x} \times (1/\mathbf{y}) \\ & -(\mathbf{x} - \mathbf{y}) = \mathbf{y} - \mathbf{x} & \mathbf{x}(-\mathbf{y}) = (-\mathbf{x})\mathbf{y} = -\mathbf{x}\mathbf{y} \end{array}$$

Les opérations $+$ et \times ne sont pas distributives sur \mathbb{IR} et les simplifications entre $+$, $-$, \times et $/$ doivent être faites avec précaution car les règles habituellement utilisées avec les réels ne sont plus valables :

$$\begin{array}{l} \text{(sous distributivité)} \\ \mathbf{x}(\mathbf{y} + \mathbf{z}) \subseteq \mathbf{x}\mathbf{y} + \mathbf{x}\mathbf{z} \\ \mathbf{x} - \mathbf{y} \subseteq (\mathbf{x} + \mathbf{z}) - (\mathbf{y} + \mathbf{z}) \\ \mathbf{x} / \mathbf{y} \subseteq (\mathbf{x}\mathbf{z}) / (\mathbf{y}\mathbf{z}) \quad 0 \notin \mathbf{y} \wedge 0 \notin \mathbf{z} \end{array}$$

Les propriétés suivantes permettent d'obtenir des informations sur les milieux et rayons des résultats des opérations de l'arithmétique des intervalles :

$$\begin{aligned} \text{mid}(\mathbf{x} \pm \mathbf{y}) &= \text{mid } \mathbf{x} \pm \text{mid } \mathbf{y} \\ \text{rad}(\mathbf{x} \pm \mathbf{y}) &= \text{rad } \mathbf{x} + \text{rad } \mathbf{y} \\ \text{mid}(\mathbf{x} \mathbf{y}) &= (\text{mid } \mathbf{x}) (\text{mid } \mathbf{y}) \quad \text{si } \mathbf{x} \in \mathbb{R} \vee \mathbf{y} \in \mathbb{R} \\ \text{rad}(\mathbf{x} \mathbf{y}) &= |\mathbf{x}| (\text{rad } \mathbf{y}) \quad \text{si } \mathbf{x} \in \mathbb{R} \vee \text{mid } \mathbf{y} = 0 \end{aligned}$$

2.5.3 Extension des fonctions élémentaires

Les fonctions élémentaires à une variable sont étendues aux intervalles de la même manière en utilisant leur extension optimale. Dans la plupart des cas, des formules simples permettent de calculer l'extension en fonction des bornes des arguments. Par exemple, pour $\mathbf{x} = [\underline{\mathbf{x}}, \bar{\mathbf{x}}]$

$$\begin{aligned} \exp \mathbf{x} &= [\exp \underline{\mathbf{x}}, \exp \bar{\mathbf{x}}] \\ \ln \mathbf{x} &= [\ln \underline{\mathbf{x}}, \ln \bar{\mathbf{x}}] \\ \arccos \mathbf{x} &= [\arccos \bar{\mathbf{x}}, \arccos \underline{\mathbf{x}}] \quad \text{pour } \mathbf{x} \subseteq [-1, 1] \end{aligned}$$

Dans le cas des fonctions non monotones, par exemple x^2 ou $\cos x$, des algorithmes simples et efficaces permettent de calculer l'image d'un intervalle quelconque (Messine[28]). Comme précédemment, les extensions des fonctions élémentaires sont croissantes par rapport à l'inclusion et, à l'exception des fonctions $\sqrt[n]{x}$, localement Lipschitz continues.

Remarque. Les fonctions $\sqrt[n]{x}$ ne sont pas dérivables en 0 (plus précisément, la dérivée tend vers l'infini lorsque x tend vers 0). De fait, elles ne sont localement Lipschitz continues que dans $]0, +\infty[$.

2.6 L'extension naturelle

2.6.1 Définition

L'extension la plus simple construite à partir des extensions des fonctions élémentaires est l'extension naturelle. Elle consiste à évaluer l'expression d'une fonction pour des arguments qui sont des intervalles en utilisant l'arithmétique des intervalles et les extensions des fonctions élémentaires.

Théorème II.2. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et $f(x_1, \dots, x_n)$ une expression de f mettant en jeu des fonctions élémentaires de Ω et Φ . L'expression f est étendue aux intervalles en remplaçant les opérations réelles par leur extension optimale aux intervalles. Alors, la fonction d'intervalles \mathbf{g} définie par $\mathbf{g}(\mathbf{x}) := f(\mathbf{x})$ est une extension de f , appelée l'extension naturelle de f .

Démonstration (Moore[74]). La démonstration utilise simplement le théorème II.1 : tout d'abord, constatons que $\mathbf{g}(x) = f(x)$. Puis \mathbf{g} est croissante par rapport à l'inclusion car elle est composée de fonctions croissantes par rapport à l'inclusion. Le théorème II.1 montre finalement que \mathbf{g} est une extension de f . \square

Exemple II.11. Soit la fonction $f(x, y) = x(y - x)$ et les intervalles $\mathbf{x} = [2, 3]$ et $\mathbf{y} = [-1, 1]$. Alors,

$$\text{range}(f, \mathbf{x}, \mathbf{y}) \subseteq \mathbf{x}(\mathbf{y} - \mathbf{x}) = [-12, -2]$$

2.6.2 Qualité de l'extension naturelle

Une analyse plus fine que celle de la démonstration du théorème II.2 montre que si l'expression f ne contient qu'une unique occurrence de chaque variable, alors l'extension naturelle est optimale. Lorsque les fonctions $\sqrt[n]{x}$ sont restreintes à $]0, +\infty[$, l'extension naturelle est localement Lipschitz continue, car composée des extensions des fonctions élémentaires qui sont localement Lipschitz continues, et a donc un ordre de convergence linéaire. L'exemple suivant illustre la non convergence linéaire d'une extension naturelle dans le cas où \sqrt{x} n'est pas restreinte à $]0, +\infty[$.

Exemple II.12. Considérons la fonction $f(x) = x - \sqrt{\text{abs } x}$ et la suite d'intervalles $\mathbf{x}^{(k)} = [-1/k, 1/k] \subseteq [-1, 1]$ définie pour $k > 0$. Alors,

$$f^{\mathbb{IR}}(\mathbf{x}) = \text{range}(f, \mathbf{x}_k) = [-1/k - \sqrt{1/k}, 1/k - \sqrt{1/k}]$$

et l'extension naturelle donne

$$\begin{aligned} f(\mathbf{x}_k) &= \mathbf{x}_k - \sqrt{\text{abs } \mathbf{x}_k} \\ &= [-1/k, 1/k] - \sqrt{[0, 1/k]} \\ &= [-1/k - \sqrt{1/k}, 1/k] \end{aligned}$$

Ainsi,

$$\text{dist}(f^{\mathbb{I}\mathbb{R}}(\mathbf{x}_k), f(\mathbf{x}_k)) = \sqrt{1/k}$$

Finalemnt, $\text{wid } \mathbf{x} = 2/k$ et nous obtenons

$$\lim_{k \rightarrow +\infty} \frac{\text{dist}(f^{\mathbb{I}\mathbb{R}}(\mathbf{x}_k), f(\mathbf{x}_k))}{\text{wid } \mathbf{x}} = +\infty$$

Ce qui prouve que cette extension naturelle n'a pas un ordre de convergence linéaire.

2.7 L'extension de la valeur moyenne

2.7.1 Définition

Une seconde extension très utilisée est construite en appliquant le théorème de la valeur moyenne. Dans le cas d'une fonction continûment dérivable à une variable $f : \mathbb{R} \rightarrow \mathbb{R}$, l'extension de la valeur moyenne est obtenue de la manière suivante : le théorème de la valeur moyenne indique que quelque soient $\mathbf{x} \in \mathbb{I}\mathbb{R}$, $x \in \mathbf{x}$ et $\tilde{x} \in \mathbf{x}$ il existe $c \in \mathbf{x}$ tel que

$$f(x) = f(\tilde{x}) + (x - \tilde{x})f'(c)$$

Ainsi,

$$\text{range}(f, \mathbf{x}) \subseteq \text{range}(m, \mathbf{x}, \mathbf{x})$$

avec $m(x, c) = f(\tilde{x}) + (x - \tilde{x})f'(c)$. Finalemnt, en considérant une extension \mathbf{f}' de f' et l'extension naturelle de la fonction m évaluée en $(\mathbf{x}, \mathbf{x})^T$,

$$\text{range}(f, \mathbf{x}) \subseteq f(\tilde{x}) + (\mathbf{x} - \tilde{x})\mathbf{f}'(\mathbf{x})$$

L'extension $\mathbf{h}(\mathbf{x}) = f(\tilde{x}) + (\mathbf{x} - \tilde{x})\mathbf{f}'(\mathbf{x})$ est appelée l'extension de la valeur moyenne de f . Dans le cas général d'une fonction à n variables, l'expression de l'extension de la valeur moyenne est la suivante :

$$\mathbf{h}(\mathbf{x}) = f(c(\mathbf{x})) + \mathbf{f}'(\mathbf{x}) \times (\mathbf{x} - c(\mathbf{x}))$$

où \mathbf{f}' est une extension de $f'(x) \in \mathbb{R}^{1 \times n}$ et $c : \mathbb{I}\mathbb{R} \rightarrow \mathbb{R}$ vérifie $c(\mathbf{x}) \in \mathbf{x}$ (par exemple $c(\mathbf{x}) = \text{mid } \mathbf{x}$). L'extension de la valeur moyenne peut être généralisée de plusieurs manières : par exemple, un point $\tilde{x} \notin \mathbf{x}$ peut-être utilisé en remplaçant $\mathbf{f}'(\mathbf{x})$ par $\mathbf{f}'(\mathbf{x} \vee \tilde{x})$. L'extension \mathbf{f}' de la dérivée de f peut-être avantageusement remplacée par les "slopes" (voir par exemple Krawczyk[73], Neumaier[8], Zuhe[85] ou Kearfott[79]).

2.7.2 Qualité de l'extension de la valeur moyenne

L'extension de la valeur moyenne a plusieurs avantages sur l'extension naturelle :

- l'extension de la valeur moyenne a un ordre de convergence quadratique (sous l'hypothèse que l'extension \mathbf{f}' du gradient utilisée soit localement Lipschitz continue). Cette propriété fut d'abord conjecturée dans Moore[74], puis prouvée par de nombreux auteurs de différentes manières (voir par exemple Hansen[12], Alefeld[31] et Stahl[100], la thèse de Stahl présentant une importante bibliographie sur le sujet).
- l'extension de la valeur moyenne permet de linéariser une fonction non linéaire de manière compatible avec l'approximation extérieure de l'image d'un intervalle. Cela permet d'appliquer aux fonctions non-linéaires les algorithmes dédiés aux fonctions linéaires.

L'ordre de convergence quadratique donne un important avantage à l'extension de la valeur moyenne sur l'extension naturelle lorsque ces extensions sont utilisées sur de petits intervalles.

2.8 Extensions des fonctions continues à valeurs vectorielles

Les extensions des fonctions continues (ou continûment dérivables lorsque l'extension de la valeur moyenne est utilisée) $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ sont aisément construites en utilisant les extensions des composantes $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ de f : si les fonctions d'intervalles $\mathbf{g}_i : \mathbb{IR}^n \rightarrow \mathbb{IR}$ sont des extensions des fonctions f_i , alors la fonction $\mathbf{g} = (\mathbf{g}_i) : \mathbb{IR}^n \rightarrow \mathbb{IR}^m$ est une extension de f . Il est alors aisé de constater que l'ordre de convergence de \mathbf{g} est égal au minimum des ordres de convergences des extensions \mathbf{g}_i .

2.9 Autres extensions

Peut être encore citée l'extension de Taylor qui est basée sur un développement de Taylor de la fonction f . Dans le cas d'une fonction à une variable, le développement suivant est utilisé : il existe $\xi \in x \vee \tilde{x}$ tel que

$$f(x) = f(\tilde{x}) + \sum_{k=1}^n \left(\frac{(x - \tilde{x})^k}{k!} f^{(k)}(\tilde{x}) \right) + \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - \tilde{x})^{(n+1)}$$

Il ne reste plus qu'à majorer le reste de la série pour obtenir un intervalle qui contient $f(x)$ (voir Neumaier[10] ou Hansen[18]). Plus l'ordre de l'extension de Taylor est élevé, plus elle sera précise. Néanmoins, la majoration du reste demande le calcul de la $n^{\text{ième}}$ dérivée, ce qui est compliqué lorsque des fonctions à plusieurs variables sont utilisées. L'extension de Bernstein est basée sur l'écriture d'un polynôme dans la base des polynômes de Bernstein (voir par exemple Hong[44] ou Garloff[34]). Elle donne de bonnes approximations de l'image d'une fonction, dans le cas restreint des polynômes.

2.10 Opérations sur les vecteurs et matrices d'intervalles

Grâce à l'arithmétique des intervalles, les opérations habituelles entre vecteurs et matrices réels peuvent être étendues aux intervalles. L'extension naturelle permet de donner les interprétations suivantes de ces opérations :

- $\mathbf{x} + \mathbf{y} = \mathbf{z}$ avec $\mathbf{z} = \{x + y \in \mathbb{R}^n \mid x \in \mathbf{x}, y \in \mathbf{y}\}$;
- $\mathbf{A} + \mathbf{B} = \mathbf{C}$ avec $\mathbf{C} = \{A + B \in \mathbb{R}^{n \times m} \mid A \in \mathbf{A}, B \in \mathbf{B}\}$;
- $\mathbf{Ax} = \mathbf{z}$ avec $\mathbf{z} = \square \{Ax \in \mathbb{R}^n \mid A \in \mathbf{A}, x \in \mathbf{x}\}$;
- $\mathbf{A} \times \mathbf{B} = \mathbf{C}$ avec $\mathbf{C} = \square \{AB \in \mathbb{R}^{n \times m} \mid A \in \mathbf{A}, B \in \mathbf{B}\}$.

Les restrictions sur les dimensions des objets mis en jeu sont les mêmes que dans le cas réel.

Chapitre 3

Intervalles généralisés

Les intervalles généralisés ont été introduits dans Ortolfo[42] et Kaucher[20] (voir aussi Kaucher[21] et Shary[96]) et permettent d'améliorer les deux défauts suivants des intervalles classiques :

- l'intersection d'intervalles disjoints n'est pas définie car l'ensemble vide n'est pas considéré comme un intervalle.
- l'addition d'intervalles n'est pas un groupe : en particulier, il existe des équations du type $\mathbf{a} + \mathbf{x} = \mathbf{b}$ (où deux intervalles sont égaux si leurs bornes sont égales) qui n'ont pas de solution \mathbf{x} . En effet, lorsqu'on additionne deux intervalles, leurs rayons s'ajoutent. Ainsi, si $\text{rad } \mathbf{b} < \text{rad } \mathbf{a}$ alors aucun intervalle \mathbf{x} ne pourra résoudre l'équation précédente. Par exemple, si l'équation $[-1, 1] + \mathbf{x} = [7, 12]$ a pour solution $\mathbf{x} = [8, 11]$, l'équation $[-1, 1] + \mathbf{x} = [9, 10]$ n'a aucune solution dans \mathbb{IR} . Le même problème sera partiellement résolu pour la multiplication d'intervalles.

Il ne semble pas qu'il y ait des motivations non mathématiques pour l'introduction des intervalles généralisés dans les années 70. Il faudra attendre des travaux ultérieures (Gardenyev[16], Chapitre 4, et Shary[93], Chapitre 5) pour voir apparaître des applications utilisant les intervalles généralisés.

3.1 Les intervalles généralisés

3.1.1 Définition des intervalles généralisés

Les intervalles généralisés sont des intervalles dont les bornes ne sont pas contraintes à être ordonnées, par exemple $[-1, 1]$ ou $[1, -1]$ sont des intervalles généralisés. L'ensemble des intervalles généralisés est noté \mathbb{KIR} et est composé de trois sous-ensembles caractéristiques :

- L'ensemble des intervalles propres, dont les bornes sont ordonnées dans l'ordre croissant. Les intervalles propres sont identifiés aux intervalles classiques, et ainsi aux ensembles de réels correspondants. L'ensemble des intervalles propres est ainsi noté $\mathbb{IR} = \{[a, b] | a \leq b\}$.
- L'ensemble des intervalles impropres, dont les bornes sont ordonnées dans l'ordre décroissant. L'ensemble des intervalles impropres est noté $\overline{\mathbb{IR}} = \{[a, b] | a \geq b\}$.
- L'ensemble des intervalles dégénérés $[a, a]$ avec $a \in \mathbb{R}$. Les intervalles dégénérés sont à la fois propres et impropres et sont identifiés aux nombres réels.

Ainsi, étant donné un ensemble de réels $\{x \in \mathbb{R} | a \leq x \leq b\}$, deux intervalles généralisés peuvent être construits : $[a, b]$ et $[b, a]$. Il sera utile de passer de l'un à l'autre en utilisant

les opérations suivantes :

- l'opération dual : $\text{dual}[a, b] = [b, a]$;
- la projection propre : $\text{pro}[a, b] = [\min\{a, b\}, \max\{a, b\}] \in \mathbb{IR}$;
- la projection impropre : $\text{imp}[a, b] = [\max\{a, b\}, \min\{a, b\}] \in \overline{\mathbb{IR}}$.

Les définitions classiques de borne inférieure, borne supérieure, milieu, rayon et largeur sont étendues aux intervalles généralisés en conservant les mêmes formules : étant donné un intervalle généralisé $\mathbf{x} = [\underline{\mathbf{x}}, \overline{\mathbf{x}}]$,

$$\inf \mathbf{x} = \underline{\mathbf{x}} \quad ; \quad \sup \mathbf{x} = \overline{\mathbf{x}}$$

$$\text{mid } \mathbf{x} = \frac{1}{2}(\overline{\mathbf{x}} + \underline{\mathbf{x}}) \quad ; \quad \text{rad } \mathbf{x} = \frac{1}{2}(\overline{\mathbf{x}} - \underline{\mathbf{x}}) \quad ; \quad |\mathbf{x}| = \max\{|\underline{\mathbf{x}}|, |\overline{\mathbf{x}}|\}$$

La largeur est définie par $\text{wid } \mathbf{x} = 2 \text{rad } \mathbf{x}$. Le rayon et la largeur sont positifs pour les intervalles propres et négatifs pour les intervalles impropres (et nuls pour les intervalles dégénérés). Finalement, il sera utile de partitionner l'ensemble des intervalles généralisés de la manière suivante :

$$\mathbb{KR} = \mathcal{P} \cup (-\mathcal{P}) \cup \mathcal{Z} \cup (\text{dual } \mathcal{Z})$$

avec

$$\begin{aligned} \mathcal{P} &= \{\mathbf{x} \in \mathbb{KR} \mid 0 \leq \underline{\mathbf{x}} \wedge 0 \leq \overline{\mathbf{x}}\} \\ -\mathcal{P} &= \{\mathbf{x} \in \mathbb{KR} \mid 0 \geq \underline{\mathbf{x}} \wedge 0 \geq \overline{\mathbf{x}}\} \\ \mathcal{Z} &= \{\mathbf{x} \in \mathbb{KR} \mid \underline{\mathbf{x}} \leq 0 \leq \overline{\mathbf{x}}\} \\ \text{dual } \mathcal{Z} &= \{\mathbf{x} \in \mathbb{KR} \mid \underline{\mathbf{x}} \geq 0 \geq \overline{\mathbf{x}}\} \end{aligned}$$

Exemple II.13. Nous avons par exemple $[1, 2] \in \mathcal{P}$, $[2, 1] \in \mathcal{P}$, $[-2, -1] \in -\mathcal{P}$, $[-1, -2] \in -\mathcal{P}$, $[-1, 1] \in \mathcal{Z}$ et $[1, -1] \in \text{dual } \mathcal{Z}$.

3.1.2 Inclusion, intersection et réunion

Les intervalles généralisés sont partiellement ordonnés par une inclusion qui prolonge l'inclusion entre intervalles classiques. Étant donnés deux intervalles généralisés $\mathbf{x} = [\underline{\mathbf{x}}, \overline{\mathbf{x}}]$ et $\mathbf{y} = [\underline{\mathbf{y}}, \overline{\mathbf{y}}]$, l'inclusion est définie de la manière suivante :

$$\mathbf{x} \subseteq \mathbf{y} \iff \underline{\mathbf{y}} \leq \underline{\mathbf{x}} \wedge \overline{\mathbf{x}} \leq \overline{\mathbf{y}}$$

Exemple II.14. Les inclusions suivantes peuvent être vérifiées grâce à la définition précédente : $[-1, 1] \subseteq [-2, 2]$, $[2, -2] \subseteq [1, 3]$ et $[2, -2] \subseteq [1, -1]$. L'ensemble \mathcal{Z} peut être défini par $\mathcal{Z} = \{\mathbf{x} \in \mathbb{KR} \mid [0, 0] \subseteq \mathbf{x}\}$ et $(\text{dual } \mathcal{Z})$ par $(\text{dual } \mathcal{Z}) = \{\mathbf{x} \in \mathbb{KR} \mid \mathbf{x} \subseteq [0, 0]\}$

Cette inclusion est liée à l'opération dual de la manière suivante :

$$\mathbf{x} \subseteq \mathbf{y} \iff (\text{dual } \mathbf{x}) \supseteq (\text{dual } \mathbf{y})$$

L'intersection et la réunion d'intervalles sont alors définies de la même manière que pour les intervalles classiques : ces opérations correspondent respectivement à la borne inférieure et à la borne supérieure de l'ordre partiel de l'inclusion. Elles sont définies pour un ensemble borné $E \subseteq \mathbb{KR}$ par

$$\begin{aligned} (\forall \mathbf{z} \in \mathbb{KR}) (\mathbf{z} \supseteq (\bigvee E)) &\iff \forall \mathbf{x} \in E, \mathbf{z} \supseteq \mathbf{x} \\ \text{et } (\forall \mathbf{z} \in \mathbb{KR}) (\mathbf{z} \subseteq (\bigwedge E)) &\iff \forall \mathbf{x} \in E, \mathbf{z} \subseteq \mathbf{x}. \end{aligned}$$

Quand $E = \{\mathbf{x}, \mathbf{y}\}$ contient seulement deux éléments, les notations habituelles $\mathbf{x} \vee \mathbf{y}$ et $\mathbf{x} \wedge \mathbf{y}$ sont utilisées. Les caractérisations suivantes sont des conséquences directes de ces définitions. Elle sont formellement identique aux caractérisations des intersection et réunion des intervalles classiques.

$$\begin{aligned}\bigwedge E &= [\sup\{\inf \mathbf{x} | \mathbf{x} \in E\} , \inf\{\sup \mathbf{x} | \mathbf{x} \in E\}] \\ \bigvee E &= [\inf\{\inf \mathbf{x} | \mathbf{x} \in E\} , \sup\{\sup \mathbf{x} | \mathbf{x} \in E\}].\end{aligned}$$

Dans le cas de où E contient deux intervalles $\mathbf{x} = [\underline{\mathbf{x}}, \bar{\mathbf{x}}]$ et $\mathbf{y} = [\underline{\mathbf{y}}, \bar{\mathbf{y}}]$, nous obtenons les expressions suivantes :

$$\begin{aligned}\mathbf{x} \wedge \mathbf{y} &= [\sup\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\} , \inf\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}] \\ \mathbf{x} \vee \mathbf{y} &= [\inf\{\underline{\mathbf{x}}, \underline{\mathbf{y}}\} , \sup\{\bar{\mathbf{x}}, \bar{\mathbf{y}}\}].\end{aligned}$$

Dans le contexte des intervalles généralisés, il est important d'utiliser deux symboles différents pour l'intersection d'intervalles et pour l'intersection ensembliste. Par exemple, $[0, 1] \cap [2, 3] = \emptyset$ alors que $[0, 1] \wedge [2, 3] = [2, 1]$.

Exemple II.15. *L'union d'intervalles entre deux intervalles dégénérés est identique au cas des intervalles classiques : $[1, 1] \vee [2, 2] = [1, 2]$, qui aussi écrit $1 \vee 2 = [1, 2]$. Par contre, l'intersection d'intervalles entre deux intervalles dégénérés est maintenant définie et donne l'intervalle impropre suivant : $[1, 1] \wedge [2, 2] = [2, 1]$, qui est aussi écrit $1 \wedge 2 = [2, 1]$.*

La relation entre l'inclusion et l'opération dual est transférée sur les opérations d'intersection et de réunion et nous obtenons les relations suivantes :

$$\begin{aligned}\bigwedge E &= \text{dual}(\bigvee \{\text{dual } \mathbf{x} \mid \mathbf{x} \in E\}) \\ \bigvee E &= \text{dual}(\bigwedge \{\text{dual } \mathbf{x} \mid \mathbf{x} \in E\}).\end{aligned}$$

Soit, quand E contient deux intervalles \mathbf{x} et \mathbf{y}

$$\begin{aligned}\mathbf{x} \wedge \mathbf{y} &= \text{dual}((\text{dual } \mathbf{x}) \vee (\text{dual } \mathbf{y})) \\ \mathbf{x} \vee \mathbf{y} &= \text{dual}((\text{dual } \mathbf{x}) \wedge (\text{dual } \mathbf{y})).\end{aligned}$$

Exemple II.16. *En reprenant l'exemple précédent, et en remarquant que l'opération dual ne change pas un intervalle dégénéré,*

$$1 \wedge 2 = \text{dual}((\text{dual } 1) \vee (\text{dual } 2)) = \text{dual}(1 \vee 2) = \text{dual}[1, 2] = [2, 1].$$

3.1.3 Distance entre intervalles généralisés

La distance entre deux intervalles généralisés $\mathbf{x} \in \mathbb{KR}$ et $\mathbf{y} \in \mathbb{KR}$ est définie formellement de la même manière que pour les intervalles classiques :

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \max\{|\underline{\mathbf{x}} - \underline{\mathbf{y}}|, |\bar{\mathbf{x}} - \bar{\mathbf{y}}|\}$$

Il est montré dans Kaucher[20][21] que \mathbb{KR} devient alors un espace métrique complet.

3.2 Vecteurs et matrices d'intervalles généralisés

Quand des vecteurs d'intervalles généralisés ou des matrices d'intervalles généralisés sont considérés, les mêmes règles que dans le cadre des intervalles classiques sont utilisées. L'ensemble des vecteurs d'intervalles généralisés, qui seront appelés simplement intervalles généralisés, est noté $\mathbb{K}\mathbb{R}^n$ et l'ensemble des matrices d'intervalles généralisés $\mathbb{K}\mathbb{R}^{n \times m}$. Étant donné un ensemble d'indices I avec $\text{card } I = n$ et un intervalle généralisé $\mathbf{x}_I \in \mathbb{K}\mathbb{R}^n$, les fonctions suivantes permettent de construire les sous-ensembles de I correspondant aux composantes propres et impropres de \mathbf{x} :

- $P(\mathbf{x}_I) = \{i \in I \mid \mathbf{x}_i \in \mathbb{I}\mathbb{R}\}$
- $I(\mathbf{x}_I) = \{i \in I \mid \mathbf{x}_i \notin \mathbb{I}\mathbb{R}\}$

Remarque. *Le fait de compter les indices des composantes dégénérées comme propres est une convention.*

Étant donné $E \subseteq \mathbb{R}^n$, il sera utile de noter l'ensemble d'intervalles généralisés

$$\{\mathbf{x} \in \mathbb{K}\mathbb{R}^n \mid \text{pro } \mathbf{x} \subseteq E\}$$

par $\mathbb{K}E$. La distance de $\mathbb{K}\mathbb{R}$ est étendue à $\mathbb{K}\mathbb{R}^n$ de la manière suivante : étant donnés $\mathbf{x}, \mathbf{y} \in \mathbb{K}\mathbb{R}^n$,

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \max_{k \in [1..n]} \text{dist}(\mathbf{x}_k, \mathbf{y}_k)$$

Cette distance est formellement identique à celle des vecteurs d'intervalles classiques. En particulier, étant donnés une suite d'intervalles $\mathbf{x}^{(k)} \in \mathbb{K}\mathbb{R}^n$ et un intervalle $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, les trois conditions suivantes sont équivalentes :

1. $\mathbf{x}^{(k)}$ converge vers \mathbf{x} .
2. pour chaque $i \in [1..n]$, les suites $\mathbf{x}_i^{(k)}$ convergent vers \mathbf{x}_i .
3. les deux suites $\inf \mathbf{x}^{(k)}$ et $\sup \mathbf{x}^{(k)}$ convergent respectivement vers $\inf \mathbf{x}$ et $\sup \mathbf{x}$.

3.3 Fonctions d'intervalles généralisés

Comme dans le contexte des extensions aux intervalles classiques, la distance entre intervalles généralisés permet d'introduire la continuité locale de Lipschitz :

Définition II.9. *Une fonction $\mathbf{g} : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{R}^m$ est localement Lipschitz continue si et seulement si pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n$, il existe un réel $\gamma > 0$ tel que pour tout $\mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}$ et $\mathbf{y} \in \mathbb{K}\mathbf{x}^{\text{ref}}$,*

$$\text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y})) \leq \gamma \text{dist}(\mathbf{x}, \mathbf{y})$$

Il est facile de vérifier que dual, pro, imp sont localement Lipschitz continus. La croissance par rapport à l'inclusion d'une fonction d'intervalles généralisés est définie comme dans le contexte des fonctions d'intervalles classiques.

3.4 Extensions des fonctions continues aux intervalles généralisés

Les extensions des fonctions continues aux intervalles généralisés ne sont pas construites de la même manière que les extensions aux intervalles classiques : elle sont définies dans

Kaucher[21] en généralisant l'expression de l'extension optimale aux intervalles classiques

$$f^{\mathbb{IR}}(\mathbf{x}) = \bigvee_{x \in \mathbf{x}} f(x) = [\inf_{x \in \mathbf{x}} f(x), \sup_{x \in \mathbf{x}} f(x)],$$

cf. expression (2.1) page 30. Les extensions sont définies dans Kaucher[21] dans le cas restreint des fonctions à valeurs réelles, i.e. $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Ainsi, d'une part une fonction réelle $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ne possède qu'une unique extension aux intervalles généralisés qui n'est pas calculable en générale. D'autre part, les extensions aux intervalles généralisés ne sont pas associées à une interprétation par rapport aux fonctions réelles qu'elles étendent.

3.4.1 Cas général

Étant donné une fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$, son extension aux intervalles généralisés est la fonction $f^K : \mathbb{KR}^n \rightarrow \mathbb{KR}$ définie par

$$f^K(\mathbf{x}) = \bigwedge_{x_1 \in \text{pro } \mathbf{x}_1}^{\mathbf{x}_1} \cdots \bigwedge_{x_n \in \text{pro } \mathbf{x}_n}^{\mathbf{x}_n} f(x),$$

où

$$\bigwedge^{\mathbf{x}} = \begin{cases} \bigvee & \text{si } \mathbf{x} \in \mathbb{IR} \\ \bigwedge & \text{sinon} \end{cases}.$$

Exemple II.17. *Considérons une fonction à trois variables $f(x_1, x_2, x_3)$ et trois intervalles généralisés $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{KR}$. L'expression de $f^K(\mathbf{x})$ pour différentes qualités propre/impropre des arguments est donnée ci-dessous :*

– si $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{IR}$ alors

$$f^K(\mathbf{x}) = \bigvee_{x \in \mathbf{x}} f(x) = [\min_{x \in \mathbf{x}} f(x), \max_{x \in \mathbf{x}} f(x)]$$

– si $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \overline{\mathbb{IR}}$ alors

$$f^K(\mathbf{x}) = \bigwedge_{x \in \text{pro } \mathbf{x}} f(x) = [\max_{x \in \text{pro } \mathbf{x}} f(x), \min_{x \in \text{pro } \mathbf{x}} f(x)]$$

– si $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{IR}$ et $\mathbf{x}_3 \in \overline{\mathbb{IR}}$ alors

$$\begin{aligned} f^K(\mathbf{x}) &= \bigvee_{x_1 \in \mathbf{x}_1} \bigvee_{x_2 \in \mathbf{x}_2} \bigwedge_{x_3 \in \text{pro } \mathbf{x}_3} f(x) \\ &= [\min_{x_1 \in \mathbf{x}_1} \min_{x_2 \in \mathbf{x}_2} \max_{x_3 \in \text{pro } \mathbf{x}_3} f(x), \max_{x_1 \in \mathbf{x}_1} \max_{x_2 \in \mathbf{x}_2} \min_{x_3 \in \text{pro } \mathbf{x}_3} f(x)] \end{aligned}$$

– si $\mathbf{x}_1, \mathbf{x}_3 \in \mathbb{IR}$ et $\mathbf{x}_2 \in \overline{\mathbb{IR}}$ alors

$$\begin{aligned} f^K(\mathbf{x}) &= \bigvee_{x_1 \in \mathbf{x}_1} \bigwedge_{x_2 \in \text{pro } \mathbf{x}_2} \bigvee_{x_3 \in \mathbf{x}_3} f(x) \\ &= [\min_{x_1 \in \mathbf{x}_1} \max_{x_2 \in \text{pro } \mathbf{x}_2} \min_{x_3 \in \mathbf{x}_3} f(x), \max_{x_1 \in \mathbf{x}_1} \min_{x_2 \in \text{pro } \mathbf{x}_2} \max_{x_3 \in \mathbf{x}_3} f(x)] \end{aligned}$$

– si $\mathbf{x}_2, \mathbf{x}_3 \in \mathbb{IR}$ et $\mathbf{x}_1 \in \overline{\mathbb{IR}}$ alors

$$\begin{aligned} f^K(\mathbf{x}) &= \bigwedge_{x_1 \in \text{pro } \mathbf{x}_1} \bigvee_{x_2 \in \mathbf{x}_2} \bigvee_{x_3 \in \mathbf{x}_3} f(x) \\ &= [\max_{x_1 \in \text{pro } \mathbf{x}_1} \min_{x_2 \in \mathbf{x}_2} \min_{x_3 \in \mathbf{x}_3} f(x), \min_{x_1 \in \text{pro } \mathbf{x}_1} \max_{x_2 \in \mathbf{x}_2} \max_{x_3 \in \mathbf{x}_3} f(x)] \end{aligned}$$

Ainsi, dans le cas où tous les arguments sont propres, seules des réunions sont utilisées et l'on retrouve l'extension optimale aux intervalles classiques : pour $\mathbf{x} \in \mathbb{IR}^n$,

$$f^K(\mathbf{x}_1, \dots, \mathbf{x}_n) = \bigvee_{x_1 \in \mathbf{x}_1} \cdots \bigvee_{x_n \in \mathbf{x}_n} f(x) = \text{range}(f, \mathbf{x}_1, \dots, \mathbf{x}_n) = f^{\mathbb{IR}}(\mathbf{x}).$$

De même, si tous les arguments sont impropres, seules des intersections sont calculées : pour $\mathbf{x} \in \overline{\mathbb{IR}}^n$,

$$f^K(\mathbf{x}_1, \dots, \mathbf{x}_n) = \bigwedge_{x_1 \in \text{pro } \mathbf{x}_1} \cdots \bigwedge_{x_n \in \text{pro } \mathbf{x}_n} f(x) = \text{dual}(\text{range}(f, \mathbf{x}_1, \dots, \mathbf{x}_n)).$$

Lorsque des arguments propres et impropres sont mis en jeux, les calculs deviennent plus compliqués. L'exemple suivant illustre cette situation.

Exemple II.18. *Considérons la fonction $f(x, y) = x^2 + y^2$ et les intervalles propres $\mathbf{x} = [-3, 3]$ et $\mathbf{y} = [4, 6]$. Alors,*

$$\begin{aligned} f^K(\mathbf{x}, \text{dual } \mathbf{y}) &= \bigvee_{x \in \mathbf{x}} \bigwedge_{y \in \mathbf{y}} f(x, y) \\ &= [\min_{x \in \mathbf{x}} \max_{y \in \mathbf{y}} f(x, y), \max_{x \in \mathbf{x}} \min_{y \in \mathbf{y}} f(x, y)] \\ &= [\min_{x \in \mathbf{x}} x^2 + \bar{\mathbf{y}}^2, \max_{x \in \mathbf{x}} x^2 + \underline{\mathbf{y}}^2] \\ &= [0 + \bar{\mathbf{y}}^2, \bar{\mathbf{x}}^2 + \underline{\mathbf{y}}^2] \\ &= [36, 9 + 16] = [36, 25] \end{aligned}$$

Il est démontré dans Kaucher[21] que les extensions de fonctions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ localement Lipschitz continues sont localement Lipschitz continues.

3.4.2 Arithmétique de Kaucher

Dans le cas des opérations de l'arithmétique réelle, i.e. $f(x, y) = x \circ y$ avec $\circ \in \{+, -, \times, \div\}$, les extensions conduisent à l'arithmétique de Kaucher : $\mathbf{x} \circ \mathbf{y} = f^K(\mathbf{x}, \mathbf{y})$. Grâce aux fortes propriétés de monotonie des fonctions réelles mises en jeux, des expressions formelles peuvent être dérivées de la définition originale des extensions :

1. $\mathbf{x} + \mathbf{y} = [\underline{\mathbf{x}} + \underline{\mathbf{y}}, \bar{\mathbf{x}} + \bar{\mathbf{y}}]$
2. $\mathbf{x} - \mathbf{y} = [\underline{\mathbf{x}} - \bar{\mathbf{y}}, \bar{\mathbf{x}} - \underline{\mathbf{y}}] = \mathbf{x} + (-\mathbf{y})$ où $-\mathbf{y} = [-\bar{\mathbf{y}}, -\underline{\mathbf{y}}]$.
3. La multiplication de Kaucher $\mathbf{x} \times \mathbf{y}$ est décrite dans le tableau 3.1 (page 43).
4. La division de Kaucher \mathbf{x}/\mathbf{y} est définie pour des intervalles généralisés \mathbf{x} et \mathbf{y} qui vérifient $0 \notin (\text{pro } \mathbf{y})$. Elle peut être caractérisée par $\mathbf{x}/\mathbf{y} = \mathbf{x} \times (1/\mathbf{y})$ où $1/\mathbf{y} = [1/\bar{\mathbf{y}}, 1/\underline{\mathbf{y}}]$.

Comme dans le cas général des extensions, l'arithmétique de Kaucher coïncide avec l'arithmétique des intervalles classiques lorsqu'elle est restreinte aux intervalles propres. Cependant, l'arithmétique de Kaucher a de meilleures propriétés algébriques que l'arithmétique des intervalles classiques : l'addition de Kaucher est un groupe d'élément neutre $[0, 0]$. L'opposé d'un intervalle \mathbf{x} est $-\text{dual } \mathbf{x}$, i.e.

$$\mathbf{x} + (-\text{dual } \mathbf{x}) = [0, 0]$$

TAB. 3.1 – La multiplication de Kaucher

$\mathbf{x} \times \mathbf{y}$	$\mathbf{y} \in \mathcal{P}$	$\mathbf{y} \in \mathcal{Z}$	$\mathbf{y} \in -\mathcal{P}$	$\mathbf{y} \in \text{dual } \mathcal{Z}$
$\mathbf{x} \in \mathcal{P}$	$[\underline{\mathbf{x}} \underline{\mathbf{y}}, \bar{\mathbf{x}} \bar{\mathbf{y}}]$	$[\bar{\mathbf{x}} \underline{\mathbf{y}}, \bar{\mathbf{x}} \bar{\mathbf{y}}]$	$[\bar{\mathbf{x}} \underline{\mathbf{y}}, \underline{\mathbf{x}} \bar{\mathbf{y}}]$	$[\underline{\mathbf{x}} \underline{\mathbf{y}}, \underline{\mathbf{x}} \bar{\mathbf{y}}]$
$\mathbf{x} \in \mathcal{Z}$	$[\underline{\mathbf{x}} \bar{\mathbf{y}}, \bar{\mathbf{x}} \bar{\mathbf{y}}]$	$[\min\{\underline{\mathbf{x}} \bar{\mathbf{y}}, \bar{\mathbf{x}} \underline{\mathbf{y}}\}, \max\{\underline{\mathbf{x}} \underline{\mathbf{y}}, \bar{\mathbf{x}} \bar{\mathbf{y}}\}]$	$[\bar{\mathbf{x}} \underline{\mathbf{y}}, \underline{\mathbf{x}} \bar{\mathbf{y}}]$	0
$\mathbf{x} \in -\mathcal{P}$	$[\underline{\mathbf{x}} \bar{\mathbf{y}}, \bar{\mathbf{x}} \underline{\mathbf{y}}]$	$[\underline{\mathbf{x}} \bar{\mathbf{y}}, \underline{\mathbf{x}} \underline{\mathbf{y}}]$	$[\bar{\mathbf{x}} \bar{\mathbf{y}}, \underline{\mathbf{x}} \bar{\mathbf{y}}]$	$[\bar{\mathbf{x}} \bar{\mathbf{y}}, \bar{\mathbf{x}} \underline{\mathbf{y}}]$
$\mathbf{x} \in \text{dual } \mathcal{Z}$	$[\underline{\mathbf{x}} \underline{\mathbf{y}}, \bar{\mathbf{x}} \underline{\mathbf{y}}]$	0	$[\bar{\mathbf{x}} \bar{\mathbf{y}}, \underline{\mathbf{x}} \bar{\mathbf{y}}]$	$[\max\{\underline{\mathbf{x}} \underline{\mathbf{y}}, \bar{\mathbf{x}} \bar{\mathbf{y}}\}, \min\{\underline{\mathbf{x}} \bar{\mathbf{y}}, \bar{\mathbf{x}} \underline{\mathbf{y}}\}]$

Exemple II.19. *Considérons deux intervalles généralisés $\mathbf{a}, \mathbf{b} \in \mathbb{KR}$. Alors l'équation*

$$\mathbf{a} + \mathbf{x} = \mathbf{b}$$

a une unique solution $\mathbf{x} = \mathbf{b} - \text{dual } \mathbf{a}$.

La multiplication de Kaucher, lorsqu'elle est restreinte aux intervalles dont les deux bornes sont strictement positives, est un groupe d'élément neutre $[1, 1]$. L'inverse d'un intervalle $\mathbf{x} > 0$ est $1/(\text{dual } \mathbf{x})$, i.e.

$$\mathbf{x} \times (1/\text{dual } \mathbf{x}) = [1, 1]$$

Une propriété utile de l'arithmétique de Kaucher est la croissance de ses opérations par rapport à l'inclusion : quelque soient $\circ \in \{+, \times, -, \div\}$ et $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathbb{KR}$,

$$\mathbf{x} \subseteq \mathbf{x}' \text{ et } \mathbf{y} \subseteq \mathbf{y}' \implies (\mathbf{x} \circ \mathbf{y}) \subseteq (\mathbf{x}' \circ \mathbf{y}')$$

De plus, les relations suivantes lient les opérations de l'arithmétique de Kaucher à la distance entre intervalles généralisés : quelque soient $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathbb{KR}$,

$$\begin{aligned} \text{dist}(\mathbf{x}\mathbf{y}, \mathbf{x}\mathbf{y}') &\leq |\mathbf{x}| \text{dist}(\mathbf{y}, \mathbf{y}') \\ \text{dist}(\mathbf{x} + \mathbf{y}, \mathbf{x}' + \mathbf{y}') &\leq \text{dist}(\mathbf{x}, \mathbf{x}') + \text{dist}(\mathbf{y}, \mathbf{y}') \end{aligned}$$

Les opérations de l'arithmétique de Kaucher vérifient :

$$(\text{dual } \mathbf{x}) \circ (\text{dual } \mathbf{y}) = \text{dual}(\mathbf{x} \circ \mathbf{y})$$

Cette propriété sera appelée commutation avec l'opération dual : en considérant $f(x) = x_1 \circ x_2$ et $\mathbf{x} \in \mathbb{KR}^2$, la propriété précédente s'écrit $f^K(\text{dual } \mathbf{x}) = \text{dual } f^K(\mathbf{x})$. Finalement, les trois lois suivantes de "faible distributivité" seront utiles :

– distributivité conditionnelle (SIGLA/X[38]) :

$$\mathbf{x} \times \mathbf{y} + (\text{imp } \mathbf{x}) \times \mathbf{z} \subseteq \mathbf{x} \times (\mathbf{y} + \mathbf{z}) \subseteq \mathbf{x} \times \mathbf{y} + (\text{pro } \mathbf{x}) \times \mathbf{z}.$$

– sous-distributivité : si $\mathbf{x} \in \mathbb{IR}$ alors $\mathbf{x} \times (\mathbf{y} + \mathbf{z}) \subseteq \mathbf{x} \times \mathbf{y} + \mathbf{x} \times \mathbf{z}$;

– super-distributivité : si $\mathbf{x} \in \overline{\mathbb{IR}}$ alors $\mathbf{x} \times (\mathbf{y} + \mathbf{z}) \supseteq \mathbf{x} \times \mathbf{y} + \mathbf{x} \times \mathbf{z}$;

– distributivité : si $x \in \mathbb{R}$ alors $x \times (\mathbf{y} + \mathbf{z}) = x \times \mathbf{y} + x \times \mathbf{z}$.

D'autres lois de distributivité plus détaillées peuvent être trouvées dans Popova[23].

3.4.3 Extension des fonctions élémentaires aux intervalles généralisés

Les extensions des fonctions élémentaires n'ont pas été beaucoup étudiées. Malgré tout, voici quelques propriétés intéressantes qui, si elles semblent nouvelles, sont néanmoins faciles à obtenir. Les expressions des extensions des fonctions à une variable sont facilement obtenues : soit $f : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue,

$$\begin{aligned} \mathbf{x} \in \mathbb{I}\mathbb{R} &\implies f^K(\mathbf{x}) = \text{range}(f, \mathbf{x}) \\ \mathbf{x} \in \overline{\mathbb{I}\mathbb{R}} &\implies f^K(\mathbf{x}) = \text{dual}(\text{range}(f, \text{pro } \mathbf{x})) \end{aligned}$$

Par exemple, dans les cas de \exp et \ln , les expressions des extensions sont les suivantes :

$$\begin{aligned} \exp^K \mathbf{x} &= [\exp(\inf \mathbf{x}), \exp(\sup \mathbf{x})] \\ \ln^K \mathbf{x} &= [\ln(\inf \mathbf{x}), \ln(\sup \mathbf{x})] \text{ pour } \mathbf{x} > 0 \end{aligned}$$

De plus, il est facile de vérifier grâce à ces expressions que quelque soient $\mathbf{x}, \mathbf{y} \in \mathbb{K}\mathbb{R}$,

$$\exp^K(\mathbf{x} + \mathbf{y}) = (\exp^K \mathbf{x})(\exp^K \mathbf{y}) \text{ et } \ln^K \exp^K \mathbf{x} = \mathbf{x}.$$

Si en plus $\mathbf{x} > 0$ et $\mathbf{y} > 0$ alors

$$\ln^K(\mathbf{xy}) = \ln^K \mathbf{y} + \ln^K \mathbf{x}.$$

Donc les bijections réciproques \ln^K et \exp^K transforment le groupe additif en groupe multiplicatif et inversement, comme dans le cas du corps des réels.

3.5 Développements supplémentaires

Une extension de Taylor est définie dans Kaucher[21] : étant donné $g : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continûment dérivable, $\mathbf{x}, \mathbf{y} \in \mathbb{K}\mathbb{R}$ et $\tilde{x} \in \text{pro } \mathbf{x}$

$$g(\tilde{x}) + (g')^K(\mathbf{x}) \times (\mathbf{x} - \tilde{x}).$$

Cette extension de Taylor au premier ordre a été généralisée dans le même article à un ordre quelconque et à des fonctions possédant des paramètres. Ces extensions n'ont pas été utilisées car elles ne présentent pas de bonne interprétation par rapport à la fonction réelle qu'elles étendent.

L'extension de la valeur moyenne que nous proposerons dans le chapitre 16 (page 171) a une expression proche de l'extension de Taylor proposée dans Kaucher[21]. Néanmoins, les différences entre les deux extensions permettent à l'extension définie dans la suite d'avoir des interprétations utiles par rapport à la fonction qu'elle étend.

Un algorithme semblable à l'algorithme de Newton par intervalle (voir par exemple Hansen[18]) utilisant les intervalles généralisés a été proposé dans Gudenberg[50]. Il permet de construire une approximation intérieure de l'ensemble

$$\{x \in \mathbb{R} \mid (\forall a_1 \in \mathbf{a}_1)(\exists a_2 \in \mathbf{a}_2)(a_2 = g(x, a_1))\}$$

où $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ est continûment dérivable et strictement monotone par rapport à x quelque soit $a \in \mathbf{a}_1$. Cet algorithme n'a pas été utilisé ni développé plus avant, en particulier les conditions très restrictives de son utilisation n'ont pas été affaiblies.

3.6 Liens entre les intervalles généralisés et les intervalles classiques

Les opérations de l'arithmétique de Kaucher peuvent être exprimées en utilisant uniquement les intervalles classiques. Pour cela, de nouvelles opérations sur les intervalles classiques sont introduites. Ces nouvelles opérations sont appelées "inner operations" ou "nonstandard operations" (voir Markov[87][88][63] et Dimitrova[65]). L'addition non standard, notée \oplus , est définie de la manière suivante :

$$\mathbf{x} \oplus \mathbf{y} = [\min\{\inf \mathbf{x} + \sup \mathbf{y}, \sup \mathbf{x} + \inf \mathbf{y}\}, \max\{\inf \mathbf{x} + \sup \mathbf{y}, \sup \mathbf{x} + \inf \mathbf{y}\}].$$

La définition de la multiplication non standard \otimes est donnée dans Dimitrova[65]. L'addition non standard a été étudiée dans Markov[88][89][63] où elle est mise en relation avec l'extension de l'addition d'intervalle en groupe. Ces opérations définies, l'arithmétique de Kaucher peut être exprimée en termes d'opérations mettant en jeu des intervalles classiques.

3.6.1 L'addition

L'addition de Kaucher peut être exprimée au moyen de l'addition de l'arithmétique des intervalles et de l'addition non standard : étant donnés $\mathbf{x}, \mathbf{y} \in \mathbb{KR}$, la projection propre du résultat de l'addition est tout d'abord déterminée de la manière suivante,

$$\text{pro}(\mathbf{x} + \mathbf{y}) = \begin{cases} (\text{pro } \mathbf{x}) + (\text{pro } \mathbf{y}) & \text{si } (\text{rad } \mathbf{x})(\text{rad } \mathbf{y}) \geq 0 \\ (\text{pro } \mathbf{x}) \oplus (\text{pro } \mathbf{y}) & \text{si } (\text{rad } \mathbf{x})(\text{rad } \mathbf{y}) < 0 \end{cases}$$

Remarque. La condition $(\text{rad } \mathbf{x})(\text{rad } \mathbf{y}) \geq 0$ veut dire que les deux intervalles généralisés \mathbf{x} et \mathbf{y} ont la même qualité propre/impropre.

Ensuite, la qualité propre/impropre de $\mathbf{x} + \mathbf{y}$ est la même que celle de l'argument qui a la plus grande largeur en valeur absolue : cela peut être exprimé de manière équivalente en disant que $\text{rad}(\mathbf{x} + \mathbf{y})$ a le même signe que $(\text{rad } \mathbf{x}) + (\text{rad } \mathbf{y})$, ou encore,

- $\mathbf{x} + \mathbf{y}$ est propre si et seulement si $(\text{rad } \mathbf{x}) + (\text{rad } \mathbf{y}) \geq 0$.
- $\mathbf{x} + \mathbf{y}$ est impropre si et seulement si $(\text{rad } \mathbf{x}) + (\text{rad } \mathbf{y}) \leq 0$.

Exemple II.20. Soient $\mathbf{x} = [2, -2]$ et $\mathbf{y} = [2, 3]$. Alors $\text{rad } \mathbf{x} = -4$ et $\text{rad } \mathbf{y} = 1$. Ainsi,

$$\text{pro}(\mathbf{x} + \mathbf{y}) = (\text{pro } \mathbf{x}) \oplus (\text{pro } \mathbf{y}) = [1, 4]$$

De plus, $(\text{rad } \mathbf{x}) + (\text{rad } \mathbf{y}) = -3$ et ainsi $\mathbf{x} + \mathbf{y}$ est impropre. Finalement, nous concluons que $\mathbf{x} + \mathbf{y} = [4, 1]$, ce qui correspond au résultat de l'arithmétique de Kaucher.

3.6.2 La multiplication

La multiplication de Kaucher peut être obtenue en utilisant la multiplication de l'arithmétique des intervalles et la multiplication non standard dans les cas particuliers où les projections propres des intervalles mis en jeu ne contiennent pas 0 : étant donnés $\mathbf{x}, \mathbf{y} \in \mathbb{KR}$ tels que $0 \notin \text{pro } \mathbf{x}$ et $0 \notin \text{pro } \mathbf{y}$,

$$\text{pro}(\mathbf{x} \times \mathbf{y}) = \begin{cases} (\text{pro } \mathbf{x}) \times (\text{pro } \mathbf{y}) & \text{si } (\text{rad } \mathbf{x})(\text{rad } \mathbf{y}) \geq 0 \\ (\text{pro } \mathbf{x}) \otimes (\text{pro } \mathbf{y}) & \text{si } (\text{rad } \mathbf{x})(\text{rad } \mathbf{y}) < 0 \end{cases}$$

La qualité propre/impropre de $\mathbf{x} \times \mathbf{y}$ est cette fois obtenue en utilisant la fonction χ définie dans Dimitrova[65] pour les intervalles généralisés dont la projection propre ne contient pas 0 de la manière suivante :

$$\chi(\mathbf{x}) = \begin{cases} \inf \mathbf{x} / \sup \mathbf{x} & \text{si } \inf \mathbf{x} > 0 \wedge \sup \mathbf{x} > 0 \\ \sup \mathbf{x} / \inf \mathbf{x} & \text{si } \inf \mathbf{x} < 0 \wedge \sup \mathbf{x} < 0 \end{cases}$$

Remarque. La fonction χ fut originellement proposée dans le contexte des intervalles classique dans [41]. Son expression originale est

$$\chi(\mathbf{x}) = \begin{cases} \inf \mathbf{x} / \sup \mathbf{x} & \text{si } |\inf \mathbf{x}| \leq |\sup \mathbf{x}| \\ \sup \mathbf{x} / \inf \mathbf{x} & \text{sinon} \end{cases}$$

Il est aisé de constater que l'expression proposée dans Dimitrova[65] est équivalente à cette dernière lorsque des intervalles propres sont mis en jeux.

La condition permettant de prévoir la qualité propre/impropre du résultat de la multiplication de Kaucher qui est proposée dans Dimitrova[65] est la suivante :

- $\mathbf{x} \times \mathbf{y}$ et \mathbf{x} ont la même qualité propre/impropre si $\chi(\mathbf{x}) \leq \chi(\mathbf{y})$.
- $\mathbf{x} \times \mathbf{y}$ et \mathbf{y} ont la même qualité propre/impropre si $\chi(\mathbf{x}) > \chi(\mathbf{y})$.

Nous avons trouvé un contre exemple à cette caractérisation :

Exemple II.21. Considérons $\mathbf{x} = [1, 2]$ et $\mathbf{y} = [10, 4]$. Alors $\chi(\mathbf{x}) = 1/2 = 0.5$ et $\chi(\mathbf{y}) = 10/4 = 2.5$. Ainsi, la caractérisation proposée dans Dimitrova[65] prévoit que $\mathbf{x} \times \mathbf{y}$ et \mathbf{x} ont la même qualité propre/impropre. Mais $\mathbf{x} \times \mathbf{y} = [10, 8]$ qui est impropre.

Nous proposons donc la condition suivante :

Proposition II.3. Soient $\mathbf{x}, \mathbf{y} \in \mathbb{KR}$ tels que $\text{pro } \mathbf{x} > 0$ et $\text{pro } \mathbf{y} > 0$. Alors

- $\mathbf{x} \times \mathbf{y} \in \mathbb{IR}$ si et seulement si $\chi(\mathbf{x}) \times \chi(\mathbf{y}) \leq 1$.
- $\mathbf{x} \times \mathbf{y} \in \overline{\mathbb{IR}}$ si et seulement si $\chi(\mathbf{x}) \times \chi(\mathbf{y}) \geq 1$.

Démonstration. Par définition de \ln^K , les intervalles $\mathbf{x} \times \mathbf{y}$ et $\ln^K(\mathbf{x} \times \mathbf{y})$ ont la même qualité propre/impropre. En utilisant les propriétés de \ln^K , nous obtenons

$$\ln^K(\mathbf{x} \times \mathbf{y}) = \ln^K \mathbf{x} + \ln^K \mathbf{y}$$

Ainsi, les intervalles $\mathbf{x} \times \mathbf{y}$ et $\ln^K \mathbf{x} + \ln^K \mathbf{y}$ ont la même qualité propre/impropre. Celle-ci peut-être déterminée en appliquant les résultats obtenus dans le cadre de l'addition sur $\ln^K \mathbf{x} + \ln^K \mathbf{y}$. Ainsi, $\ln^K \mathbf{x} + \ln^K \mathbf{y}$ est propre si et seulement si $\text{wid } \ln^K \mathbf{x} + \text{wid } \ln^K \mathbf{y} \geq 0$. Or, en appliquant l'expression de \ln^K , nous obtenons

$$\begin{aligned} \text{wid } \ln^K \mathbf{x} + \text{wid } \ln^K \mathbf{y} &= \ln \sup \mathbf{x} - \ln \inf \mathbf{x} + \ln \sup \mathbf{y} - \ln \inf \mathbf{y} \\ &= \ln \frac{\sup \mathbf{x}}{\inf \mathbf{x}} + \ln \frac{\sup \mathbf{y}}{\inf \mathbf{y}}. \end{aligned}$$

Ainsi, $\ln^K \mathbf{x} + \ln^K \mathbf{y}$ est propre si et seulement si

$$\ln \frac{1}{\chi(\mathbf{x})} + \ln \frac{1}{\chi(\mathbf{y})} \leq 0.$$

C'est à dire $\ln(\chi(\mathbf{x})) + \ln(\chi(\mathbf{y})) \geq 0$. Finalement, en appliquant \exp aux deux membres de l'équation (\exp étant croissante) nous obtenons la condition présentée dans l'énoncé : $\chi(\mathbf{x}) \times \chi(\mathbf{y}) \geq 1$. Le seconde condition est obtenue de manière similaire. \square

Exemple II.22. Soient $\mathbf{x} = [1, 2]$ et $\mathbf{y} = [10, 4]$. Alors $\text{rad } \mathbf{x} = 1$ et $\text{rad } \mathbf{y} = -6$ donc

$$\text{pro}(\mathbf{x} \times \mathbf{y}) = \mathbf{x} \otimes \mathbf{y} = [8, 10]$$

De plus, $\chi(\mathbf{x}) = 0.5$ et $\chi(\mathbf{y}) = 2.5$ donc $\chi(\mathbf{x}) \times \chi(\mathbf{y}) \geq 1$ et finalement nous concluons que $\mathbf{x} \times \mathbf{y}$ est impropre. Nous avons donc montré que $\mathbf{x} \times \mathbf{y} = [10, 8]$ en n'effectuant que des opérations sur des intervalles classiques, ce qui correspond bien au résultat de la multiplication de Kaucher, i.e. $\mathbf{x} \times \mathbf{y} = [1 \times 10, 2 \times 8]$.

Chapitre 4

Intervalles modaux

Les intervalles modaux ont été introduits dans Gardenes[16]. La formulation actuelle de la théorie est décrite dans SIGLA/X[37][38]. La théorie des intervalles modaux propose une généralisation des intervalles classiques ainsi que des extensions des fonctions continues dont l'interprétation améliore l'interprétation offerte par les extensions aux intervalles classiques. Les applications de ces interprétations enrichies sont prometteuses (voir par exemple Vehí[49], Armengol[45], Sainz[62][60][61] et Herrero[69][70]). Malgré tout, la théorie des intervalles modaux reste confidentielle : seule les chercheurs de l'université de Gironne utilisent les intervalles modaux et produisent des publications sur le sujet.

Nous verrons en fin de chapitre que les intervalles modaux et les intervalles généralisés sont isomorphes. Cet isomorphisme concerne toutes les opérations définies dans ces deux structures, i.e. l'inclusion, l'intersection et la réunion et l'arithmétique. Il est cependant à noter que si l'arithmétique des intervalles modaux coïncide (à un isomorphisme près) avec l'arithmétique de Kaucher, elles sont obtenues par des voix différentes. Ainsi, nous voyons la situation de la manière suivante : les intervalles modaux permettent de donner une interprétation à l'arithmétique de Kaucher, alors que les intervalles généralisés offre des expressions efficaces (pour les calculs et pour certaines démonstrations) à l'arithmétique des intervalles modaux.

4.1 Les intervalles modaux

Dans un souci de clarté, les intervalles modaux sont maintenant présentés indépendamment des intervalles généralisés. Le fort lien entre ces deux structures (qui sont isomorphes) sera présenté dans la dernière section. Ainsi, la présentation faite ici est légèrement différente de celle faite habituellement par les auteurs de la théorie des intervalles modaux : l'introduction des intervalles généralisés au début de la théorie est nécessaire pour certaines démonstrations proposées par les auteurs. L'absence de démonstrations dans la description de la théorie proposée ici nous permet de faire une présentation indépendante des intervalles généralisés.

Notation. *La théorie des intervalles modaux est maintenant présentée avec des notations différentes de celle choisies par les auteurs de la théorie dans SIGLA/X[37, 38]. Ce choix est motivé par la nécessité de conserver des notations homogènes dans cette thèse. Un intervalle modal sera donc noté en utilisant des lettres grasses surmontée d'un point :*

notations	intervalles classiques	intervalles modaux
SIGLA/X[37]	X'	X
dissertation	\mathbf{x}	$\dot{\mathbf{x}}$

De plus, les propositions quantifiées sont écrites de la manière suivante dans SIGLA/X[37] :

$$U(x, \mathbf{x})E(y \in \mathbf{y})P(x, y)$$

Elles seront écrites en conservant la notation choisie dans cette thèse, i.e.

$$(\forall x \in \mathbf{x})(\exists y \in \mathbf{y})(P(x, y))$$

Quelques notations supplémentaires différant des notations proposées dans SIGLA/X[37, 38] seront introduites dans ce chapitre. Ces nouvelles notations seront mises en valeur dans des paragraphes similaires à celui-ci, intitulés **Notation**.

4.1.1 Définition des intervalles modaux

Un intervalle modal est un couple intervalle/quantificateur. Par exemple,

$$([-1, 1], \exists) \text{ et } ([-1, 1], \forall)$$

sont des intervalles modaux, respectivement appelés intervalle existentiel (ou propre) et universel (ou impropre). L'ensemble des intervalles modaux est noté $I^*(\mathbb{R})$. Les fonctions suivantes permettent de manipuler les intervalles modaux : soit $\dot{\mathbf{x}} = (\mathbf{x}, \mathbf{Q})$, alors les opérations $\text{Set}(\dot{\mathbf{x}})$ et $\text{Mod}(\dot{\mathbf{x}})$ sont définies de la manière suivante : $\text{Set}(\dot{\mathbf{x}}) = \mathbf{x}$ et $\text{Mod}(\dot{\mathbf{x}}) = \mathbf{Q}$. L'opération Dual inverse le quantificateur associé à un intervalle modal.

Notation. Pour permettre l'utilisation des notations habituelles des propositions quantifiées, le quantificateur dépendant de la modalité d'un intervalle modal, noté $Q(x, \dot{\mathbf{x}})P(x)$ dans SIGLA/X[37], est noté $(\mathbf{Q}(\dot{\mathbf{x}})x \in \text{Set}(\dot{\mathbf{x}}))(P(x))$. Il est défini par $\mathbf{Q}(\dot{\mathbf{x}}) = \text{Mod}(\dot{\mathbf{x}})$. Ainsi, si $\text{Mod}(\dot{\mathbf{x}}) = \forall$ alors

$$(\mathbf{Q}(\dot{\mathbf{x}})x \in \text{Set}(\dot{\mathbf{x}}))(P(x)) \iff (\forall x \in \text{Set}(\dot{\mathbf{x}}))(P(x));$$

si $\text{Mod}(\dot{\mathbf{x}}) = \exists$ alors

$$(\mathbf{Q}(\dot{\mathbf{x}})x \in \text{Set}(\dot{\mathbf{x}}))(P(x)) \iff (\exists x \in \text{Set}(\dot{\mathbf{x}}))(P(x)).$$

Remarque. Lorsque $\text{Set}(\dot{\mathbf{x}})$ est dégénéré, l'égalité $([a, a], \forall) = ([a, a], \exists)$ est imposée implicitement dans la théorie des intervalles modaux (lemmes I.2.3, I.2.4 et I.2.5 de SIGLA/X[37]). La formalisation de cette idée devrait passer par la définition des intervalles modaux comme les classes d'équivalence obtenues en utilisant la relation d'équivalence

$$\dot{\mathbf{x}}\mathcal{R}\dot{\mathbf{y}} \iff (\dot{\mathbf{x}} = \dot{\mathbf{y}} \vee (\exists a \in \mathbb{R})(\text{Set}(\dot{\mathbf{x}}) = \text{Set}(\dot{\mathbf{y}}) = [a, a])),$$

i.e. $I^*(\mathbb{R}) = (\mathbb{IR} \times \{\forall, \exists\})/\mathcal{R}$. En suivant la présentation de la théorie faite dans SIGLA/X[37], les intervalles modaux $([a, a], \forall)$ et $([a, a], \exists)$ seront simplement identifiés. Le plus important est que dans ce cas, les deux propositions quantifiées

$$\forall x \in [a, a], P(x) \text{ et } \exists x \in [a, a], P(x)$$

sont équivalentes quelque soit le prédicat P (ces deux propositions quantifiées étant équivalentes à $P(a)$). Ainsi, la proposition quantifiée $(\mathbf{Q}(\dot{\mathbf{x}})x \in \text{Set}(\dot{\mathbf{x}}))(P(x))$ a un sens bien défini dans le cas où $\text{Set}(\dot{\mathbf{x}})$ est dégénéré.

Notation. L'intervalle modal $([a, a], \exists) = ([a, a], \forall)$ sera aussi noté a .

Si un intervalle classique est caractérisé par l'ensemble des points qu'il contient, un intervalle modal est caractérisé par l'ensemble des prédicats qu'il valide : l'ensemble des prédicats réels, i.e. des fonctions de \mathbb{R} dans $\{0, 1\}$, est noté $\text{Pred}(\mathbb{R})$. Parmi ces prédicats, ceux qui vérifient

$$\{x \in \mathbb{R} \mid P(x)\} \in \mathbb{IR}$$

ont un rôle privilégié (ces prédicats sont appelés "interval-predicats" dans SIGLA/X[37]).

Notation. Étant donné un intervalle $\mathbf{x} \in \mathbb{IR}$, le prédicat $x \in \mathbf{x}$ est noté $P_{\mathbf{x}}(x)$.

L'ensemble des "interval-predicats" est noté

$$\begin{aligned} \text{Pred}^*(\mathbb{R}) &= \{P \in \text{Pred}(\mathbb{R}) \mid \{x \in \mathbb{R} \mid P(x)\} \in \mathbb{IR}\} \\ &= \{P \in \text{Pred}(\mathbb{R}) \mid (\exists \mathbf{x} \in \mathbb{IR})(P = P_{\mathbf{x}})\}. \end{aligned}$$

L'ensemble des "interval-predicats" validés par un intervalle modal est alors défini de la manière suivante : soit un intervalle modal $\dot{\mathbf{x}} = (\mathbf{x}, \mathbf{Q})$,

$$\text{Pred}^*(\dot{\mathbf{x}}) = \left\{ P \in \text{Pred}^*(\mathbb{R}) \mid (\mathbf{Q}(\dot{\mathbf{x}})x \in \mathbf{x})(P(x)) \right\}.$$

Exemple II.23. Soit $\dot{\mathbf{x}} = ([-1, 1], \exists)$. Alors,

$$\text{Pred}^*(\dot{\mathbf{x}}) = \{P \in \text{Pred}^*(\mathbb{R}) \mid (\exists x \in [-1, 1])(P(x))\}.$$

Exemple II.24. Soit $\dot{\mathbf{x}} = ([-1, 1], \forall)$. Alors,

$$\text{Pred}^*(\dot{\mathbf{x}}) = \{P \in \text{Pred}^*(\mathbb{R}) \mid (\forall x \in [-1, 1])(P(x))\}$$

En généralisant les deux exemples précédents, l'ensemble $\text{Pred}^*(\dot{\mathbf{x}})$ est caractérisé de la manière suivante : soit un intervalle modal $\dot{\mathbf{x}} = (\mathbf{x}, \mathbf{Q})$,

– si $\mathbf{Q} = \forall$ alors

$$\begin{aligned} \text{Pred}^*(\dot{\mathbf{x}}) &= \{P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid (\forall x \in \mathbf{x})(P_{\mathbf{z}}(z))\} \\ &= \{P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid (\forall x \in \mathbf{x})(x \in \mathbf{z})\} \\ &= \{P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid \mathbf{x} \subseteq \mathbf{z}\} \end{aligned}$$

– si $\mathbf{Q} = \exists$ alors

$$\begin{aligned} \text{Pred}^*(\dot{\mathbf{x}}) &= \{P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid (\exists x \in \mathbf{x})(P_{\mathbf{z}}(z))\} \\ &= \{P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid (\exists x \in \mathbf{x})(x \in \mathbf{z})\} \\ &= \left\{ P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid \mathbf{x} \cap \mathbf{z} \neq \emptyset \right\} \end{aligned}$$

Remarque. Un deuxième ensemble de prédicats $\text{Copred}^*(\dot{\mathbf{x}})$ est associé aux intervalles modaux. Néanmoins, ces "interval-copredicats" ne sont pas définis explicitement dans cette présentation de la théorie des intervalles modaux (voir SIGLA/X[37] pour la définition des "interval-copredicats").

L'inclusion entre intervalles modaux peut être¹³ définie de la manière suivante : soient deux intervalles modaux $\dot{\mathbf{x}}, \dot{\mathbf{y}} \in I^*(\mathbb{R})$,

$$\dot{\mathbf{x}} \subseteq \dot{\mathbf{y}} \iff \text{Pred}^*(\dot{\mathbf{x}}) \subseteq \text{Pred}^*(\dot{\mathbf{y}})$$

Les trois exemples suivants présentent les différentes situations rencontrées pour l'inclusion de deux intervalles modaux. Le premier exemple présente le cas de deux intervalles modaux existentiels.

Exemple II.25. Soient $\dot{\mathbf{x}} = ([-1, 1], \exists)$ et $\dot{\mathbf{y}} = ([-2, 2], \exists)$. Alors

$$\text{Pred}^*(\dot{\mathbf{x}}) = \left\{ P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid \mathbf{z} \cap [-1, 1] \neq \emptyset \right\}$$

et

$$\text{Pred}^*(\dot{\mathbf{y}}) = \left\{ P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid \mathbf{z} \cap [-2, 2] \neq \emptyset \right\}.$$

La condition $\mathbf{z} \cap [-1, 1] \neq \emptyset$ est plus contraignante que $\mathbf{z} \cap [-2, 2] \neq \emptyset$ car $[-1, 1] \subseteq [-2, 2]$. Ainsi, nous avons $\text{Pred}^*(\dot{\mathbf{x}}) \subseteq \text{Pred}^*(\dot{\mathbf{y}})$ et finalement $\dot{\mathbf{x}} \subseteq \dot{\mathbf{y}}$.

Le second exemple présente le cas de deux intervalles modaux universels.

Exemple II.26. Soient $\dot{\mathbf{x}} = ([-1, 1], \forall)$ et $\dot{\mathbf{y}} = ([-2, 2], \forall)$. Alors

$$\text{Pred}^*(\dot{\mathbf{x}}) = \{ P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid [-1, 1] \subseteq \mathbf{z} \} \quad \text{et} \quad \text{Pred}^*(\dot{\mathbf{y}}) = \{ P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid [-2, 2] \subseteq \mathbf{z} \}$$

La condition $[-2, 2] \subseteq \mathbf{z}$ est plus contraignante que $[-1, 1] \subseteq \mathbf{z}$ car $[-1, 1] \subseteq [-2, 2]$. Ainsi, nous avons $\text{Pred}^*(\dot{\mathbf{y}}) \subseteq \text{Pred}^*(\dot{\mathbf{x}})$ et finalement, $\dot{\mathbf{y}} \subseteq \dot{\mathbf{x}}$.

Ces deux premiers exemples montrent que les intervalles existentiels jouent le rôle des intervalles classiques car $(\mathbf{x}, \exists) \subseteq (\mathbf{y}, \exists) \iff \mathbf{x} \subseteq \mathbf{y}$. Enfin, le troisième exemple présente le cas d'un intervalle modal existentiel et d'un intervalle modal universel.

Exemple II.27. Soient $\dot{\mathbf{x}} = ([-1, 1], \forall)$ et $\dot{\mathbf{y}} = ([0, 2], \exists)$. Alors

$$\text{Pred}^*(\dot{\mathbf{x}}) = \{ P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid [-1, 1] \subseteq \mathbf{z} \} \quad \text{et} \quad \text{Pred}^*(\dot{\mathbf{y}}) = \left\{ P_{\mathbf{z}} \in \text{Pred}^*(\mathbb{R}) \mid \mathbf{z} \cap [0, 2] \neq \emptyset \right\}$$

La condition $[-1, 1] \subseteq \mathbf{z}$ étant plus contraignante que $\mathbf{z} \cap [-1, 1] \neq \emptyset$, nous avons $\text{Pred}^*(\dot{\mathbf{x}}) \subseteq \text{Pred}^*(\dot{\mathbf{y}})$. Finalement, $\dot{\mathbf{x}} \subseteq \dot{\mathbf{y}}$.

Le lemme suivant généralise les trois exemples précédents.

Lemme II.1 (SIGLA/X[37], Lem I.2.4). Soient $\dot{\mathbf{x}}, \dot{\mathbf{y}} \in I^*(\mathbb{R})$. Alors,

$$\dot{\mathbf{x}} \subseteq \dot{\mathbf{y}} \iff \begin{cases} \text{if } \text{Mod}(\dot{\mathbf{x}}) = \text{Mod}(\dot{\mathbf{y}}) = \exists & \text{then } \text{Set}(\dot{\mathbf{x}}) \subseteq \text{Set}(\dot{\mathbf{y}}) \\ \text{if } \text{Mod}(\dot{\mathbf{x}}) = \text{Mod}(\dot{\mathbf{y}}) = \forall & \text{then } \text{Set}(\dot{\mathbf{x}}) \supseteq \text{Set}(\dot{\mathbf{y}}) \\ \text{if } \text{Mod}(\dot{\mathbf{x}}) = \forall \wedge \text{Mod}(\dot{\mathbf{y}}) = \exists & \text{then } \text{Set}(\dot{\mathbf{x}}) \cap \text{Set}(\dot{\mathbf{y}}) \neq \emptyset \\ \text{if } \text{Mod}(\dot{\mathbf{x}}) = \exists \wedge \text{Mod}(\dot{\mathbf{y}}) = \forall & \text{then } \text{Set}(\dot{\mathbf{x}}) = \text{Set}(\dot{\mathbf{y}}) = [a, a] \end{cases}$$

¹³La définition proposée dans SIGLA/X[37] fait intervenir les ensembles de prédicats $\text{Pred}(\dot{\mathbf{x}})$. La définition proposée ici est équivalente et permet une présentation simplifiée qui ne nécessite pas l'introduction des ensembles de prédicats $\text{Pred}(\dot{\mathbf{x}})$.

4.1.2 Union et intersection

L'intersection et la réunion des intervalles modaux sont définies comme étant la borne inférieure et la borne supérieure de l'inclusion. Elle sont notées respectivement par \bigwedge et \bigvee . Formellement, soit un ensemble d'intervalles modaux $E \subseteq I^*(\mathbb{R})$. Alors,

$$\begin{aligned} (\forall \dot{\mathbf{z}} \in I^*(\mathbb{R})) (\dot{\mathbf{z}} \supseteq (\bigvee E) &\iff \forall \dot{\mathbf{x}} \in E, \dot{\mathbf{z}} \supseteq \dot{\mathbf{x}}) \\ \text{et } (\forall \dot{\mathbf{z}} \in I^*(\mathbb{R})) (\dot{\mathbf{z}} \subseteq (\bigwedge E) &\iff \forall \dot{\mathbf{x}} \in E, \dot{\mathbf{z}} \subseteq \dot{\mathbf{x}}). \end{aligned}$$

Lorsque E contient deux éléments $\dot{\mathbf{x}}$ et $\dot{\mathbf{y}}$, $\bigwedge E$ et $\bigvee E$ sont respectivement notés $\dot{\mathbf{x}} \bigwedge \dot{\mathbf{y}}$ et $\dot{\mathbf{x}} \bigvee \dot{\mathbf{y}}$. Les formules explicites de ces opérations sont facilement obtenues en utilisant l'isomorphisme entre les intervalles modaux et les intervalles généralisés (voir Section 4.3). Les interprétations suivantes sont obtenues sans difficulté sans l'introduction des intervalles généralisés : soient deux intervalles modaux existentiels $\dot{\mathbf{x}} = (\mathbf{x}, \exists)$ et $\dot{\mathbf{y}} = (\mathbf{y}, \exists)$ alors,

$$\mathbf{x} \bigcap \mathbf{y} \neq \emptyset \implies (\mathbf{x}, \exists) \bigwedge (\mathbf{y}, \exists) = (\mathbf{x} \bigcap \mathbf{y}, \exists) \quad (4.1)$$

$$\mathbf{x} \bigcap \mathbf{y} = \emptyset \implies (\mathbf{x}, \exists) \bigwedge (\mathbf{y}, \exists) = (\mathbf{z}, \exists) \quad (4.2)$$

où \mathbf{z} est le plus petit intervalle qui vérifie $\mathbf{z} \bigcap \mathbf{x} \neq \emptyset$ et $\mathbf{z} \bigcap \mathbf{y} \neq \emptyset$. Cet intervalle s'écrit $[\sup \mathbf{x}, \inf \mathbf{y}]$ si $\mathbf{x} \leq \mathbf{y}$ ou $[\sup \mathbf{y}, \inf \mathbf{x}]$ si $\mathbf{y} \leq \mathbf{x}$.

4.2 Extensions des fonctions continues aux intervalles modaux

4.2.1 Extensions sémantiques

La proposition suivante donne une formulation équivalente pour la définition d'une extension faible aux intervalles classiques, i.e. une fonction d'intervalles \mathbf{g} qui vérifie $\text{range}(f, \mathbf{x}) \subseteq \mathbf{g}(\mathbf{x})$. C'est cette formulation de la définition d'une extension faible aux intervalles classiques qui va être généralisée pour conduire aux extensions sémantiques aux intervalles modaux. Il est aisé de constater que l'inclusion $\text{range}(f, \mathbf{x}) \subseteq \mathbf{z}$ est vraie si et seulement si

$$(\forall \mathbf{u} \in \mathbb{I}\mathbb{R}^n) (\mathbf{u} \bigcap \mathbf{x} \neq \emptyset \implies \text{range}(f, \mathbf{u}) \bigcap \mathbf{z} \neq \emptyset)$$

Exprimée en utilisant la sémantique des prédicats de la théorie des intervalles modaux, nous obtenons la proposition suivante :

Lemme II.2 (SIGLA/X[37], Lem II.2.1). *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$ et $\mathbf{z} \in \mathbb{I}\mathbb{R}$. Alors l'inclusion $\text{range}(f, \mathbf{x}) \subseteq \mathbf{z}$ est vraie si et seulement si*

$$(\forall \mathbf{u} \in \mathbb{I}\mathbb{R}^n) (P_{\mathbf{u}} \in \text{Pred}^*(\mathbf{x}, \exists) \implies P_{\text{range}(f, \mathbf{u})} \in \text{Pred}^*(\mathbf{z}, \exists))$$

Remarque. *Le lemme précédent est en fait plus général que le lemme II.2.1 de SIGLA/X[37], mais sa démonstration est similaire.*

Les extensions sémantiques d'une fonction réelle sont alors définies en remplaçant la condition $\text{range}(f, \mathbf{x}) \subseteq \mathbf{g}(\mathbf{x})$ par la condition équivalente donnée par la proposition précédente et en prenant en compte les intervalles modaux universels en plus des existentiels :

Définition II.10 (SIGLA/X[37], Def II.2.2). Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue et $\dot{\mathbf{g}} : I^*(\mathbb{R}^n) \longrightarrow I^*(\mathbb{R})$ une fonction d'intervalles modaux. La fonction d'intervalles modaux $\dot{\mathbf{g}}$ est une extension sémantique aux intervalles modaux¹⁴ de f en $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$ si et seulement si

$$(\forall \mathbf{u} \in \mathbb{R}^n)(P_{\mathbf{u}} \in \text{Pred}^*(\dot{\mathbf{x}}) \implies P_{\text{range}(f, \mathbf{u})} \in \text{Pred}^*(\dot{\mathbf{g}}(\dot{\mathbf{x}})))$$

La fonction d'intervalles modaux $\dot{\mathbf{g}}$ est une extension sémantique de f si elle est une extension sémantique de f pour tout $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$.

Ainsi, lorsque $\dot{\mathbf{x}}$ et $\mathbf{g}(\dot{\mathbf{x}})$ sont existentiels, i.e. $\dot{\mathbf{x}} = (\mathbf{x}, \exists)$ et $\mathbf{g}(\dot{\mathbf{x}}) = (\mathbf{z}, \exists)$, l'interprétation d'une extension sémantique correspond à l'interprétation d'une extension aux intervalles classiques : en effet, la proposition II.2 montre alors que $\text{range}(f, \mathbf{x}) \subseteq \mathbf{z}$ (les intervalles classiques sont donc identifiés aux intervalles modaux existentiels). Lorsque $\dot{\mathbf{x}}$ et $\mathbf{g}(\dot{\mathbf{x}})$ ne sont plus existentiels, de nouvelles interprétations sont offertes par la définition des extensions sémantiques. La description de ces nouvelles interprétations et les conditions pour qu'une fonction d'intervalles modaux soit une extension sémantique sont l'objet de la prochaine sous-section.

4.2.2 Fonctions sémantiques

Deux fonctions jouent un rôle central dans la théorie des intervalles modaux. Étant donnés une fonction continue $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et un intervalle modal $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$, la fonction *-sémantique f^* et la fonction **sémantique f^{**} sont définies de la manière suivante : soit $\dot{\mathbf{x}} = ((\mathbf{x}_1, \mathbf{Q}_1), \dots, (\mathbf{x}_n, \mathbf{Q}_n))$,

$$f^*(\dot{\mathbf{x}}) = \bigvee_{x_P \in \mathbf{x}_P} \bigwedge_{x_I \in \mathbf{x}_I} f(x) \quad \text{et} \quad f^{**}(\dot{\mathbf{x}}) = \bigwedge_{x_I \in \mathbf{x}_I} \bigvee_{x_P \in \mathbf{x}_P} f(x)$$

où P et I sont les ensembles d'indices correspondant respectivement au composantes propres et impropres de $\dot{\mathbf{x}}$.

Remarque. L'isomorphisme entre les intervalles modaux et les intervalles généralisés offre une procédure de calcul pour l'intersection et la réunion d'intervalles modaux (voir section 4.3 page 61). Il permet ainsi d'obtenir une expression de f^* utilisant les quantités suivantes :

$$\min_{x_P \in \mathbf{x}_P} \max_{x_I \in \mathbf{x}_I} f(x) \quad ; \quad \max_{x_P \in \mathbf{x}_P} \min_{x_I \in \mathbf{x}_I} f(x)$$

Une expression similaire est disponible pour f^{**} .

L'exemple suivant illustre le calcul de f^{**} .

Exemple II.28. Considérons la fonction $f(x, y) = x^2 + y^2$ et les intervalles modaux $\dot{\mathbf{x}} = ([-4, 5], \exists)$ et $\dot{\mathbf{y}} = ([4, 6], \forall)$. Alors, par définition de f^{**} ,

$$\begin{aligned} f^{**}(\dot{\mathbf{x}}, \dot{\mathbf{y}}) &= \bigwedge_{y \in \text{Set}(\dot{\mathbf{y}})} \bigvee_{x \in \text{Set}(\dot{\mathbf{x}})} f(x, y) \\ &= \bigwedge_{y \in \text{Set}(\dot{\mathbf{y}})} ([0 + y^2, (\bar{\mathbf{x}})^2 + y^2], \exists) \\ &= \bigwedge_{y \in \text{Set}(\dot{\mathbf{y}})} ([(\bar{\mathbf{y}})^2, (\bar{\mathbf{x}})^2 + (\underline{\mathbf{y}})^2], \exists) \end{aligned}$$

¹⁴"Semantic interval extension" dans SIGLA/X[37].

La dernière égalité étant vraie car l'intersection modale d'intervalles existentiels correspond à l'intersection de leur domaine lorsque cette dernière n'est pas vide, cf. Équation (4.1) (page 53). Finalement, $f^{**} = ([36, 41], \exists)$.

Le théorème suivant, appelé le théorème sémantique pour f^* , relie la fonction sémantique f^* aux extensions sémantiques de f et donne une interprétation à ces dernières.

Théorème II.3 (SIGLA/X[37], Thm II.4.1). *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$ et $\dot{\mathbf{z}} \in I^*(\mathbb{R})$. Alors les trois conditions suivantes sont équivalentes :*

1. $f^*(\dot{\mathbf{x}}) \subseteq \dot{\mathbf{z}}$;
2. $(\forall \mathbf{u} \in \mathbb{I}\mathbb{R}^n) (P_{\mathbf{u}} \in \text{Pred}^*(\dot{\mathbf{x}}) \implies P_{\text{range}(f, \mathbf{u})} \in \text{Pred}^*(\dot{\mathbf{z}}))$;
3. $(\forall x_P \in \text{Set}(\dot{\mathbf{x}})_P) (\mathbf{Q}^{\dot{\mathbf{z}}})z \in \text{Set}(\dot{\mathbf{z}}) (\exists x_I \in \text{Set}(\mathbf{x})_I) (f(x) = z)$.

L'équivalence 1. \iff 2. montre que, d'une part la fonction sémantique f^* est une extension sémantique, et d'autre part que c'est la plus petite au sens de l'inclusion modale. L'équivalence 2. \iff 3. donne une interprétation des extensions sémantiques par rapport à la fonction réelle originale f et montre que cette interprétation n'est possible que pour les extensions sémantiques de f . Les calculs de l'exemple suivant sont pris de l'exemple II.3.1 de SIGLA/X[37].

Exemple II.29. *Considérons la fonction $g(x) = (x_1 + x_2)^2$ et les intervalles modaux $\dot{\mathbf{x}}_1 = ([-1, 1], \exists)$ et $\dot{\mathbf{x}}_2 = ([-1, 1], \forall)$. Alors,*

$$g^*(\dot{\mathbf{x}}) = \bigvee_{x_1 \in \text{Set}(\dot{\mathbf{x}}_1)} \bigwedge_{x_2 \in \text{Set}(\dot{\mathbf{x}}_2)} g(x).$$

Il est montré dans SIGLA/X[37] (exemple II.3.1 page 182) qu'alors $g^*(\dot{\mathbf{x}}) = ([0, 1], \forall)$. Ainsi, la proposition quantifiée suivante est vraie :

$$(\forall x_1 \in [-1, 1]) (\forall z \in [0, 1]) (\exists x_2 \in [-1, 1]) (f(x) = z)$$

De plus, en remarquant que $g^*(\dot{\mathbf{x}}) \subseteq ([0.1, 0.9], \forall)$, la proposition quantifiée suivante est validée par l'implication 1. \implies 3. du théorème sémantique pour f^* :

$$(\forall x_1 \in [-1, 1]) (\forall z \in [0.1, 0.9]) (\exists x_2 \in [-1, 1]) (f(x) = z)$$

De la même manière, en remarquant que $g^*(\dot{\mathbf{x}}) \subseteq ([0.9, 2], \exists)$, la proposition quantifiée suivante est vraie :

$$(\forall x_1 \in [-1, 1]) (\exists z \in [0.9, 2]) (\exists x_2 \in [-1, 1]) (f(x) = z)$$

Le théorème suivant, appelé le théorème sémantique pour f^{**} , est le dual du théorème précédent dédié à la fonction sémantique f^{**} .

Théorème II.4 (SIGLA/X[37], Thm II.4.2). *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$ et $\dot{\mathbf{z}} \in I^*(\mathbb{R})$. Alors les trois conditions suivantes sont équivalentes :*

1. $\mathbf{z} \subseteq f^*(\dot{\mathbf{x}})$;
2. $(\forall x_I \in \text{Set}(\mathbf{x})_I) (\mathbf{Q}^{\text{Dual}\dot{\mathbf{z}}})z \in \text{Set}(\dot{\mathbf{z}}) (\exists x_P \in \text{Set}(\dot{\mathbf{x}})_P) (f(x) = z)$.

Remarque. Par définition du quantificateur modal, $\mathbf{Q}^{\text{Dual}\dot{\mathbf{z}}}$ est le quantificateur inverse de $\mathbf{Q}^{\dot{\mathbf{z}}}$.

Ce théorème joue le même rôle que le précédent, dans le contexte de f^{**} . Cependant, le point 2. ne correspond plus aux extensions sémantiques (il faudrait définir deux types d'extensions sémantiques, mais ce n'est pas fait dans la théorie des intervalles modaux).

Exemple II.30. *Considérons la fonction $g(x) = (x_1 + x_2)^2$ et les intervalles modaux $\dot{\mathbf{x}}_1 = ([-1, 1], \exists)$ et $\dot{\mathbf{x}}_2 = ([-1, 1], \forall)$. Alors,*

$$g^{**}(\dot{\mathbf{x}}) = \bigwedge_{x_2 \in \text{Set}(\dot{\mathbf{x}}_2)} \bigvee_{x_1 \in \text{Set}(\dot{\mathbf{x}}_1)} g(x).$$

*Il est montré dans SIGLA/X[37] (exemple II.3.1 page 182) qu'alors $g^{**}(\dot{\mathbf{x}}) = ([0, 1], \exists)$. Ainsi, la proposition quantifiée suivante est vraie :*

$$(\forall x_2 \in [-1, 1]) (\forall z \in [0, 1]) (\exists x_1 \in [-1, 1]) (f(x) = z)$$

*De plus, en remarquant que $([0.1, 0.9], \exists) \subseteq g^{**}(\dot{\mathbf{x}})$, la proposition quantifiée suivante est validée par l'implication 1. \implies 3. du théorème sémantique pour f^{**} :*

$$(\forall x_1 \in [-1, 1]) (\forall z \in [0.1, 0.9]) (\exists x_2 \in [-1, 1]) (f(x) = z)$$

*De la même manière, en remarquant que $([0.9, 2], \forall) \subseteq g^{**}(\dot{\mathbf{x}})$, la proposition quantifiée suivante est vrai :*

$$(\forall x_1 \in [-1, 1]) (\exists z \in [0.9, 2]) (\exists x_2 \in [-1, 1]) (f(x) = z)$$

Les propriétés suivantes sont démontrées dans SIGLA/X[37] :

- $f^*(\dot{\mathbf{x}}) \subseteq f^{**}(\dot{\mathbf{x}})$
- $\dot{\mathbf{x}} \subseteq \dot{\mathbf{y}} \implies f^*(\dot{\mathbf{x}}) \subseteq f^*(\dot{\mathbf{y}})$
- $\dot{\mathbf{x}} \subseteq \dot{\mathbf{y}} \implies f^{**}(\dot{\mathbf{x}}) \subseteq f^{**}(\dot{\mathbf{y}})$

Une fonction f vérifiant $f^*(\dot{\mathbf{x}}) = f^{**}(\dot{\mathbf{x}})$ est dite "JM-commutable" pour $\dot{\mathbf{x}}$ (cette expression viens de "J(oin)M(eet)" car l'égalité $f^*(\dot{\mathbf{x}}) = f^{**}(\dot{\mathbf{x}})$ s'obtient en commutant les opérations d'intersection et de réunion dans les expressions respectives des fonctions sémantiques).

4.2.3 Les * et ** extensions rationnelles

Étant donné une expression f pour une fonction f , les extensions rationnelles de f sont obtenues en remplaçant les opérations réelles dans f par les fonctions * ou ** sémantiques correspondantes. Ainsi, deux extensions rationnelles sont définies : fR^* et fR^{**} .

Exemple II.31 (SIGLA/X[37], Ex II.6.1). *Soit la fonction*

$$f(x) = x_1x_2 + g(x_1, x_2) \quad \text{avec} \quad g(x_1, x_2) = (x_1 + x_2)^2.$$

Alors,

- $fR^*(\dot{\mathbf{x}}) = \dot{\mathbf{x}}_1 \times^* \dot{\mathbf{x}}_2 +^* g^*(\dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2)$;
- $fR^{**}(\dot{\mathbf{x}}) = \dot{\mathbf{x}}_1 \times^{**} \dot{\mathbf{x}}_2 +^* g^{**}(\dot{\mathbf{x}}_1, \dot{\mathbf{x}}_2)$.

Remarque. *Des arrondis extérieurs et intérieurs des extensions rationnelles sont définis dans SIGLA/X[37]. Ces arrondis sont compatibles avec les interprétations enrichies offertes par les extensions aux intervalles modaux. Néanmoins, ils ne sont pas présentés en détail ici car une procédure d'arrondi similaire est présentée dans la section 12.5 de la partie IV (page 142) dans le contexte de la reformulation de la théorie des intervalles modaux. De manière générale, les arrondis de la théorie des intervalles modaux vérifient $\dot{\mathbf{z}} \subseteq \text{Out}(\dot{\mathbf{z}})$ et $\text{Inn}(\dot{\mathbf{z}}) \subseteq \dot{\mathbf{z}}$.*

En général $fR^*(\dot{\mathbf{x}}) \neq fR^{**}(\dot{\mathbf{x}})$. Les deux théorèmes suivants présentent les situations dans lesquelles les extensions rationnelles sont interprétables. Le théorème suivant s'appelle "** interpretability of modal rational functions*".

Théorème II.5 (SIGLA/X[37], Thm II.6.1). *Soit la fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par une composition de fonctions continues et $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$. Si d'une part $\text{Out}(fR^*(\dot{\mathbf{x}}))$ est définie et d'autre part si chaque composante universelle de $\dot{\mathbf{x}}$ a une unique incidence dans l'expression de f , alors $f^*(\dot{\mathbf{x}}) \subseteq \text{Out}(fR^*(\dot{\mathbf{x}}))$, et ainsi le théorème sémantique pour f^* s'applique.*

Remarque. *Il faut faire attention que les composantes universelles de $\dot{\mathbf{x}}$ correspondent aux variables qui sont quantifiées existentiellement dans la proposition quantifiée qui correspond à l'interprétation offerte par le théorème sémantique pour f^* .*

Exemple II.32. *Soit $f(x) = x_1(x_2 - x_1)$ et $\dot{\mathbf{x}}_1 = ([2, 4], \forall)$ et $\dot{\mathbf{x}}_2 = ([1, 2], \exists)$. Alors*

$$fR(\dot{\mathbf{x}}) = \dot{\mathbf{x}}_1 \times^* (\dot{\mathbf{x}}_2 -^* \dot{\mathbf{x}}_1)$$

n'est pas interprétable grâce au théorème sémantique II.3 car $\dot{\mathbf{x}}_1$ est universel et a deux occurrences dans l'expression de f . Il faudra utiliser le théorème II.7 pour obtenir un intervalle modal interprétable, cf. Exemple II.36.

Exemple II.33. *Soit $f(x) = x_1(x_2 - x_1)$ et $\dot{\mathbf{x}}_1 = ([2, 4], \exists)$ et $\dot{\mathbf{x}}_2 = ([1, 2], \forall)$. Cette fois,*

$$fR^*(\dot{\mathbf{x}}) = \dot{\mathbf{x}}_1 \times^* (\dot{\mathbf{x}}_2 -^* \dot{\mathbf{x}}_1) = ([-8, -2], \exists)$$

est interprétable grâce au théorème sémantique II.3 : l'intervalle universel $\dot{\mathbf{x}}_2$ a une unique occurrence dans l'expression de f et le théorème II.5 s'applique montrant que $f^(\dot{\mathbf{x}}) \subseteq fR^*(\dot{\mathbf{x}})$. Ainsi le théorème sémantique II.3 s'applique et la proposition quantifiée suivante est vraie :*

$$(\forall x_1 \in [2, 4]) (\exists z \in [-8, -2]) (\exists x_2 \in [1, 2]) (f(x) = z)$$

Le théorème suivant s'appelle "*** interpretability of modal rational functions*" et est similaire au précédent, mais s'applique à fR^{**} .

Théorème II.6 (SIGLA/X[37], Thm II.6.2). *Soit la fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par une composition de fonctions continues et $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$. Si d'une part $\text{Out}(fR^*(\dot{\mathbf{x}}))$ est définie et d'autre part si chaque composante existentielle de $\dot{\mathbf{x}}$ a une unique incidence dans l'expression de f , alors $\text{Inn}(fR^{**}(\dot{\mathbf{x}})) \subseteq f^{**}(\dot{\mathbf{x}})$, et ainsi le théorème sémantique pour f^{**} s'applique.*

Remarque. *Cette fois, les composantes existentielles de $\dot{\mathbf{x}}$ correspondent aux variables quantifiées existentiellement dans l'interprétation offerte par le théorème sémantique pour f^{**} . Ainsi, dans les deux théorèmes précédents, ce sont les multi-occurrences des variables quantifiées existentiellement dans l'interprétation qui empêchent l'interprétation des extensions rationnelles.*

Exemple II.34. *Soit $f(x) = x_1(x_2 - x_1)$ et $\dot{\mathbf{x}}_1 = ([2, 4], \forall)$ et $\dot{\mathbf{x}}_2 = ([1, 2], \exists)$ ce qui donne*

$$fR^{**}(\dot{\mathbf{x}}) = \dot{\mathbf{x}}_1 \times^{**} (\dot{\mathbf{x}}_2 -^{**} \dot{\mathbf{x}}_1) = ([-8, -2], \forall)$$

*L'intervalle existentiel $\dot{\mathbf{x}}_2$ a une unique occurrence dans l'expression de f et le théorème II.6 s'applique montrant que $fR^{**}(\dot{\mathbf{x}}) \subseteq f^{**}(\dot{\mathbf{x}})$. Ainsi le théorème sémantique II.4 s'applique et la proposition quantifiée suivante est vraie :*

$$(\forall x_1 \in [2, 4]) (\exists z \in [-8, -2]) (\exists x_2 \in [1, 2]) (f(x) = z)$$

Remarque. Avec des calculs différents et l'utilisation de théorèmes différents, les deux exemples II.33 et II.34 valident finalement la même proposition quantifiée.

Exemple II.35. Soit $f(x) = x_1(x_2 - x_1)$ et $\dot{\mathbf{x}}_1 = ([2, 4], \exists)$ et $\dot{\mathbf{x}}_2 = ([1, 2], \forall)$. Cette fois, $fR(\dot{\mathbf{x}})$ n'est pas interprétable grâce au théorème sémantique II.4 car $\dot{\mathbf{x}}_1$ est existentiel et a deux occurrences dans l'expression de f . Il faudra utiliser le théorème II.8 pour obtenir un intervalle modal interprétable, cf. Exemple II.37.

4.2.4 L'extension rationnelle

En général, une fonction f est donnée par une expression utilisant les opérations $+$, $-$, \times et $/$ et des fonctions élémentaires à une variable. Toutes ces fonctions étant "JM-commutables", les deux extensions rationnelles fR^* et fR^{**} coïncident. Ainsi, ces fonctions sémantiques peuvent être notées sans référence au type d'extension, i.e. $\circ := \circ^* = \circ^{**}$. Alors, l'extension rationnelle fR est définie par

$$fR(\dot{\mathbf{x}}) = fR^*(\dot{\mathbf{x}}) = fR^{**}(\dot{\mathbf{x}})$$

Ainsi, si toutes les variables ont une unique occurrence dans l'expression de f , les deux théorèmes II.5 et II.6 s'appliquent et les deux inclusions suivantes sont vérifiées :

$$f^*(\dot{\mathbf{x}}) \subseteq fR(\dot{\mathbf{x}}) \subseteq f^{**}(\dot{\mathbf{x}})$$

Si de plus f est "JM-commutable" pour $\dot{\mathbf{x}}$ (ce qui n'est pas nécessairement vrai même si f est composée de fonctions "JM-commutable", voir par l'exemple II.6.3 de SIGLA/X[37]) alors

$$f^*(\dot{\mathbf{x}}) = fR(\dot{\mathbf{x}}) = f^{**}(\dot{\mathbf{x}}) \quad (4.3)$$

Dans les cas non traités par les théorèmes II.5 et II.6 (par exemple les exemples II.32 et II.35) les deux théorèmes suivants permettent de modifier le calcul de fR pour le rendre interprétable.

Théorème II.7 (SIGLA/X[37], Thm II.6.7). Soit la fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par une composition de fonctions continues et $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$. $fR(\dot{\mathbf{x}}T^*)$ est obtenu en modifiant le calcul de $fR(\dot{\mathbf{x}})$ de la manière suivante : pour chaque composante universelle de $\dot{\mathbf{x}}$, toutes les occurrences sauf une sont dualisées. Alors $f^*(\dot{\mathbf{x}}) \subseteq fR(\dot{\mathbf{x}}T^*)$, et ainsi le théorème sémantique pour f^* s'applique.

Remarque. La preuve de ce théorème proposée dans SIGLA/X[37] est fautive (Goldsztejn[3]) et les auteurs de la théorie des intervalles modaux ont confirmé que la preuve proposée dans SIGLA/X[37] devait être corrigée. A ce jour, aucune preuve alternative n'a été publiée.

L'exemple suivant reprend l'exemple II.32 qui ne pouvait pas être interprété directement par le théorème II.5.

Exemple II.36. Soit $f(x) = x_1(x_2 - x_1)$ et $\dot{\mathbf{x}}_1 = ([2, 4], \forall)$ et $\dot{\mathbf{x}}_2 = ([1, 2], \exists)$. L'intervalle universel $\dot{\mathbf{x}}_1$ a deux occurrences dans l'expression de f et le théorème II.7 doit être utilisé à la place du théorème II.5. $fR^*(\dot{\mathbf{x}}T^*)$ peut être calculée de deux manières (en fonction de l'occurrence de $\dot{\mathbf{x}}_1$ qui est dualisée) :

$$- fR(\dot{\mathbf{x}}T^*) = (\text{Dual } \dot{\mathbf{x}}_1)(\dot{\mathbf{x}}_2 - \dot{\mathbf{x}}_1) = -4;$$

$$- fR(\dot{\mathbf{x}}T^*) = \dot{\mathbf{x}}_1(\dot{\mathbf{x}}_2 - \text{Dual}(\dot{\mathbf{x}}_1)) = ([-6, 0], \exists).$$

Dans les deux cas, le théorème II.7 montre que $f^*(\dot{\mathbf{x}}) \subseteq fR(\dot{\mathbf{x}}T^*)$ et ainsi que les deux propositions quantifiées suivantes sont vraies :

$$(\forall x_2 \in [1, 2])(\exists x_1 \in [2, 4])(f(x) = -4)$$

$$(\forall x_2 \in [1, 2])(\exists z \in [-6, 0])(\exists x_1 \in [2, 4])(f(x) = z)$$

La première apporte plus d'informations et peut ainsi être considérée comme meilleure.

Le théorème suivant correspond au théorème précédent mais appliqué à f^{**} .

Théorème II.8 (SIGLA/X[37], Thm II.6.8). *Soit la fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par une composition de fonctions continues et $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$. $fR(\dot{\mathbf{x}}T^{**})$ est obtenu en modifiant le calcul de $fR(\dot{\mathbf{x}})$ de la manière suivante : pour chaque composante existentielle de $\dot{\mathbf{x}}$, toutes les occurrences sauf une sont dualisées. Alors $fR(\dot{\mathbf{x}}T^{**}) \subseteq f^{**}(\dot{\mathbf{x}})$, et ainsi le théorème sémantique pour f^{**} s'applique.*

Remarque. *La preuve de ce théorème se base sur la preuve du précédent et demande donc d'être aussi corrigée.*

L'exemple suivant reprend l'exemple II.35 qui ne pouvait pas être interprété directement par le théorème II.6.

Exemple II.37. *Soit $f(x) = x_1(x_2 - x_1)$ et $\dot{\mathbf{x}}_1 = ([2, 4], \exists)$ et $\dot{\mathbf{x}}_2 = ([1, 2], \forall)$. L'intervalle existentiel $\dot{\mathbf{x}}_1$ a deux occurrences dans l'expression de f et le théorème II.8 doit être utilisé à la place du théorème II.6. $fR^*(\dot{\mathbf{x}}T^{**})$ peut être calculé de deux manières :*

$$- fR(\dot{\mathbf{x}}T^{**}) = (\text{Dual } \dot{\mathbf{x}}_1)(\dot{\mathbf{x}}_2 - \dot{\mathbf{x}}_1) = -4;$$

$$- fR(\dot{\mathbf{x}}T^{**}) = \dot{\mathbf{x}}_1(\dot{\mathbf{x}}_2 - \text{Dual}(\dot{\mathbf{x}}_1)) = ([-6, 0], \forall).$$

Dans les deux cas, le théorème II.7 montre que $fR(\dot{\mathbf{x}}T^{**}) \subseteq f^{**}(\dot{\mathbf{x}})$ et ainsi que les deux propositions quantifiées suivantes sont vraie :

$$(\forall x_2 \in [1, 2])(\exists x_1 \in [2, 4])(f(x) = -4)$$

$$(\forall x_2 \in [1, 2])(\exists z \in [-6, 0])(\exists x_1 \in [2, 4])(f(x) = z)$$

La première apporte plus d'informations et peut ainsi être considérée comme meilleure.

Remarque. *Encore une fois, avec des calculs différents et l'utilisation de théorèmes différents, les deux exemples II.36 et II.37 valident finalement les mêmes propositions quantifiées.*

4.2.5 "Coercion to optimality"

La définition de l'optimalité est différente de celle adoptée pour les intervalles classiques.

Définition II.11 (SIGLA/X[37], Def III.2.1). *Une expression d'intervalles modaux fR est optimale si pour tout $\dot{\mathbf{x}} \in I^*(\mathbb{R}^n)$ pour lequel $fR(\dot{\mathbf{x}})$ est définie, les conditions suivantes sont vérifiées :*

1. $f^*(\dot{\mathbf{x}}) = f^{**}(\dot{\mathbf{x}})$

2. $fR(\dot{\mathbf{x}}) = f^*(\dot{\mathbf{x}})$

Remarque. Si la seconde condition est intuitive, la première l'est moins. La justification de cette définition de l'optimalité n'est pas donnée dans la théorie des intervalles modaux et nous n'avons pas trouver d'explication. Nous proposerons une définition de l'optimalité qui sera plus intuitive et plus représentative de la qualité d'une extension.

Avec cette définition de l'optimalité, $f^* = f^{**}$ est l'extension optimale des fonction $f \in \{+, -, \times, /\}$ ainsi que de toute fonction à une variable. Dans le cas des extensions aux intervalles classiques, il suffit qu'une expression contienne une unique occurrence de chaque variable pour que l'extension naturelle correspondante soit optimale. Une condition supplémentaire sur la structure de l'expression de la fonction doit être ajoutée pour obtenir l'optimalité dans le contexte des intervalles modaux : la "tree-optimality" (voir SIGLA/X[37][38]). Ainsi, si une expression est "tree-optimal" et contient une unique occurrence de chaque variable alors $fR(\dot{\mathbf{x}})$ est optimale. Finalement, l'étude de la monotonie de la fonction considérée permet d'améliorer le calcul de $fR(\dot{\mathbf{x}})$ grâce aux théorèmes "*-partially optimal coercion", "**-partially optimal coercion" (et d'autre théorèmes, voir SIGLA/X[37][38] pour les énoncés de ces théorèmes). L'exemple suivant reprend l'exemple II.36 en améliorant le résultat obtenu par l'utilisation du théorème de "*-partially optimal coercion".

Exemple II.38. Soit $f(x) = x_1(x_2 - x_1)$ et $\dot{\mathbf{x}}_1 = ([2, 4], \forall)$ et $\dot{\mathbf{x}}_2 = ([1, 2], \exists)$. La fonction f est totalement monotone par rapport à x_1 , i.e. elle est décroissante quelque soit la valeur de $x_2 \in [1, 2]$. Ainsi le calcul suivant

$$fR^*(\dot{\mathbf{x}}DT^*) = \dot{\mathbf{x}}_1(\dot{\mathbf{x}}_2 - \dot{\mathbf{x}}_1) = ([-8, -2], \forall)$$

est interprétable, i.e. le théorème "*-partially optimal coercion" montre que $f^*(\dot{\mathbf{x}}) \subseteq fR^*(\dot{\mathbf{x}}DT^*)$. Finalement la proposition suivante est validée par le théorème sémantique pour f^* :

$$(\forall x_2 \in [1, 2])(\forall z \in [-8, -2])(\exists x_1 \in [2, 4])(f(x) = z).$$

Cette proposition quantifiée est meilleur que celle obtenue dans l'exemple II.36. Un calcul direct de $f^*(\dot{\mathbf{x}})$ montre que, dans ce cas, $f^*(\dot{\mathbf{x}}) = fR^*(\dot{\mathbf{x}}DT^*)$.

4.2.6 Fonctions à valeur vectorielle

Dans le cas des fonctions continues $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$, il n'y a pas de définition d'extension sémantique de f aux intervalles modaux. Mais les extensions des composantes f_i de f permettent de valider des propositions quantifiées en relation avec le prédicat $f(x) = z$. En particulier, le théorème suivant, appelé "interpretability of n -dimensional computations", étend le théorème "coercion to *-interpretability" aux fonctions vectorielles.

Remarque. L'énoncé du prochain théorème est présenté en anglais et en utilisant les notations proposées dans SIGLA/X[37]. Sa formulation proposée dans SIGLA/X[37] manque de précision et sa traduction risquerait de changer le sens de l'énoncé du théorème.

Théorème II.9 (SIGLA/X[37], Thm II.7.1). Let $X \in I^*(\mathbb{R}^m)$ and $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ defined by $f(x) := (f_1(x), \dots, f_n(x))$ continuous in X' . Let $(f_1R(A_1), \dots, f_nR(A_n)) \subseteq (B_1, \dots, B_n)$ be a system of interpretable outer-rounded computations of $(f_1^*(X_1), \dots, f_n^*(X_n))$ where (X_1, \dots, X_n) are defined from X so to have the same proper components of X , and with each improper component of X transformed into its dual in each X_j excepting only one. In this case,

$$U(x_P, X'_P)Q^*(z, B)E(x_I, X'_I), z = f(x_P, x_I)$$

with $X = (X_P, X_I)$ and $Q^*(z, B)$ being the sequence of the prefixes $Q(z_1, B_1), \dots, Q(z_n, B_n)$ with the ones corresponding to the universal delimiters heading the sequence.

Remarque. L'intervalle classique X' , défini à partir de l'intervalle modal X , est $\text{Set}(X)$. Les notations $U(x, X')$, $E(x, X')$ et $Q(x, X)$ correspondent respectivement à $\forall x \in X'$, $\exists x \in X'$ et $Q^{(\dot{z})} Xx \in X'$.

Un théorème similaire est obtenu en utilisant f^{**} au lieu de f^* . L'exemple suivant est repris de SIGLA/X[37].

Exemple II.39 (SIGLA/X[37], Ex II.7.1). Soient $f(x) = (x_1 + x_2 + x_3, x_1x_2 - x_1)^T$ et $\dot{\mathbf{x}}_1 = ([1, 6], \forall)$, $\dot{\mathbf{x}}_2 = ([2, 6], \forall)$ et $\dot{\mathbf{x}}_3 = ([-3, 2], \exists)$. Alors, les calculs suivants sont effectués :

$$\begin{aligned} f_1R(\dot{\mathbf{a}}_1) &= \text{Dual}(\dot{\mathbf{x}}_1) + \dot{\mathbf{x}}_2 + \dot{\mathbf{x}}_3 = ([4, 10], \exists) \\ f_2R(\dot{\mathbf{a}}_2) &= \dot{\mathbf{x}}_1 \times \text{Dual}(\dot{\mathbf{x}}_2) - \dot{\mathbf{x}}_3 = ([5, 6], \forall). \end{aligned}$$

Ainsi, le premier calcul valide la proposition quantifiée

$$(\forall x_3 \in [-3, 2]) (\forall x_1 \in [1, 6]) (\exists z_1 \in [4, 10]) (\exists x_2 \in [2, 6]) (f_1(x) = z_1)$$

et le deuxième calcul la proposition quantifiée

$$(\forall x_3 \in [-3, 2]) (\forall x_2 \in [2, 6]) (\forall z_2 \in [5, 6]) (\exists x_1 \in [1, 6]) (f_2(x) = z_2).$$

Le théorème précédent valide finalement la proposition quantifiée

$$(\forall x_3 \in [-3, 2]) (\forall z_2 \in [5, 6]) (\exists z_1 \in [4, 10]) (\exists x_1 \in [1, 6]) (\exists x_2 \in [2, 6]) (f(x) = z).$$

4.3 Lien entre les intervalles modaux et intervalles généralisés

Les intervalles modaux sont isomorphes avec les intervalles généralisés. Cet isomorphisme occupe une place importante dans la théorie des intervalles modaux où les intervalles généralisés sont introduits comme une notation pour les intervalles modaux (appelée "canonical notation" dans SIGLA/X[37]) :

$$[a, b] := \begin{cases} ([a, b], \exists) & \text{si } a \leq b \\ ([b, a], \forall) & \text{si } a \geq b \end{cases}$$

Cet isomorphisme est compatible avec l'inclusion des deux ensembles. D'une part, cela donne une interprétation à l'inclusion des intervalles généralisés : par exemple, $[1, -1] \subseteq [0, 2]$ peut être interprété par $[-1, 1] \cap [0, 2] \neq \emptyset$. D'autre part, cela permet de calculer aisément les opérations d'intersection et de réunion d'intervalles modaux. Par exemple, les expressions respectives des fonctions sémantiques donnent

$$\begin{aligned} f^*(\mathbf{x}) &= \bigvee_{x_P \in \mathbf{x}_P} \bigwedge_{x_I \in (\text{pro } \mathbf{x}_I)} f(x) \\ &= \left[\min_{x_P \in \mathbf{x}_P} \max_{x_I \in (\text{pro } \mathbf{x}_I)} f(x), \max_{x_P \in \mathbf{x}_P} \min_{x_I \in (\text{pro } \mathbf{x}_I)} f(x) \right] \end{aligned}$$

et

$$\begin{aligned} f^{**}(\mathbf{x}) &= \bigwedge_{x_I \in (\text{pro } \mathbf{x}_I)} \bigvee_{x_P \in \mathbf{x}_P} f(x) \\ &= \left[\max_{x_I \in (\text{pro } \mathbf{x}_I)} \min_{x_P \in \mathbf{x}_P} f(x), \min_{x_I \in (\text{pro } \mathbf{x}_I)} \max_{x_P \in \mathbf{x}_P} f(x) \right]. \end{aligned}$$

Il est remarquable qu'aucune des deux fonctions sémantiques modales ne coïncident en général avec les extensions introduites par Kaucher, sauf dans le cas des fonctions réelles à une variable où une seule opération (intersection ou réunion) est effectuée. Dans le cas des fonctions à plusieurs variables, les opérations intersections et réunions sont en général effectuées sur les mêmes variables mais calculées dans un ordre différent. Dans le cas particulier des fonctions à deux variables, les égalités suivantes sont vérifiées par définitions des différentes extensions mises en jeu :

- si $\mathbf{x}, \mathbf{y} \in \mathbb{IR}$ ou $\mathbf{x}, \mathbf{y} \in \overline{\mathbb{IR}}$ alors $f^*(\mathbf{x}, \mathbf{y}) = f^{**}(\mathbf{x}, \mathbf{y}) = f^K(\mathbf{x}, \mathbf{y})$;
- si $\mathbf{x} \in \mathbb{IR}$ et $\mathbf{y} \in \overline{\mathbb{IR}}$ alors $f^*(\mathbf{x}, \mathbf{y}) = f^K(\mathbf{x}, \mathbf{y})$;
- si $\mathbf{x} \in \overline{\mathbb{IR}}$ et $\mathbf{y} \in \mathbb{IR}$ alors $f^{**}(\mathbf{x}, \mathbf{y}) = f^K(\mathbf{x}, \mathbf{y})$.

Ainsi, dans le cas des fonctions à deux variables, lorsque $f^* = f^{**}$, i.e. lorsque f est "JM-commutable", alors $f^*(\mathbf{x}, \mathbf{y}) = f^{**}(\mathbf{x}, \mathbf{y}) = f^K(\mathbf{x}, \mathbf{y})$ quel que soient $\mathbf{x}, \mathbf{y} \in \mathbb{KR}$. L'arithmétique de Kaucher et l'arithmétique modale coïncident donc. Comme dans le cas de l'inclusion, cette coïncidence permet d'une part de donner une interprétation à l'arithmétique de Kaucher et d'autre part de donner une formulation utile pour l'arithmétique des intervalles modaux.

4.4 Conclusion

La théorie des intervalle modaux présente plusieurs contributions importantes : d'une part, elle a enrichi les interprétations offertes par les extensions aux intervalles classiques tout en montrant qu'une extension naturelle modifiée (l'extension rationnelle aux intervalles modaux) peut être utilisée avec ces interprétations enrichies. D'autre part, une procédure d'arrondi simple et efficace permet de conserver le caractère garanti des calculs effectués, sur le modèle de la théorie des intervalles classiques. Si des applications prometteuses de ces interprétations enrichies ont été mises en valeur, il reste du travail à effectuer avant de les appliquer à des situations réalistes, cf. Herrero[69] où les intervalles modaux sont appliqués à la résolution d'un problème pratique, mais seulement après simplification de ce problème. De nouveaux développements doivent être faits pour avancer dans cette direction. Pour cela, nous proposons dans la Partie IV une nouvelle formulation de la théorie des intervalles modaux, qui tout en conservant les aspects positifs de la théorie, la simplifie et permet donc d'introduire de nouveaux concepts plus aisément. En particulier, une extension de la valeur moyenne aux intervalles généralisée sera proposée.

Chapitre 5

Systemes d'intervalles

De nombreuses situations réelles peuvent être modéliser par une fonction $f(a, x) = b$ où a sont les paramètres du système, x les variables et b les sorties. Associer des intervalles aux paramètres et aux sorties permet de modéliser des incertitudes, eg. incertitudes de mesure ou valeurs admissibles des commandes du système. Ainsi, un système d'intervalles est défini dans Shary[96] par les données suivantes :

- une fonction $f : \mathbb{R}^p \times \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Dans l'expression $f(a, x) = b$, x sont les variables, a les paramètres et b les sorties du système ;
- un vecteur d'intervalles $\mathbf{a} \in \mathbb{IR}^p$ dans lequel les paramètres a sont contraints à être ;
- un vecteur d'intervalles $\mathbf{b} \in \mathbb{IR}^m$ dans lequel les sorties b sont contraintes à être.

La fonction f sera considérée continue. De plus, le domaine de définition de f peut être un sous-ensemble de $\mathbb{R}^p \times \mathbb{R}^n$ mais les fonctions considérées seront définies dans $\mathbb{R}^p \times \mathbb{R}^n$, ce qui simplifiera l'exposé. Dans le cas des systèmes linéaires, les paramètres seront présentés sous la forme d'une matrice $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ et les intervalles liés à ces paramètres sous la forme d'une matrice d'intervalles $\mathbf{A} = (\mathbf{a}_{ij}) \in \mathbb{IR}^{m \times n}$. Ce chapitre présente les différents ensembles de solutions définis pour les systèmes d'intervalles et les techniques d'approximation de ces ensembles de solutions.

L'utilisation des intervalles classiques pour l'étude de ces systèmes d'intervalles est tout d'abord présentée. Ensuite, nous verrons que les intervalles généralisés et les intervalles modaux permettent de prendre en compte des quantifications plus riches des paramètres et sorties du système d'intervalle.

5.1 "United Solution set"

5.1.1 Définition

Le "united solution set" d'un système d'intervalle est l'ensemble des solutions du système pour une valeur quelconque des paramètres a et des sorties b dans leurs intervalles respectifs. Il est défini formellement de la manière suivante :

Définition II.12. Soit un système d'intervalle défini par f , \mathbf{a} et \mathbf{b} . Son "united solution set" est noté $\Sigma(f, \mathbf{a}, \mathbf{b})$ et est défini par

$$\Sigma(f, \mathbf{a}, \mathbf{b}) = \{x \in \mathbb{R}^n | (\exists a \in \mathbf{a}) (\exists b \in \mathbf{b}) (f(a, x) = b)\}.$$

Lorsque des systèmes linéaires sont considérés, la notation suivante sera utilisée :

Notation. Soient $\mathbf{A} \in \mathbb{IR}^{n \times m}$ et $\mathbf{b} \in \mathbb{IR}^m$, le "united solution set" correspondant est noté

$$\Sigma(\mathbf{A}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\exists A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b)\}$$

Les techniques dédiées à l'approximation extérieure des "united solution sets" sont maintenant présentées. L'approximation intérieure des "united solution sets" n'est généralement pas considérée dans le cadre de la théorie des intervalles classiques.

5.1.2 Approximation extérieure des système d'intervalles linéaires

Parmi d'autres, nous présentons maintenant deux opérateurs qui permettent de réduire une approximation extérieure initiale d'un "united-solution set" linéaire lorsque $m = n$. Ainsi, partant d'une approximation initiale $\mathbf{x}^{(0)}$ du "united solution set", une suite d'approximations extérieures $\mathbf{x}^{(k)}$ peut-être construite en appliquant itérativement les opérateurs présentés. L'itération peut être arrêtée lorsque l'amélioration entre deux itérations est insuffisante.

Un système d'intervalles linéaire peut être préconditionné de la manière suivante : soit $C \in \mathbb{R}^{n \times n}$ une matrice régulière, alors l'inclusion suivante est vérifiée :

$$\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \Sigma(C\mathbf{A}, C\mathbf{b})$$

Ainsi, une approximation extérieure du "united-solution set" préconditionné $\Sigma(C\mathbf{A}, C\mathbf{b})$ sera une approximation extérieure du "united solution set" original $\Sigma(\mathbf{A}, \mathbf{b})$. Une matrice de préconditionnement fréquemment utilisée est l'inverse du milieu de la matrice d'intervalles \mathbf{A} , i.e. $C = (\text{mid } \mathbf{A})^{-1}$. En général, un préconditionnement efficace améliore le comportement des méthodes de résolution utilisées (voir par exemple Neumaier[8], Kearfott[81] et Hansen[18]).

Opérateur de Krawczyk pour systèmes linéaires

L'opérateur de Krawczyk pour systèmes d'intervalles linéaires est défini de la manière suivante :

$$K^l(\mathbf{A}, \mathbf{b}, \mathbf{x}) = (\mathbf{b} - (\mathbf{A} - I)\mathbf{x}) \bigcap \mathbf{x}$$

Dans sa version préconditionnée, il devient $K^l(C\mathbf{A}, C\mathbf{b}, \mathbf{x})$. L'opérateur de Krawczyk pour système d'intervalles linéaire vérifie

$$\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \mathbf{x} \implies \Sigma(\mathbf{A}, \mathbf{b}) \subseteq K^l(\mathbf{A}, \mathbf{b}, \mathbf{x})$$

De nombreuses propriétés supplémentaires (eg. vitesse de convergence, qualité de la limite du procédé itératif) de cet opérateur sont décrites dans Neumaier[8].

Opérateur de Gauss-Seidel

L'opérateur de Gauss-Seidel dans sa version actuelle fut proposé dans Hansen[19]. Il donne toujours de meilleurs résultats que l'opérateur de Krawczyk et il lui sera donc préféré. L'opérateur de Gauss-Seidel est d'abord défini dans le cas $n = 1$: soit $\mathbf{a} = [\underline{\mathbf{a}}, \bar{\mathbf{a}}] \in \mathbb{IR}$ et $\mathbf{b} = [\underline{\mathbf{b}}, \bar{\mathbf{b}}] \in \mathbb{IR}$. Le "united solution set" correspondant est

$$\Sigma(\mathbf{a}, \mathbf{b}) = \{x \in \mathbb{R} \mid (\exists a \in \mathbf{a})(\exists b \in \mathbf{b})(ax = b)\}$$

Alors, l'opérateur de Gauss-Seidel est défini de la manière suivante :

$$\Gamma(\mathbf{a}, \mathbf{b}, \mathbf{x}) = \square(\Sigma(\mathbf{a}, \mathbf{b}) \bigcap \mathbf{x}) \quad (5.1)$$

L'expression suivante pour l'opérateur peut-être trouvée dans Neumaier[8] :

$$\Gamma(\mathbf{a}, \mathbf{b}, \mathbf{x}) = \begin{cases} \mathbf{b}/\mathbf{a} \bigcap \mathbf{x} & \text{si } 0 \notin \mathbf{a} \\ \square(\mathbf{x} \setminus \mathbf{b}/\underline{\mathbf{a}}, \mathbf{b}/\bar{\mathbf{a}}) & \text{si } \mathbf{b} > 0 \wedge 0 \in \mathbf{a} \\ \square(\mathbf{x} \setminus \mathbf{b}/\bar{\mathbf{a}}, \mathbf{b}/\underline{\mathbf{a}}) & \text{si } \mathbf{b} < 0 \wedge 0 \in \mathbf{a} \\ \mathbf{x} & \text{si } 0 \in \mathbf{a} \wedge 0 \in \mathbf{b} \end{cases}$$

L'opérateur de Gauss-Seidel est alors défini de la manière suivante dans le cas général $n > 1$:

$$\Gamma(\mathbf{A}, \mathbf{b}, \mathbf{x}) = \mathbf{y} \quad \text{avec} \quad \mathbf{y}_i = \Gamma(\mathbf{A}_{ii}, \mathbf{b}_i - \sum_{k \neq i} \mathbf{A}_{ik} \mathbf{x}_k, \mathbf{x}_i)$$

Remarque. L'opérateur de Gauss-Seidel est généralement défini avec une optimisation des calculs effectués : en effet, lorsque l'on calcule \mathbf{y}_i , les calculs de \mathbf{y}_j avec $j < i$ ont déjà été effectués et ils peuvent être utilisés à la place des \mathbf{x}_j , ce qui accélère la convergence de la suite $\mathbf{x}^{(k+1)} = \Gamma(\mathbf{A}, \mathbf{b}, \mathbf{x}^{(k)})$ sans aucun désavantage.

L'opérateur de Gauss-Seidel pour système d'intervalles linéaire vérifie

$$\Sigma(\mathbf{A}, \mathbf{b}) \bigcap \mathbf{x} \subseteq \Gamma(\mathbf{A}, \mathbf{b}, \mathbf{x})$$

De nombreuses propriétés supplémentaires (eg. vitesse de convergence, qualité de la limite du procédé itératif, comparaison avec l'opérateur de Krawczyk) de cet opérateur sont décrites dans Neumaier[8].

Opérateur de Gauss-Seidel simplifié

C'est l'opérateur simplifié de Gauss-Seidel qui fut proposé avant sa version actuelle dans Ris[29]. Cette version est présentée ici car c'est elle qui sera généralisée à des quantifications plus riches des paramètres dans la prochaine section. Il est défini en restreignant la définition 5.1 au cas où $0 \notin \mathbf{a}$: soient $\mathbf{A} \in \mathbb{R}^{n \times n}$ et $\mathbf{b} \in \mathbb{R}^n$, l'opérateur simplifié de Gauss-Seidel est noté $\Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x})$ et défini si $0 \notin A_{ii}$ pour tout $i \in [1..n]$. Il peut être exprimé en utilisant la notation matricielle suivante :

$$\Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x}) = \mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})\mathbf{x}) \bigcap \mathbf{x},$$

où $\mathbf{D} = \text{diag}(1/\mathbf{A}_{11}, \dots, 1/\mathbf{A}_{nn})$ et \mathbf{L} et \mathbf{U} sont respectivement la partie inférieure hors diagonale ($\mathbf{L}_{ij} = \mathbf{A}_{ij}$ si $j < i$ et $\mathbf{L}_{ij} = 0$ sinon) et la partie supérieure hors diagonale ($\mathbf{U}_{ij} = \mathbf{A}_{ij}$ si $j > i$ et $\mathbf{U}_{ij} = 0$ sinon) de la matrice \mathbf{A} .

5.1.3 Système d'intervalles non linéaires

Les opérateurs définis dans la sous-section précédentes sont généralisés pour des "united solution sets" non linéaires dans le cas particulier où $\mathbf{a} = a \in \mathbb{R}^p$ est un intervalle dégénéré, i.e. il n'y a pas d'incertitude sur les paramètres, et $\mathbf{b} = 0$. C'est à dire lorsque $\Sigma(f, \mathbf{a}, \mathbf{b})$ est l'ensemble des solutions de l'équation $f(a, x) = 0$. Dans ce cas, l'équation

$f(a, x) = 0$ est notée $f(x) = 0$ en abandonnant la référence aux paramètres et $\Sigma(f, \mathbf{a}, \mathbf{b})$ est noté

$$\Sigma(f, 0) = \{x \in \mathbb{R}^n \mid f(x) = 0\}.$$

L'équation $f(x) = 0$ est étudiée dans le cas particulier $m = n$, c'est à dire lorsqu'il y a autant d'équations que de variables. L'approximation de $\Sigma(f, 0)$ est faite localement, c'est à dire que $\Sigma(f, 0) \cap \mathbf{x}$ est considéré, où $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$ est un intervalle donné. Les opérateurs suivants permettent de borner $\Sigma(f, 0) \cap \mathbf{x}$, et ainsi de réduire l'intervalle \mathbf{x} sans perdre de solutions de l'équation $f(x) = 0$. De plus, ils permettent aussi de prouver l'existence d'une solution dans la boîte considérée. Seules les définitions et interprétations des opérateurs sont décrites dans les sous-sections suivantes. De nombreuses propriétés complémentaires (eg. vitesse de convergence, qualité de la limite du procédé itératif, comparaisons entre les différents opérateurs) sont décrites dans Neumaier[8].

L'opérateur de Newton par intervalle

Le "united solution set" $\Sigma(f, 0)$ est localement linéarisé dans \mathbf{x} en utilisant le théorème de la valeur moyenne : étant donné $\tilde{x} \in \mathbf{x}$,

$$\Sigma(f, 0) \cap \mathbf{x} \subseteq \{x \in \mathbb{R}^n \mid (\exists A \in \mathbf{A})(f(\tilde{x}) + A(x - \tilde{x}) = 0)\},$$

où \mathbf{A} est une matrice d'intervalles qui vérifie $\mathbf{A} \supseteq \{f'(x) \mid x \in \mathbf{x}\}$. Le changement de variable $y = x - \tilde{x}$ est alors effectué pour obtenir l'inclusion équivalente suivante :

$$\Sigma(f, 0) \cap \mathbf{x} \subseteq \{y + \tilde{x} \in \mathbb{R}^n \mid (\exists A \in \mathbf{A})(Ay = -f(\tilde{x}))\}.$$

Ainsi, si \mathbf{y} est une approximation extérieure du "united solution set" linéaire

$$\Sigma(\mathbf{A}, -f(\tilde{x})) \tag{5.2}$$

alors $\mathbf{y} + \tilde{x}$ est une approximation extérieure de $\Sigma(f, 0) \cap \mathbf{x}$. Un opérateur de Newton par intervalle est alors défini de la manière suivante :

$$N(f, \mathbf{x}, \tilde{x}) = \tilde{x} + \mathbf{y} \text{ avec } \mathbf{y} \supseteq \Sigma(\mathbf{A}, -f(\tilde{x})) \text{ et } \mathbf{A} \supseteq \{f'(x) \mid x \in \mathbf{x}\}$$

Les opérateurs de Newton vérifient donc

$$\Sigma(f, 0) \cap \mathbf{x} \subseteq N(f, \mathbf{x}, \tilde{x})$$

L'opérateur de Newton est généralement couplé avec un processus de préconditionnement : l'équation (5.2) est alors remplacé par $\Sigma(C\mathbf{A}, -Cf(\tilde{x}))$ où $C \in \mathbb{R}^{n \times n}$ est régulière. Les deux opérateurs suivants sont des cas particuliers de l'opérateur de Newton.

Opérateur de Krawczyk pour systèmes non-linéaires

L'opérateur de Krawczyk pour systèmes non-linéaires, noté $K(f, \mathbf{x}, \tilde{x})$, a été introduit dans Krawczyk[72]. Il est obtenu en utilisant l'opérateur de Krawczyk dédié aux systèmes linéaires pour construire une approximation extérieure du "united solution set" linéaire (5.2). Pour cela, il est nécessaire d'avoir un intervalle initial pour l'approximation de (5.2). Un tel intervalle est disponible car seul $\Sigma(f, 0) \cap \mathbf{x}$ est approximé. Ainsi l'intervalle $\mathbf{x} - \tilde{x}$ peut-être utilisé :

$$K(f, \mathbf{x}, \tilde{x}) = \tilde{x} + K^l(C\mathbf{A}, -Cf(\tilde{x}), \mathbf{x} - \tilde{x}).$$

L'expression explicite de l'opérateur de Krawczyk est

$$\begin{aligned} K(f, \mathbf{x}, \tilde{x}) &= \tilde{x} + (-Cf(\tilde{x}) - (C\mathbf{A} - I)(\mathbf{x} - \tilde{x})) \bigcap (\mathbf{x} - \tilde{x}) \\ &= (\tilde{x} - Cf(\tilde{x}) - (C\mathbf{A} - I)(\mathbf{x} - \tilde{x})) \bigcap \mathbf{x}. \end{aligned}$$

Remarque. Une autre expression est proposée dans *Bliek[13]* :

$$K(f, \mathbf{x}, \tilde{x}) = \tilde{x} - K^l(C\mathbf{A}, Cf(\tilde{x}), \mathbf{x} - \tilde{x}).$$

Ces deux expressions coïncident lorsque $\tilde{x} = \text{mid } \mathbf{x}$. L'expression originale proposée dans *Krawczyk[72]* est l'expression explicite donnée ci-dessus.

Opérateur de Hansen-Sengupta

L'opérateur de Hansen-Sengupta a été proposé dans *Hansen[19]*. Il est noté $H(f, \mathbf{x}, \tilde{x})$ et est obtenu en utilisant l'opérateur de Gauss-Seidel pour construire une approximation extérieure du "united solution set" (5.2). Encore une fois, l'intervalle $\mathbf{x} - \tilde{x}$ est utilisé comme intervalle initial :

$$H(f, \mathbf{x}, \tilde{x}) = \tilde{x} + \Gamma(C\mathbf{A}, -Cf(\tilde{x}), \mathbf{x} - \tilde{x}),$$

Preuve d'existence de solution d'un système d'équation

L'opérateur de Newton, en plus de permettre de borner les solutions de l'équation $f(x) = 0$, peut aussi permettre de prouver l'existence de solution pour cette équation :

$$N(f, \mathbf{x}, \tilde{x}) \subseteq \text{int } \mathbf{x} \implies (\exists x \in \mathbf{x})(f(x) = 0)$$

De plus, l'inclusion $N(f, \mathbf{x}, \tilde{x}) \subseteq \text{int } \mathbf{x}$ prouve aussi qu'il existe une *unique* solution dans la boîte \mathbf{x} . Ceci s'applique donc en particuliers aux opérateurs de Krawczyk et de Hansen-Sengupta. La preuve d'existence de solution nécessite généralement un préconditionné avec $C \approx (\text{mid } \mathbf{A})^{-1}$.

Les ensembles de solutions présentés jusqu'à présent ne sont définis qu'en utilisant des quantificateurs existentiels. Cependant, les applications rencontrées peuvent nécessiter une modélisation plus fine par différents quantificateurs, eg. un intervalle peut représenter une incertitude de mesure ou un domaine de valeurs admissibles pour une commande. Ces deux types d'intervalles nécessitent alors des quantifications différentes. Les "AE-solution sets" sont maintenant introduits pour offrir une plus grande liberté de modélisation.

5.2 "AE-Solution set"

Les "AE-solution sets" généralisent les "united solution sets" en donnant plus de liberté pour la quantification des paramètres (voir *Shary[93][96]* et *Jirstrand[57]* pour des applications de l'approximation des "AE-solution sets"). Dans ce contexte, les intervalles généralisés remplacent les intervalles classiques : d'une part ils vont permettre de modéliser les quantificateurs associés à une variable grâce à leur qualité propre ou impropre. D'autre part, ils sont un outil d'analyse important pour l'approximation des "AE-solution sets".

Nous avons choisi une convention différente de celle choisie dans Shary[96] pour l'association d'un quantificateur à la qualité propre/impropre d'un intervalle généralisé, cf. Définition II.13. Le choix de cette nouvelle convention semble apporter de nombreux avantages. Le plus important étant une homogénéisation des notations et des expressions des algorithmes dédiés aux "united solution set" et aux "AE-solution sets".

5.2.1 Définition

Les "AE-solution sets" généralisent les "united solution sets" en autorisant une quantification arbitraire des paramètres du système, avec la contrainte que les quantificateurs universels précèdent toujours les quantificateurs existentiels (cette contrainte étant la source de la dénomination A(II)E(xist)). Ainsi, nous devons maintenant associer à chaque paramètre et à chaque sortie un domaine et un quantificateur (l'ordre d'apparition des quantificateurs étant alors fixé à la permutation près des quantificateurs de même type entre eux). Les intervalles généralisés sont donc très bien adaptés pour la modélisation de ces données : la qualité propre/impropre d'un intervalle généralisé est utilisée pour modéliser le type de quantificateur associé à une variable et la projection propre de l'intervalle généralisé représente le domaine associé à la variable. La convention "propre \leftrightarrow existentiel" et "impropre \leftrightarrow universel" est choisie pour généraliser les notations liées aux "united solution sets". Les "AE-solution sets" sont formellement définis de la manière suivante, en généralisant la définition et la notation associées au "united solution sets" :

Définition II.13 (Shary[96]). Soit $f : \mathbb{R}^p \times \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue, $\mathbf{a} \in \mathbb{K}\mathbb{R}^p$ et $\mathbf{b} \in \mathbb{K}\mathbb{R}^m$. Le "AE-solution set" correspondant est alors noté $\Sigma(f, \mathbf{a}, \mathbf{b})$ et est défini par

$$\Sigma(f, \mathbf{a}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\forall a_P \in \mathbf{a}_P) (\forall b_{P'} \in \mathbf{b}_{P'}) (\exists b_{I'} \in \mathbf{b}_{I'}) (\exists a_I \in \mathbf{a}_I), f(a, x) = b\}$$

où $P = P(\mathbf{a})$, $I = I(\mathbf{a})$, $P' = P(\mathbf{b})$ et $I' = I(\mathbf{b})$ sont les ensembles d'indices des composantes propres et impropres des intervalles \mathbf{a} et \mathbf{b} .

Remarque. Les conventions et notations ont été changées par rapport à celles adoptées dans Shary[96] : ces deux notations sont liées par la relation

$$\Sigma(f, \mathbf{a}, \mathbf{b}) = \Xi(f, \text{dual } \mathbf{a}, \mathbf{b}).$$

Cette nouvelle convention permet d'homogénéiser les écritures des "united solution sets" et des "AE-solution sets".

Ainsi, lorsque les intervalles mis en jeu sont propres, $\Sigma(f, \mathbf{a}, \mathbf{b})$ correspond au "united solution set" du système. Lorsqu'un intervalle impropre est utilisé à la place d'un intervalle propre, le quantificateur correspondant est inversé. Les cas particuliers suivants sont plus souvent rencontrés : étant donnés $\mathbf{a} \in \mathbb{I}\mathbb{R}^p$ et $\mathbf{b} \in \mathbb{I}\mathbb{R}^m$,

- "united solution set" : $\Sigma(f, \mathbf{a}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\exists a \in \mathbf{a}) (\exists b \in \mathbf{b}) (f(a, x) = b)\}$;
- "tolerable solution set" : $\Sigma(f, \text{dual } \mathbf{a}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\forall a \in \mathbf{a}) (\exists b \in \mathbf{b}) (f(a, x) = b)\}$;
- "controllable solution set" : $\Sigma(f, \mathbf{a}, \text{dual } \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\forall b \in \mathbf{b}) (\exists a \in \mathbf{a}) (f(a, x) = b)\}$.

Du fait que les quantificateurs universels sont plus restrictifs que les quantificateurs existentiels, nous avons trivialement quelques soient $\mathbf{a} \in \mathbb{K}\mathbb{R}^p$ et $\mathbf{b} \in \mathbb{K}\mathbb{R}^m$,

$$\Sigma(f, \mathbf{a}, \mathbf{b}) \subseteq \Sigma(\text{pro } \mathbf{a}, \text{pro } \mathbf{b})$$

Les mêmes notations simplifiées introduites dans le cadre des "united solution sets" sont utilisées pour les "AE-solution sets" linéaires.

Exemple II.40 (Shary[96]). *Considérons le système d'intervalles linéaire défini par*

$$\mathbf{A} = \begin{pmatrix} [2, 4] & [-2, -1] \\ [-1, 2] & [2, 4] \end{pmatrix} \quad \text{et} \quad \mathbf{b} = \begin{pmatrix} [-2, 2] \\ [-2, 2] \end{pmatrix}$$

La figure 5.1 page 70 (prise de Shary[96]) illustre les ensembles suivants :

- le "united solution set" $\Sigma(\mathbf{A}, \mathbf{b}) = \{x \in \mathbb{R}^2 \mid (\exists A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b)\}$ est noté Ξ_{uni} sur la figure 5.1 ;
- le "tolerable solution set" $\Sigma(\text{dual } \mathbf{A}, \mathbf{b}) = \{x \in \mathbb{R}^2 \mid (\forall A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b)\}$ est noté Ξ_{tol} sur la figure 5.1 ;
- le "controllable solution set" $\Sigma(\mathbf{A}, \text{dual } \mathbf{b}) = \{x \in \mathbb{R}^2 \mid (\forall b \in \mathbf{b})(\exists A \in \mathbf{A})(Ax = b)\}$ est noté Ξ_{ctr} sur la figure 5.1 et est vide pour cet exemple.

La figure 5.2 illustre d'autres "AE-solution sets" qui peuvent être obtenus en utilisant \mathbf{A} et \mathbf{b} . Les "AE-solution sets" sont notés

$$\Xi \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \quad (5.3)$$

sur la figure 5.2. Le quantificateur Q_{ij} correspond au paramètre A_{ij} et le quantificateur Q_k au paramètre b_k . Ainsi, le "AE-solution set" (5.3) correspond au "AE-solution set" $\Sigma(\mathbf{A}', \mathbf{b}')$ où \mathbf{A}'_{ij} est propre si $Q_{ij} = \exists$ et impropre si $Q_{ij} = \forall$ et de même pour \mathbf{b}' .

L'approximation des "AE-solution sets" non-linéaires n'est pas encore très développée. Les deux prochaines sous-sections présentent les méthodes dédiées aux "AE-solution sets" linéaires. La dernière sous-section présente des méthodes alternatives et décrit leur champ d'application (qui comprend des cas particuliers d'approximations d'"AE-solution sets" non-linéaires).

5.2.2 Le théorème fondamental de caractérisation des "AE-solution sets" linéaires

Le théorème suivant est fondamental pour l'approximation des "AE-solution sets" linéaires. Il donne une condition nécessaire et suffisante pour qu'un vecteur appartienne à un "AE-solution set".

Théorème II.10 (Shary[96]). *Soit $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$. Alors,*

$$x \in \Sigma(\mathbf{A}, \mathbf{b}) \iff (\text{dual } \mathbf{A})x \subseteq \mathbf{b}.$$

Remarque. *A cause de la convention différente adoptée dans cette thèse, le théorème précédent s'énonce $x \in \Xi(\mathbf{A}, \mathbf{b}) \iff \mathbf{Ax} \subseteq \mathbf{b}$ dans Shary[96] et Goldsztejn[5].*

Ce théorème généralise et reformule dans le langage des intervalles généralisés les conditions connues d'appartenance aux "united solution sets" (voir Neumaier[8]) "tolerable solution sets" (voir Neumaier[6]) et "controlable solution sets" (voir Shary[86]). Par exemple, si \mathbf{A} et \mathbf{b} sont propres, i.e. $\Sigma(\mathbf{A}, \mathbf{b})$ est un "united solution set", alors

$$(\text{dual } \mathbf{A})x \subseteq \mathbf{b} \iff (\mathbf{Ax}) \cap \mathbf{b} \neq \emptyset \iff (\exists A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b).$$

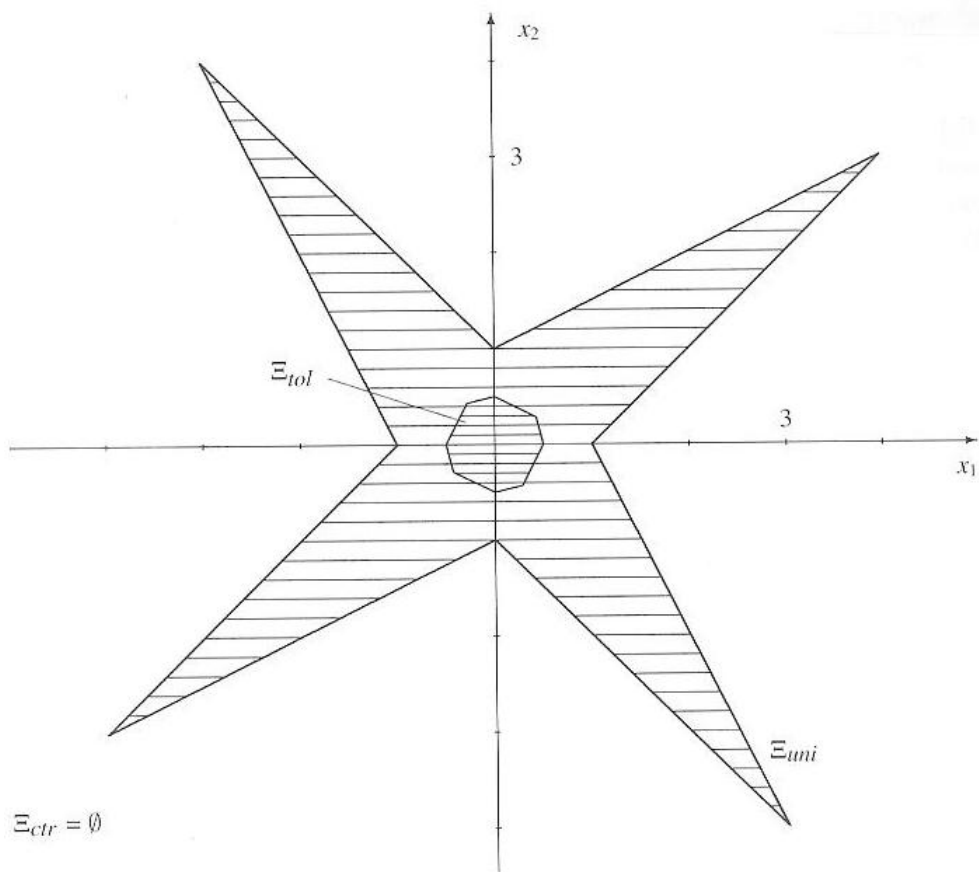


FIG. 5.1 – "United solution set", "tolerable solution set" et "controlable solution set" de l'exemple II.40.

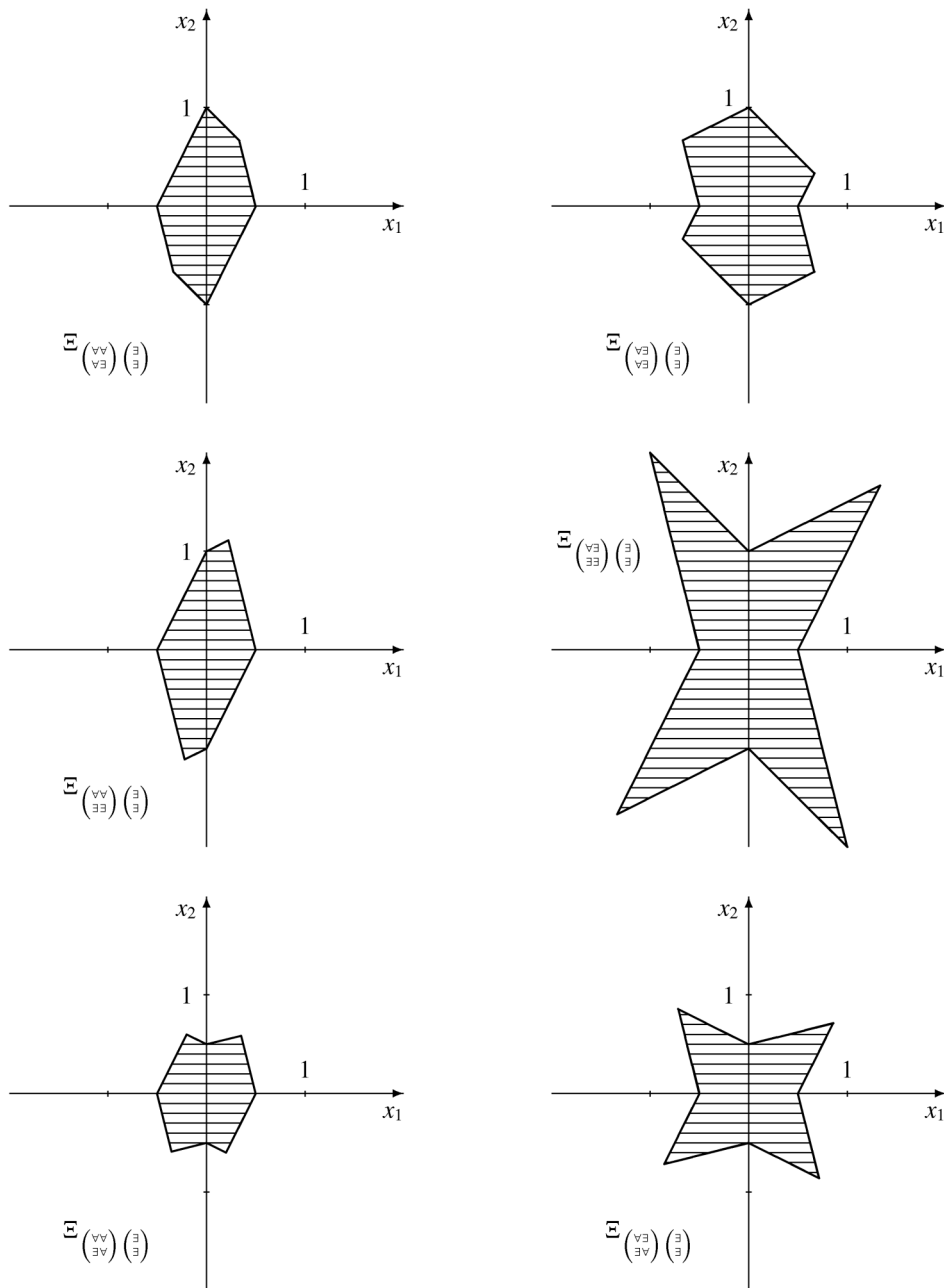


FIG. 5.2 – Autres "AE-solution sets" de l'exemple II.40.

5.2.3 Une approche formelle pour l'approximation des "AE-solution sets" linéaires

Une approche formelle pour l'approximation des "AE-solution sets" consiste à proposer des équations auxiliaires d'intervalles dans l'espace des intervalles généralisés et la relation de leurs solutions avec les "AE-solution sets" (Kupriyanova[54], Shary[93, 92, 96]). Des algorithmes dédiés permettent enfin de construire les solutions de ces équations, et ainsi de construire des approximations intérieures et extérieures des "AE-solution sets".

Équations auxiliaires d'intervalles

Quatre équations auxiliaires d'intervalles sont maintenant présentées avec la relation entre leurs solutions et les "AE-solution sets" mis en jeux. L'équation auxiliaire d'intervalles (5.4) est dédiée à l'approximation intérieure des "AE-solution sets". Une démonstration simple de ce théorème est présentée qui met en valeur l'importance du théorème II.10 (page 69).

Théorème II.11 (Shary[96], Sainz[60]). *Soient $\mathbf{A} \in \mathbb{KR}^{n \times n}$ et $\mathbf{b} \in \mathbb{KR}^n$. Alors toute solution propre $\mathbf{x} \in \mathbb{IR}^n$ de l'équation*

$$(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b} \quad (5.4)$$

est une approximation intérieure de $\Sigma(\mathbf{A}, \mathbf{b})$, i.e. $\mathbf{x} \subseteq \Sigma(\mathbf{A}, \mathbf{b})$.

Démonstration (Shary[96]). Soit un intervalle propre $\mathbf{x} \in \mathbb{IR}^n$ solution de l'équation (5.4) et $x \in \mathbf{x}$. Alors, par la croissance des opérations de l'arithmétique de Kaucher par rapport à l'inclusion, nous avons

$$(\text{dual } \mathbf{A})x \subseteq (\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b},$$

c'est à dire, $x \in \Sigma(\mathbf{A}, \mathbf{b})$ par le théorème II.10. □

Remarque. *L'équation (5.4) est écrite $\mathbf{A}\mathbf{x} = \mathbf{b}$ dans Shary[96] et Goldsztejn[5]. L'écriture proposée ici met en valeur le fait que pour construire une approximation intérieure d'un "united solution set", i.e. lorsque \mathbf{A} et \mathbf{b} sont propres, il faut considérer une matrice d'intervalles impropres.*

L'équation (5.4) a une unique solution si (pro \mathbf{A}) est à diagonal strictement dominante (voir Shary[96], Sainz[60], Markov[90]). Cette condition suffisante d'existence et d'unicité sera affinée dans le chapitre 7.1 de la partie III. Le prochain théorème présente l'équation auxiliaire d'intervalles (5.5) qui est dédiée à l'approximation extérieure des "AE-solution sets".

Théorème II.12 (Shary[96]). *Soient $\mathbf{A} \in \mathbb{KR}^{n \times n}$ et $\mathbf{b} \in \mathbb{KR}^n$. Supposons de plus que¹⁵*

$$\rho(|\mathbf{A} - I|) < 1.$$

Alors, l'équation suivante a une unique solution $\mathbf{x} \in \mathbb{KR}^n$:

$$\mathbf{x} = \mathbf{b} - (\mathbf{A} - I)\mathbf{x}. \quad (5.5)$$

De plus, si la solution \mathbf{x} est propre alors $\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \mathbf{x}$; sinon $\Sigma(\mathbf{A}, \mathbf{b}) = \emptyset$.

¹⁵Le rayon spectral $\rho(M)$ d'une matrice réelle M est le module maximum de ses valeurs propres. Il ne sera pas calculé explicitement dans cette thèse.

Le théorème précédent correspond au théorème 7.3 de Shary[96]. Son énoncé a été légèrement modifié. En particulier, les nouvelles conventions proposées dans cette thèse ont été utilisées : l'équation (5.5) s'écrit $\mathbf{x} = (I - \text{dual } \mathbf{A})\mathbf{x} + \mathbf{b}$ dans Shary[96] et Goldsztejn[5]. Ces nouvelles conventions permettent de voir un lien direct entre l'équation (5.5) et l'opérateur de Krawczyk dédié à l'approximation extérieure des "united solution sets" linéaires. Dans le cas spécifique des "united solution sets", i.e. $\mathbf{A} \in \mathbb{IR}^{n \times n}$ et $\mathbf{b} \in \mathbb{IR}^n$, des conditions plus générales sont proposées dans Mayer et Warnke[33] pour l'interprétation des solutions de l'équation (5.5). L'équation auxiliaire d'intervalles (5.6) est une version préconditionnée de (5.5). L'inclusion suivante est prouvée dans Shary[96] : étant donné $\mathbf{A} \in \mathbb{KR}^{n \times n}$, $\mathbf{b} \in \mathbb{KR}^n$ et $C \in \mathbb{R}^{n \times n}$ régulière,

$$\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \Sigma(C\mathbf{A}, C\mathbf{b}).$$

La matrice de préconditionnement C est appliquée directement à $\mathbf{A} \in \mathbb{KR}^{n \times n}$ et $\mathbf{b} \in \mathbb{KR}^n$ (ainsi, les quantificateurs peuvent être différents entre $\Sigma(\mathbf{A}, \mathbf{b})$ et $\Sigma(C\mathbf{A}, C\mathbf{b})$). Comme le "AE-solution set" préconditionné contient le "AE-solution set" original, une approximation extérieure du premier est aussi une approximation extérieure du second. Ainsi, en appliquant le théorème II.12 au "AE-solution set" préconditionné $\Sigma(C\mathbf{A}, C\mathbf{b})$, le théorème suivant est obtenu :

Théorème II.13 (Shary[96]). *Soient $\mathbf{A} \in \mathbb{KR}^{n \times n}$ et $\mathbf{b} \in \mathbb{KR}^n$. Supposons de plus que*

$$\rho(|C\mathbf{A} - I|) < 1.$$

Alors, l'équation suivante a une unique solution $\mathbf{x} \in \mathbb{KR}^n$:

$$\mathbf{x} = C\mathbf{b} - (C\mathbf{A} - I)\mathbf{x}. \quad (5.6)$$

De plus, si la solution \mathbf{x} est propre alors $\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \mathbf{x}$; sinon $\Sigma(\mathbf{A}, \mathbf{b}) = \emptyset$.

Remarque. *Le théorème précédant correspond au théorème 7.10 de Shary[96]. Son énoncé a été modifié de manière similaire à l'énoncé du théorème II.12.*

Finalement, il est prouvé dans Shary[96] que si $(\text{pro } \mathbf{A})$ est fortement régulière, la condition $\rho(|(\text{mid } \mathbf{A})^{-1}\mathbf{A} - I|) < 1$ est satisfaite, et le théorème II.13 peut-être utilisé avec $C = (\text{mid } \mathbf{A})^{-1}$ pour construire une approximation extérieure de $\Sigma(\mathbf{A}, \mathbf{b})$. Les matrices fortement régulières étant plus souvent rencontrées en pratique que les matrices vérifiant $\rho(|\mathbf{A} - I|) < 1$, ce protocole de préconditionnement élargit de manière significative le champ d'application de cette méthode.

Une quatrième équation auxiliaire d'intervalles dédiée à l'approximation extérieure est proposée dans Sainz[61] : étant donnés $\mathbf{A} \in \mathbb{KR}^{n \times n}$ et $\mathbf{b} \in \mathbb{KR}^n$,

$$\bigwedge_{i \in [1..n]} (\text{dual } \mathbf{A}_{ii})\mathbf{x}_i + \sum_{\substack{j \in [1..n] \\ j \neq i}} (\text{dual } \mathbf{A}_{ij})(\text{dual } \mathbf{x}_j) = \mathbf{b}_i. \quad (5.7)$$

Il est prouvé dans Sainz[61] qu'une solution de cette équation est une approximation extérieure de $\Sigma(\mathbf{A}, \mathbf{b})$.

Résolution des équations auxiliaires d'intervalles

Plusieurs techniques permettent de résoudre les équations présentées précédemment :

1. l'équation auxiliaire d'intervalles est réécrite sous la forme équivalente $\mathbf{x} = F(\mathbf{x})$ en utilisant les relations de l'arithmétique de Kaucher. Ensuite, la suite $\mathbf{x}^{(k+1)} = F(\mathbf{x}^{(k)})$ est construite. Si celle-ci converge, sa limite vérifie $\mathbf{x} = F(\mathbf{x})$, et vérifie donc l'équation initiale.
2. l'équation auxiliaire d'intervalles de $\mathbb{K}\mathbb{R}^n$ est transformée en une équation de \mathbb{R}^{2n} en explicitant les formules de l'arithmétique de Kaucher par rapport aux bornes des intervalles. Ensuite, des techniques dédiées aux équations réelles peuvent être utilisées pour résoudre l'équation de \mathbb{R}^{2n} (voir Shary[92][94]).
3. l'équation auxiliaire d'intervalles est transformée en un problème d'optimisation dont l'optimum est la solution de l'équation (voir Sainz[60]).

Seule la première classe de méthodes est maintenant présentée en détail : les trois théorèmes suivants donnent des conditions suffisantes pour la convergence des suites construites à partir de l'écriture en point fixe des équations auxiliaires d'intervalles. Le théorème suivant permet de résoudre l'équation auxiliaire d'intervalle (5.4).

Théorème II.14 (Sainz[60], Markov[65]). Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$. Supposons que (pro \mathbf{A}) est à diagonale strictement dominante. Alors la suite $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$, définie par

$$\mathbf{x}^{(k+1)} = \mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x}^{(k)})),$$

où $\mathbf{D} = \text{diag}(1/\mathbf{A}_{11}, \dots, 1/\mathbf{A}_{nn})$ et \mathbf{L} et \mathbf{U} sont respectivement la partie inférieure et la partie supérieure de la matrice \mathbf{A} , convergence vers l'unique solution de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ quelque soit $\mathbf{x}^{(0)} \in \mathbb{K}\mathbb{R}^n$.

Il arrive souvent de rencontrer des matrices à diagonale non strictement dominante dans les applications. Ainsi, le théorème précédent ne pourra pas être utilisé en général dans ces cas. Le procédé de préconditionnement proposé dans la partie III de la thèse apportera une solution partielle à ce problème. Le théorème suivant permet de résoudre l'équation auxiliaire d'intervalle (5.6) (et (5.5) lorsque $C = I$).

Théorème II.15 (Shary[96]). Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $C \in \mathbb{R}^{n \times n}$ telles que $\rho(|I - C\mathbf{A}|) < 1$ et $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$. Alors la suite $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$, définie par

$$\mathbf{x}^{(k+1)} = C\mathbf{b} - (C\mathbf{A} - I)\mathbf{x}^{(k)},$$

converge vers l'unique solution de l'équation $\mathbf{x} = C\mathbf{b} - (C\mathbf{A} - I)\mathbf{x}$ quelque soit $\mathbf{x}^{(0)} \in \mathbb{K}\mathbb{R}^n$.

Grâce aux nouvelles notations utilisées dans cette thèse, l'écriture sous forme de point fixe de l'équation (5.6) peut être rapprochée de l'opérateur de Krawczyk dédié à l'approximation extérieur des "united solution sets" linéaires. Lorsqu'elle est mise sous forme de point fixe, l'équation (5.7) devient :

$$\bigwedge_{i \in [1..n]} \mathbf{x}_i = \frac{1}{\mathbf{A}_{ii}} (\mathbf{b}_i - \sum_{\substack{j \in [1..n] \\ j \neq i}} \mathbf{A}_{ij} \mathbf{x}_j). \quad (5.8)$$

Grâce aux nouvelles notations introduites dans cette thèse, l'équation précédente peut-être rapprochée de l'opérateur de Gauss-Seidel, et de sa généralisation qui est présentée dans la prochaine sous-section. Il est démontré dans Sainz[60] que la suite alors obtenue converge si pro \mathbf{A} est à diagonale strictement dominante.

5.2.4 Extension de l'approche classique pour l'approximation extérieure des "AE-solution sets" linéaires

L'opérateur de Gauss-Seidel, dans sa forme simplifiée, a été généralisé dans Shary[95] pour l'approximation extérieure des "AE-solution sets" linéaires. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$ et $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$. L'opérateur généralisé de Gauss-Seidel est noté $\Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x})$ et défini si $0 \notin (\text{pro } \mathbf{A}_{ii})$ pour tout $i \in [1..n]$ par

$$\Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x}) = \mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})\mathbf{x}) \wedge \mathbf{x},$$

où $\mathbf{D} = \text{diag}(1/\mathbf{A}_{11}, \dots, 1/\mathbf{A}_{nn})$ et \mathbf{L} et \mathbf{U} sont respectivement la partie inférieure et la partie supérieure de la matrice \mathbf{A} . Grâce aux nouvelles notations introduites dans cette thèse, l'expression de l'opérateur généralisé de Gauss-Seidel est la même que celle de l'opérateur classique de Gauss-Seidel (dans sa version simplifiée). Les propriétés suivantes sont prouvées dans Shary[95][96] :

- $\Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x}) \subseteq \mathbf{x}$;
- si $\Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x}) \in \mathbb{I}\mathbb{R}^n$ alors $\Sigma(\mathbf{A}, \mathbf{b}) \cap \mathbf{x} \subseteq \Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x})$;
- si $\Gamma^s(\mathbf{A}, \mathbf{b}, \mathbf{x}) \notin \mathbb{I}\mathbb{R}^n$ alors $\Sigma(\mathbf{A}, \mathbf{b}) = \emptyset$.

Remarque. *L'équation d'intervalles (5.7) et de l'opérateur généralisé de Gauss-Seidel ont été découverts indépendamment. Leur rapprochement a été fait dans Goldsztejn[5]. Les conditions d'utilisation de l'opérateur généralisé de Gauss-Seidel sont moins contraignantes que pour l'équation intervalles (5.7).*

5.2.5 Autres méthodes

L'approximation de "AE-solution sets" non-linéaires est d'actualité. Quatre méthodes sont maintenant présentées qui s'attaquent à ce problème.

L'élimination de quantificateurs

L'élimination de quantificateur est une méthode formelle qui permet de transformer un système quantifié en un système non quantifié équivalent (voir par exemple Davenport[46] ou Collins[15]). Elle s'applique à des ensembles de solutions plus généraux que les "AE-solution set", où les quantificateurs ne sont pas contraints à être ordonnés. Elle peut donc s'appliquer également aux "AE-solution sets". Par exemple, le AE-solution set

$$\Sigma = \{(x_1, x_2)^T \in \mathbb{R}^2 \mid (\forall x_4 \in [-1, 1]) (\exists x_3 \in [-3, 3]) (\sum x_i^2 = 9)\}$$

est transformé, après simplification, en

$$\Sigma = \{(x_1, x_2)^T \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 8\}.$$

Cette dernière expression peut alors être traitée au moyen de méthodes numériques classiques si nécessaire. L'élimination de quantificateurs a deux principaux défauts : premièrement, elle a une complexité importante qui réduit son champ d'application à des problèmes très simples. Deuxièmement, elle est restreinte à des contraintes polynomiales.

"Continuous first order constraint satisfaction"

Dans Ratschan[75][83], un algorithme spécifique "branch and prune" a été mis au point pour approximer les contraintes quantifiées mettant en jeu des inégalités. En modélisant une égalité par deux inégalités, cet algorithme permet de construire une approximation extérieure des "AE-solutions sets" aussi précise que voulue (l'approximation intérieure des "AE-solution sets" ne peut pas être faite grâce à cette technique). Une comparaison avec les méthodes mises au point dans cette thèse sera effectuée (voir Chapitre 18 page 191).

"E-constraints"

Dans Ratschan[77], un algorithme spécifique "branch and prune" permet de d'approximer intérieurement et extérieurement les ensembles de solutions de la forme suivante :

$$\{x \in \mathbb{R}^n \mid (\exists y \in \mathbf{y})(f(x, y) = 0 \wedge g_1(x, y) \geq 0 \wedge \dots \wedge g_k(x, y) \geq 0)\}$$

où y est un vecteur de dimension quelconque. La méthode est généralisée sans difficulté à une disjonction de telles contraintes, appelées "E-constraint" dans Ratschan[77]. Si cette méthode est une des seule disponible pour traiter des contraintes quantifiées avec égalité, elle est restreinte à de petits problème dans sa version actuelle.

Contraintes universellement quantifiées

Dans le cadre d'inégalités avec une variable quantifiée universellement, une méthode efficace a été proposée dans Benhamou[27] et Christie[56] (il s'agit d'une généralisation de la méthode proposée dans Collavizza[40] consistant à prouver une contrainte en prouvant que sa négation est fausse). Elle permet de construire des approximations intérieures et extérieures des (x_1, \dots, x_n) vérifiant la contrainte

$$(\forall t \in \mathbf{t})(f_1(x_1, \dots, x_n, t) \geq 0 \wedge \dots \wedge f_m(x_1, \dots, x_n, t) \geq 0)$$

A cause des performance de l'algorithme proposé, cette technique est restreinte à une unique variable quantifiée universellement. Il ne semble pas que cette méthode ait été comparée avec les méthodes précédentes.

"Quantified set inversion algorithm"

Dans Herrero[69][70], les intervalles modaux sont utilisés pour construire des approximations intérieures et extérieures aux "AE-solution sets" contenant une fonction à valeur réelle, i.e. $m = 1$ dans la définition II.13. La méthode proposée consiste en l'imbrication de deux algorithmes de "branch and prune" : un premier algorithme permet de décider si une boîte \mathbf{x} est incluse dans le "AE-solution set" ou si elle est à l'extérieur de celui-ci. Cet algorithme est utilisé dans un autre algorithme de "branch and prune" pour construire une approximation du "AE-solution set". Cette technique est facilement étendue aux cas $m > 1$ dans les cas suivants :

- dans le cas de l'approximation extérieure des "AE-solution sets" ;
- dans le cas de l'approximation intérieure, si chaque variable quantifiée existentiellement n'apparaît que dans une seule équation.

Aucune comparaison n'est possible car aucun autre algorithme numérique n'a été proposé pour résoudre ce type de problème. Cependant, l'imbrication de deux algorithmes de bisection laisse penser que les performances peuvent être améliorées.

5.2.6 Conclusion

A notre connaissance, au vue des méthodes décrites précédemment le problème d'approximation intérieure des "AE-solution sets" non-linéaires dont la fonction est à valeurs vectorielles est un problème ouvert (lorsqu'il ne se ramène pas à un problème plus simple, i.e. si des variables quantifiées existentiellement sont partagées par plusieurs équations). De même, le problème plus simple qui consiste à calculer une approximation intérieure de l'image d'un intervalle par une fonction à valeurs vectorielles $f = (f_1, \dots, f_n)$, dont certaines variables apparaissent dans l'expression de plusieurs composantes f_i , n'est pas résolu aujourd'hui à notre connaissance.

La nouvelle formulation de la théorie des intervalles modaux va permettre d'introduire une nouvelle extension de la valeur moyenne aux intervalles généralisés. Celle-ci va permettre de construire des algorithmes pour ces problèmes mal résolus aujourd'hui. Par exemple, dans le Chapitre 17 (page 187) une approximation intérieure de range (f, \mathbf{x}) , où

$$f(x) = \begin{pmatrix} x_1^6 + x_2^6 + x_1x_2 \\ x_1^6 - x_2^6 - x_1x_2 \end{pmatrix}$$

et $\mathbf{x} = ([0.99, 1.01], [0.99, 1.01])^T$, est calculée par une simple évaluation de la nouvelle extension de la valeur moyenne.

Troisième partie

Nouveaux développements pour les
"AE-solution sets linéaires"

Chapitre 6

Introduction

Cette partie de la thèse présente deux contributions autour de l'approximation des "AE-solution sets" linéaires :

- un nouveau procédé de préconditionnement à droite (Goldsztein[5]) ;
- une amélioration de l'opérateur généralisé de Gauss-Seidel.

Préconditionnement à droite des AE-solution sets

Les trois premiers chapitres sont dédiés à la présentation d'un procédé de préconditionnement à droite pour les "AE-solution sets" linéaires. Dans le contexte des "united solution sets", préconditionner un "united solution set" consiste à considérer un "united solution set" auxiliaire, sur lequel les méthodes de résolution ont un meilleur comportement. Le préconditionnement le plus utilisé est le *préconditionnement à gauche* qui consiste à étudier le "united solution set" auxiliaire suivant : étant donné $\mathbf{A} \in \mathbb{IR}^{n \times n}$, $\mathbf{b} \in \mathbb{IR}^n$ et $C \in \mathbb{R}^{n \times n}$, le préconditionnement à gauche de $\Sigma(\mathbf{A}, \mathbf{b})$ conduit au "united solution set" suivant :

$$\Sigma(C\mathbf{A}, C\mathbf{b}).$$

Ce préconditionnement est adapté à l'estimation extérieure des "united solution sets" à cause de l'inclusion suivante :

$$\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \Sigma(C\mathbf{A}, C\mathbf{b}).$$

Ainsi, une approximation extérieure du "united solution set" préconditionné est aussi une approximation extérieure du "united solution set" original :

$$\Sigma(C\mathbf{A}, C\mathbf{b}) \subseteq \mathbf{x} \implies \Sigma(\mathbf{A}, \mathbf{b}) \subseteq \mathbf{x}.$$

Une matrice de préconditionnement C souvent utilisée est l'inverse du milieu de \mathbf{A} , i.e. $C = (\text{mid } \mathbf{A})^{-1}$. Lorsque les intervalles de \mathbf{A} sont assez petits, la matrice d'intervalles préconditionnée $(\text{mid } \mathbf{A})^{-1}\mathbf{A}$ est proche de la matrice identité et les méthodes de résolution ont un bon comportement pour construire une approximation extérieure de $\Sigma(C\mathbf{A}, C\mathbf{b})$ (voir Neumaier[8] pour une présentation détaillée du préconditionnement à gauche).

Le préconditionnement à droite est une alternative au préconditionnement à gauche proposée dans le contexte de l'approximation extérieure des "united solution sets" linéaires dans Neumaier[7]. Il consiste à considérer un "united solution sets" auxiliaire où le produit $\mathbf{A}C$ apparaît, au lieu du produit CA . Il est prouvé dans Neumaier[7] qu'un

tel préconditionnement permet de construire une approximation extérieure sous la forme d'un parallélépipède, aussi appelé boîte penchée¹⁶, i.e. l'image d'un intervalle de \mathbb{R}^n par une application linéaire :

$$\{Cu \in \mathbb{R}^n \mid u \in \mathbf{u}\}$$

Il est aussi prouvé dans Neumaier[7] que de telles approximations peuvent être beaucoup plus précises que les approximations par des intervalles.

L'approximation des "AE-solution set" tire aussi bénéfice d'une procédure de préconditionnement. Un préconditionnement à gauche compatible avec l'approximation extérieure des "AE-solution sets" a été proposé dans Shary[96]. Il y est montré que quelque soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$ et $C \in \mathbb{R}^{n \times n}$ régulière,

$$\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \Sigma(C\mathbf{A}, C\mathbf{b})$$

Remarque. Si \mathbf{A} est composée d'éléments propres et impropres, il se peut que les qualités propres/impropres des éléments de $C\mathbf{A}$ soient différentes de celles des éléments de \mathbf{A} . Ainsi, le préconditionnement proposé dans Shary[96] peut changer les quantificateurs associés aux paramètres du système d'intervalles. Il en est de même pour \mathbf{b} et $C\mathbf{b}$.

Jusqu'à maintenant, aucun préconditionnement n'était disponible pour l'approximation intérieure des "AE-solution sets" (ni dans le cas particulier des "united solution sets"). L'approximation intérieure a bien besoin d'un préconditionnement particulier : en effet, le "AE-solution set" préconditionné doit cette fois être contenu à l'intérieur du "AE-solution set" original, i.e.

$$\Sigma(\mathbf{A}', \mathbf{b}') \subseteq \Sigma(\mathbf{A}, \mathbf{b}),$$

où $\Sigma(\mathbf{A}', \mathbf{b}')$ est le "AE-solution set" préconditionné. Ainsi, une approximation intérieure du "AE-solution set" préconditionné sera aussi une approximation intérieure du "AE-solution set" original :

$$\mathbf{x} \subseteq \Sigma(\mathbf{A}', \mathbf{b}') \implies \mathbf{x} \subseteq \Sigma(\mathbf{A}, \mathbf{b}).$$

Les préconditionnements proposés dans cette thèse sont des préconditionnements à droite qui vont permettre de construire des approximations intérieures et extérieures d'un "AE-solution set" linéaire sous la forme de boîtes penchées. La figure 6.1 (page 83) illustre ces constructions : un "AE-solution set" Σ et ses deux approximations \mathbf{s} et \mathbf{s}' y sont représentés. Les avantages de ces préconditionnements seront étudiés théoriquement et sur des exemples académiques.

Amélioration de l'opérateur généralisé de Gauss-Seidel

L'opérateur généralisé de Gauss-Seidel défini dans Shary[95][96] est en fait une généralisation de l'opérateur simplifié de Gauss-Seidel, i.e. celui initialement proposé dans Ris[29]. Le chapitre 8 présente une généralisation de l'opérateur de Gauss-Seidel sous sa forme actuelle, i.e. celui proposé dans Hansen[19]. Cela permettra d'utiliser l'opérateur généralisé de Gauss-Seidel avec des matrices d'intervalles non-régulières.

¹⁶"Skew box" en anglais.

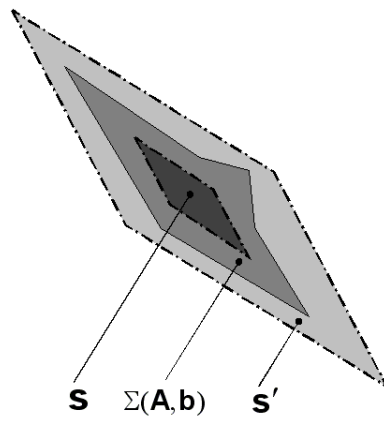


FIG. 6.1 – Un "AE-solution set" avec ses approximations intérieures et extérieures sous forme de parallélogrammes.

Chapitre 7

Préconditionnement à droite des "AE-solution sets" linéaires

Ce chapitre présente un nouveau procédé de preconditionnement pour les "AE-solution sets" linéaires. Il permet de construire des approximations intérieures et extérieures d'un "AE-solution set" linéaire sous la forme de boîtes penchées. Ce preconditionnement présente des avantages importants dans le cas de l'approximation intérieure : d'une part, dans certaines situations où l'approximation intérieure "AE-solution sets" linéaires était difficile, voir impossible, sans preconditionnement, le preconditionnement à droite permet de construire une approximation intérieure aisément. D'autre part, le preconditionnement accélère les algorithmes utilisés pour la construction des approximation intérieures.

La Section 7.1 présente une contribution moins importante : il s'agit une amélioration de la condition suffisante pour l'existence et l'unicité de la solution de l'équation d'intervalle liée à l'approximation intérieure des "AE-solution sets" linéaires. Cette nouvelle condition est aussi une condition suffisante pour la convergence des algorithmes vers la solution de l'équation. Elle sera utilisée dans les sections suivantes.

7.1 Nouvelle condition suffisante d'existence et d'unicité pour les solutions de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$

Il a été montré par plusieurs auteurs (Markov[90], Sainz[60]) que la suite

$$\mathbf{x}^{(k+1)} = \mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x}^{(k)})),$$

où $\mathbf{D} = \text{diag}(1/\mathbf{A}_{11}, \dots, 1/\mathbf{A}_{nn})$ et \mathbf{L} et \mathbf{U} sont respectivement la partie inférieure et la partie supérieure de la matrice \mathbf{A} , converge vers l'unique solution de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ si la matrice d'intervalle $\text{pro } \mathbf{A}$ est à diagonale strictement dominante, cf. Théorème II.14 page 74. Dans ce chapitre, cette condition de convergence est affaiblie : il est montré que la suite ainsi obtenue converge vers l'unique solution de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ si la matrice $\text{pro } \mathbf{A}$ est une H-matrice (voir définition d'une H-matrice page 26). Cette nouvelle condition généralise bien la condition précédemment établie car toute matrice à diagonale strictement dominante est une H-matrice, cf. Neumaier[8]. Nous aurons ainsi prouvé que l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ a une unique solution si la matrice $\text{pro } \mathbf{A}$ est une H-matrice.

7.1.1 Caractérisation des H-matrices

La proposition suivante donne une condition nécessaire et suffisante pour qu'une matrice d'intervalle soit une H-matrice. Étant donné un vecteur $s \in \mathbb{R}^n$, la matrice diagonale dont les entrées sont s_i est notée $\text{diag}(s)$.

Proposition III.1. *Soit $\mathbf{A} \in \mathbb{IR}^{n \times n}$. Alors,*

1. *pour tout $s \in \mathbb{R}^n$ tel que $s > 0$, $\langle \mathbf{A} \text{diag}(s) \rangle = \langle \mathbf{A} \rangle \text{diag}(s)$;*
2. *$\mathbf{A} \in \mathbb{IR}^{n \times n}$ est une H-matrice si et seulement si il existe $s \in \mathbb{R}^n$, $s > 0$, tel que $\mathbf{A} \text{diag}(s)$ est à diagonale strictement dominante.*

Démonstration. 1. Le calcul des deux produits matriciels est aisé car $\text{diag}(s)$ est diagonale. Il reste à appliquer la proposition proposition 1.6.1 de Neumaier[8] de la manière suivante : $\langle \mathbf{A}_{ii} s_i \rangle = \langle \mathbf{A}_{ii} \rangle |s_i| = \langle \mathbf{A}_{ii} \rangle s_i$ et $|\mathbf{A}_{ij} s_j| = |\mathbf{A}_{ij}| |s_j| = |\mathbf{A}_{ij}| s_j$, où $|s_k| = s_k$ parce que $s_k > 0$.

2. Premièrement, remarquons que les entrées hors de la diagonale de $\langle \mathbf{A} \rangle$ sont négatives par définition de la matrice de comparaison. Donc, par définition d'une M-matrice, \mathbf{A} est une H-matrice si et seulement si $\exists s > 0$, $\langle \mathbf{A} \rangle s > 0$. Le vecteur s peut alors être écrit sous la forme du produit $\text{diag}(s)e$ où $e = (1, \dots, 1)^T$. Donc, \mathbf{A} est une H-matrice si et seulement si $\exists s > 0$, $\langle \mathbf{A} \rangle \text{diag}(s)e > 0$. En appliquant le premier cas de la proposition, \mathbf{A} est une H-matrice si et seulement si $\exists s > 0$, $\langle \mathbf{A} \text{diag}(s) \rangle e > 0$, i.e. si et seulement si $\exists s > 0$ tel que $\mathbf{A} \text{diag}(s)$ est à diagonale strictement dominante. \square

Cette condition n'est pas très utile pour tester si une matrice d'intervalles est une H-matrice : en effet, il est en général difficile de déterminer le vecteur s pour des dimensions plus grandes que 3. Néanmoins, cette condition sera utile pour la démonstration proposée dans la section suivante.

7.1.2 Nouvelle condition d'existence et d'unicité

Le lemme suivant sera utile pour la démonstration du Théorème III.1.

Lemme III.1. *Soient $s \in \mathbb{R}^n$ tel que $s > 0$, $\mathbf{M} \in \mathbb{KR}^{n \times n}$ et $\mathbf{z} \in \mathbb{KR}^n$. La matrice diagonale $\text{diag}(s)$ est notée S . Alors, $(\mathbf{M}S)\mathbf{z} = \mathbf{M}(S\mathbf{z})$ et $(S\mathbf{M})\mathbf{z} = S(\mathbf{M}\mathbf{z})$. Cela implique en particulier que la fonction continue $\mathbf{x} \mapsto S\mathbf{x}$ est injective.*

Démonstration. Remarquons simplement que S étant diagonale, les produits matriciels sont très simples et qu'aucune distribution ne doit être faite entre $+$ et \times . Maintenant, la fonction $\mathbf{x} \mapsto S\mathbf{x}$ est continue comme c'est prouvé dans Shary[92] et est injective parce que $S\mathbf{x} = S\mathbf{y} \implies S^{-1}(S\mathbf{x}) = S^{-1}(S\mathbf{y})$ qui implique finalement $\mathbf{x} = \mathbf{y}$ grâce aux propriétés prouvées précédemment. \square

Le prochain théorème présente une condition de convergence améliorée par rapport au Théorème II.14 page 74.

Théorème III.1. *Soient $\mathbf{A} \in \mathbb{KR}^{n \times n}$ et $\mathbf{b} \in \mathbb{KR}^n$. Supposons que $(\text{pro } \mathbf{A})$ soit une H-matrice. Alors la suite $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$, définie par*

$$\mathbf{x}^{(k+1)} = \mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x}^{(k)})),$$

où $\mathbf{D} = \text{diag}(1/\mathbf{A}_{11}, \dots, 1/\mathbf{A}_{nn})$ et \mathbf{L} et \mathbf{U} sont respectivement la partie inférieure et la partie supérieure de la matrice \mathbf{A} , convergence vers l'unique solution de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ quelque soit $\mathbf{x}^{(0)} \in \mathbb{K}\mathbb{R}^n$.

Démonstration. La matrice $(\text{pro } \mathbf{A})$ étant une H-matrice, la Proposition III.1 prouve qu'il existe $s > 0$ tel que $(\text{pro } \mathbf{A})S$ est à diagonale strictement dominante, où $S := \text{diag}(s)$. Définissons $\mathbf{A}' := \mathbf{A}S$ (qui est équivalent à $\mathbf{A}'S^{-1} = (\mathbf{A}S)S^{-1} = \mathbf{A}$ par le Lemme III.1). Grâce à la simplicité du produit avec une matrice diagonale, il est facile de vérifier que $(\text{pro } \mathbf{A}') = (\text{pro } \mathbf{A})S$. Par conséquent, $(\text{pro } \mathbf{A}')$ est à diagonale strictement dominante. Donc, nous pouvons utiliser le Théorème II.14 qui prouve que la suite $(\mathbf{y}^{(k)})_{k \in \mathbb{N}}$, définie par

$$\mathbf{y}^{(k+1)} = \mathbf{D}'(\mathbf{b} - (\mathbf{L}' + \mathbf{U}')(\text{dual } \mathbf{y}^{(k)})),$$

converge vers l'unique solution de l'équation $(\text{dual } \mathbf{A}')\mathbf{y} = \mathbf{b}$ quelque soit $\mathbf{y}^{(0)}$, où

$$\mathbf{D}' = \text{diag}(1/\mathbf{A}'_{11}, \dots, 1/\mathbf{A}'_{nn})$$

et \mathbf{L}' et \mathbf{U}' sont respectivement la partie inférieure et la partie supérieure de la matrice \mathbf{A}' . Choisissons $\mathbf{y}^{(0)} = S^{-1}\mathbf{x}^{(0)}$. Nous allons maintenant prouver par récurrence que $\mathbf{y}^{(k)} = S^{-1}\mathbf{x}^{(k)}$ pour tout $k \in \mathbb{N}$. Pour cela, supposons que $\mathbf{y}^{(k)} = S^{-1}\mathbf{x}^{(k)}$. Donc, par définition de $\mathbf{y}^{(k+1)}$ nous avons

$$\begin{aligned} \mathbf{y}^{(k+1)} &= \mathbf{D}'(\mathbf{b} - (\mathbf{L}' + \mathbf{U}')(\text{dual}(S^{-1}\mathbf{x}^{(k)}))) \\ &= \mathbf{D}'(\mathbf{b} - (\mathbf{L}' + \mathbf{U}')(S^{-1}(\text{dual } \mathbf{x}^{(k)}))), \end{aligned}$$

la deuxième égalité étant vraie car $(\text{dual } S^{-1}) = S^{-1}$. En utilisant le Lemme III.1 nous obtenons

$$\mathbf{y}^{(k+1)} = \mathbf{D}'(\mathbf{b} - ((\mathbf{L}' + \mathbf{U}')S^{-1})(\text{dual } \mathbf{x}^{(k)})).$$

Alors, à cause de la simplicité du produit par une matrice diagonale, il est facile de conclure que $\mathbf{A}'S^{-1} = \mathbf{A}$ implique $(\mathbf{L}' + \mathbf{U}')S^{-1} = (\mathbf{L} + \mathbf{U})$ et $\mathbf{D}' = \mathbf{D}S^{-1} = S^{-1}\mathbf{D}$. Nous obtenons donc

$$\mathbf{y}^{(k+1)} = (S^{-1}\mathbf{D})(\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x}^{(k)})).$$

En appliquant le Lemme III.1 nous obtenons

$$\mathbf{y}^{(k+1)} = S^{-1}(\mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x}^{(k)}))).$$

ce qui implique finalement $\mathbf{y}^{(k+1)} = S^{-1}\mathbf{x}^{(k+1)}$ grâce à la définition récursive de $\mathbf{x}^{(k+1)}$. Nous avons finalement prouvé que $\mathbf{y}^{(k)} = S^{-1}\mathbf{x}^{(k)}$ pour tout $k \in \mathbb{N}$. Maintenant, en multipliant à gauche les deux termes de cette égalité par S et en appliquant le Lemme III.1, nous obtenons $S\mathbf{y}^{(k)} = S(S^{-1}\mathbf{x}^{(k)}) = \mathbf{x}^{(k)}$, qui est valable pour tout $k \in \mathbb{N}$. La fonction S étant continue, la suite $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$ est l'image par une fonction continue de la suite convergente $(\mathbf{y}^{(k)})_{k \in \mathbb{N}}$, ce qui implique qu'elle converge aussi. De plus, les limites $\mathbf{x}^{(\infty)}$ et $\mathbf{y}^{(\infty)}$ vérifient aussi $\mathbf{y}^{(\infty)} = S^{-1}\mathbf{x}^{(\infty)}$. Maintenant, comme $(\text{dual } \mathbf{A}')\mathbf{y}^{(\infty)} = \mathbf{b}$, nous avons $(\text{dual } \mathbf{A}')(S^{-1}\mathbf{x}^{(\infty)}) = \mathbf{b}$, c'est à dire, par définition de \mathbf{A}' , $(\text{dual } \mathbf{A}S)(S^{-1}\mathbf{x}^{(\infty)}) = \mathbf{b}$. Cela implique finalement

$$(\text{dual } \mathbf{A})\mathbf{x}^{(\infty)} = \mathbf{b}$$

grâce au Lemme III.1. Jusqu'à maintenant, nous avons prouvé que la suite $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$ définie dans l'énoncé du théorème converge vers une solution de l'équation $(\text{dual } \mathbf{A})\mathbf{x} =$

b. Il reste donc à montrer que cette solution est l'*unique* solution de cette équation. Supposons que $\tilde{\mathbf{x}} \neq \mathbf{x}^{(\infty)}$ satisfasse aussi $(\text{dual } \mathbf{A})\tilde{\mathbf{x}} = \mathbf{b}$ et prouvons que cela conduit à une absurdité. Considérons alors $\tilde{\mathbf{y}} := S^{-1}\tilde{\mathbf{x}}$ qui est différent de $\mathbf{y}^{(\infty)} = S^{-1}\mathbf{x}^{(\infty)}$ car la fonction $\mathbf{x} \mapsto S^{-1}\mathbf{x}$ est injective par le Lemme III.1. Cependant, $(\text{dual } \mathbf{A}')\tilde{\mathbf{y}} = (\text{dual } \mathbf{A}S)(S^{-1}\tilde{\mathbf{x}}) = (\text{dual } \mathbf{A})\tilde{\mathbf{x}}$ qui est finalement égale à \mathbf{b} par supposition. Donc, $\mathbf{y}^{(\infty)}$ et $\tilde{\mathbf{y}}$ sont deux solutions différentes de $(\text{dual } \mathbf{A}')\mathbf{y} = \mathbf{b}$, ce qui contredit finalement l'unicité de la solution de cette équation imposée par le Théorème II.14. \square

L'exemple suivant illustre la condition de convergence prouvée par le théorème précédent.

Exemple III.1. Soit $\mathbf{A} = \begin{pmatrix} 150 & 1 \\ 50 & -50 \end{pmatrix} \in \mathbb{IR}^{2 \times 2}$ et $\mathbf{b} = \begin{pmatrix} [10, 20] \\ [-10, -20] \end{pmatrix} \in \mathbb{KR}^2$. La matrice d'intervalles \mathbf{A} est une H-matrice (utiliser la Proposition III.1 avec $s = (0.01, 1)^T$) mais n'est pas à diagonale strictement dominante. L'application du processus d'itération du Théorème III.1 conduit à une suite qui converge vers $\mathbf{x}^{(\infty)} = [0.0, 0.4] \times [10, -40]$. On peut alors vérifier que $(\text{dual } \mathbf{A})\mathbf{x}^{(\infty)} = \mathbf{b}$.

Le corollaire suivant du Théorème III.1 est une version améliorée du théorème 6.7 de Shary[96].

Corollaire III.1. Soient $\mathbf{A} \in \mathbb{KR}^{n \times n}$ telle que $(\text{pro } \mathbf{A})$ est une H-matrice et $\mathbf{b} \in \mathbb{KR}^n$. Alors l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ a une unique solution.

Démonstration. Ce corollaire est une version simplifiée du Théorème III.1. \square

7.2 Préconditionnement à droite des "AE-solution sets" linéaires

Le preconditionnement des "AE-solution sets" linéaires à droite est maintenant présenté. Il va permettre de construire des approximations intérieures et extérieures sous la forme de boîte penchées.

7.2.1 Sur l'utilisation de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ pour l'approximation intérieure des "AE-solution sets"

L'approximation intérieure du "AE-solution sets" $\Sigma(\mathbf{A}, \mathbf{b})$ est faite en construisant des solutions de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ (cf. 5.2 page 67). Trois raisons peuvent rendre la résolution de cette équation inefficace :

- l'équation a une solution qui n'est pas propre, et donc qui ne peut pas être interprétée comme une approximation intérieure ;
- l'équation n'a pas de solution ;
- l'équation a une infinité de solutions. Dans ce cas, les méthodes présentées dans la section 5.2 page 67 ne permettent pas en générale de construire une des solutions.

Ces trois situations sont rencontrées dans des cas très simples de "AE-solution sets" non-vides. L'exemple suivant illustre le premier cas.

Exemple III.2. Considérons $A = \begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix}$, $\mathbf{b} \in \mathbb{IR}^2$ et le "united-solution set" correspondant

$$\Sigma(A, \mathbf{b}) = \{x \in \mathbb{R}^2 \mid (\exists b \in \mathbf{b})(Ax = b)\}.$$

L'approximation intérieure de ce "united solution set" passe par la résolution de l'équation $A\mathbf{x} = \mathbf{b}$. Dans ce cas, la matrice est triangulaire et $0 \neq A_{ii}$. Ainsi, la solution de cette équation peut être calculée simplement de la manière suivante :

$$\begin{aligned} \mathbf{x}_2 &= \mathbf{b}_2, \\ \mathbf{x}_1 &= 0.5(\mathbf{b}_1 - 3 \text{ dual } \mathbf{x}_2) = 0.5(\mathbf{b}_1 - 3 \text{ dual } \mathbf{b}_2). \end{aligned}$$

Par conséquent, la solution \mathbf{x} est propre si et seulement si $\text{rad } \mathbf{b}_1 \geq 3 \text{ rad } \mathbf{b}_2$. Dans ce cas, la résolution de l'équation $A\mathbf{x} = \mathbf{b}$ conduit à une approximation intérieure du "united solution set" considéré. Cependant, si $\text{rad } \mathbf{b}_1 < 3 \text{ rad } \mathbf{b}_2$, la solution de l'équation n'est pas propre et ainsi n'est d'aucune utilité pour la construction d'une approximation intérieure. Dans cette situation, la technique de bisection des paramètres¹⁷ a été proposée dans Shary[96]. Elle consiste à étudier un "united solution set" auxiliaire inclus dans le "united solution set" original. Ce "united solution set" auxiliaire est obtenu en considérant $\mathbf{b}' \subseteq \mathbf{b}$ au lieu de \mathbf{b} . Par exemple, si $\mathbf{b} = ([-2, 2], [-10, 10])^T$, donc l'équation $A\mathbf{x} = \mathbf{b}$ n'a pas de solution propre, on peut considérer $\mathbf{b}' = ([-2, 2], [k, k + 1])^T$ avec $k \in [-10..9]$. Alors, $\text{rad } \mathbf{b}'_1 \geq 3 \text{ rad } \mathbf{b}'_2$ et ainsi chaque équation $A\mathbf{x} = \mathbf{b}'$ avec $k \in [-10..9]$ a une solution propre \mathbf{x} qui est une approximation intérieure du "united solution set" original. Finalement, le "united solution set" original est approximé intérieurement par 20 intervalles.

L'exemple suivant illustre les deux dernières situations.

Exemple III.3. Considérons $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, $\mathbf{b} \in \mathbb{IR}^2$ et le "united solution set" correspondant

$$\Sigma(A, \mathbf{b}) = \{x \in \mathbb{R}^2 \mid (\exists b \in \mathbf{b})(Ax = b)\}.$$

Comme dans l'exemple précédent, nous étudions l'équation $A\mathbf{x} = \mathbf{b}$ pour construire une approximation intérieure de ce "united solution set". Cependant, toutes les méthodes de résolution présentées précédemment divergent lorsqu'elles sont appliquées à ce "united solution set". Pour analyser ce comportement, nous suivons l'idée proposée dans Shary[92] : nous réécrivons l'équation d'intervalles sous la forme d'une équation réelle dans \mathbb{R}^4 en explicitant les opérations de l'arithmétique de Kaucher. Dans ce cas, la matrice d'intervalles A étant dégénérée, nous obtenons une équation linéaire de \mathbb{R}^4 :

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\bar{\mathbf{b}}_1 \\ -\bar{\mathbf{b}}_2 \\ \underline{\mathbf{b}}_1 \\ \underline{\mathbf{b}}_2 \end{pmatrix},$$

où $\mathbf{x}_1 = [u, v]$ et $\mathbf{x}_2 = [y, z]$. La matrice correspondant à ce système d'équations linéaires est singulière (vérifier que $LA = L1 + L3 - L2$) ce qui explique la divergence des méthodes de résolution. Il y a donc deux cas possibles :

- le membre de droite est compatible avec la dépendance des équations, c'est à dire $\text{rad } \mathbf{b}_1 = \text{rad } \mathbf{b}_2$. Dans ce cas, il y a une infinité de solutions à l'équation $A\mathbf{x} = \mathbf{b}$. Si par exemple $\mathbf{b}_1 = \mathbf{b}_2 = [-1, 1]$, ces solutions peuvent être décrites par la paramétrisation suivante (communication personnelle de S.P. Shary) :

$$\mathbf{x}(t) = \frac{1}{2} \begin{pmatrix} [t - 1, 1 - t] \\ [-t - 1, t + 1] \end{pmatrix}.$$

¹⁷"Squeezing of parameters" en anglais.

Dans cette infinité de solutions, celles qui satisfont $-1 \leq t \leq 1$ sont propres, et ainsi sont des approximations intérieures du "united solution set" étudié.

- le membre de droite n'est pas compatible avec la dépendance des équations, i.e. $\text{rad } \mathbf{b}_1 \neq \text{rad } \mathbf{b}_2$. Dans ce cas, l'équation $\mathbf{A}\mathbf{x} = \mathbf{b}$ n'a pas de solution. Donc, la technique de bisection des paramètres doit être appliquée avant d'obtenir une infinité de solutions.

Comme illustré par ces deux exemples, la technique de bisection des paramètres est parfois nécessaire pour construire des approximations intérieures des "AE-solution sets" à partir de l'équation (dual \mathbf{A}) $\mathbf{x} = \mathbf{b}$. Cependant, cette technique est difficile à utiliser dans des situations plus compliquées car nous n'avons pas d'information sur les modifications à opérer pour que l'équation (dual \mathbf{A}) $\mathbf{x} = \mathbf{b}$ ait une solution propre. Une recherche exhaustive est proscrite à cause de sa complexité et aucune heuristique n'a été proposée jusqu'à aujourd'hui. Le preconditionnement proposé dans la prochaine section va apporter une solution partielle à ces problèmes.

7.2.2 Équations auxiliaires d'intervalles preconditionnées à droite

Étant donnés $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$, $C \in \mathbb{R}^{n \times n}$ régulière et $\tilde{x} \in \mathbb{R}^n$, les relations entre les deux équations preconditionnées à droite suivantes et le "AE-solution set" $\Sigma(\mathbf{A}, \mathbf{b})$ sont maintenant étudiées :

$$((\text{pro } \mathbf{A})C)\mathbf{u} = \mathbf{b} - \mathbf{A}\tilde{x}; \quad (7.1)$$

$$\mathbf{u} = (I - (\text{pro } \mathbf{A})C)\mathbf{u} + \mathbf{b} - \mathbf{A}\tilde{x}. \quad (7.2)$$

Remarque. Les deux équations d'intervalles précédentes sont écrites différemment dans Goldsztejn[5] à cause du changement de convention pour l'écriture des "AE-solution sets". Ces nouvelles écritures ont l'avantage de ne faire intervenir que des intervalles propres lorsqu'elles sont utilisées pour approximer des "united solution sets", i.e. $\mathbf{A} \in \mathbb{I}\mathbb{R}^{n \times n}$ et $\mathbf{b} \in \mathbb{I}\mathbb{R}^n$.

Les paramètres C et \tilde{x} peuvent être choisis quelconques, mais les choix naturels $C = (\text{mid } \mathbf{A})^{-1}$ et $\tilde{x} = (\text{mid } \mathbf{A})^{-1}(\text{mid } \mathbf{b})$ seront étudiés plus en détails à la fin de la section et dans la prochaine section. Les deux propositions suivantes présentent des résultats intermédiaires qui seront utilisés pour l'interprétation des solutions des deux équations auxiliaires (7.1) et (7.2).

Proposition III.2. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$ et $C \in \mathbb{R}^{n \times n}$ régulière.

1. Si \mathbf{A} est impropre alors $\{Cu \mid u \in \Sigma(\mathbf{A}C, \mathbf{b})\} \subseteq \Sigma(\mathbf{A}, \mathbf{b})$.
2. Si \mathbf{A} est propre alors $\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \{Cu \mid u \in \Sigma(\mathbf{A}C, \mathbf{b})\}$.

Démonstration. (1) Considérons un vecteur $u \in \mathbb{R}^n$. D'après le Théorème II.10 page 69,

$$u \in \Sigma(\mathbf{A}C, \mathbf{b}) \implies \text{dual}(\mathbf{A}C)u \subseteq \mathbf{b}.$$

En remarquant que $\text{dual}(\mathbf{A}C)$ est propre car \mathbf{A} est impropre, nous pouvons utiliser la sous-distributivité et nous obtenons

$$(\text{dual } \mathbf{A})(Cu) \subseteq (\text{dual } \mathbf{A}C)u.$$

Finalement, nous obtenons $(\text{dual } \mathbf{A})(Cu) \subseteq \mathbf{b}$, qui implique $Cu \in \Sigma(\mathbf{A}, \mathbf{b})$ par le Théorème II.10. En résumé, $u \in \Sigma(\mathbf{AC}, \mathbf{b})$ implique $Cu \in \Sigma(\mathbf{A}, \mathbf{b})$, ce qui est équivalent à l'énoncé de la proposition.

(2) La démonstration du deuxième énoncé est faite de manière similaire en utilisant la super-distributivité : on peut alors vérifier que $Cu \in \Sigma(\mathbf{A}, \mathbf{b})$ implique $u \in \Sigma(\mathbf{AC}, \mathbf{b})$. \square

Le lemme suivant sera utilisé dans la démonstration de la Proposition III.3.

Lemme III.2. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $\mathbf{a} \in \mathbb{K}\mathbb{R}$. Alors,

1. pour tout $x \in \mathbb{R}$ et $\tilde{x} \in \mathbb{R}$, $(\text{imp } \mathbf{a})(x - \tilde{x}) + \mathbf{a}\tilde{x} \subseteq \mathbf{a}x \subseteq (\text{pro } \mathbf{a})(x - \tilde{x}) + \mathbf{a}\tilde{x}$;
2. pour tout $x \in \mathbb{R}^n$ et $\tilde{x} \in \mathbb{R}^n$, $(\text{imp } \mathbf{A})(x - \tilde{x}) + \mathbf{A}\tilde{x} \subseteq \mathbf{A}x \subseteq (\text{pro } \mathbf{A})(x - \tilde{x}) + \mathbf{A}\tilde{x}$.

Démonstration. (1) La distributivité conditionnelle appliquée au produit $\mathbf{a}((x - \tilde{x}) + \tilde{x})$ donne les inclusions suivantes :

$$(\text{imp } \mathbf{a})(x - \tilde{x}) + \mathbf{a}\tilde{x} \subseteq \mathbf{a}((x - \tilde{x}) + \tilde{x}) \subseteq (\text{pro } \mathbf{a})(x - \tilde{x}) + \mathbf{a}\tilde{x}.$$

Il reste juste à remarquer que $\mathbf{a}((x - \tilde{x}) + \tilde{x}) = \mathbf{a}x$.

(2) Grâce au premier cas, $\forall i \in [1..n]$ et $\forall j \in [1..n]$,

$$(\text{imp } \mathbf{A}_{ij})(x_j - \tilde{x}_j) + \mathbf{A}_{ij}\tilde{x}_j \subseteq \mathbf{A}_{ij}x_j \subseteq (\text{pro } \mathbf{A}_{ij})(x_j - \tilde{x}_j) + \mathbf{A}_{ij}\tilde{x}_j.$$

Ensuite, en utilisant la croissance de l'addition par rapport à l'inclusion, nous pouvons sommer les inclusions précédentes pour $j \in [1..n]$. Nous obtenons donc $\forall i \in [1..n]$,

$$\begin{aligned} \sum_{j \in [1..n]} (\text{imp } \mathbf{A}_{ij})(x_j - \tilde{x}_j) + \sum_{j \in [1..n]} \mathbf{A}_{ij}\tilde{x}_j \\ \subseteq \\ \sum_{j \in [1..n]} \mathbf{A}_{ij}x_j \\ \subseteq \\ \sum_{j \in [1..n]} (\text{pro } \mathbf{A}_{ij})(x_j - \tilde{x}_j) + \sum_{j \in [1..n]} \mathbf{A}_{ij}\tilde{x}_j. \end{aligned}$$

Cela correspond finalement aux inclusions écrites sous forme de produits matricielles dans l'énoncé. \square

Vient alors le dernier résultat intermédiaire. Le translaté d'un ensemble $E \subseteq \mathbb{R}^n$ par un vecteur $\tilde{x} \in \mathbb{R}^n$ est noté $\tilde{x} + E$.

Proposition III.3. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$ et $\tilde{x} \in \mathbb{R}^n$. Alors,

$$\tilde{x} + \Sigma(\text{imp } \mathbf{A}, \mathbf{b} - \mathbf{A}\tilde{x}) \subseteq \Sigma(\mathbf{A}, \mathbf{b}) \subseteq \tilde{x} + \Sigma(\text{pro } \mathbf{A}, \mathbf{b} - \mathbf{A}\tilde{x}).$$

Démonstration. Considérons d'abord la première inclusion et un quelconque x vérifiant

$$x \in \tilde{x} + \Sigma(\text{imp } \mathbf{A}, \mathbf{b} - \mathbf{A}\tilde{x}).$$

Remarquons que cette appartenance est exprimée de manière équivalente par

$$(x - \tilde{x}) \in \Sigma(\text{imp } \mathbf{A}, \mathbf{b} - \mathbf{A}\tilde{x}).$$

Le Théorème II.10 prouve alors que $(\text{pro } \mathbf{A})(x - \tilde{x}) \subseteq \mathbf{b} - \mathbf{A}\tilde{x}$. Ensuite, en additionnant $(\text{dual } \mathbf{A})\tilde{x}$ à chaque membre de l'inclusion et en remarquant que $\mathbf{A}\tilde{x} - (\text{dual } \mathbf{A})\tilde{x} = 0$, nous obtenons $(\text{pro } \mathbf{A})(x - \tilde{x}) + (\text{dual } \mathbf{A})\tilde{x} \subseteq \mathbf{b}$. Finalement, le Lemme III.2 (appliqué avec $(\text{dual } \mathbf{A})$) montre que $(\text{dual } \mathbf{A})x \subseteq \mathbf{b}$, ce qui implique $x \in \Sigma(\mathbf{A}, \mathbf{b})$ par le Théorème II.10. La seconde inclusion est démontrée de la même manière. \square

Les relations entre les équations auxiliaires (7.1) et (7.2) et le "AE-solution set" $\Sigma(\mathbf{A}, \mathbf{b})$ sont maintenant présentées. Le prochain théorème prouve que la résolution de l'équation (7.1) permet de construire une boîte penchée incluse dans le "AE-solution set" $\Sigma(\mathbf{A}, \mathbf{b})$.

Théorème III.2. *Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$, $C \in \mathbb{R}^{n \times n}$ régulière et $\tilde{x} \in \mathbb{R}^n$. Soit $\mathbf{u} \in \mathbb{I}\mathbb{R}^n$ une solution propre de l'équation*

$$((\text{pro } \mathbf{A})C)\mathbf{u} = \mathbf{b} - \mathbf{A}\tilde{x}.$$

Alors, la boîte penchée $\{\tilde{x} + Cu \mid u \in \mathbf{u}\}$ est incluse dans $\Sigma(\mathbf{A}, \mathbf{b})$.

Démonstration. Comme \mathbf{u} est une solution propre de l'équation $(\text{dual } \mathbf{A}')\mathbf{u} = \mathbf{b}'$, où $\mathbf{A}' := (\text{imp } \mathbf{A})C$ et $\mathbf{b}' := \mathbf{b} - \mathbf{A}\tilde{x}$, le Théorème II.11 (page 72) prouve l'inclusion suivante :

$$\mathbf{u} \subseteq \Sigma(\mathbf{A}', \mathbf{b}').$$

Alors, \mathbf{A}' étant impropre, la Proposition III.2 peut-être utilisée pour démontrer

$$\{Cu \mid u \in \Sigma(\mathbf{A}', \mathbf{b}')\} \subseteq \Sigma(\text{imp } \mathbf{A}, \mathbf{b}').$$

En utilisant l'inclusion $\mathbf{u} \subseteq \Sigma(\mathbf{A}', \mathbf{b}')$, nous obtenons :

$$\{Cu \mid u \in \mathbf{u}\} \subseteq \Sigma(\text{imp } \mathbf{A}, \mathbf{b}').$$

En traduisant ces deux sous-ensembles de \mathbb{R}^n du vecteur \tilde{x} , nous obtenons l'inclusion suivante :

$$\{\tilde{x} + Cu \mid u \in \mathbf{u}\} \subseteq \tilde{x} + \Sigma(\text{imp } \mathbf{A}, \mathbf{b}').$$

Finalement, la Proposition III.3 démontre que $\tilde{x} + \Sigma(\text{imp } \mathbf{A}, \mathbf{b}') \subseteq \Sigma(\mathbf{A}, \mathbf{b})$, ce qui conclut la démonstration. \square

Le prochain théorème prouve que la résolution de l'équation (7.2) permet de construire une boîte penchée qui contient le "AE-solution set" $\Sigma(\mathbf{A}, \mathbf{b})$.

Théorème III.3. *Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$, $C \in \mathbb{R}^{n \times n}$ régulière et $\tilde{x} \in \mathbb{R}^n$. Supposons que*

$$\rho(|I - (\text{pro } \mathbf{A})C|) < 1.$$

Alors, L'équation suivante a une unique solution $\mathbf{u} \in \mathbb{K}\mathbb{R}^n$:

$$\mathbf{u} = (I - (\text{pro } \mathbf{A})C)\mathbf{u} + \mathbf{b} - \mathbf{A}\tilde{x}.$$

De plus, si la solution \mathbf{u} est propre alors $\{\tilde{x} + Cu \mid u \in \mathbf{u}\} \subseteq \Sigma(\mathbf{A}, \mathbf{b})$; si la solution \mathbf{u} n'est pas propre alors $\Sigma(\mathbf{A}, \mathbf{b}) = \emptyset$.

Démonstration. L'intervalle \mathbf{u} est une solution de l'équation $\mathbf{u} = (I - \mathbf{A}')\mathbf{u} + \mathbf{b}'$ où $\mathbf{A}' = (\text{pro } \mathbf{A})C$ et $\mathbf{b}' = \mathbf{b} - \mathbf{A}\tilde{x}$. Comme $\rho(|I - \mathbf{A}'|) < 1$, le Théorème II.12 (page 72) montre que l'équation a une unique solution. Supposons que cette solution soit propre. Le Théorème II.12 (page 72) montre alors que

$$\Sigma(\mathbf{A}', \mathbf{b}') \subseteq \mathbf{u}.$$

Alors, \mathbf{A}' étant propre, nous pouvons appliquer la Proposition III.2 (page 90) qui montre que

$$\Sigma(\text{pro } \mathbf{A}, \mathbf{b}') \subseteq \{Cu \mid u \in \Sigma(\mathbf{A}', \mathbf{b}')\}.$$

En utilisant l'inclusion $\Sigma(\mathbf{A}', \mathbf{b}') \subseteq \mathbf{u}$, nous obtenons

$$\Sigma(\text{pro } \mathbf{A}, \mathbf{b}') \subseteq \{Cu \mid u \in \mathbf{u}\}.$$

En translatant ces deux sous-ensemble de \mathbb{R}^n du vecteur \tilde{x} , nous obtenons l'inclusion suivante :

$$\tilde{x} + \Sigma(\text{pro } \mathbf{A}, \mathbf{b}') \subseteq \{\tilde{x} + Cu \mid u \in \mathbf{u}\}.$$

Finalement, la Proposition III.3 (page 91) démontre que $\Sigma(\mathbf{A}, \mathbf{b}) \subseteq \tilde{x} + \Sigma(\text{pro } \mathbf{A}, \mathbf{b}')$, ce qui conclut cette partie de la preuve. Supposons maintenant que \mathbf{u} ne soit pas propre. Alors, le Théorème II.12 (page 72) démontre que $\Sigma(\mathbf{A}', \mathbf{b}') = \emptyset$. Donc, en utilisant la Proposition III.2 (page 90) $\Sigma(\text{pro } \mathbf{A}, \mathbf{b}')$ est aussi vide car $\{Cu \mid u \in \emptyset\} = \emptyset$. Finalement, en utilisant la Proposition III.3 (page 91) nous pouvons conclure que $\Xi(\mathbf{A}, \mathbf{b})$ est vide. \square

Comme dans le cas du preconditionnement à gauche, les matrices fortement régulières vont permettre de décrire une classe importante de "AE-solution sets" qui peuvent être estimés grâce aux deux équations preconditionnées à droite qui ont été introduites. Avant cela, il faut montrer que les matrices fortement régulières, qui ont été introduites dans le contexte du preconditionnement à gauche, peuvent aussi être utilisées dans le contexte du preconditionnement à droite. La proposition suivante est similaire à la proposition 4.1.1 de Neumaier[8] mais dédiée au preconditionnement à droite.

Proposition III.4. *Soit $\mathbf{A} \in \mathbb{R}^{n \times n}$ telle que $\text{mid } \mathbf{A}$ est régulière. Alors les conditions suivantes sont équivalentes :*

1. \mathbf{A} est fortement régulière ;
2. \mathbf{A}^T est fortement régulière ;
3. $\mathbf{A}(\text{mid } \mathbf{A})^{-1}$ est régulière ;
4. $\rho((\text{rad } \mathbf{A}) | (\text{mid } \mathbf{A})^{-1}) < 1$;
5. il existe un vecteur $u \in \mathbb{R}^n$ strictement positif tel que $\|I - \mathbf{A}(\text{mid } \mathbf{A})^{-1}\|_v < 1$;
6. $\mathbf{A}(\text{mid } \mathbf{A})^{-1}$ est une H-matrice.

Démonstration. (1 \Leftrightarrow 2) : idem proposition 4.1.1 de Neumaier[8].

(2 \Leftrightarrow 3) : Grâce au premier cas, \mathbf{A} est fortement régulière si et seulement si \mathbf{A}^T est fortement régulière. Par définition, cela est vrai si et seulement si $(\text{mid } \mathbf{A}^T)^{-1} \mathbf{A}^T$ est régulière. Donc, comme $(\text{mid } \mathbf{A}^T)^{-1} = ((\text{mid } \mathbf{A})^{-1})^T$, cela est vrai si et seulement si $((\text{mid } \mathbf{A})^{-1})^T \mathbf{A}^T$ est régulière. Une matrice d'intervalles est régulière si et seulement si sa transposée est régulière, donc \mathbf{A} est fortement régulière si et seulement si $((\text{mid } \mathbf{A})^{-1})^T \mathbf{A}^T$ est régulière. Cela est finalement vrai si et seulement si $\mathbf{A}(\text{mid } \mathbf{A})^{-1}$ est régulière.

(3 \Rightarrow 4), (4 \Rightarrow 5), (5 \Rightarrow 6) et (6 \Rightarrow 3) sont prouvées de manière similaire à Neumaier[8], en changeant la définition de \mathbf{B}_0 en $\mathbf{B}_0 = \mathbf{A}\hat{\mathbf{A}}^{-1}$. \square

A partir de maintenant et dans le chapitre suivant, le preconditionnement sera effectué en utilisant $C = (\text{mid } \mathbf{A})^{-1}$ et $\tilde{x} = (\text{mid } \mathbf{A})^{-1}(\text{mid } \mathbf{b})$. La proposition précédente montre que les "AE-solution sets" $\Sigma(\mathbf{A}, \mathbf{b})$ tels que $(\text{pro } \mathbf{A})$ est fortement régulière peuvent être preconditionnés à droite en utilisant les deux équations auxiliaires d'intervalles proposées dans cette section :

Approximation intérieure

Si $\text{pro } \mathbf{A}$ est fortement régulière, la Proposition III.4 prouve que

$$(\text{pro } \mathbf{A})(\text{mid } \mathbf{A})^{-1}$$

est une H-matrice. Dans ce cas, le Théorème III.1 montre que l'unique solution de l'équation (7.1) peut-être calculée facilement en écrivant l'équation sous forme de point fixe. Donc, les problèmes d'existence d'une infinité de solutions ou de pas de solution du tout ne peuvent pas être rencontrés en utilisant l'équation preconditionnée à droite si $\text{pro } \mathbf{A}$ est fortement régulière. Comme dans le cas de l'équation non-preconditionnée, il se peut que le "AE-solution set" soit non-vide et que l'unique solution de l'équation preconditionnée à droite ne soit pas propre. Dans le cas où la matrice \mathbf{A} est dégénérée, ce problème ne se rencontre pas : l'équation (7.1) est alors simplifiée de la manière suivante :

$$\mathbf{u} = \mathbf{b} - A\tilde{x}.$$

La solution \mathbf{u} est alors toujours propre car $A\tilde{x} = (\text{mid } \mathbf{b}) \subseteq \mathbf{b}$ est toujours vrai (c'est à dire $0 \subseteq \mathbf{b} - A\tilde{x}$ avec $\tilde{x} = (\text{mid } \mathbf{A})^{-1} \text{mid } \mathbf{b}$). De plus, la boîte penchée

$$\{\tilde{x} + A^{-1}u | u \in \mathbf{b} - \text{mid } \mathbf{b}\} = \{A^{-1}u | u \in \mathbf{b}\}$$

est exactement le "AE-solution set" approximé. Cependant, la situation est plus compliquée si la matrice \mathbf{A} n'est pas dégénérée, comme c'est illustré dans la section 7.3.2 où des expérimentations sont faites.

Approximation extérieure

Si $\text{pro } \mathbf{A}$ est fortement régulière, la Proposition III.4 montre que

$$\rho(|I - (\text{pro } \mathbf{A})(\text{mid } \mathbf{A})^{-1}|) < 1.$$

Dans ce cas, le Théorème II.15 montre que l'équation (7.2) a une unique solution qui peut être calculée par itération de l'équation mise sous forme de point fixe. Dans le cas où la matrice est dégénérée, la solution de l'équation est

$$\mathbf{u} = \mathbf{b} - A\tilde{x},$$

ce qui conduit aussi au "united solution set" exact.

7.3 Sur la qualité du preconditionnement à droite

Cette section présente deux résultats théoriques sur le preconditionnement à droite et des expérimentations sur des exemples académiques de "AE-solution sets" linéaires. Ces études sont conduites en utilisant le preconditionnement $C = (\text{mid } \mathbf{A})^{-1}$ et $\tilde{x} = (\text{mid } \mathbf{A})^{-1}(\text{mid } \mathbf{b})$

7.3.1 Étude théorique

Cette section présente deux résultats théoriques qui mettent en valeur certains avantages du préconditionnement à droite :

- lorsque les approximations intérieures et extérieures sont obtenues, la distance de Hausdorff entre les deux est majorée ;
- lorsque la matrice \mathbf{A} est propre (par exemple dans le cas des "united solution sets" et "controllable solution sets") il est montré que l'approximation extérieure grâce à l'équation préconditionnée à droite est toujours meilleure que l'approximation extérieure grâce à l'équation préconditionnée à gauche ;

Pour un vecteur strictement positif $v \in \mathbb{R}^n$, les normes suivantes sont définies (voir par exemple Neumaier[8]) :

$$\|x\|_v = \max_{i \in [1..n]} \frac{|x_i|}{v_i} \quad \text{and} \quad \|M\|_v = \max_{i \in [1..n]} \sum_{j \in [1..n]} |M_{ij}| \frac{v_j}{v_i}.$$

Remarquons que $\|x\| = \|x\|_v$ pour $v = (1, \dots, 1)^T$ et que $\|x\| \leq \|v\| \|x\|_v$. Rappelons que la distance de Hausdorff entre $E \subseteq \mathbb{R}^n$ et $E' \subseteq \mathbb{R}^n$ est définie par

$$\text{dist}_H(E, E') = \max \left\{ \max_{x \in E} \min_{x' \in E'} \|x - x'\|, \max_{x' \in E'} \min_{x \in E} \|x - x'\| \right\},$$

et qu'il s'agit de la distance utilisée pour les intervalles.

Résultats préliminaires

Le lemme suivant donne une majoration de la distance de Hausdorff entre deux boîtes penchées dans la situation spécifique qui sera rencontrée dans les développements à venir.

Lemme III.3. *Soient $M \in \mathbb{R}^{n \times n}$, $\tilde{x} \in \mathbb{R}^n$, $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^n$ tels que $\mathbf{u} \subseteq \mathbf{u}'$. Notons $\{\tilde{x} + Mu | u \in \mathbf{u}\}$ et $\{\tilde{x} + Mu' | u' \in \mathbf{u}'\}$ respectivement par \mathbf{s} et \mathbf{s}' . Alors,*

$$\text{dist}_H(\mathbf{s}, \mathbf{s}') \leq \|M\| \text{dist}_H(\mathbf{u}, \mathbf{u}').$$

Démonstration. Par définition de la distance de Hausdorff,

$$\text{dist}_H(\mathbf{s}, \mathbf{s}') = \max \left\{ \max_{x \in \mathbf{s}} \min_{x' \in \mathbf{s}'} \|x - x'\|, \max_{x' \in \mathbf{s}'} \min_{x \in \mathbf{s}} \|x - x'\| \right\}.$$

Comme $\mathbf{s} \subseteq \mathbf{s}'$ parce que $\mathbf{u} \subseteq \mathbf{u}'$, nous avons $\max_{x \in \mathbf{s}} \min_{x' \in \mathbf{s}'} \|x - x'\| = 0$ et ainsi

$$\text{dist}_H(\mathbf{s}, \mathbf{s}') = \max_{x' \in \mathbf{s}'} \min_{x \in \mathbf{s}} \|x - x'\|.$$

En utilisant la définition d'une boîte penchée, l'égalité précédente s'écrit

$$\text{dist}_H(\mathbf{s}, \mathbf{s}') = \max_{u' \in \mathbf{u}'} \min_{u \in \mathbf{u}} \|Mu + \tilde{x} - (Mu' + \tilde{x})\| = \max_{u' \in \mathbf{u}'} \min_{u \in \mathbf{u}} \|M(u - u')\|.$$

Maintenant, comme $\|M(u - u')\| \leq \|M\| \|u - u'\|$ et parce qu'effectuer un minimum ou un maximum sur des valeurs plus grandes donne un résultat plus grand, nous obtenons

$$\text{dist}_H(\mathbf{s}, \mathbf{s}') \leq \max_{u' \in \mathbf{u}'} \min_{u \in \mathbf{u}} \|M\| \|u - u'\|.$$

Finalement, l'énoncé du lemme est obtenu en factorisant le nombre positif $\|M\|$ et en remarquant que $\text{dist}_H(\mathbf{u}, \mathbf{u}') = \max_{u' \in \mathbf{u}'} \min_{u \in \mathbf{u}} \|u - u'\|$ parce que $\mathbf{u} \subseteq \mathbf{u}'$. \square

Les deux lemmes suivants donnent des propriétés sur les milieux et rayons des opérations sur les intervalles.

Lemme III.4. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $\mathbf{B} \in \mathbb{K}\mathbb{R}^{n \times n}$. Notons $\text{mid } \mathbf{A}$ et $\text{mid } \mathbf{B}$ respectivement par $\hat{\mathbf{A}}$ et $\hat{\mathbf{B}}$. Alors,

- $\text{mid}(\mathbf{A} \pm \mathbf{B}) = \hat{\mathbf{A}} \pm \hat{\mathbf{B}}$;
- $\text{rad}(\mathbf{A} \pm \mathbf{B}) = (\text{rad } \mathbf{A}) + (\text{rad } \mathbf{B})$;
- $\text{mid}(\mathbf{A}\mathbf{B}) = \hat{\mathbf{A}}\hat{\mathbf{B}}$ si \mathbf{A} ou \mathbf{B} est dégénérée ;
- $\text{mid}(\mathbf{A}\mathbf{B}) = 0$ si $\hat{\mathbf{A}} = 0$ ou $\hat{\mathbf{B}} = 0$, où 0 est la matrice nulle.

Si \mathbf{A} et \mathbf{B} sont propres,

- $\text{rad}(\mathbf{A}\mathbf{B}) = |\mathbf{A}|(\text{rad } \mathbf{B})$ si \mathbf{A} est dégénérée ou $\hat{\mathbf{B}} = 0$;
- $\text{rad}(\mathbf{A}\mathbf{B}) = (\text{rad } \mathbf{A})|\mathbf{B}|$ si \mathbf{B} est dégénérée ou $\hat{\mathbf{A}} = 0$.

Démonstration. Les trois premières relations sont des conséquences directes de l'arithmétique de Kaucher. Remarquons que les rayons mis en jeu dans la seconde relation peuvent être négatifs : par exemple $[-1, 1] + [5, 0] = [4, 1]$ et la relation est satisfaite car $2 + (-5) = -3$. La quatrième relation est vraie car le produit de deux intervalles est centré sur 0 lorsque l'un des deux au moins est centré sur 0. La cinquième, restreinte aux matrices propres, est prise de Neumaier[8] (proposition 3.1.12) et la sixième est démontrée de manière similaire à la cinquième. \square

Remarque. Les propriétés énoncées par le lemme précédent sont aussi valables pour l'addition de vecteurs d'intervalles généralisés, ainsi que pour la multiplication entre matrices et vecteurs d'intervalles généralisés.

Lemme III.5. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$ tels que $\text{mid } \mathbf{A} = I$, $0 \notin (\text{pro } \mathbf{A}_{ii})$ pour tout $i \in [1..n]$ et $\text{mid } \mathbf{b} = 0$. Alors toute solution \mathbf{x} de l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ vérifie $(\text{mid } \mathbf{x}) = 0$.

Démonstration. Comme $0 \notin \mathbf{A}_{ii}$, l'équation $(\text{dual } \mathbf{A})\mathbf{x} = \mathbf{b}$ s'écrit de manière équivalente :

$$\mathbf{x} = \mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x})).$$

où \mathbf{D} , \mathbf{L} et \mathbf{U} sont définies comme dans le Théorème II.11. Alors, comme $\mathbf{L} + \mathbf{U}$ est centrée sur 0, le Lemme III.4 prouve que $(\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x})$ est aussi centré sur 0. Alors, comme \mathbf{b} est centré sur 0, $\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x})$ est aussi centré sur 0 et finalement $\mathbf{D}(\mathbf{b} - (\mathbf{L} + \mathbf{U})(\text{dual } \mathbf{x})) = \mathbf{x}$ est aussi centré sur 0. \square

Finalement, le prochain lemme donne une majoration sur la distance entre les solutions respectives de deux équations linéaires réelles qui seront rencontrées dans la suite.

Lemme III.6. Soient $\Delta \in \mathbb{R}^{n \times n}$ non-négative, $v \in \mathbb{R}^n$ strictement positif tels que $\|\Delta\|_v < 1$, et $\beta \in \mathbb{R}^n$. Si $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$ vérifient respectivement

$$(I - \Delta)x = \beta \quad \text{et} \quad (I + \Delta)y = \beta$$

alors

$$\|x - y\| \leq 2\|v\| \frac{\|\Delta\|_v \|\beta\|_v}{1 - \|\Delta\|_v^2}.$$

Démonstration. Ce lemme repose sur les séries de Neumann (voir Neumaier[8]) : si $\|\Delta\|_v < 1$ (ce qui implique $\| -\Delta \|_v < 1$) alors $I - \Delta$ et $I + \Delta$ sont inversibles et leurs inverses s'expriment de la manière suivante :

$$(I - \Delta)^{-1} = \sum_{k=0}^{\infty} \Delta^k \quad \text{et} \quad (I + \Delta)^{-1} = \sum_{k=0}^{\infty} (-1)^k \Delta^k.$$

Maintenant, comme Δ est non-négative, nous avons

$$\sum_{k=0}^{\infty} |(-1)^k \Delta^k| = \sum_{k=0}^{\infty} |\Delta^k| = \sum_{k=0}^{\infty} \Delta^k.$$

La dernière série étant convergente, $\sum_{k=0}^{\infty} \Delta^k$ et $\sum_{k=0}^{\infty} (-1)^k \Delta^k$ sont absolument convergentes. Par conséquent, nous pouvons utiliser l'associativité et la commutativité pour réarranger les termes de la série $(I - \Delta)^{-1} - (I + \Delta)^{-1}$:

$$(I - \Delta)^{-1} - (I + \Delta)^{-1} = 2\Delta + 2\Delta^3 + 2\Delta^5 + \dots = 2\Delta \sum_{k=0}^{\infty} \Delta^{2k}.$$

Par hypothèse $\|\Delta\|_v < 1$, donc la somme géométrique $\sum_{k=0}^{\infty} \left((\|\Delta\|_v)^2 \right)^k$ converge et

$$\sum_{k=0}^{\infty} \left((\|\Delta\|_v)^2 \right)^k = \frac{1}{1 - (\|\Delta\|_v)^2}.$$

Nous avons donc

$$\|(I - \Delta)^{-1} - (I + \Delta)^{-1}\|_v \leq 2\|\Delta\|_v \frac{1}{1 - (\|\Delta\|_v)^2}.$$

Alors, $x - y = ((I - \Delta)^{-1} - (I + \Delta)^{-1})\beta$ et l'inégalité suivante est vraie :

$$\begin{aligned} \|x - y\|_v &= \|((I - \Delta)^{-1} - (I + \Delta)^{-1})\beta\|_v \\ &\leq 2\|\Delta\|_v \frac{1}{1 - (\|\Delta\|_v)^2} \|\beta\|_v. \end{aligned}$$

Finalement, l'énoncé est obtenu grâce à $\|x - y\| \leq \|v\| \|x - y\|_v$. □

Majoration de la distance de Hausdorff entre les approximations intérieures et extérieures

Le prochain théorème donne un résultat général sur la qualité des approximations par les boîtes penchées construites grâce aux équations (7.1) et (7.2). On peut noter que ce résultat n'est valable que lorsque une approximation intérieure est disponible, c'est à dire lorsque l'équation (7.1) a une solution propre. Un résultat intéressant serait de trouver une condition suffisante pour obtenir une solution propre à cette équation, mais cela reste à faire.

Théorème III.4. *Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$ telle que $\text{pro } \mathbf{A}$ est fortement régulière et $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$. Considérons les équations (7.1) et (7.2) avec*

$$C = \hat{\mathbf{A}}^{-1} \quad \text{et} \quad \tilde{x} = \hat{\mathbf{A}}^{-1}(\text{mid } \mathbf{b}),$$

où $\hat{\mathbf{A}} = \text{mid } \mathbf{A}$. Supposons de plus que les solutions respectives \mathbf{u} et \mathbf{u}' de ces équations sont propres. Alors, quelque soit le vecteur strictement positif $v \in \mathbb{R}^n$ vérifiant¹⁸

$$\|I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}\|_v < 1,$$

la distance de Hausdorff entre les approximations intérieures $\mathbf{s} := \{\tilde{x} + Cu | u \in \mathbf{u}\}$ et extérieures $\mathbf{s}' := \{\tilde{x} + Cu | u \in \mathbf{u}'\}$ est majorée de la manière suivante :

$$\text{dist}_H(\mathbf{s}, \mathbf{s}') \leq 2 \|v\| \|\hat{\mathbf{A}}^{-1}\| \frac{\|I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}\|_v \|\mathbf{b} - \mathbf{A}\tilde{x}\|_v}{1 - (\|I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}\|_v)^2}.$$

Démonstration. Les intervalles propres \mathbf{u} et \mathbf{u}' vérifient respectivement les équations (7.1) et (7.2), c'est à dire

$$\begin{aligned} ((\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1})\mathbf{u} &= \mathbf{b} - \mathbf{A}\tilde{x}, \\ \mathbf{u}' &= (I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1})\mathbf{u}' + \mathbf{b} - \mathbf{A}\tilde{x}. \end{aligned}$$

Comme $\tilde{x} = \hat{\mathbf{A}}^{-1}(\text{mid } \mathbf{b})$, nous pouvons utiliser le Lemme III.4 qui montre que $\text{mid}(\mathbf{b} - \mathbf{A}\tilde{x}) = 0$. De plus, le Lemme III.4 implique

$$\text{mid}((\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}) = I \quad \text{et} \quad \text{mid}(I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}) = 0.$$

Nous concluons finalement que $\text{mid } \mathbf{u} = 0$ grâce au Lemme III.5, et que $\text{mid } \mathbf{u}' = 0$ grâce au Lemme III.4. Maintenant, remarquons que $\text{rad}(\mathbf{b} - \mathbf{A}\tilde{x}) = |\mathbf{b} - \mathbf{A}\tilde{x}|$, parce que $\mathbf{b} - \mathbf{A}\tilde{x}$ est centré sur 0, et notons ce vecteur réel $\beta := |\mathbf{b} - \mathbf{A}\tilde{x}|$. Ensuite, en appliquant rad aux deux équations et en utilisant le Lemme III.4 (\mathbf{u} et \mathbf{u}' étant propres et centrés sur 0) nous obtenons les deux équations suivantes :

$$\begin{aligned} |(\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{u}) &= \beta, \\ \text{rad } \mathbf{u}' &= \Delta(\text{rad } \mathbf{u}') + \beta, \end{aligned}$$

où $\Delta := |I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}| = \text{rad}((\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1})$ par le Lemme III.4. Maintenant, par la proposition 3.1.10 de Neumaier[8], nous avons

$$|(\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}| = I + \text{rad}((\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}) = I + \Delta.$$

La situation est résumée par les deux équations réelles suivantes :

$$\begin{aligned} (I + \Delta)(\text{rad } \mathbf{u}) &= \beta, \\ (I - \Delta)(\text{rad } \mathbf{u}') &= \beta. \end{aligned}$$

Nous pouvons maintenant utiliser le Lemme III.6 : pour tout vecteur strictement positif v tel que $\|\Delta\|_v < 1$, nous avons

$$\|(\text{rad } \mathbf{u}) - (\text{rad } \mathbf{u}')\| \leq 2 \frac{\|\Delta\|_v}{1 - \|\Delta\|_v^2} \|\beta\|_v \|v\|.$$

Comme \mathbf{u} et \mathbf{u}' sont centrés sur 0, nous avons $\text{dist}_H(\mathbf{u}, \mathbf{u}') = \|(\text{rad } \mathbf{u}) - (\text{rad } \mathbf{u}')\|$. L'inégalité proposée dans l'énoncé du théorème est finalement obtenue en utilisant le Lemme III.3 qui prouve que $\text{dist}_H(\mathbf{s}, \mathbf{s}') \leq \|\hat{\mathbf{A}}^{-1}\| \text{dist}_H(\mathbf{u}, \mathbf{u}')$. \square

¹⁸Il existe bien de tels vecteurs v car $\text{pro } \mathbf{A}$ est fortement régulière, voir Proposition III.4.

Remarque. En pratique, le théorème précédent peut être utilisé avec $v = (1, \dots, 1)^T$, et donc $\|v\| = 1$ et $\|\cdot\|_v = \|\cdot\|$. Cela peut être fait à condition que $\|I - (\text{pro } \mathbf{A})(\text{mid } \mathbf{A})^{-1}\| < 1$, c'est à dire que à condition que $(\text{pro } \mathbf{A})(\text{mid } \mathbf{A})^{-1}$ soit à diagonale strictement dominante. Cette condition est plus restrictive que "pro \mathbf{A} est fortement régulière" mais est néanmoins souvent satisfaite.

Comme dans le cas de la majoration proposée dans Neumaier[7], si

$$\|I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}\| \leq \epsilon \quad \text{et} \quad \|\mathbf{b} - \mathbf{A}\tilde{\mathbf{x}}\| \leq \epsilon,$$

où ϵ peut être choisi tel que $\epsilon \ll 1$ si les rayon de pro \mathbf{A} et pro \mathbf{b} sont "assez petits", la majoration proposée devient

$$\text{dist}_H(\mathbf{s}, \mathbf{s}') \leq 2 \|\hat{\mathbf{A}}^{-1}\| \frac{\epsilon^2}{1 - \epsilon^2} \approx 2 \|\hat{\mathbf{A}}^{-1}\| \epsilon^2.$$

Cela prouve que l'approximation par les boites penchées construite par la résolution des équations proposées peut être très précise (convergence quadratique avec les rayons des intervalles mis en jeu). Une extension intéressante de ce théorème serait de considérer une matrice de préconditionnement C différente de $(\text{mid } \mathbf{A})^{-1}$, certainement vérifiant $\|I - (\text{pro } \mathbf{A})C\|_v < 1$.

Comparaison du préconditionnement à gauche et à droite dans certains cas d'approximations extérieures

Le prochain théorème prouve que, dans le cas de l'approximation extérieure d'un "AE-solution set" $\Sigma(\mathbf{A}, \mathbf{b})$ tel que \mathbf{A} est propre, et lorsque $C = (\text{mid } \mathbf{A})^{-1}$ et $\tilde{\mathbf{x}} = (\text{mid } \mathbf{A})^{-1}(\text{mid } \mathbf{b})$, le préconditionnement à droite est toujours plus précis que le préconditionnement à gauche, la convergence de l'algorithme étant similaire dans les deux cas de préconditionnement.

Théorème III.5. Soient $\mathbf{A} \in \mathbb{R}^{n \times n}$ telle que pro \mathbf{A} est fortement régulière et $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$ tels que $\Sigma(\mathbf{A}, \mathbf{b}) \neq \emptyset$. Considérons les équations (5.6) (page 73) et (7.2) (page 90) avec

$$C = (\text{mid } \mathbf{A})^{-1} \quad \text{et} \quad \tilde{\mathbf{x}} = (\text{mid } \mathbf{A})^{-1}(\text{mid } \mathbf{b}).$$

Notons les uniques solutions propres de ces équations respectivement par \mathbf{x} et \mathbf{u} . Alors,

$$\square \{ \tilde{\mathbf{x}} + Cu | u \in \mathbf{u} \} = \mathbf{x}.$$

Ce qui implique trivialement $\{ \tilde{\mathbf{x}} + Cu | u \in \mathbf{u} \} \subseteq \mathbf{x}$, c'est à dire l'approximation par la boite penchée est plus précise que l'approximation par l'intervalle.

Démonstration. Premièrement, remarquons que les solutions des équations (5.6) et (7.2) sont uniques car pro \mathbf{A} est fortement régulière, et sont propres car le "AE-solution set" est supposé non-vide (utiliser les théorèmes II.13 et III.3). Maintenant, en reprenant la preuve du Théorème III.4, nous savons que mid \mathbf{u} et rad \mathbf{u} satisfont les relations suivantes :

$$\begin{aligned} \text{mid } \mathbf{u} &= 0, \\ \text{rad } \mathbf{u} &= \text{rad}((\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1})(\text{rad } \mathbf{u}) + \text{rad}(\mathbf{b} - \mathbf{A}\tilde{\mathbf{x}}). \end{aligned}$$

En remarquant que $(\text{rad pro } \mathbf{A}) = |\text{rad } \mathbf{A}|$ et en utilisant le Lemme III.4, nous obtenons

$$\text{rad}((\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}) = |\text{rad } \mathbf{A}| |\hat{\mathbf{A}}^{-1}|.$$

Nous utilisons maintenant l'hypothèse que \mathbf{A} est propre : nous pouvons utiliser le Lemme III.4 pour obtenir

$$\text{rad}(\mathbf{b} - \mathbf{A}\tilde{x}) = \text{rad } \mathbf{b} + |\text{rad } \mathbf{A}| |\tilde{x}|.$$

Par conséquent, $\text{rad } \mathbf{u}$ vérifie

$$\text{rad } \mathbf{u} = |\text{rad } \mathbf{A}| |\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{u}) + \text{rad } \mathbf{b} + |\text{rad } \mathbf{A}| |\tilde{x}|.$$

Maintenant, multiplions à gauche les deux membres de l'égalité par $|\hat{\mathbf{A}}^{-1}|$ pour obtenir

$$|\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{u}) = |\hat{\mathbf{A}}^{-1}| |\text{rad } \mathbf{A}| |\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{u}) + |\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{b}) + |\hat{\mathbf{A}}^{-1}| |\text{rad } \mathbf{A}| |\tilde{x}|.$$

Notons $\mathbf{y} := \square \{\tilde{x} + \hat{\mathbf{A}}^{-1}\mathbf{u} | \mathbf{u} \in \mathbf{u}\}$. Il est alors facile de vérifier que

$$\mathbf{y} = \tilde{x} + \hat{\mathbf{A}}^{-1}\mathbf{u}.$$

Par conséquent, $\text{mid } \mathbf{y} = \tilde{x}$ et $\text{rad } \mathbf{y} = |\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{u})$. Nous obtenons donc l'équation suivante pour $\text{rad } \mathbf{y}$:

$$\text{rad } \mathbf{y} = |\hat{\mathbf{A}}^{-1}| |\text{rad } \mathbf{A}|(\text{rad } \mathbf{y}) + |\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{b}) + |\hat{\mathbf{A}}^{-1}| |\text{rad } \mathbf{A}| |\tilde{x}|.$$

Recherchons maintenant l'équation vérifiée par $\text{rad } \mathbf{x}$. En remarquant que $|\mathbf{x}| = (\text{rad } \mathbf{x}) + |\text{mid } \mathbf{x}| = (\text{rad } \mathbf{x}) + |\tilde{x}|$, on peut vérifier qu'en appliquant les règles données par le Lemme III.4 à l'équation d'intervalles (5.6), on obtient les équations suivantes :

$$\begin{aligned} \text{mid } \mathbf{x} &= (\text{mid } \mathbf{A})^{-1}(\text{mid } \mathbf{b}) = \tilde{x}, \\ \text{rad } \mathbf{x} &= |\hat{\mathbf{A}}^{-1}| |\text{rad } \mathbf{A}|(\text{rad } \mathbf{x}) + |\hat{\mathbf{A}}^{-1}|(\text{rad } \mathbf{b}) + |\hat{\mathbf{A}}^{-1}| |\text{rad } \mathbf{A}| |\tilde{x}|. \end{aligned}$$

Ainsi, $\text{mid } \mathbf{y} = \text{mid } \mathbf{x}$. De plus, $\text{rad } \mathbf{y}$ et $\text{rad } \mathbf{x}$ vérifient la même équation. Finalement, comme $|\hat{\mathbf{A}}^{-1}| |\text{rad } \mathbf{A}| = |I - \hat{\mathbf{A}}^{-1}(\text{pro } \mathbf{A})|$, cette dernière ayant un rayon spectral inférieur à 1 parce que $\text{pro } \mathbf{A}$ est fortement régulière, le théorème de Schröder prouve que l'équation partagée par $\text{rad } \mathbf{y}$ et $\text{rad } \mathbf{x}$ a une unique solution. Donc, $\text{mid } \mathbf{x} = \text{mid } \mathbf{y}$ et $\text{rad } \mathbf{x} = \text{rad } \mathbf{y}$ ce qui implique $\mathbf{x} = \mathbf{y}$. \square

7.3.2 Expérimentations

Les algorithmes présentés ont été codés en C/C++ : nous avons pris comme base la bibliothèque PROFIL/BIAS (Knueppel[68]) et nous l'avons adaptée pour l'arithmétique de Kaucher. Les temps de calculs ne sont pas indiqués car le nombre d'itérations des algorithmes est plus significative pour la comparaison des preconditionnements proposés. Les équations auxiliaires d'intervalles suivantes sont considérées :

- l'équation non preconditionnée (5.4) (page 72) dédiée à l'approximation intérieure. Sa solution sera notée \mathbf{x}^I quand elle est disponible.
- l'équation preconditionnée à gauche (5.6) (page 73) dédiée à l'approximation extérieure. Sa solution sera notée \mathbf{x}^O quand elle est disponible.
- l'équation preconditionnée à droite (7.1) (page 90) dédiée à l'approximation intérieure. Sa solution sera notée \mathbf{u}^I quand elle est disponible, et la boîte penchée correspondante \mathbf{s}^I .

- l'équation préconditionnée à droite (7.2) (page 90) dédiée à l'approximation extérieure. Sa solution sera notée \mathbf{u}^O quand elle est disponible, et la boîte penchée correspondante \mathbf{s}^O .

Les équations sont préconditionnées avec

$$C = (\text{mid } \mathbf{A})^{-1} \quad \text{and} \quad \tilde{x} = (\text{mid } \mathbf{A})^{-1}(\text{mid } \mathbf{b}).$$

La surface (hypervolume) des boîtes penchées est calculée en utilisant la formule classique $\text{vol}\{Mu|u \in \mathbf{u}\} = (\det M)(\text{vol } \mathbf{u})$ (qui généralise la formule qui permet de calculer le volume d'un intervalle, obtenue en considérant $M = I$).

Le prochain exemple illustre le fait que l'approximation en utilisant des boîtes penchées peut être plus précise qu'en utilisant des intervalles.

Exemple III.4. *Considérons le "united solution set" $\Sigma(\mathbf{A}, \mathbf{b})$ avec*

$$\mathbf{A} = \begin{pmatrix} 0.4490 \pm 8.432E^{-4} & , & 0.4385 \pm 8.578E^{-4} \\ 0.4385 \pm 8.062E^{-4} & , & 0.6510 \pm 6.098E^{-4} \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 5.657 \pm 1.029E^{-3} \\ 0.6119 \pm 1.583E^{-3} \end{pmatrix},$$

tiré de Beaumont[66][67]. Chacune des équations étudiées a une unique solution qui est propre :

$$\begin{aligned} \mathbf{x}^I &\approx ([0.9968, 1.0018], [0.2645, 0.2690])^T, \\ \mathbf{x}^O &\approx ([0.9740, 1.0246], [0.2458, 0.2877])^T, \\ \mathbf{u}^I &\approx ([-2.06E^{-3}, 2.06E^{-3}], [-2.51E^{-3}, 2.51E^{-3}])^T, \\ \mathbf{u}^O &\approx ([-2.13E^{-3}, 2.13E^{-3}], [-2.58E^{-3}, 2.58E^{-3}])^T. \end{aligned}$$

Il est remarquable que \mathbf{x}^I ait été atteinte après 134 itérations alors que \mathbf{u}^I a été atteinte après seulement 6 itérations. Cela s'explique parce qu'une fois préconditionnée, l'équation mise sous forme de point fixe a un coefficient de contraction plus faible.

La comparaison des distances de Hausdorff suivantes montre que l'approximation par des boîtes penchées est plus précise que l'approximation par des intervalles :

$$\text{dist}_H(\mathbf{x}^O, \mathbf{x}^I) \approx 2.3E^{-2} \quad ; \quad \text{dist}_H(\mathbf{s}^O, \mathbf{s}^I) \approx 7.9E^{-4}.$$

La majoration de la distance de Hausdorff entre les deux boîtes penchées, donnée par le Théorème III.4, est $9.4E^{-4}$, ce qui est précis. L'approximation optimale d'intervalles de \mathbf{s}^O est

$$\square \mathbf{s}^O = \tilde{x} + (\text{mid } \mathbf{A})^{-1} \mathbf{u}^O,$$

qui est égale à \mathbf{x}^O (aux erreurs d'arrondi près). Le Théorème III.5 est donc confirmé. Les différentes approximations du "united solution set" de cet exemple sont représentées sur la figure 7.1 (les deux boîtes penchées sont trop proches pour être discernées). Ce graphique montre que, sur cet exemple, l'approximation par des boîtes penchées est beaucoup plus précise que l'approximation par des intervalles.

Les deux prochains exemples illustrent le fait que l'équation préconditionnée à droite dédiée à l'approximation intérieure permet de construire des approximations intérieures dans des situations où la résolution de l'équation non préconditionnée ne le permet pas.

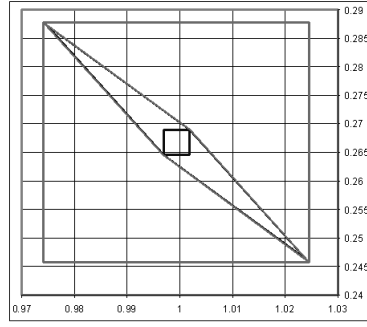


FIG. 7.1 – Le "united solution set" de l'Exemple III.4 avec ses différentes approximations.

Exemple III.5. *Considérons le "tolerable solution set" $\Sigma(\text{dual } \mathbf{A}, \mathbf{b})$ où*

$$\mathbf{A} = \begin{pmatrix} \mathbf{a} & \mathbf{a} & 0 & 0 & 0 & 0 \\ \mathbf{a} & \mathbf{a} & \mathbf{a} & 0 & 0 & 0 \\ 0 & \mathbf{a} & \mathbf{a} & \mathbf{a} & 0 & 0 \\ 0 & 0 & \mathbf{a} & \mathbf{a} & \mathbf{a} & 0 \\ 0 & 0 & 0 & \mathbf{a} & \mathbf{a} & \mathbf{a} \\ 0 & 0 & 0 & 0 & \mathbf{a} & \mathbf{a} \end{pmatrix} \quad \text{et} \quad \mathbf{b} = \begin{pmatrix} [0.9, 1.1] \\ [-1.1, -0.9] \\ [0.9, 1.1] \\ [-1.1, -0.9] \\ [0.9, 1.1] \\ [-1.1, 0.9] \end{pmatrix}.$$

et $\mathbf{a} = [0.999, 1.001]$. Pour construire une approximation intérieure de $\Sigma(\text{dual } \mathbf{A}, \mathbf{b})$, nous pouvons considérer l'équation $\mathbf{A}\mathbf{x} = \mathbf{b}$ (où $\text{dual dual } \mathbf{A} = \mathbf{A}$). L'écriture sous forme de point fixe de l'équation conduit à une suite divergente. Par contre, la méthode sous-différentielle de Newton proposée dans Shary[92][94] calcule une solution de cette équation. Cependant, cette solution n'est pas propre et ne peut donc pas être utilisée pour construire une approximation intérieure du "tolerable solution set" considéré. Il faudrait donc utiliser la technique de bisection des paramètres pour obtenir un "AE-solution set" auxiliaire $\mathbf{A}'\mathbf{x} = \mathbf{b}'$ qui aurait une solution propre. Cependant, la technique de bisection des paramètres est difficile à appliquer car nous n'avons aucune information sur les conditions qui nous permettraient d'obtenir une solution propre. L'équation preconditionnée à droite $\mathbf{A}\mathbf{C} = \mathbf{b} - \mathbf{A}\tilde{\mathbf{x}}$ (où $\text{pro dual } \mathbf{A} = \mathbf{A}$) a la solution propre suivante :

$$\mathbf{u}^I \approx \left([-0.0981, 0.0981], [-0.0959, 0.0959], [-0.0941, 0.0941], \right. \\ \left. [-0.0941, 0.0941], [-0.0959, 0.0959], [-0.0981, 0.0981] \right)^T.$$

Cette solution est calculée en 6 itérations de l'équation mise sous forme de point fixe. De plus, la majoration de la distance de Hausdorff entre les boîtes penchées intérieures et extérieures donnée par le Théorème III.4 est

$$2 \|\hat{\mathbf{A}}^{-1}\| \frac{\|I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}\| \|\mathbf{b} - (\text{dual } \mathbf{A})\tilde{\mathbf{x}}\|}{1 - (\|I - (\text{pro } \mathbf{A})\hat{\mathbf{A}}^{-1}\|)^2} \approx 1.5E^{-2}.$$

L'hypervolume du "tolerable solution" est supérieur à l'hypervolume de son approximation intérieure, ce dernier étant

$$\text{vols}^I = \det(C) \text{vol } \mathbf{u}^I \approx (0.19)^6.$$

Donc la boîte penchée intérieure est précise (car $1.5E^{-2} \ll 0.19$, en comparant une longueur avec une longueur). Finalement, l'approximation extérieure \mathbf{x}^O est cette fois strictement incluse dans $\square \mathbf{s}^O$, ce qui prouve que le Théorème III.5 ne peut pas être généralisé aux "tolerable solution sets". Cependant, l'hypervolume de la boîte penchée extérieure est $(0.20)^6$ qui est très inférieur à celui de la \mathbf{x}^O qui est $(0.65)^6$. Par conséquent, les deux approximations sont complémentaires, bien que la boîte penchée extérieure semble donner une description plus fine "tolerable solution set".

Exemple III.6. Considérons le "tolerable solution set" de l'exemple précédent où \mathbf{A}_{16} est changé en -1 . Dans ce cas, toutes les méthodes de résolution divergent lorsqu'elles sont utilisées pour calculer une solution de l'équation $\mathbf{Ax} = \mathbf{b}$. C'est certainement la conséquence du fait que l'équation $\mathbf{Ax} = \mathbf{b}$ a une infinité de solutions ou pas de solution du tout. Construire une approximation intérieure à partir de la résolution de l'équation $\mathbf{Ax} = \mathbf{b}$ en utilisant la technique de bisection des paramètres semble être une tâche impossible sur cet exemple.

L'équation préconditionnée à droite n'est pas sensible à cette altération de l'Exemple III.5. Sa solution est maintenant

$$\mathbf{u}^I \approx \left([-0.0982, 0.0982], [-0.0960, 0.0960], [-0.0942, 0.0942], \right. \\ \left. [-0.0943, 0.0943], [-0.0963, 0.0963], [-0.0985, 0.0985] \right)^T,$$

qui est propre. Cette solution est calculée après le même nombre d'itérations et la qualité de l'approximation est la même que dans l'exemple précédent.

Le prochain exemple décrit une situation où l'équation préconditionnée à droite dédiée à l'approximation intérieure a un moins bon comportement que l'équation non préconditionnée.

Exemple III.7. Considérons le "united solution set" $\Sigma(\mathbf{A}, \mathbf{b})$ où

$$\mathbf{A} = \begin{pmatrix} [2, 3] & [0, 1] \\ [1, 2] & [2, 3] \end{pmatrix} \quad \text{et} \quad \mathbf{b} = \begin{pmatrix} [-1, 1] \\ [-10, 10] \end{pmatrix}.$$

L'équation d'intervalle $\mathbf{Ax} = \mathbf{b}$ a la solution propre suivante :

$$\mathbf{x}^I = \left([-0.5, 0.5], [-4.75, 4.75] \right)^T,$$

qui est donc une approximation intérieure du "united solution set". L'équation préconditionnée à droite a une solution qui n'est pas propre, et donc qui ne peut pas être interprétée comme une approximation intérieure.

Finalement, un comportement surprenant des deux équations d'intervalles dédiées à l'approximation intérieure est obtenu en modifiant l'exemple précédent :

Exemple III.8. Considérons le "united solution set" de l'exemple précédent où \mathbf{A} est modifiée en gardant inchangé son centre et en multipliant son rayon par $q = 0.1$, i.e.

$$\mathbf{A} = \begin{pmatrix} [2.45, 2.55] & [0.45, 0.55] \\ [1.45, 1.55] & [2.45, 2.55] \end{pmatrix}.$$

Cette fois, l'équation d'intervalle $\mathbf{Ax} = \mathbf{b}$ a une solution qui n'est pas propre alors que l'équation préconditionnée à droite a une solution propre. Choisir un coefficient $0 \leq q < 0.1$ ne change plus la qualité propre/impropre des solutions correspondantes.

7.4 Conclusion

Le preconditionnement à droite que nous avons proposé présente des avantages dans le cas de l'approximation intérieure des "AE-solution sets" linéaires : l'utilisation du preconditionnement par $C = (\text{mid } \mathbf{A})^{-1}$ permet d'obtenir la convergence des algorithmes vers une approximation intérieure à condition que la matrice d'intervalles mise en jeu soit fortement régulière. Il s'agit d'une amélioration importante par rapport à la condition de convergence des algorithmes non preconditionnés, qui nécessitent des matrices à diagonale strictement dominante (ou des H-matrices) sous peine de non convergence. Ainsi, nous avons proposé des exemples où le preconditionnement à droite semble être nécessaire pour l'obtention d'une approximation intérieure. Enfin, le preconditionnement à droite accélère la convergence des algorithmes : sur certains exemples testés, l'algorithme preconditionné converge vingt fois plus vite que la version non preconditionnée.

Le preconditionnement à droite pour l'approximation extérieure des "AE-solution sets" linéaires présente les mêmes avantages que ceux décrits dans Neumaier[7] dans le cas restreint des "united solution sets" linéaires : les approximations par des boîtes penchées est plus précise que l'approximation par des vecteurs d'intervalles pour des algorithmes d'une complexité identique qu'avec le preconditionnement à gauche.

Chapitre 8

Amélioration de l'opérateur généralisé de Gauss-Seidel

Ce chapitre présente une amélioration de l'opérateur généralisé de Gauss-Seidel Γ^s proposé dans Shary[95][96] dans le contexte de l'approximation extérieure des "AE-solution sets" linéaires. L'opérateur de Gauss-Seidel y a été généralisé sous sa forme simplifiée. Une version dédiée à l'approximation extérieure des "AE-solution sets" de l'opérateur de Gauss-Seidel Γ proposé dans Hansen[19] est maintenant proposée, ce qui permet d'appliquer l'opérateur de Gauss-Seidel aux "AE-solution sets" dont la matrice n'est pas régulière.

8.1 L'opérateur généralisé de Gauss-Seidel (cas $n = 1$)

La démarche suivie est la même que dans Neumaier[8] (voir sous-section 5.1.2 page 64) : l'opérateur Γ est tout d'abord défini dans le cas $n = 1$.

Proposition III.5. *Soient $\mathbf{a}, \mathbf{b} \in \mathbb{K}\mathbb{R}$ et $\mathbf{x} \in \mathbb{I}\mathbb{R}$. Définissons*

$$\Gamma(\mathbf{a}, \mathbf{b}, \mathbf{x}) := \square \Sigma(\mathbf{a}, \mathbf{b}) \cap \mathbf{x}$$

Alors,

$$\Gamma(\mathbf{a}, \mathbf{b}, \mathbf{x}) = \begin{cases} (\mathbf{b}/\mathbf{a}) \wedge \mathbf{x} & \text{si } 0 \notin \text{pro } \mathbf{a}, \\ \square(\mathbf{x} \setminus [\min\{0, \underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\}, \max\{0, \underline{\mathbf{b}}/\overline{\mathbf{a}}, \overline{\mathbf{b}}/\underline{\mathbf{a}}\}]) & \text{si } 0 \subseteq \mathbf{a}, \\ [\max\{\underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\}, \min\{\overline{\mathbf{b}}/\underline{\mathbf{a}}, \underline{\mathbf{b}}/\overline{\mathbf{a}}\}] \wedge \mathbf{x} & \text{si } \mathbf{a} \subseteq 0, \end{cases}$$

Remarques.

1. L'expression donnée si-dessus contient un abus d'écriture : les premiers et troisièmes cas utilisent des intervalles généralisés. Ainsi, un résultat impropre est interprété comme l'ensemble vide. Le second cas ne met en jeu que des intervalles propres. Dans ce cas, nous considérons que $\square \emptyset = \emptyset$, même si cette écriture n'est en fait pas définie car l'ensemble vide n'est pas considéré comme un intervalle.
2. Comme dans l'expression donnée dans Neumaier[8], les cas où une des bornes de \mathbf{a} est nulle ne sont pas explicitement décrits dans l'expression précédente. Ces cas peuvent se déduire de ceux présentés ici.

3. Si \mathbf{a} et \mathbf{b} sont propres, la deuxième expression est équivalente à celles données dans Neumaier[8]. Par exemple, si $\mathbf{b} > 0$ alors, en comparant les signes des bornes des intervalles mis en jeux, nous obtenons

$$] \min\{0, \underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\}, \max\{0, \underline{\mathbf{b}}/\overline{\mathbf{a}}, \overline{\mathbf{b}}/\underline{\mathbf{a}}\} [=] \underline{\mathbf{b}}/\underline{\mathbf{a}}, \underline{\mathbf{b}}/\overline{\mathbf{a}} [$$

Si $0 \in \mathbf{b}$, une comparaison des signes des bornes des intervalles mis en jeu montre que

$$] \min\{0, \underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\}, \max\{0, \underline{\mathbf{b}}/\overline{\mathbf{a}}, \overline{\mathbf{b}}/\underline{\mathbf{a}}\} [=]0, 0[= \emptyset,$$

et ainsi que $\Gamma(\mathbf{a}, \mathbf{b}, \mathbf{x}) = \mathbf{x}$.

4. Dans la troisième expression, si $0 \notin \mathbf{b}$ alors une comparaison des signes des bornes des intervalles mis en jeux montre que

$$[\max\{\underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\}, \min\{\overline{\mathbf{b}}/\underline{\mathbf{a}}, \underline{\mathbf{b}}/\overline{\mathbf{a}}\}]$$

est impropre, et ainsi que $\Gamma(\mathbf{a}, \mathbf{b}, \mathbf{x}) = \emptyset$.

La preuve de cette proposition est faite au cas par cas.

Cas où $0 \notin \text{pro } \mathbf{a}$

En utilisant le Théorème II.10 (page 69) nous obtenons

$$x \in \Sigma(\mathbf{a}, \mathbf{b}) \iff (\text{dual } \mathbf{a})x \subseteq \mathbf{b}.$$

Il suffit alors de multiplier les deux membres de l'inclusion par $1/\mathbf{a}$ pour obtenir

$$x \in \Sigma(\mathbf{a}, \mathbf{b}) \iff x \subseteq \mathbf{b}/\mathbf{a}.$$

Ainsi, si \mathbf{b}/\mathbf{a} est impropre alors $(\mathbf{b}/\mathbf{a}) \wedge \mathbf{x}$ est aussi impropre et $\Sigma(\mathbf{a}, \mathbf{b})$ est vide. Si \mathbf{b}/\mathbf{a} est propre alors les nombres qui sont à la fois dans $\Sigma(\mathbf{a}, \mathbf{b})$ et dans \mathbf{x} sont les nombres qui sont dans $(\mathbf{b}/\mathbf{a}) \wedge \mathbf{x} = (\mathbf{b}/\mathbf{a}) \cap \mathbf{x}$.

Cas où $0 \subseteq \mathbf{a}$

Le Théorème II.10 montre que

$$x \in \Sigma(\mathbf{a}, \mathbf{b}) \iff (\text{dual } \mathbf{a})x \subseteq \mathbf{b}.$$

Si $0 \leq x$ alors $(\text{dual } \mathbf{a})x = [\overline{\mathbf{a}}x, \underline{\mathbf{a}}x]$. Ainsi,

$$0 \leq x \wedge x \in \Sigma(\mathbf{a}, \mathbf{b}) \iff \underline{\mathbf{b}} \leq \overline{\mathbf{a}}x \wedge \underline{\mathbf{a}}x \leq \overline{\mathbf{b}}. \quad (8.1)$$

Comme $\underline{\mathbf{a}} \leq 0 \leq \overline{\mathbf{a}}$, ces inégalités s'écrivent $\underline{\mathbf{b}}/\overline{\mathbf{a}} \leq x \wedge x \geq \overline{\mathbf{b}}/\underline{\mathbf{a}}$. De la même manière, si $x \leq 0$ alors $(\text{dual } \mathbf{a})x = [\underline{\mathbf{a}}x, \overline{\mathbf{a}}x]$ et

$$x \leq 0 \wedge x \in \Sigma(\mathbf{a}, \mathbf{b}) \iff \underline{\mathbf{b}} \leq \underline{\mathbf{a}}x \wedge \overline{\mathbf{a}}x \leq \overline{\mathbf{b}} \quad (8.2)$$

Comme $\underline{\mathbf{a}} \leq 0 \leq \overline{\mathbf{a}}$, ces inégalités s'écrivent $\underline{\mathbf{b}}/\underline{\mathbf{a}} \geq x \wedge x \leq \overline{\mathbf{b}}/\overline{\mathbf{a}}$. En regroupant les deux cas précédents, nous avons finalement

$$\begin{aligned} x \in \Sigma(\mathbf{a}, \mathbf{b}) &\iff x \leq \min\{0, \underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\} \vee \max\{0, \underline{\mathbf{b}}/\overline{\mathbf{a}}, \overline{\mathbf{b}}/\underline{\mathbf{a}}\} \leq x \\ &\iff x \in] -\infty, \min\{0, \underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\}] \cup [\max\{0, \underline{\mathbf{b}}/\overline{\mathbf{a}}, \overline{\mathbf{b}}/\underline{\mathbf{a}}\}, +\infty[\end{aligned}$$

Et donc,

$$x \in \Sigma(\mathbf{a}, \mathbf{b}) \cap \mathbf{x} \iff \mathbf{x} \setminus] \min\{0, \underline{\mathbf{b}}/\underline{\mathbf{a}}, \overline{\mathbf{b}}/\overline{\mathbf{a}}\}, \max\{0, \underline{\mathbf{b}}/\overline{\mathbf{a}}, \overline{\mathbf{b}}/\underline{\mathbf{a}}\} [$$

Cas où $\mathbf{a} \subseteq 0$

Les relations (8.1) et (8.2) sont toujours valables dans ce cas. Mais cette fois l'intervalle \mathbf{a} vérifie $\bar{\mathbf{a}} \leq 0 \leq \underline{\mathbf{a}}$ et nous obtenons donc les relations suivantes :

$$\begin{aligned} 0 \leq x \wedge x \in \Sigma(\mathbf{a}, \mathbf{b}) &\iff \underline{\mathbf{b}/\bar{\mathbf{a}}} \geq x \wedge x \leq \bar{\mathbf{b}}/\underline{\mathbf{a}} \\ x \leq 0 \wedge x \in \Sigma(\mathbf{a}, \mathbf{b}) &\iff \underline{\mathbf{b}/\bar{\mathbf{a}}} \leq x \wedge x \geq \bar{\mathbf{b}}/\underline{\mathbf{a}}. \end{aligned}$$

En regroupant ces inégalités, nous obtenons

$$\begin{aligned} x \in \Sigma(\mathbf{a}, \mathbf{b}) &\iff \max\{\underline{\mathbf{b}/\bar{\mathbf{a}}}, \bar{\mathbf{b}}/\underline{\mathbf{a}}\} \leq x \leq 0 \vee 0 \leq x \leq \min\{\underline{\mathbf{b}/\bar{\mathbf{a}}}, \bar{\mathbf{b}}/\underline{\mathbf{a}}\} \\ &\iff x \in [\max\{\underline{\mathbf{b}/\bar{\mathbf{a}}}, \bar{\mathbf{b}}/\underline{\mathbf{a}}\}, \min\{\underline{\mathbf{b}/\bar{\mathbf{a}}}, \bar{\mathbf{b}}/\underline{\mathbf{a}}\}] \end{aligned}$$

8.2 L'opérateur généralisé de Gauss-Seidel (cas général)

L'opérateur généralisé de Gauss-Seidel est alors construit de la façon suivante : soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $\mathbf{x} \in \mathbb{I}\mathbb{R}^{n \times n}$,

$$\Gamma(\mathbf{A}, \mathbf{b}, \mathbf{x}) = \mathbf{y} \text{ avec } \mathbf{y}_i = \Gamma(\mathbf{A}_{ii}, \mathbf{b}_i - \sum_{j \neq i} \mathbf{A}_{ij} \mathbf{x}_j, \mathbf{x}_i)$$

La proposition suivante montre que l'opérateur généralisé de Gauss-Seidel peut être utilisé pour l'approximation extérieure des "AE-solution sets" linéaires.

Théorème III.6. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^{n \times n}$ et $\mathbf{x} \in \mathbb{I}\mathbb{R}^{n \times n}$. Alors,

$$\Sigma(\mathbf{A}, \mathbf{b}) \cap \mathbf{x} \subseteq \Gamma(\mathbf{A}, \mathbf{b}, \mathbf{x}) \subseteq \mathbf{x}$$

Démonstration. La seconde inclusion est triviale, étant donné que par définition Γ dans le cas $n = 1$ nous avons $\mathbf{y}_i \subseteq \mathbf{x}_i$. Considérons maintenant la première inclusion. Par le Théorème II.10,

$$x \in \Sigma(\mathbf{A}, \mathbf{b}) \iff (\text{dual } \mathbf{A})x \subseteq \mathbf{b}.$$

Maintenant, en explicitant chaque ligne de l'équation précédente,

$$x \in \Sigma(\mathbf{A}, \mathbf{b}) \iff \bigwedge_{i \in [1..n]} (\text{dual } \mathbf{A}_{ii})x_i + \sum_{j \neq i} (\text{dual } \mathbf{A}_{ij})x_j \subseteq \mathbf{b}_i.$$

En utilisant les propriétés de groupe l'arithmétique de Kaucher, nous obtenons

$$x \in \Sigma(\mathbf{A}, \mathbf{b}) \iff \bigwedge_{i \in [1..n]} (\text{dual } \mathbf{A}_{ii})x_i \subseteq \mathbf{b}_i - \sum_{j \neq i} \mathbf{A}_{ij}x_j.$$

En utilisant la croissance par rapport à l'inclusion des opérations de l'arithmétique de Kaucher, nous obtenons

$$x \in \Sigma(\mathbf{A}, \mathbf{b}) \implies \bigwedge_{i \in [1..n]} (\text{dual } \mathbf{A}_{ii})x_i \subseteq \mathbf{b}_i - \sum_{j \neq i} \mathbf{A}_{ij}x_j.$$

En utilisant encore une fois le Théorème II.10, nous obtenons

$$x \in \Sigma(\mathbf{A}, \mathbf{b}) \implies \bigwedge_{i \in [1..n]} x_i \in \Sigma(\mathbf{A}_{ii}, \mathbf{b}_i - \sum_{j \neq i} \mathbf{A}_{ij}x_j).$$

Finalement, en utilisant la Proposition III.5, nous obtenons

$$x \in \mathbf{x} \wedge x \in \Sigma(\mathbf{A}, \mathbf{b}) \implies \bigwedge_{i \in [1..n]} x_i \in \Gamma(\mathbf{A}_{ii}, \mathbf{b}_i - \sum_{j \neq i} \mathbf{A}_{ij} \mathbf{x}_j, \mathbf{x}_i).$$

Ce qui conclue la preuve. □

Le nouvel opérateur généralisé de Gauss-Seidel est illustré par l'exemple suivant.

Exemple III.9. *Considérons le "AE-solution set" suivant :*

$$\Sigma(\mathbf{A}, \mathbf{b}) \text{ avec } \mathbf{A} = \begin{pmatrix} 1 & [-1, 1] \\ 2 & [1, -1] \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} [1, -1] \\ [-1, 1] \end{pmatrix}$$

et l'intervalle $\mathbf{x} = ([1, 2], [1, 2])^T$. Nous avons alors

$$\Gamma(\mathbf{A}, \mathbf{b}, \mathbf{x}) = \begin{pmatrix} \Gamma(1, [1, -1] - [-1, 1] \times \mathbf{x}_2) \\ \Gamma([1, -1], [-1, 1] - 2 \times \mathbf{x}_1) \end{pmatrix}$$

Finalement, $\Gamma([1, -1], [-1, 1] - 2 \times \mathbf{x}_1) = \emptyset$ et l'opérateur généralisé de Gauss-Seidel permet finalement de montrer que $\Sigma(\mathbf{A}, \mathbf{b}) \cap \mathbf{x} = \emptyset$. La version proposée dans Shary[96] ne permet pas de conclure dans ce cas.

8.3 Conclusion

Le nouvel opérateur généralisé de Gauss-Seidel ne présente que des avantages sur la version précédente : en effet, lorsque la matrice mise en jeu ne contient que des intervalles de signe constant sur sa diagonale, les deux versions de l'opérateurs sont identiques. Mais lorsque des intervalles contenant 0 sont présents sur la diagonale, la nouvelle version est encore capable d'effectuer des réductions alors que l'ancienne version est impuissante.

Dans le cas des "united-solution sets", il est prouvé que l'opérateur de Gauss-Seidel dans sa version améliorée est toujours plus efficace que l'opérateur de Krawczyk (voir par exemple Neumaier[8]). Il serait intéressant de comparer ces deux opérateurs dans le cas plus général des "AE-solution sets" maintenant qu'une version améliorée de l'opérateur de Gauss-Seidel a été proposée.

Quatrième partie

AE-extensions des fonctions réelles continues

Chapitre 9

Introduction

Cette partie de la thèse présente une nouvelle formulation de la théorie des intervalles modaux (Goldsztein[2, 1]). Le Chapitre 10 donne une présentation informelle de la nouvelle formulation de la théorie des intervalles modaux (Goldsztein[4]). La motivation principale de la reformulation de la théorie des intervalles modaux est d'obtenir les résultats principaux de la théorie en utilisant une construction semblable à la théorie des intervalles classiques. Nous décrivons d'abord les principaux points qui seront conservés ou modifiés entre la théorie originale des intervalles modaux et sa reformulation.

Les éléments conservés

L'interprétation des extensions sémantiques

L'interprétation des extensions sémantiques est conservée. Ainsi, la nouvelle formulation offrira les mêmes interprétations que la théorie des intervalles modaux. Ces interprétations enrichissent les interprétations de la théorie des intervalles classiques : lorsqu'elle est restreinte aux intervalles propres, la reformulation de théorie des intervalles modaux coïncide exactement avec la théorie des intervalles classiques.

La composition des extensions sémantiques

L'un des résultats principaux de la théorie des intervalles modaux est d'avoir montré que, sous certaines conditions, la composition des extensions sémantiques est une extension sémantique de la fonction réelle correspondant à cette composition. Cette propriété est utilisée dans l'interprétation des extensions rationnelles aux intervalles modaux et est à rapprocher de l'extension naturelle de la théorie des intervalles classiques. Cette propriété jouera aussi un rôle central dans la nouvelle formulation de la théorie, où elle permettra de définir les AE-extensions naturelles.

L'arrondi des extensions

La théorie des intervalles modaux propose une procédure d'arrondi qui utilise les intervalles généralisés. L'arrondi extérieur des intervalles généralisés permet en effet d'effectuer soit un arrondi intérieur soit un arrondi extérieur en fonction du besoin (besoin différent généralement qu'un quantificateur universel ou existentiel est mis en jeu). Cette

procédure d'arrondi est fondamentale pour une implémentation efficace des extensions et sera présentée en détail dans la nouvelle formulation de la théorie.

Les éléments modifiés

Association d'un quantificateur à un intervalle classique

L'association d'un intervalle classique et d'un quantificateur est abandonnée. Cette association crée en effet deux besoins contradictoires : d'une part, pour rester proche de la théorie des intervalles classiques, il faut qu'un des deux cas (\mathbf{x}, \exists) ou (\mathbf{x}, \forall) soit identifié aux intervalles classiques¹⁹. D'autre part, l'interprétation des extensions aux intervalles classiques doit faire partie des interprétations offertes par la nouvelle formulation de la théorie des intervalles modaux. Cependant, ces deux contraintes impliquent que les quantificateurs couplés aux intervalles classiques ne peuvent pas correspondre aux quantificateurs présents dans les propositions quantifiées correspondant aux interprétations. En effet, dans le cas des extensions aux intervalles classiques, un seul type d'intervalle est rencontré alors que deux types de quantificateurs sont mis en jeu : $(\forall x \in \mathbf{x})(\exists z \in \mathbf{z})(z = f(x))$.

Remarque. *Une extension de la théorie des intervalles classiques où des quantificateurs seraient couplés aux intervalles classiques et où les extensions des fonctions auraient des interprétations dans lesquelles les quantificateurs correspondent aux quantificateurs présents dans les couples a été explorée. Mais une telle théorie crée plus de difficultés qu'elle n'en résout : parmi d'autres difficultés, les additions et multiplications perdent leur propriétés intéressantes pour devenir contre-intuitives, une distance est difficile à définir, les procédures d'arrondi sont difficiles à définir...*

Redondance entre les * et ** extension modales

Comme l'ont montré les exemples II.32, II.33, II.34, II.35, II.36 et II.37 les * et ** extensions sont redondantes. Bien que ce ne soit pas explicitement expliqué dans la théorie des intervalles modaux, il semble que l'introduction des deux types d'extensions soit faite pour définir la "JM-commutativité", qui joue un rôle central dans la théorie des intervalles modaux. La "JM-commutativité" est avantageusement remplacée par une autre notion dans la nouvelle formulation : la commutation avec l'opérateur dual. Cette notion est plus intuitive et permettra de ne pas définir deux types d'extensions.

L'optimalité des extensions aux intervalles modaux

L'optimalité des extensions rationnelles aux intervalles modaux n'est pas intuitive. De plus, il semble qu'elle ne reflète pas exactement la précision d'une extension : en effet, une extension rationnelle qui vérifierait $fR(\dot{\mathbf{x}}) = f^*(\dot{\mathbf{x}})$ n'est pas optimale si $f^*(\dot{\mathbf{x}}) \neq f^{**}(\dot{\mathbf{x}})$, alors qu'il n'existe aucune extension qui soit plus précise qu'elle. Ainsi, l'optimalité est remplacée par une notion plus intuitive et plus représentative de la qualité d'une extension : la minimalité par rapport à l'inclusion. De plus, cette nouvelle notion va permettre d'introduire le concept important d'ordre de convergence d'une extension. La minimalité

¹⁹Dans la théorie des intervalles modaux, ce sont les intervalles existentiels qui sont identifiés aux intervalles classiques.

et l'ordre de convergence ainsi introduits généraliserons les notions correspondantes de la théorie des intervalles classiques.

Chapitre 10

Présentation informelle de la nouvelle formulation de la théorie des intervalles modaux

10.1 Introduction

La théorie des intervalles classiques est utilisée dans de nombreuses situations pour calculer de manière rigoureuse en prenant en compte des incertitudes. Elle conduit généralement à des approximations extérieures d'ensembles définis grâce à des quantificateurs existentiels. L'objet fondamental de la théorie des intervalles est l'extension des fonctions réelles. Elles permettent de construire des sur-ensembles de l'image de la fonction mise en jeux. Mais elle peuvent aussi être interprétées en utilisant des propositions quantifiées :

Exemple IV.1. *L'extension naturelle de la fonction $f(x) = x^2 - x$ évaluée pour l'intervalle $\mathbf{x} = [2, 3]$ conduit à $\mathbf{z} := \mathbf{x}^2 - \mathbf{x} = [1, 7]$ qui est interprété par $\{f(x) \mid x \in \mathbf{x}\} \subseteq [1, 7]$. Une interprétation alternative est la proposition quantifiée suivante :*

$$(\forall x \in \mathbf{x})(\exists z \in \mathbf{z})(z = f(x)).$$

En partant de cette dernière interprétation des extensions des fonctions réelles aux intervalles classiques, la théorie des intervalles modaux a réussi à construire des extensions ayant des interprétations plus riches : une plus grande liberté du choix des quantificateurs associés aux variables est proposée, tout en conservant les avantages des extensions aux intervalles décrits en introduction de la partie IV. Comme dans le cadre de la théorie des intervalles classiques, l'évaluation d'une extension modale est une question centrale et la problématique du pessimisme introduit par l'évaluation d'intervalles est cruciale pour l'applicabilité de la théorie. Il reste beaucoup de travail à faire dans cette direction. Une étape importante serait un procédé de linéarisation, comme une extension de la valeur moyenne, compatible avec les interprétations enrichies proposées par la théorie des intervalles modaux.

Une présentation informelle de la reformulation de la théorie des intervalles modaux est maintenant proposée. Ce cadre théorique permet d'une part de mieux comprendre les mécanismes sous-jacents de la théorie. D'autre part, il permet d'introduire une extension de la valeur moyenne aux intervalles généralisés (qui sera appelée l'AE-extension de la valeur moyenne) compatible avec les interprétations enrichies de la théorie des intervalles modaux.

10.2 Description du problème à résoudre

Les extensions des fonctions réelles aux intervalles classiques permettent de résoudre le problème suivant : étant donné f et $\mathbf{x} \in \mathbb{IR}^n$, nous voulons déterminer un intervalle \mathbf{z} qui vérifie la proposition quantifiée

$$(\forall x \in \mathbf{x})(\exists z \in \mathbf{z})(z = f(x)).$$

Ce problème est généralisé de la façon suivante : étant donné f , $\mathbf{x} \in \mathbb{IR}^n$ et un quantificateur \mathbf{Q}_{x_k} pour chaque variable x_k , nous voulons déterminer un quantificateur \mathbf{Q} et un intervalle \mathbf{z} qui vérifient la proposition quantifiée

$$(\forall x_A \in \mathbf{x}_A)(\mathbf{Q}z \in \mathbf{z})(\exists x_E \in \mathbf{x}_E)(z = f(x)), \quad (10.1)$$

où $A = \{k \mid \mathbf{Q}_{x_k} = \forall\}$ et $E = \{k \mid \mathbf{Q}_{x_k} = \exists\}$.

Remarque. Les ensembles d'indices A et E peuvent être vides. La proposition quantifiée précédente est alors facilement adaptée. De plus, le bloc de quantificateur $(\mathbf{Q}z \in \mathbf{z})$ est écrit au centre de la proposition pour que l'ordre AE (A (ll) E (xists)) soit conservé quelque soit le quantificateur \mathbf{Q} .

Remarque. Cette classe de problèmes pourrait être décrite plus simplement en se restreignant aux expressions $f(x) = 0$. En effet, $f(x) = z$ peut s'écrire $g(x, z) = 0$ avec $g(x, z) = f(x) - z$. L'expression (10.1) est choisie pour deux raisons : d'une part elle permet de généraliser le problème résolu par la théorie des intervalles classiques, ce qui permettra de construire des méthodes de résolutions semblables dans les deux cas. D'autre part, l'expression (10.1) va permettre de composer la résolution de problèmes élémentaires pour résoudre des problèmes plus difficiles. Cette composition serait beaucoup compliquée à écrire avec la formulation alternative de ces problèmes.

Exemple IV.2. Considérons la fonction $f(x) = x_1 + x_2$, l'intervalle $\mathbf{x} = ([-1, 1], [4, 8])$ et les quantificateurs $\mathbf{Q}_{x_1} = \forall$ et $\mathbf{Q}_{x_2} = \exists$. Nous voulons donc déterminer un quantificateur \mathbf{Q} et un intervalle \mathbf{z} qui vérifient

$$(\forall x_1 \in [-1, 1])(\mathbf{Q}z \in \mathbf{z})(\exists x_2 \in [4, 8])(z = x_1 + x_2) \quad (10.2)$$

On peut vérifier que $\mathbf{Q} = \forall$ et $\mathbf{z} = [5, 7]$ est une solution de notre problème. Trivialement, d'autres solutions peuvent alors être construites, eg. $\mathbf{Q} = \forall$ et $\mathbf{z} = [5.5, 6.5]$ parce que $[5.5, 6.5] \subseteq [5, 7]$, ou $\mathbf{Q} = \exists$ et $\mathbf{z} = [6.5, 9]$ parce que $[6.5, 9] \cap [5, 7] \neq \emptyset$.

Dans le cas des fonctions à valeurs vectorielles, i.e. $f = (f_1, \dots, f_m)$, m intervalles \mathbf{z}_k et m quantificateurs \mathbf{Q}_{z_k} doivent être déterminés. La proposition quantifiée qui doit être validée est alors

$$(\forall x_A \in \mathbf{x}_A)(\forall z_{A'} \in \text{pro } \mathbf{z}_{A'})(\exists z_{E'} \in \mathbf{z}_{E'})(\exists x_E \in \text{pro } \mathbf{x}_E)(z = f(x))$$

où $A = \{k \mid \mathbf{Q}_{x_k} = \forall\}$, $E = \{k \mid \mathbf{Q}_{x_k} = \exists\}$, $A' = \{k \mid \mathbf{Q}_{z_k} = \forall\}$ et $E' = \{k \mid \mathbf{Q}_{z_k} = \exists\}$.

10.3 Intervalles généralisés et quantificateurs

A partir d'un ensemble de réels $\{x \in \mathbb{R} \mid a \leq x \leq b\}$, deux intervalles généralisés peuvent être construits, i.e. l'intervalle propre $[a, b]$ et l'intervalle impropre $[b, a]$. Ainsi, la qualité propre/impropre d'un intervalle généralisé peut-être utilisée pour modéliser un quantificateur alors que la projection propre de cet intervalle généralisé peut-être conservée pour modéliser le domaine associé à une variable. Nous choisissons la convention suivante pour retrouver l'interprétation classique lorsque les intervalles mis en jeux sont propres :

- Si \mathbf{x}_k est propre, i.e. $\underline{\mathbf{x}} \leq \bar{\mathbf{x}}$, alors $\mathbf{Q}_{x_k} = \forall$. Si \mathbf{x}_k n'est pas propre, i.e. $\underline{\mathbf{x}} > \bar{\mathbf{x}}$, alors $\mathbf{Q}_k = \exists$.
- Si \mathbf{z}_k est propre alors $\mathbf{Q}_{z_k} = \exists$. Si \mathbf{z}_k n'est pas propre alors $\mathbf{Q}_{z_k} = \forall$.

Nous sommes maintenant en position pour reformuler le problème présenté en section 10.2 dans le langage des intervalles généralisés : étant donné f et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, nous voulons construire un intervalle $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$ qui vérifie

$$(\forall x_P \in \mathbf{x}_P)(\forall z_{I'} \in \text{pro } \mathbf{z}_{I'}) (\exists z_{P'} \in \mathbf{z}_{P'}) (\exists x_I \in \text{pro } \mathbf{x}_I) (z = f(x)) \quad (10.3)$$

où $P = P(\mathbf{x})$, $I = I(\mathbf{x})$, $P' = P(\mathbf{z})$ et $I' = I(\mathbf{z})$. Un tel intervalle \mathbf{z} est dit interprétable par rapport à f et \mathbf{x} (ou (f, \mathbf{x}) -interprétable). Ainsi, si \mathbf{x} est propre et si l'intervalle (f, \mathbf{x}) -interprétable \mathbf{z} est propre, la proposition précédente devient

$$(\forall x \in \mathbf{x})(\exists z \in \mathbf{z})(z = f(x)),$$

ce qui correspond à l'interprétation classique. Dans le cas de fonctions $f : \mathbb{R}^n \rightarrow \mathbb{R}$, la proposition quantifiée (10.3) s'écrit

$$(\forall x_P \in \mathbf{x}_P)(\mathbf{Q}^{(\mathbf{z})} z \in \text{pro } \mathbf{z})(\exists x_I \in \text{pro } \mathbf{x}_I)(z = f(x))$$

où $\mathbf{Q}^{(\mathbf{z})} = \exists$ si $\mathbf{z} \in \mathbb{I}\mathbb{R}$, $\mathbf{Q}^{(\mathbf{z})} = \forall$ sinon.

Exemple IV.3. *L'exemple IV.1 s'exprime maintenant en disant que $[1, 7]$ est $(f, [2, 3])$ -interprétable (l'interprétation des extensions aux intervalles classiques est obtenue lorsque les intervalles mis en jeux sont propres). L'exemple IV.2 s'exprime maintenant en disant que les intervalles $[7, 5]$, $[6.5, 5.5]$ et $[6.5, 9]$ sont $(f, ([-1, 1], [8, 4]))$ -interprétables.*

On peut remarquer que les trois intervalles de l'exemple précédent sont liés par la relation $[7, 5] \subseteq [6.5, 5.5] \subseteq [6.5, 9]$. En général, l'inclusion entre intervalles généralisés peut être utilisée pour comparer la précision des intervalles (f, \mathbf{x}) -interprétables (voir Section 12.3 page 135). Ainsi, si deux intervalles (f, \mathbf{x}) -interprétables \mathbf{z} et \mathbf{z}' sont liés par l'inclusion $\mathbf{z} \subseteq \mathbf{z}'$, l'intervalle \mathbf{z} apporte plus d'information que l'intervalle \mathbf{z}' et il sera donc considéré comme plus précis.

10.4 L'arithmétique de Kaucher

Nous définissons maintenant une arithmétique des intervalles généralisés qui résout le problème présenté dans la section 10.2 dans le cas des fonctions élémentaires : étant donné une fonction élémentaire f et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, où $n = 1$ ou 2 pour les fonctions élémentaires, nous définissons $f(\mathbf{x})$ comme étant le "meilleur" intervalle (f, \mathbf{x}) -interprétable.

Remarque. Le meilleur intervalle (f, \mathbf{x}) -interprétable est le plus petit au sens de l'inclusion des intervalles généralisés. Cela revient à calculer le plus large domaine lorsque $Q = \forall$ ou le plus petit domaine lorsque $Q = \exists$ (voir Section 12.3 page 135).

Grâce à la simplicité des fonctions élémentaires considérées, nous pouvons calculer leur extension formellement (voir Chapitre 13 page 145 et Chapitre 14 page 151) :

- $\mathbf{x} + \mathbf{y} = [\underline{\mathbf{x}} + \underline{\mathbf{y}}, \overline{\mathbf{x}} + \overline{\mathbf{y}}]$ et $\mathbf{x} - \mathbf{y} = [\underline{\mathbf{x}} - \overline{\mathbf{y}}, \overline{\mathbf{x}} - \underline{\mathbf{y}}]$.
- Si toutes les bornes des intervalles mis en jeu sont positives, $\mathbf{x} \times \mathbf{y} = [\underline{\mathbf{x}} \times \underline{\mathbf{y}}, \overline{\mathbf{x}} \times \overline{\mathbf{y}}]$ et $\mathbf{x}/\mathbf{y} = [\underline{\mathbf{x}}/\overline{\mathbf{y}}, \overline{\mathbf{x}}/\underline{\mathbf{y}}]$.
- pour les fonctions élémentaires à une variable, pro $f(\mathbf{x}) = \{f(x) \mid x \in \text{pro } \mathbf{x}\}$ et $f(\mathbf{x})$ et \mathbf{x} ont la même qualité propre/impropre, eg. $\sqrt{\mathbf{x}} = [\sqrt{\underline{\mathbf{x}}}, \sqrt{\overline{\mathbf{x}}}]$ pour $\underline{\mathbf{x}}, \overline{\mathbf{x}} \geq 0$.

Dans le cas des fonctions élémentaires, i.e. $+$, $-$, \times , $/$ et les fonctions à une variable, les extensions définie ici coïncident avec l'arithmétique de Kaucher bien qu'elles soient définies de manière différentes (voir Chapitre 14 page 151). Les interprétations des extensions aux intervalles généralisés peuvent être facilement retrouvées de la manière suivante : lorsque tous les intervalles mis en jeu sont propres, leur interprétation correspond à l'interprétation des extensions aux intervalles classiques :

Exemple IV.4. L'addition de Kaucher $[-1, 2] + [7, 8] = [6, 10]$ est interprétée par

$$(\forall x_1 \in [-1, 2]) (\forall x_2 \in [7, 8]) (\exists z \in [6, 10]) (z = x_1 + x_2).$$

Ensuite, lorsqu'une qualité propre/impropre est changée, le quantificateur correspondant est lui aussi changé :

Exemple IV.5. L'addition de Kaucher $[-1, 2] + [8, 7] = [7, 9]$ est interprétée par

$$(\forall x_1 \in [-1, 2]) (\exists z \in [7, 9]) (\exists x_2 \in [7, 8]) (z = x_1 + x_2).$$

L'addition de Kaucher $[-1, 2] + [8, 4] = [7, 6]$ est interprétée par

$$(\forall x_1 \in [-1, 2]) (\forall z \in [6, 7]) (\exists x_2 \in [4, 8]) (z = x_1 + x_2).$$

10.5 Une interprétation simplifiée de l'évaluation d'intervalles généralisés

Une importante contribution de la théorie des intervalles modaux est d'avoir montré que le problème de la section 10.2 peut-être décomposé en sous-problèmes similaires mais plus simples en suivant la structure de la fonction. Il s'agit là du même mécanisme que celui utilisé pour l'extension naturelle aux intervalles classiques. Ainsi, sous certaines restrictions, l'évaluation d'intervalles généralisés conduit à un intervalle qui est interprétable : considérons une expression $\mathbf{f}(x_1, \dots, x_n)$ d'une fonction continue f où chaque variable a une unique occurrence. Alors, l'intervalle $\mathbf{f}(\mathbf{x})$, calculé en utilisant l'arithmétique de Kaucher, est (f, \mathbf{x}) -interprétable (voir Section 15.1 page 159 pour plus de détails).

Exemple IV.6. Considérons la fonction $f(x, y, u) = u(y-x)$, $\mathbf{x}, \mathbf{u} \in \mathbb{IR}$ et $\mathbf{y} \in \overline{\mathbb{IR}}$. Ainsi, $Q_x = Q_u = \forall$ et $Q_y = \exists$. Calculons $\mathbf{t} = \mathbf{x} + \mathbf{y}$ et $\mathbf{z} = \mathbf{u} \times \mathbf{t}$. Du fait que des AE-extensions des fonctions élémentaires $+$ et \times sont utilisées, les proposition quantifiées suivantes sont vraies. Si \mathbf{t} est propre, nous obtenons les deux propositions quantifiées suivantes :

$$(\forall x \in \mathbf{x}) (\exists t \in \mathbf{t}) (\exists y \in \text{pro } \mathbf{y}) (t = x + y)$$

$$(\forall u \in \mathbf{u})(\forall t \in \mathbf{t})(\mathbf{Q}^{(\mathbf{z})} z \in \mathbf{z})(z = u t)$$

Si \mathbf{t} est impropre, nous obtenons les deux propositions quantifiées suivantes :

$$(\forall x \in \mathbf{x})(\forall t \in \text{pro } \mathbf{t})(\exists y \in \text{pro } \mathbf{y})(t = x + y)$$

$$(\forall u \in \mathbf{u})(\mathbf{Q}^{(\mathbf{z})} z \in \mathbf{z})(\exists t \in \text{pro } \mathbf{t})(z = u t)$$

Dans les deux cas, nous concluons que la proposition quantifiée suivante est vraie :

$$(\forall x \in \mathbf{x})(\forall u \in \mathbf{u})(\mathbf{Q}^{(\mathbf{z})} z \in \mathbf{z})(\exists y \in \text{pro } \mathbf{y})(\exists t \in \text{pro } \mathbf{t})(z = u t \wedge t = x + y),$$

Cette dernière proposition quantifiée implique

$$(\forall x \in \mathbf{x})(\forall u \in \mathbf{u})(\mathbf{Q}^{(\mathbf{z})} z \in \mathbf{z})(\exists y \in \text{pro } \mathbf{y})(z = f(x, y, u))$$

car $(\exists t \in \text{pro } \mathbf{t})(z = u t \wedge t = x + y)$ implique $z = f(x, y, u)$. Ainsi, \mathbf{z} est $(f, \mathbf{x}, \mathbf{y}, \mathbf{u})$ -interprétable. Considérons une application numérique : $\mathbf{x} = [-1, 1]$, $\mathbf{u} = [1, 2]$ et $\mathbf{y} = [2, -2]$ conduisent à $\mathbf{z} = [1, -1]$ qui est interprété par

$$(\forall x \in [-1, 1])(\forall u \in [1, 2])(\forall z \in [1, -1])(\exists y \in [-2, 2])(z = f(u, x, y)).$$

Si l'expression de la fonction contient plusieurs occurrences d'une même variable, un problème auxiliaire dans lequel chaque variable a une unique occurrence doit être traité. Par exemple, si la fonction $f(x, y) = x(y - x)$ est considérée, nous devons utiliser la fonction auxiliaire $\tilde{f}(x_1, x_2, y) = x_1(y - x_2)$ pour pouvoir utiliser le raisonnement précédent. La situation n'est pas aussi simple que dans le cas des extensions aux intervalles classiques : avant d'être utilisée avec l'arithmétique de Kaucher, l'expression \tilde{f} doit être modifiée. En reprenant l'exemple précédent, nous obtiendrions les deux nouvelles expressions

$$g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = \mathbf{x}_1(\mathbf{y} - \text{pro } \mathbf{x}_2) \quad \text{et} \quad g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = (\text{pro } \mathbf{x}_1)(\mathbf{y} - \mathbf{x}_2),$$

qui permettent toutes deux de construire des intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables (voir Section 15.2 page 162 pour plus de détails).

L'évaluation d'intervalles généralisés va maintenant être utilisée pour construire une nouvelle AE-extension similaire à l'extension de la valeur moyenne dans la théorie des intervalles classiques.

10.6 Une extension de la valeur moyenne aux intervalles généralisés

Considérons une fonction f continûment dérivable, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ et $\Delta_i \in \mathbb{IR}$ vérifiant

$$\left\{ \frac{\partial f}{\partial x_i}(x) \mid x \in \text{pro } \mathbf{x} \right\} \subseteq \Delta_i.$$

De tels intervalles Δ_i peuvent être calculés en utilisant une extension des dérivées partielles de f . Alors, l'intervalle suivant est (f, \mathbf{x}) -interprétable :

$$f^{MV}(\mathbf{x}) := f(\tilde{x}) + \sum_{k=1}^n \Delta_k \times (\mathbf{x}_k - \tilde{x}_k) \quad \text{où} \quad \tilde{x} \in \text{pro } \mathbf{x}.$$

Justifions cet énoncé : d'une part, $\mathbf{z} := f^{MV}(\mathbf{x})$ est l'évaluation d'intervalles généralisés de la fonction $g(x, \delta) = f(\tilde{x}) + \sum \delta_k(x_k - \tilde{x}_k)$ pour $\Delta \in \mathbb{IR}^n$ et $\mathbf{x} \in \mathbb{KR}^n$. Comme l'expression de cette fonction n'a qu'une occurrence de chaque variable, la proposition quantifiée suivante est vraie :

$$(\forall \delta \in \Delta)(\forall x_A \in \mathbf{x}_A)(\mathbf{Q}^{(\mathbf{z})} z \in \text{pro } \mathbf{z})(\exists x_E \in \text{pro } \mathbf{x}_E)(z = g(x, \delta)),$$

où $A = \{k \mid \mathbf{x}_k \in \mathbb{IR}\}$ et $E = \{k \mid \mathbf{x}_k \notin \mathbb{IR}\}$. D'autre part, le théorème de la valeur moyenne implique la proposition quantifiée suivante :

$$(\forall x \in \mathbf{x})(\exists \delta \in \Delta)(f(x) = g(x, \delta)).$$

Il peut être prouvé (voir Chapitre 16 page 171) que la conjonction de ces deux dernières propositions quantifiées implique la proposition quantifiée

$$(\forall x_A \in \mathbf{x}_A)(\mathbf{Q}^{(\mathbf{z})} z \in \mathbf{z})(\exists x_E \in \mathbf{x}_E)(z = f(x)).$$

Ainsi, l'intervalle $f^{MV}(\mathbf{x})$ est (f, \mathbf{x}) -interprétable. La fonction d'intervalles $f^{MV}(\mathbf{x})$ est appelée une extension de la valeur moyenne à cause de sa similitude avec l'extension de la valeur moyenne classique.

Exemple IV.7. *Considérons $f(x) = x_1(x_1 - x_2)$, $\mathbf{x}_1 = [4, 2]$ et $\mathbf{x}_2 = [0, 1]$, de telle sorte que $\mathbf{Q}_{x_1} = \exists$ et $\mathbf{Q}_{x_2} = \forall$. Nous pouvons utiliser $\tilde{x} = \text{mid } \mathbf{x}$, $\Delta_1 = 2(\text{pro } \mathbf{x}_1) - \mathbf{x}_2 = [3, 8]$ et $\Delta_2 = -(\text{pro } \mathbf{x}_1) = [-4, -2]$. Alors, $f^{MV}(\mathbf{x}) = [8.5, 6.5]$ et la proposition quantifiée suivante est donc vraie :*

$$(\forall x_2 \in [0, 1])(\forall z \in [6.5, 8.5])(\exists x_1 \in [2, 4])(z = f(x)).$$

Bien que cette extension de la valeur moyenne ne calcule pas l'intervalle (f, \mathbf{x}) -interprétable optimal, elle possède un ordre de convergence quadratique, i.e. la distance à l'intervalle (f, \mathbf{x}) -interprétable optimal décrit en $(\text{wid } \mathbf{x})^2$ (voir Section 12.4 page 140 et Section 16.4 page 181).

10.7 Cas des fonctions à valeurs vectorielles

La situation est plus compliquée dans le cas des fonctions à valeurs vectorielles. Illustrons cette situation sur un exemple. Considérons $f = (f_1, f_2) : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, $\mathbf{a} \in \mathbb{IR}$ et $\mathbf{x} \in \overline{\mathbb{IR}}^2$, de telle sorte que $\mathbf{Q}_a = \forall$ et $\mathbf{Q}_{x_1} = \mathbf{Q}_{x_2} = \exists$. Nous voulons calculer des intervalles généralisés \mathbf{z}_1 et \mathbf{z}_2 qui vérifient la proposition quantifiée suivante :

$$(\forall a \in \mathbf{a}) \left(\begin{array}{l} \mathbf{Q}^{(\mathbf{z}_1)} z_1 \in \text{pro } \mathbf{z}_1 \\ \mathbf{Q}^{(\mathbf{z}_2)} z_2 \in \text{pro } \mathbf{z}_2 \end{array} \right) (\exists x_1 \in \text{pro } \mathbf{x}_1)(\exists x_2 \in \text{pro } \mathbf{x}_2)(z = F(a, x)), \quad (10.4)$$

où les blocs $(\mathbf{Q}^{(\mathbf{z}_k)} z_k \in \text{pro } \mathbf{z}_k)$ sont placés de telle sorte que les quantificateurs universels précèdent les quantificateurs existentiels. Une façon d'obtenir une telle interprétation en utilisant les extensions \mathbf{f}_k des composantes f_k de f est d'effectuer les calculs suivants : $\mathbf{z}_1 = \mathbf{f}_1(\mathbf{a}, \mathbf{x}_1, \text{pro } \mathbf{x}_2)$ et $\mathbf{z}_2 = \mathbf{f}_2(\mathbf{a}, \text{pro } \mathbf{x}_1, \mathbf{x}_2)$. Ainsi les deux propositions quantifiées suivantes sont vraies :

$$\begin{aligned} & (\forall a \in \mathbf{a})(\forall x_2 \in \text{pro } \mathbf{x}_2)(\mathbf{Q}^{(\mathbf{z}_1)} z_1 \in \text{pro } \mathbf{z}_1)(\exists x_1 \in \text{pro } \mathbf{x}_1)(z_1 = f_1(a, x)) \\ & (\forall a \in \mathbf{a})(\forall x_1 \in \text{pro } \mathbf{x}_1)(\mathbf{Q}^{(\mathbf{z}_2)} z_2 \in \text{pro } \mathbf{z}_2)(\exists x_2 \in \text{pro } \mathbf{x}_2)(z_2 = f_2(a, x)) \end{aligned}$$

Il peut alors être prouvé que ces deux dernières propositions quantifiées impliquent la proposition quantifiée (10.4). L'autre choix $\mathbf{z}_1 = \mathbf{f}_1(\mathbf{a}, \text{pro } \mathbf{x}_1, \mathbf{x}_2)$ et $\mathbf{z}_2 = \mathbf{f}_2(\mathbf{a}, \mathbf{x}_1, \text{pro } \mathbf{x}_2)$, qui conduirait aux interprétations

$$\begin{aligned} & (\forall a \in \mathbf{a}) (\forall x_1 \in \text{pro } \mathbf{x}_1) (\mathbf{Q}^{(\mathbf{z}_1)} z_1 \in \text{pro } \mathbf{z}_1) (\exists x_2 \in \text{pro } \mathbf{x}_2) (z_1 = f_1(a, x)) \\ & (\forall a \in \mathbf{a}) (\forall x_2 \in \text{pro } \mathbf{x}_2) (\mathbf{Q}^{(\mathbf{z}_2)} z_2 \in \text{pro } \mathbf{z}_2) (\exists x_1 \in \text{pro } \mathbf{x}_1) (z_2 = f_2(a, x)) \end{aligned}$$

serait aussi correct. Cependant, les calculs naïfs $\mathbf{z}_1 = \mathbf{f}_1(\mathbf{a}, \mathbf{x}_1, \mathbf{x}_2)$ et $\mathbf{z}_2 = \mathbf{f}_2(\mathbf{a}, \mathbf{x}_1, \mathbf{x}_2)$ conduiraient aux interprétations suivantes :

$$\begin{aligned} & (\forall a \in \mathbf{a}) (\mathbf{Q}_1 z_1 \in \text{pro } \mathbf{z}_1) (\exists x_1 \in \text{pro } \mathbf{x}_1) (\exists x_2 \in \text{pro } \mathbf{x}_2) (z_1 = f_1(a, x)) \\ & (\forall a \in \mathbf{a}) (\mathbf{Q}_2 z_2 \in \text{pro } \mathbf{z}_2) (\exists x_1 \in \text{pro } \mathbf{x}_1) (\exists x_2 \in \text{pro } \mathbf{x}_2) (z_2 = f_2(a, x)) \end{aligned}$$

La conjonction de ces deux propositions quantifiées n'implique pas en général (10.4).

Chapitre 11

AE-extensions des relations réelles

Ce chapitre présente les AE-extensions des relations réelles. Il s'agit d'un *outil théorique* qui va simplifier la définition des AE-extensions des fonctions continues dans le chapitre suivant et l'expression de certaines propriétés qui seront utilisées par la suite.

Les AE-extensions des relations réelles permettent de travailler sur des propositions quantifiées (sous la forme AE, i.e. dont les quantificateurs universels précèdent les quantificateurs existentiels) en utilisant des intervalles généralisés à la place des couples quantificateur/intervalles, donc sans utiliser explicitement les quantificateurs. De nombreuses propriétés des propositions quantifiées seront alors plus facilement décrites et utilisées. En particulier, l'inclusion entre intervalles généralisés a un fort lien avec la sémantique des AE-extensions des relations, et va permettre d'exprimer facilement des propriétés importantes de ces dernières.

11.1 L'inclusion entre intervalles généralisés

Quand elle est appliquée aux intervalles propres, l'inclusion entre intervalles généralisés coïncide avec l'inclusion des intervalles classiques. Elle a donc une interprétation en termes d'ensembles de nombres réels : si $\mathbf{x}, \mathbf{y} \in \mathbb{IR}$ alors

$$\mathbf{x} \subseteq \mathbf{y} \iff \{x \in \mathbb{R} \mid \inf \mathbf{x} \leq x \leq \sup \mathbf{x}\} \subseteq \{y \in \mathbb{R} \mid \inf \mathbf{y} \leq y \leq \sup \mathbf{y}\}$$

Une interprétation en termes d'ensembles de nombres réels (c'est à dire qui ne met en jeu que des intervalles propres) est maintenant présentée pour les autres cas de l'inclusion entre intervalles généralisés. Le lemme suivant montre que l'inclusion entre un intervalle propre et un intervalle impropre est liée à l'intersection entre les ensembles de nombres réels mis en jeu.

Lemme IV.1. *Soient $\mathbf{x} \in \mathbb{IR}$ et $\mathbf{y} \in \mathbb{IR}$. Alors,*

$$\mathbf{x} \cap \mathbf{y} \neq \emptyset \iff (\text{dual } \mathbf{x}) \subseteq \mathbf{y}$$

Démonstration. L'intersection entre deux intervalles est vide si et seulement si l'un est strictement plus petit que l'autre. Ce qui conduit à

$$\begin{aligned} \mathbf{x} \cap \mathbf{y} \neq \emptyset &\iff \neg(\mathbf{y} < \mathbf{x} \vee \mathbf{x} < \mathbf{y}) \\ &\iff \neg(\bar{\mathbf{y}} < \underline{\mathbf{x}} \vee \bar{\mathbf{x}} < \underline{\mathbf{y}}) \\ &\iff \bar{\mathbf{y}} \geq \underline{\mathbf{x}} \wedge \bar{\mathbf{x}} \geq \underline{\mathbf{y}} \end{aligned}$$

Il reste juste à remarquer que $(\text{dual } \mathbf{x}) = [\bar{\mathbf{x}}, \underline{\mathbf{x}}]$ et en utilisant l'expression de l'inclusion nous obtenons finalement $(\text{dual } \mathbf{x}) \subseteq \mathbf{y} \iff \underline{\mathbf{y}} \leq \bar{\mathbf{x}} \wedge \underline{\mathbf{x}} \leq \bar{\mathbf{y}}$. \square

Remarque. Ce lemme est implicitement donné dans la théorie des intervalles modaux (lemmes I.2.4 et I.2.8) et utilisé dans Hansen[24] et Shary[92].

La proposition suivante donne l'interprétation des différents cas rencontrés dans l'inclusion de deux intervalles généralisés.

Proposition IV.1. Soient $\mathbf{x}, \mathbf{y} \in \mathbb{KR}$. Alors, $\mathbf{x} \subseteq \mathbf{y}$ est équivalent à la disjonction des conditions suivantes :

1. $\mathbf{x} \in \mathbb{IR} \wedge \mathbf{y} \in \mathbb{IR} \wedge \mathbf{x} \subseteq \mathbf{y}$
2. $\mathbf{x} \in \overline{\mathbb{IR}} \wedge \mathbf{y} \in \mathbb{IR} \wedge (\text{pro } \mathbf{x}) \cap \mathbf{y} \neq \emptyset$
3. $\mathbf{x} \in \overline{\mathbb{IR}} \wedge \mathbf{y} \in \overline{\mathbb{IR}} \wedge (\text{pro } \mathbf{x}) \supseteq (\text{pro } \mathbf{y})$
4. $\mathbf{x} \in \mathbb{IR} \wedge \mathbf{y} \in \overline{\mathbb{IR}} \wedge \mathbf{x} = \mathbf{y} = [x, x]$

Démonstration. Démontrons tout d'abord que l'inclusion $\mathbf{x} \subseteq \mathbf{y}$ implique la disjonction énoncée. Étant donné deux intervalles généralisés \mathbf{x} et \mathbf{y} , l'une des quatre conditions $\mathbf{x} \in \mathbb{IR} \wedge \mathbf{y} \in \mathbb{IR}$, $\mathbf{x} \in \overline{\mathbb{IR}} \wedge \mathbf{y} \in \mathbb{IR}$, $\mathbf{x} \in \overline{\mathbb{IR}} \wedge \mathbf{y} \in \overline{\mathbb{IR}}$ ou $\mathbf{x} \in \mathbb{IR} \wedge \mathbf{y} \in \overline{\mathbb{IR}}$ est vraie. Supposons que $\mathbf{x} \in \mathbb{IR} \wedge \mathbf{y} \in \mathbb{IR}$. L'inclusion des intervalles généralisés correspond alors à l'inclusion classique. Supposons que $\mathbf{x} \in \overline{\mathbb{IR}} \wedge \mathbf{y} \in \mathbb{IR}$. Nous pouvons appliquer le Lemme IV.1 pour obtenir $(\text{pro } \mathbf{x}) \cap \mathbf{y} \neq \emptyset \iff (\text{imp } \mathbf{x}) \subseteq \mathbf{y}$. Finalement, comme \mathbf{x} est impropre, $(\text{imp } \mathbf{x}) = \mathbf{x}$. Supposons que $\mathbf{x} \in \overline{\mathbb{IR}} \wedge \mathbf{y} \in \overline{\mathbb{IR}}$. Nous avons $\mathbf{x} \subseteq \mathbf{y} \iff (\text{dual } \mathbf{x}) \supseteq (\text{dual } \mathbf{y})$ et, \mathbf{x} et \mathbf{y} étant impropres, $(\text{dual } \mathbf{x}) = (\text{pro } \mathbf{x})$ et $(\text{dual } \mathbf{y}) = (\text{pro } \mathbf{y})$. Finalement le dernier cas est une conséquence directe de la définition de l'inclusion. La seconde implication est implicitement démontrée car nous avons démontré des équivalences. \square

Remarque. Le dernier cas montre simplement qu'un intervalle impropre ne peut pas contenir un intervalle propre, sauf si tous les deux sont égaux et dégénérés.

L'exemple suivant expose une situation rencontrée dans chacun des trois premiers cas de la proposition précédente.

Exemple IV.8.

1. $[-1, 1] \subseteq [-2, 2]$ est interprété comme l'inclusion entre intervalles classiques ;
2. $[2, -2] \subseteq [1, 3]$ est interprété par $[-2, 2] \cap [1, 3] \neq \emptyset$;
3. $[2, -2] \subseteq [1, -1]$ est interprété par $[-1, 1] \subseteq [-2, 2]$;

Le prochain corollaire est une autre formulation des cas 2. et 3. de la Proposition IV.1. Il sera utile dans de nombreuses situations rencontrées par la suite.

Corollaire IV.1. Soient $\mathbf{x} \in \overline{\mathbb{IR}}$ et $\mathbf{y} \in \mathbb{KR}$. Alors,

$$\mathbf{x} \subseteq \mathbf{y} \iff (\mathbf{Q}^{(\mathbf{y})} t \in \text{pro } \mathbf{y})(\phi(t))$$

où $\phi(t)$ est définie par $\phi(t) \iff t \in \text{pro } \mathbf{x}$, $\mathbf{Q}^{(\mathbf{y})} = \exists$ si $\mathbf{y} \in \mathbb{IR}$ et $\mathbf{Q}^{(\mathbf{y})} = \forall$ sinon.

Démonstration. Supposons d'abord que $\mathbf{y} \in \mathbb{IR}$. Alors, grâce à la Proposition IV.1, $\mathbf{x} \subseteq \mathbf{y} \iff (\text{pro } \mathbf{x}) \cap \mathbf{y} \neq \emptyset$ qui est finalement équivalent à $(\exists t \in \mathbf{y})(\phi(t))$. Supposons maintenant que $\mathbf{y} \in \overline{\mathbb{IR}}$. Alors, grâce à la proposition Proposition IV.1, $\mathbf{x} \subseteq \mathbf{y} \iff (\text{pro } \mathbf{x}) \supseteq (\text{pro } \mathbf{y})$ qui est finalement équivalent à $(\forall t \in \text{pro } \mathbf{y})(\phi(t))$. \square

Cette section est terminée par une présentation informelle de l'interprétation de l'intersection et la réunion entre intervalles généralisés. Ces interprétations sont des conséquences directes de l'interprétation de l'inclusion et de la définition de ces opérations.

Exemple IV.9. *Les quatre cas suivants illustrent les différentes interprétations de l'intersection entre intervalles généralisés :*

- $[-1, 1] \wedge [0, 2] = [0, 1]$ donc l'intervalle propre $[0, 1]$ est le plus grand intervalle propre contenu dans $[-1, 1]$ et $[0, 2]$;
- $[-1, 1] \wedge [2, 3] = [2, 1]$ donc l'intervalle propre $\text{pro}[2, 1] = [1, 2]$ est le plus petit intervalle propre (le plus grand intervalle impropre est le plus petit après projection propre) qui a une intersection non vide avec $[-1, 1]$ et $[2, 3]$.
- $[-1, 1] \wedge [3, 2] = [3, 1]$ donc l'intervalle propre $\text{pro}[3, 1] = [1, 3]$ est le plus petit intervalle propre qui intersecte $[-1, 1]$ et contient $[2, 3]$.
- $[1, -1] \wedge [3, 2] = [3, -1]$ donc l'intervalle propre $\text{pro}[3, -1] = [-1, 3]$ est le plus petit intervalle qui contient $[-1, 1]$ et $[2, 3]$.

L'intersection et la réunion étant liées par

$$\mathbf{x} \vee \mathbf{y} = \mathbf{z} \iff (\text{dual } \mathbf{x}) \wedge (\text{dual } \mathbf{y}) = (\text{dual } \mathbf{z})$$

les interprétations de la réunion d'intervalles généralisés sont obtenues grâce aux interprétations de l'intersection.

11.2 AE-extensions des relations réelles

Étant donnée une relation réelle ϕ , son extension aux intervalles généralisés est maintenant définie. Ces extensions sont appelées AE-extensions à cause de la succession des quantificateurs universels (A(II)) suivis des quantificateurs existentiels (E(xists)) dans leur interprétation. Les AE-extensions des relations réelles sont introduites car elles représentent un langage utile pour la description des propositions quantifiées qui seront utilisées dans la suite ainsi que pour la description de leurs propriétés.

Définition IV.1. *Soit ϕ une relation de \mathbb{R}^n . L'AE-extension de ϕ est notée avec le même symbole ϕ et est définie pour tout $\mathbf{x} \in \mathbb{KR}^n$ par*

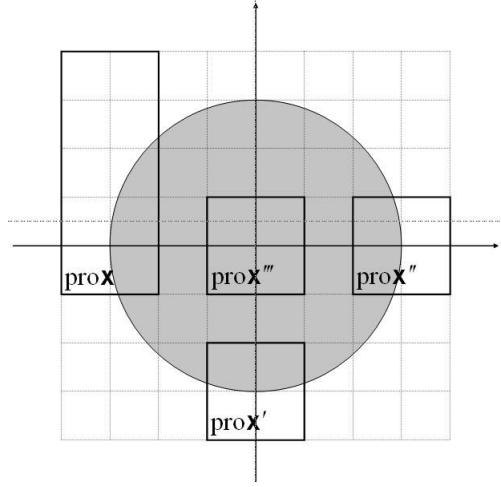
$$\phi(\mathbf{x}) \iff (\forall x_I \in \text{pro } \mathbf{x}_I) (\exists x_P \in \mathbf{x}_P) (\phi(x))$$

où $P = P(\mathbf{x})$ et $I = I(\mathbf{x})$ sont respectivement les indices des composantes propres et impropres de \mathbf{x} (vérifiant donc $P \cup I = [1..n]$ et $P \cap I = \emptyset$).

Remarques. 1. *L'abus d'écriture suivant sera autorisé tout au long de la thèse : les notations générales utilisées dans la définition IV.1 sont aussi valables dans les cas $P = \emptyset$ or $I = \emptyset$. Ces cas correspondent respectivement à $(\forall x \in \mathbf{x}) (\phi(x))$ et $(\exists x \in \mathbf{x}) (\phi(x))$.*

2. *Par définition des fonctions P et I , les intervalles dégénérés sont considérés comme propres. Cette convention n'a pas d'incidence sur l'interprétation de l'AE-extension car $(\exists x \in [a, a]) (\phi(x)) \iff (\forall x \in [a, a]) (\phi(x)) \iff \phi(a)$.*

3. *Quand elle est restreinte aux intervalles propres, la définition IV.1 devient $\phi(\mathbf{x}) \iff (\exists x \in \mathbf{x}) (\phi(x))$. Cela ne correspond pas exactement à la définition d'une extension aux intervalles classiques proposée dans Collavizza[39] et Benhamou[27], où une extension de ϕ est définie par $\phi(\mathbf{x}) \iff (\exists x \in \mathbf{x}) (\phi(x))$.*

FIG. 11.1 – Graphe de la relation $x_1^2 + x_2^2 \leq 9$.

L'exemple suivant illustre les différentes interprétations de l'AE-extension d'une relation à deux variables :

Exemple IV.10. *Considérons la relation $\phi(x) \iff x_1^2 + x_2^2 \leq 9$. Son graphe est illustré par la figure 11.1. Considérons les quatre intervalles généralisés $\mathbf{x} = ([-4, -2], [-1, 4])^T$, $\mathbf{x}' = ([1, -1], [-4, -2])^T$, $\mathbf{x}'' = ([2, 4], [1, -1])^T$ et $\mathbf{x}''' = ([1, -1], [1, -1])^T$. Les quatre intervalles propres $\text{pro } \mathbf{x}$, $\text{pro } \mathbf{x}'$ et $\text{pro } \mathbf{x}''$ et $\text{pro } \mathbf{x}'''$ sont aussi représenté dans la figure 11.1. Alors, nous pouvons vérifier sur la figure les interprétations suivantes de l'AE-extension de ϕ :*

- $\phi(\mathbf{x})$ est vraie parce que $(\exists x_1 \in \mathbf{x}_1)(\exists x_2 \in \mathbf{x}_2)(\phi(x))$ est vraie.
- $\phi(\mathbf{x}')$ est vraie parce que $(\forall x_1 \in \text{pro } \mathbf{x}'_1)(\exists x_2 \in \mathbf{x}'_2)(\phi(x))$ est vraie.
- $\phi(\mathbf{x}'')$ est vraie parce que $(\forall x_2 \in \text{pro } \mathbf{x}''_2)(\exists x_1 \in \mathbf{x}''_1)(\phi(x))$ est vraie.
- $\phi(\mathbf{x}''')$ est vraie parce que $(\forall x_1 \in \text{pro } \mathbf{x}'''_1)(\forall x_2 \in \text{pro } \mathbf{x}'''_2)(\phi(x))$ est vraie.

L'interprétation des AE-extensions est fortement liée à l'inclusion entre les intervalles généralisés. Ce lien est fondamental et est donné par la proposition suivante :

Proposition IV.2. *Soient ϕ une relation définie sur \mathbb{R}^n , $\mathbf{x}, \mathbf{y} \in \mathbb{K}\mathbb{R}^n$ tels que $\mathbf{x} \subseteq \mathbf{y}$. Alors,*

$$\phi(\mathbf{x}) \implies \phi(\mathbf{y})$$

Démonstration. Définissons les ensembles d'indices suivants :

- $P = P(\mathbf{x}) \cap P(\mathbf{y})$
- $I = I(\mathbf{x}) \cap I(\mathbf{y})$
- $E = [1..n] \setminus (P \cup I)$

Grâce à la Proposition IV.1, l'inclusion $\mathbf{x} \subseteq \mathbf{y}$ implique

- si $k \in P$ alors $\mathbf{x}_k \subseteq \mathbf{y}_k$ et donc $\mathbf{x}_P \subseteq \mathbf{y}_P$;
- si $k \in I$ alors $\text{pro } \mathbf{y}_k \subseteq \text{pro } \mathbf{x}_k$ et donc $\text{pro } \mathbf{y}_I \subseteq \text{pro } \mathbf{x}_I$
- si $k \in E$ alors $\mathbf{x}_k \in \overline{\mathbb{R}}$, $\mathbf{y}_k \in \mathbb{R}$ et $(\text{pro } \mathbf{x}_k) \cap \mathbf{y}_k \neq \emptyset$ et donc $(\text{pro } \mathbf{x}_E) \cap \mathbf{y}_E \neq \emptyset$

Supposons que les trois ensembles d'indices définis ci-dessus sont non-vides (les autres cas étant similaires et plus simples). Maintenant, par définition d'une AE-extension, $\phi(\mathbf{x})$

est vraie si et seulement si

$$(\forall x_I \in \text{pro } \mathbf{x}_I)(\forall x_E \in \text{pro } \mathbf{x}_E)(\exists x_P \in \mathbf{x}_P)(\phi(x)) \quad (11.1)$$

et $\phi(\mathbf{y})$ est vraie si et seulement si

$$(\forall x_I \in \text{pro } \mathbf{y}_I)(\exists x_E \in \mathbf{y}_E)(\exists x_P \in \mathbf{y}_P)(\phi(x)) \quad (11.2)$$

Nous devons donc prouver que (11.1) implique (11.2). Il est clair que cette implication est vraie car $\mathbf{x}_P \subseteq \mathbf{y}_P$, $\text{pro } \mathbf{y}_I \subseteq \text{pro } \mathbf{x}_I$ et $(\text{pro } \mathbf{x}_E) \cap \mathbf{y}_E \neq \emptyset$. \square

Exemple IV.11. *Considérons une relation ϕ quelconque de \mathbb{R}^3 et les intervalles*

$$\mathbf{x} = ([2, -2], [2, -2], [-1, 1])^T \quad \text{et} \quad \mathbf{y} = ([1, -1], [1, 4], [-2, 2])^T.$$

Comme $\mathbf{x} \subseteq \mathbf{y}$, la Proposition IV.2 prouve que $\phi(\mathbf{x}) \implies \phi(\mathbf{y})$. C'est en effet vrai car

$$(\forall x_1 \in [-2, 2])(\forall x_2 \in [-2, 2])(\exists x_3 \in [-1, 1])(\phi(x))$$

implique

$$(\forall x_1 \in [-1, 1])(\exists x_2 \in [1, 4])(\exists x_3 \in [-2, 2])(\phi(x))$$

quelque soit relation ϕ .

Finalement, les deux prochaines propositions sont des résultats techniques qui expriment dans le langage des AE-extensions des propriétés des propositions quantifiées qui seront utiles.

Proposition IV.3. *Soient ϕ une relation de \mathbb{R}^n et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Définissons $\mathbf{a} \in \mathbb{I}\mathbb{R}^n$ par*

$$\mathbf{a} = \square\{x \in (\text{pro } \mathbf{x}) \mid \phi(x)\}.$$

Alors, $\phi(\mathbf{x}) \implies (\text{imp } \mathbf{a}) \subseteq \mathbf{x}$ et $\phi(\mathbf{x}) \implies \phi(\mathbf{x} \wedge \mathbf{a})$.

Démonstration. Définissons les ensembles d'indices $P = P(\mathbf{x})$ et $I = I(\mathbf{x})$. Nous supposons que P et I sont non-vides, les autres cas étant similaires et plus simples. Commençons par la première implication. Considérons $k \in [1..n]$ quelconque. Par définition de \mathbf{a} , nous avons

$$x \in (\text{pro } \mathbf{x}) \wedge \phi(x) \implies x \in \mathbf{a} \implies x_k \in \mathbf{a}_k$$

Donc, $\phi(\mathbf{x})$ implique

$$(\forall x_I \in \text{pro } \mathbf{x}_I)(\exists x_P \in \mathbf{x}_P)(x_k \in \mathbf{a}_k)$$

ce qui implique

$$(\mathbb{Q}^{(\mathbf{x}_k)} x_k \in \text{pro } \mathbf{x}_k)(x_k \in \mathbf{a}_k)$$

parce que la relation $x_k \in \mathbf{a}_k$ ne fait pas intervenir les variables x_j pour $j \neq k$. Finalement, grâce au corollaire IV.1, nous avons $(\text{imp } \mathbf{a}_k) \subseteq \mathbf{x}_k$. Cette inclusion étant vraie quelque soit $k \in [1..n]$, nous avons $(\text{imp } \mathbf{a}) \subseteq \mathbf{x}$. Considérons maintenant la seconde implication. Notons $\mathbf{x} \wedge \mathbf{a}$ par \mathbf{z} . Supposons que $\phi(\mathbf{x})$ est vraie et prouvons que $\phi(\mathbf{z})$ est alors vraie. Comme $\phi(\mathbf{x})$ est vraie, le premier cas de la proposition montre que $(\text{imp } \mathbf{a}) \subseteq \mathbf{x}$. Cette inclusion est équivalente à $\text{dual } \mathbf{x} \subseteq \mathbf{a}$ parce que \mathbf{a} est propre. Explicitons les conséquences de cette inclusion composante par composante. D'une part, si \mathbf{x}_k est impropre, nous avons $(\text{pro } \mathbf{x}_k) \subseteq \mathbf{a}_k$. Comme $\mathbf{x}_k \subseteq (\text{pro } \mathbf{x}_k)$, nous avons $\mathbf{x}_k \subseteq \mathbf{a}_k$ et donc $\mathbf{z}_k = \mathbf{x}_k \in \overline{\mathbb{I}\mathbb{R}}$. D'autre

part, si \mathbf{x}_k est propre le Lemme IV.1 montre que $\mathbf{x}_k \cap \mathbf{a}_k \neq \emptyset$ et donc $\mathbf{z}_k = \mathbf{x}_k \cap \mathbf{a}_k \in \mathbb{IR}$.
Maintenant,

$$(\forall x_I \in \text{pro } \mathbf{x}_I)(\exists x_P \in \mathbf{x}_P)(\phi(x))$$

implique

$$(\forall x_I \in \text{pro } \mathbf{x}_I)(\exists x_P \in (\mathbf{x}_P \cap \mathbf{a}_P))(\phi(x))$$

parce que, par définition de \mathbf{a} , nous avons $x_P \in \mathbf{x}_P \wedge x_P \notin \mathbf{a}_P$ implique $\neg\phi(x)$ quelque soit $x_I \in \mathbf{x}_I$. Cette dernière proposition quantifiée correspond à $\phi(\mathbf{x} \wedge \mathbf{a})$ car $\mathbf{z}_P = \mathbf{x}_P \cap \mathbf{a}_P$ et $\mathbf{z}_I = \mathbf{x}_I$. \square

La prochaine proposition nécessite le lemme suivant, qui est une conséquence directe du fait que la distance entre intervalles propres correspond à la distance de Hausdorff.

Lemme IV.2. Soit $\mathbf{z}^{(k)} \in \mathbb{IR}^p$ une suite d'intervalles propres. Si cette suite converge vers $\mathbf{z} \in \mathbb{IR}^p$ alors les deux conditions suivantes sont vraies :

1. toute suite $(z^{(k)})_{k \in \mathbb{N}}$ satisfaisant $z^{(k)} \in \mathbf{z}^{(k)}$ possède au moins un point d'accumulation et tout ses points d'accumulation sont dans \mathbf{z} .
2. pour tout $z \in \mathbf{z}$, il existe une suite convergente $(z^{(k)})_{k \in \mathbb{N}}$ satisfaisant $z^{(k)} \in \mathbf{z}^{(k)}$ dont la limite est z .

Démonstration.

1. La suite $\mathbf{z}^{(k)}$ étant convergente, elle aussi bornée. Donc, toute suite $z^{(k)} \in \mathbf{z}^{(k)}$ est bornée. Cette dernière a donc un point d'accumulation par le théorème de Bolzano-Weierstrass. Par définition de la distance de Hausdorff et parce que $z^{(k)} \in \mathbf{z}^{(k)}$, nous avons pour $k \in \mathbb{N}$

$$\text{dist}(\mathbf{z}^{(k)}, \mathbf{z}) \geq \max_{z \in \mathbf{z}^{(k)}} \min_{y \in \mathbf{z}} d(z, y) \geq \min_{y \in \mathbf{z}} \text{dist}(z^{(k)}, y) \geq 0$$

Comme la suite $\mathbf{z}^{(k)}$ converge vers \mathbf{z} , nous avons aussi $\lim_{k \rightarrow \infty} \text{dist}(\mathbf{z}^{(k)}, \mathbf{z}) = 0$. Nous obtenons donc finalement

$$\lim_{k \rightarrow \infty} \min_{y \in \mathbf{z}} \text{dist}(z^{(k)}, y) = 0$$

Considérons un point d'accumulation quelconque z^* de la suite $z^{(k)}$. Par définition d'un point d'accumulation, nous avons

$$\min_{y \in \mathbf{z}} \text{dist}(z^*, y) = \lim_{k \rightarrow \infty} \min_{y \in \mathbf{z}} \text{dist}(z^{(k)}, y) = 0.$$

Finalement, comme \mathbf{z} est fermé dans \mathbb{R}^n , nous avons $z^* \in \mathbf{z}$.

2. Quelque soit $k \in \mathbb{N}$ et $z \in \mathbf{z}$, les intervalles $\mathbf{z}^{(k)}$ étant non-vides et fermés dans \mathbb{R}^n , il existe $z^{(k)} \in \mathbf{z}^{(k)}$ qui vérifie

$$\text{dist}(z^{(k)}, z) = \min_{y \in \mathbf{z}^{(k)}} \text{dist}(y, z)$$

De manière similaire au cas 1, nous pouvons montrer que

$$\lim_{k \rightarrow \infty} \min_{y \in \mathbf{z}^{(k)}} \text{dist}(y, z) = 0$$

Nous avons donc finalement montré que $\lim_{k \rightarrow \infty} \text{dist}(z^{(k)}, z) = 0$ et nous avons donc construit une suite $z^{(k)} \in \mathbf{z}^{(k)}$ qui converge vers z .

□

Remarque. La réciproque du lemme précédent est vraie, mais n'est pas démontrée ici.

Rappelons que par définition, étant donné un ensemble E et un sous-ensemble F de E , F est fermé dans E si et seulement si tout point de E qui est la limite d'une suite convergente de point de F est aussi un point de F .

Proposition IV.4. Soit ϕ une relation de \mathbb{R}^n . Si le graphe de ϕ est fermé dans \mathbb{R}^n alors le graphe de l' AE -extension de ϕ est fermé dans $\mathbb{K}\mathbb{R}^n$.

Démonstration. Considérons une suite convergente quelconque $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$ qui vérifie $\phi(\mathbf{x}^{(k)})$ pour tout $k \in \mathbb{N}$. Nous devons prouver que $\phi(\mathbf{x}^{(\infty)})$ est vraie, où $\mathbf{x}^{(\infty)}$ est la limite de la suite $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$. Considérons une sous-suite $(\mathbf{y}^{(k)})_{k \in \mathbb{N}}$ de $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$ dont les éléments ont des qualités propre/impropres constantes pour chaque composante. C'est en effet possible, car sinon il n'existerait qu'un nombre fini de $\mathbf{x}^{(k)}$ pour chacune des 2^n possibilités de qualités propres/impropres des composantes, c'est à dire un nombre fini de $\mathbf{x}^{(k)}$, ce qui est absurde. Comme une sous-suite d'une suite convergente converge vers la même limite, nous avons juste à prouver que $\phi(\mathbf{y}^{(\infty)})$ est vraie. Maintenant, les composantes de $\mathbf{y}^{(\infty)}$ ont les mêmes qualités propres/impropres que les composantes de $\mathbf{y}^{(k)}$ car $\mathbb{I}\mathbb{R}$ et $\overline{\mathbb{I}\mathbb{R}}$ sont fermés dans $\mathbb{K}\mathbb{R}$. Définissons $P = P(\mathbf{y}^{(k)}) = P(\mathbf{y}^{(\infty)})$ et $I = I(\mathbf{y}^{(k)}) = I(\mathbf{y}^{(\infty)})$. Nous supposons que P et I sont non-vides, les autres cas étant similaire et plus simples. Alors, quelque soit $k \in \mathbb{N}$, la proposition quantifiée suivante est vraie :

$$(\forall x_I \in \text{pro } \mathbf{y}_I^{(k)}) (\exists x_P \in \mathbf{y}_P^{(k)}) (\phi(x)) \quad (11.3)$$

La suite $\text{pro } \mathbf{y}^{(k)}$ convergeant vers $\text{pro } \mathbf{y}^{(\infty)}$ (car pro est continue) le Lemme IV.2 montre que pour tout $x_I \in \text{pro } \mathbf{y}_I^{(\infty)}$, il existe une suite $x_I^{(k)} \in \text{pro } \mathbf{y}_I^{(k)}$ qui converge vers x_I . Ensuite, en utilisant la proposition quantifiée (11.3), pour tout $k \in \mathbb{N}$ il existe $x_P^{(k)} \in \mathbf{y}_P^{(k)}$ tel que $\phi(x^{(k)})$. La suite $(x^{(k)})_{k \in \mathbb{N}}$ ainsi définie, qui vérifie $x^{(k)} \in \text{pro } \mathbf{y}^{(k)}$, a au moins un point d'accumulation $x^* \in \text{pro } \mathbf{y}^{(\infty)}$ par le Lemme IV.2. Le sous vecteur x_I^* est un point d'accumulation de la suite $x_I^{(k)}$ qui converge vers x_I et nous avons ainsi $x_I^* = x_I$. De plus, $\phi(x^*)$ est vraie car, par hypothèse, le graphe de ϕ est fermé dans \mathbb{R}^n . Finalement, nous avons prouvé la proposition quantifiée suivante :

$$(\forall x_I \in \text{pro } \mathbf{y}_I^{(\infty)}) (\exists x_P \in \mathbf{y}_P^{(\infty)}) (\phi(x)).$$

Et $\phi(\mathbf{y}^{(\infty)})$ est donc vraie. □

Chapitre 12

AE-extensions des fonctions continues $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

Les AE-extensions des fonctions continues sont maintenant définies, avec les notions d'optimalité et d'ordre de convergence. L'étude de l'arrondi des AE-extensions est présentée à la fin du chapitre. Les AE-extensions généralisent les extensions aux intervalles classiques en offrant des interprétations plus riches grâce à la prise en compte de la qualité propre/impropre des intervalles généralisés. Les notions d'optimalité et d'ordre de convergence définies dans le contexte des AE-extensions généralisent les définitions correspondantes dans le contexte des intervalles classiques.

12.1 Définition des AE-extensions

La définition des extensions aux intervalles classiques est tout d'abord reformulée en utilisant les AE-extensions des relations réelles. Cette nouvelle formulation de la définition pourra être appliquée aux fonctions d'intervalles généralisés, ce qui conduira à la définition des AE-extensions. Rappelons tout d'abord la définition d'une extension aux intervalles classiques.

Définition IV.2 (Neumaier[8]). Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \longrightarrow \mathbb{IR}^m$ une fonction d'intervalles. \mathbf{g} est une extension aux intervalles de f si et seulement si les deux conditions suivantes sont vérifiées :

1. $(\forall x \in \mathbb{R}^n)(\mathbf{g}(x) = f(x))$
2. $(\forall \mathbf{x} \in \mathbb{IR}^n)(\mathbf{g}(\mathbf{x}) \supseteq \text{range}(f, \mathbf{x}))$

Il est aisé de constater que l'inclusion $\text{range}(f, \mathbf{x}) \subseteq \mathbf{g}(\mathbf{x})$ est équivalente à la proposition quantifiée suivante :

$$(\forall x \in \mathbf{x})(\exists z \in \mathbf{g}(\mathbf{x}))(z = f(x))$$

Cette proposition quantifiée correspond à l'AE-extension de la relation $\phi(x, z)$ définie par $\phi(x, z) \iff z = f(x)$ évaluée en $(\text{dual } \mathbf{x}, \mathbf{g}(\mathbf{x}))$. En effet, d'une part $(\text{dual } \mathbf{x})$ étant impropre, les quantificateurs correspondants sont universels. D'autre part, $\mathbf{g}(\mathbf{x})$ étant propre, les quantificateurs correspondant sont existentiel. Nous avons donc l'équivalence suivante :

$$(\forall x \in \mathbf{x})(\exists z \in \mathbf{g}(\mathbf{x}))(z = f(x)) \iff \phi(\text{dual } \mathbf{x}, \mathbf{g}(\mathbf{x}))$$

Ainsi, la définition des extensions aux intervalles classiques peut être formulée de la manière suivante :

Définition IV.3 (Reformulation de la définition IV.2). Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{IR}^n \longrightarrow \mathbb{IR}^m$ une fonction d'intervalles. \mathbf{g} est une extension aux intervalles de f si et seulement si les deux conditions suivantes sont vérifiées :

1. $(\forall x \in \mathbb{R}^n) (\mathbf{g}(x) = f(x))$
2. $(\forall \mathbf{x} \in \mathbb{IR}^n) (\phi(\text{dual } \mathbf{x}, \mathbf{g}(\mathbf{x})))$

où la relation $\phi(x, z)$ est définie par $\phi(x, z) \iff z = f(x)$.

La définition des AE-extensions est finalement obtenue en considérant des fonctions d'intervalles généralisées à la place des fonctions d'intervalles classiques.

Définition IV.4. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{KR}^n \longrightarrow \mathbb{KR}^m$ une fonction d'intervalles généralisés. \mathbf{g} est une AE-extension de f si et seulement si les deux conditions suivantes sont vérifiées :

1. $(\forall x \in \mathbb{R}^n) (\mathbf{g}(x) = f(x))$
2. $(\forall \mathbf{x} \in \mathbb{KR}^n) (\phi(\text{dual } \mathbf{x}, \mathbf{g}(\mathbf{x})))$

où la relation $\phi(x, z)$ est définie par $\phi(x, z) \iff z = f(x)$.

La définition suivante sera utile :

Définition IV.5. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue, $\mathbf{x} \in \mathbb{KR}^n$ et $\mathbf{z} \in \mathbb{KR}^m$. L'intervalle \mathbf{z} est dit interprétable par rapport à f et \mathbf{x} (ou (f, \mathbf{x}) -interprétable) si $\phi(\text{dual } \mathbf{x}, \mathbf{z})$ est vraie, où la relation $\phi(x, z)$ est définie par $\phi(x, z) \iff z = f(x)$.

Ainsi, une fonction d'intervalles \mathbf{g} est une AE-extension de f si et seulement si pour tout $x \in \mathbb{R}^n$, $\mathbf{g}(x) = f(x)$ et pour tout $\mathbf{x} \in \mathbb{KR}^n$ l'intervalle $\mathbf{g}(\mathbf{x})$ est (f, \mathbf{x}) -interprétable. Certaines fonctions ne peuvent pas satisfaire $\mathbf{g}(x) = f(x)$ (par exemple les AE-extensions arrondies). Donc, en conservant la terminologie introduite dans Neumaier[8], une fonction d'intervalles \mathbf{g} qui vérifie seulement la deuxième condition de la définition IV.4 est appelée une AE-extension faible de f .

Informellement, quand tous les intervalles mis en jeux sont propres, la définition des AE-extensions coïncident avec la définition des extensions aux intervalles classiques. Quand un intervalle impropre remplace un intervalle propre, le quantificateur correspondant est inversé, en faisant attention à conserver l'ordre AE des quantificateurs dans la proposition quantifiée ainsi obtenue. Formellement, $\mathbf{z} \in \mathbb{KR}^m$ est (f, \mathbf{x}) -interprétable si et seulement si la proposition quantifiées suivante est vraie :

$$(\forall x_P \in \mathbf{x}_P) (\forall z_{I'} \in \text{pro } \mathbf{z}_{I'}) (\exists z_{P'} \in \mathbf{z}_{P'}) (\exists x_I \in \text{pro } \mathbf{x}_I) (z = f(x))$$

où $P = P(\mathbf{x})$, $I = I(\mathbf{x})$, $P' = P(\mathbf{z})$ et $I' = I(\mathbf{z})$ (lorsque l'un des ensembles d'indices est vide, le bloc correspondant n'est pas pris en compte). Dans le cas particulier des fonctions à valeurs réelles, i.e. $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, la proposition quantifiée précédente peut s'écrire en utilisant le quantificateur dépendant de la qualité propre/impropre qui a été défini dans le corollaire IV.1 :

$$(\forall x_P \in \mathbf{x}_P) (\mathbf{Q}^{(\mathbf{z})} z \in \text{pro } \mathbf{z}) (\exists x_I \in \text{pro } \mathbf{x}_I) (z = f(x))$$

où $Q^{(z)} = \exists$ si $\mathbf{z} \in \mathbb{IR}$, $Q^{(z)} = \forall$ sinon. Le bloc $(Q^{(z)}z \in \text{pro } \mathbf{z})$ est écrit au centre de la proposition de telle sorte que l'ordre AE soit conservé quelque soit le quantificateur $Q^{(z)}$.

Quelques propriétés générales des AE-extensions sont maintenant présentées. Premièrement, quand tous les arguments sont propres, le résultat d'une AE-extension est aussi propre. Ainsi, les AE-extensions étendent bien les extensions aux intervalles classiques à des arguments impropres.

Proposition IV.5. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue, $\mathbf{x} \in \mathbb{IR}^n$ et $\mathbf{z} \in \mathbb{KR}^m$. L'intervalle \mathbf{z} est (f, \mathbf{x}) -interprétable si et seulement si $\mathbf{z} \supseteq \square \text{range}(f, \mathbf{x})$, ce qui implique en particulier que \mathbf{z} soit propre.*

Démonstration. $\mathbf{z} \supseteq \square \text{range}(f, \mathbf{x})$ implique évidemment que \mathbf{z} soit (f, \mathbf{x}) -interprétable, comme illustré au début de la section. Il reste à prouver que si \mathbf{z} est (f, \mathbf{x}) -interprétable alors $\mathbf{z} \supseteq \square \text{range}(f, \mathbf{x})$. Supposons que $I \neq \emptyset$. Par définition, \mathbf{z} vérifie

$$(\forall x \in \mathbf{x})(\forall z_I \in \text{pro } \mathbf{z}_I)(\exists z_P \in \mathbf{z}_P)(z = f(x)),$$

ce qui implique

$$(\forall x \in \mathbf{x})(\forall z_I \in \text{pro } \mathbf{z}_I)(f_I(x) = z_I).$$

L'intervalle \mathbf{z}_I n'est pas dégénéré car les indices des composantes dégénérées sont contenus dans $P(\mathbf{z})$. Il existe donc $z_I \in \text{pro } \mathbf{z}_I$ et $z'_I \in \text{pro } \mathbf{z}_I$ qui vérifient $z_I \neq z'_I$. Donc, pour tout $x \in \mathbf{x}$ nous avons $f_I(x) = z_I$ et $f_I(x) = z'_I$, ce qui est absurde. Nous en concluons que $I = \emptyset$. Comme l'intervalle propre \mathbf{z} est (f, \mathbf{x}) -interprétable, il vérifie $(\forall x \in \mathbf{x})(\exists z \in \mathbf{z})(z = f(x))$, c'est à dire $\mathbf{z} \supseteq \square \text{range}(f, \mathbf{x})$. \square

Une question se pose : est-ce que toute fonction continue a au moins une AE-extension ? La prochaine proposition donne une réponse positive à cette question.

Proposition IV.6. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{r} : \mathbb{KR}^n \rightarrow \mathbb{IR}^m$ définie par $\mathbf{r}(\mathbf{x}) = \square \text{range}(f, \text{pro } \mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{KR}^n$. Alors, \mathbf{r} est une AE-extension de f .*

Démonstration. Remarquons d'abord que $\mathbf{r}(x) = f(x)$ pour tout $x \in \mathbb{R}^n$. Notons la relation $z = f(x)$ par $\phi(x, z)$ et considérons un $\mathbf{x} \in \mathbb{KR}^n$. Comme $\mathbf{r}(\mathbf{x}) = \square \text{range}(f, \text{pro } \mathbf{x})$, la proposition quantifiée suivante est vraie :

$$(\forall x \in \text{pro } \mathbf{x})(\exists z \in \mathbf{r}(\mathbf{x}))(z = f(x))$$

C'est à dire, $\phi(\text{imp } \mathbf{x}, \mathbf{r}(\mathbf{x}))$ est vraie. Maintenant l'inclusion $(\text{imp } \mathbf{x}) \subseteq (\text{dual } \mathbf{x})$ étant vraie, la Proposition IV.2 montre que $\phi(\text{dual } \mathbf{x}, \mathbf{r}(\mathbf{x}))$ est vraie. L'intervalle généralisé $\mathbf{r}(\mathbf{x})$ est donc (f, \mathbf{x}) -interprétable. \square

Exemple IV.12. *Soit $\mathbf{z} = \square \text{range}(f, \mathbf{x}_1, \mathbf{x}_2)$. Alors la proposition quantifiée*

$$(\forall x_1 \in \mathbf{x}_1)(\forall x_2 \in \mathbf{x}_2)(\exists z \in \mathbf{z})(z = f(x))$$

est vraie et implique

$$(\forall x_1 \in \mathbf{x}_1)(\exists x_2 \in \mathbf{x}_2)(\exists z \in \mathbf{z})(z = f(x))$$

Donc, l'intervalle \mathbf{z} est $(f, \mathbf{x}_1, \text{dual } \mathbf{x}_2)$ -interprétable.

Finalement, la proposition suivante donne une minoration (au sens de l'inclusion) pour toute AE-extension.

Proposition IV.7. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ et $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$. Si \mathbf{z} est (f, \mathbf{x}) -interprétable alors $\text{imp } \mathbf{r}(\mathbf{x}) \subseteq \mathbf{z}$, où $\mathbf{r}(\mathbf{x}) = \square \text{ range}(f, \text{pro } \mathbf{x})$.*

Démonstration. Notons la relation $z = f(x)$ par $\phi(x, z)$. Par définition, l'intervalle \mathbf{z} est (f, \mathbf{x}) -interprétable si et seulement si $\phi(\text{dual } \mathbf{x}, \mathbf{z})$. Définissons alors l'intervalle propre suivant :

$$\mathbf{y} = \square \{(x, z) \in (\text{pro } \mathbf{x}, \text{pro } \mathbf{z}) \mid \phi(x, z)\}$$

Nous avons $\mathbf{y} \subseteq (\text{pro } \mathbf{x}, \mathbf{r}(\mathbf{x}))$ car $\phi(x, z) \iff z = f(x)$ et $x \in \mathbf{x} \implies f(x) \in \mathbf{r}(\mathbf{x})$. Donc $(\text{imp } \mathbf{x}, \text{imp } \mathbf{r}(\mathbf{x})) \subseteq \text{imp } \mathbf{y}$. De plus, nous pouvons appliquer le premier cas de la Proposition IV.3 pour montrer que

$$\phi(\text{dual } \mathbf{x}, \mathbf{z}) \implies (\text{imp } \mathbf{y}) \subseteq (\text{dual } \mathbf{x}, \mathbf{z})$$

En combinant les deux dernières inclusions, nous obtenons $(\text{imp } \mathbf{x}, \text{imp } \mathbf{r}(\mathbf{x})) \subseteq (\text{dual } \mathbf{x}, \mathbf{z})$ qui implique finalement $\text{imp } \mathbf{r}(\mathbf{x}) \subseteq \mathbf{z}$ parce que $\text{imp } \mathbf{x} \subseteq \text{dual } \mathbf{x}$ est toujours vraie. \square

12.2 Interprétation des AE-extensions

Les AE-extensions ont des interprétations plus riches que les extensions classiques : elle peuvent conduire à un sous ensemble ou un sur ensemble de l'image d'une boîte par la fonction étendue ; elle peuvent aussi permettre de prouver qu'une boîte \mathbf{x} est dans la projection de $\{(x, y) \mid f(x, y) = z_0\}$. Des exemples de ces interprétations sont maintenant présentés. Dans cette section, $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ est une fonction continue et $\mathbf{x} \in \mathbb{K}\mathbb{R}$ et $\mathbf{y} \in \mathbb{K}\mathbb{R}$ sont deux intervalles. L'interprétation d'intervalles $\mathbf{z} \in \mathbb{K}\mathbb{R}$ qui sont $(f, \mathbf{x}, \mathbf{y})$ -interprétable est présentée, ainsi que la comparaison de la précision de différents intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables.

Supposons que $\mathbf{x} \in \mathbb{I}\mathbb{R}$ et $\mathbf{y} \in \mathbb{I}\mathbb{R}$

La proposition Proposition IV.5 montre que dans ce cas \mathbf{z} est $(f, \mathbf{x}, \mathbf{y})$ -interprétable si et seulement si $\mathbf{z} \supseteq \square \text{ range}(f, \mathbf{x}, \mathbf{y})$. Considérons deux intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables \mathbf{z} et \mathbf{z}' vérifiant $\mathbf{z}' \subseteq \mathbf{z}$. Donc les deux propositions quantifiées suivantes sont vraies :

$$(\forall x \in \mathbf{x})(\forall y \in \mathbf{y})(\exists z \in \mathbf{z})(f(x, y) = z)$$

$$(\forall x \in \mathbf{x})(\forall y \in \mathbf{y})(\exists z \in \mathbf{z}')(f(x, y) = z)$$

Nous pouvons remarquer que la première proposition donne moins d'information que la seconde, parce que $\mathbf{z}' \subseteq \mathbf{z}$ (et donc la première implique la seconde). L'intervalle $(f, \mathbf{x}, \mathbf{y})$ -interprétable \mathbf{z}' peut donc être considéré comme plus précis que \mathbf{z} . Dans ce cas, la comparaison entre intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables est similaire à la comparaison des extensions aux intervalles classiques.

Supposons que $\mathbf{x} \in \overline{\mathbb{IR}}$ et $\mathbf{y} \in \overline{\mathbb{IR}}$

Dans ce cas, un intervalle $(f, \mathbf{x}, \mathbf{y})$ -interprétable peut être propre ou impropre. Considérons deux intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables $\mathbf{z} \in \mathbb{IR}$ et $\mathbf{z}' \in \overline{\mathbb{IR}}$. Donc les deux propositions quantifiées suivantes sont vraies :

$$\begin{aligned} & (\exists z \in \mathbf{z})(\exists x \in \text{pro } \mathbf{x})(\exists y \in \text{pro } \mathbf{y})(f(x, y) = z) \\ & (\forall z \in \text{pro } \mathbf{z}')(\exists x \in \text{pro } \mathbf{x})(\exists y \in \text{pro } \mathbf{y})(f(x, y) = z) \end{aligned}$$

Ces deux propositions quantifiées s'interprètent de la manière suivante :

$$\mathbf{z} \cap \text{range}(f, \text{pro } \mathbf{x}, \text{pro } \mathbf{y}) \neq \emptyset$$

et

$$\text{pro } \mathbf{z}' \subseteq \text{range}(f, \text{pro } \mathbf{x}, \text{pro } \mathbf{y}).$$

Considérons maintenant un troisième intervalle $(f, \mathbf{x}, \mathbf{y})$ -interprétable $\mathbf{z}'' \in \overline{\mathbb{IR}}$ et supposons que $\mathbf{z}'' \subseteq \mathbf{z}' \subseteq \mathbf{z}$. La proposition quantifiée suivante est donc vraie :

$$(\forall z \in \text{pro } \mathbf{z}'')(\exists x \in \text{pro } \mathbf{x})(\exists y \in \text{pro } \mathbf{y})(f(x, y) = z)$$

D'une part, $\mathbf{z}'' \subseteq \mathbf{z}'$ implique $\text{pro } \mathbf{z}' \subseteq \text{pro } \mathbf{z}''$ et donc la seconde proposition quantifiée donne moins d'information que la troisième (parce que la troisième implique la seconde). D'autre part, $\mathbf{z}' \subseteq \mathbf{z}$ implique $\text{pro } \mathbf{z}' \cap \mathbf{z} \neq \emptyset$ et donc la première proposition quantifiée donne moins d'information que la seconde (parce que la seconde implique la première). L'intervalle \mathbf{z}'' peut donc être considéré comme plus précis que \mathbf{z}' , et \mathbf{z}' comme plus précis que \mathbf{z} . Encore dans ce cas, l'inclusion des intervalles généralisés permet d'ordonner les intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables selon leur précision.

Supposons que $\mathbf{x} \in \mathbb{IR}$ et $\mathbf{y} \in \overline{\mathbb{IR}}$

Dans ce cas, un intervalle $(f, \mathbf{x}, \mathbf{y})$ -interprétable peut être propre ou impropre. Considérons deux intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables $\mathbf{z} \in \mathbb{IR}$ et $\mathbf{z}' \in \overline{\mathbb{IR}}$. Les deux propositions quantifiées suivantes sont donc vraies :

$$\begin{aligned} & (\forall x \in \mathbf{x})(\exists z \in \mathbf{z})(\exists y \in \text{pro } \mathbf{y})(f(x, y) = z) \\ & (\forall x \in \mathbf{x})(\forall z \in \text{pro } \mathbf{z}')(\exists y \in \text{pro } \mathbf{y})(f(x, y) = z) \end{aligned}$$

Si la première n'offre pas d'interprétation intéressante, la seconde indique que l'intervalle \mathbf{x} est un sous-ensemble de la projection de la relation $f(x, y) = z_0$ sur l'axe x , quelque soit la valeur de $z_0 \in \text{pro } \mathbf{z}'$. Comme dans le cas précédant, si $\mathbf{z}' \subseteq \mathbf{z}$ alors $\text{pro } \mathbf{z}' \cap \mathbf{z} \neq \emptyset$ et la première proposition quantifiée donne moins d'information que la seconde. L'inclusion entre intervalles généralisés permet donc encore une fois d'ordonner les intervalles $(f, \mathbf{x}, \mathbf{y})$ -interprétables selon leur précision.

12.3 Minimalité des AE-extensions

Dans le contexte des extensions aux intervalles classiques, si deux extensions \mathbf{g} et \mathbf{h} d'une fonction continue f sont liées par la relation $\mathbf{g}(\mathbf{x}) \subseteq \mathbf{h}(\mathbf{x})$ alors \mathbf{g} est plus précise que \mathbf{h} . C'est justifié car une fois $\mathbf{g}(\mathbf{x})$ évaluée, $\mathbf{h}(\mathbf{x})$ ne donne plus d'information

supplémentaire. Ou de manière équivalente, si \mathbf{g} est une extension de f alors toute fonction d'intervalles \mathbf{h} qui vérifie $\mathbf{g}(\mathbf{x}) \subseteq \mathbf{h}(\mathbf{x})$ est aussi une extension de f . La section précédente a illustré le fait que la comparaison entre AE-extensions peut aussi être faite en utilisant l'inclusion des intervalles généralisés. Cela conduit à la définition suivante :

Définition IV.6. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Considérons deux intervalles (f, \mathbf{x}) -interprétables $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$ et $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$. Alors \mathbf{z} est plus précis que \mathbf{z}' si et seulement si $\mathbf{z} \subseteq \mathbf{z}'$; \mathbf{z} strictement plus précis que \mathbf{z}' si de plus $\mathbf{z} \neq \mathbf{z}'$.

Cette définition est justifiée par la proposition suivante, qui généralise les cas particuliers décrits dans la section précédente :

Proposition IV.8. Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Considérons un intervalle (f, \mathbf{x}) -interprétable $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$ et un intervalle $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$. Si $\mathbf{z} \subseteq \mathbf{z}'$ alors \mathbf{z}' est aussi (f, \mathbf{x}) -interprétable.

Démonstration. Définissons la relation ϕ par $\phi(x, z) \iff z = f(x)$. L'intervalle \mathbf{z} est donc (f, \mathbf{x}) -interprétable si et seulement si $\phi(\text{dual } \mathbf{x}, \mathbf{z})$ et l'intervalle \mathbf{z}' est (f, \mathbf{x}) -interprétable si et seulement si $\phi(\text{dual } \mathbf{x}, \mathbf{z}')$. Maintenant, comme $\mathbf{z} \subseteq \mathbf{z}'$ et parce que $\text{dual } \mathbf{x} \subseteq \text{dual } \mathbf{x}$, nous avons

$$(\text{dual } \mathbf{x}, \mathbf{z})^T \subseteq (\text{dual } \mathbf{x}, \mathbf{z}')^T$$

Donc, la Proposition IV.2 montre que

$$\phi(\text{dual } \mathbf{x}, \mathbf{z}) \implies \phi(\text{dual } \mathbf{x}, \mathbf{z}')$$

□

Une fois que la définition IV.6 est posée, une définition des AE-extensions minimale peut être donnée naturellement de la manière suivante : une AE-extension est minimale s'il n'existe aucune autre AE-extension qui soit strictement plus précise.

Définition IV.7. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Un intervalle (f, \mathbf{x}) -interprétable $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$ est minimal si et seulement si pour tout intervalle (f, \mathbf{x}) -interprétable $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$,

$$\mathbf{z}' \subseteq \mathbf{z} \implies \mathbf{z}' = \mathbf{z}$$

Une AE-extension $\mathbf{g} : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{R}^m$ de f est minimale si et seulement si pour tout $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ l'intervalle (f, \mathbf{x}) -interprétable $\mathbf{g}(\mathbf{x})$ est minimal.

Dans les cas particuliers où \mathbf{x} est propre ou impropre, la minimalité des AE-extensions est liée à la qualité des approximations de l'image de la fonction :

Proposition IV.9. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ et un intervalle $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$.

1. Si $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$ alors \mathbf{z} est (f, \mathbf{x}) -interprétable et minimal si et seulement si $\mathbf{z} = \square \text{range}(f, \mathbf{x})$.
2. Si $\mathbf{x} \in \overline{\mathbb{I}\mathbb{R}^n}$, \mathbf{z} est impropre et $\text{pro } \mathbf{z}$ est une approximation intérieure non-extensible de $\text{range}(f, \text{pro } \mathbf{x})$ alors \mathbf{z} est (f, \mathbf{x}) -interprétable et minimal.
3. Si $\mathbf{x} \in \overline{\mathbb{I}\mathbb{R}^n}$ et $\mathbf{z} \in \mathbb{I}\mathbb{R}^m$ alors \mathbf{z} ne peut pas être (f, \mathbf{x}) -interprétable et minimal.

Démonstration. (1) Comme c'est prouvé dans la section 12.2, si \mathbf{x} est propre alors \mathbf{z} est (f, \mathbf{x}) -interprétable si et seulement si $\text{range}(f, \mathbf{x}) \subseteq \mathbf{z}$. Ainsi, \mathbf{z} est minimal si et seulement si $\mathbf{z} = \square \text{range}(f, \mathbf{x})$.

(2) Considérons un intervalle $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$ quelconque qui soit (f, \mathbf{x}) -interprétable et qui vérifie $\mathbf{z}' \subseteq \mathbf{z}$. Nous avons juste à prouver que $\mathbf{z}' = \mathbf{z}$. A cause de l'inclusion $\mathbf{z}' \subseteq \mathbf{z}$, \mathbf{z}' est aussi impropre. Ainsi, $\text{pro } \mathbf{z}'$ est une approximation intérieure de $\text{range}(f, \text{pro } \mathbf{x})$ et $\text{pro } \mathbf{z} \subseteq \text{pro } \mathbf{z}'$. A cause de la non-extensibilité de l'approximation intérieure $\text{pro } \mathbf{z}$, nous avons $\text{pro } \mathbf{z}' = \text{pro } \mathbf{z}$, et finalement $\mathbf{z}' = \mathbf{z}$.

(3) Si \mathbf{z} est (f, \mathbf{x}) -interprétable alors la proposition quantifiée suivante est vraie :

$$(\exists z \in \mathbf{z})(\exists x \in \mathbf{x})(z = f(x))$$

Notons $\mathbf{z}' = [z, z]$ où z vérifie la proposition quantifiée précédente. L'intervalle \mathbf{z}' est (f, \mathbf{x}) -interprétable et vérifie $\mathbf{z}' \subseteq \mathbf{z}$ et $\mathbf{z}' \neq \mathbf{z}$. L'intervalle \mathbf{z} n'est donc pas minimal. \square

Remarque. *Nous n'avons pas pu décider si la réciproque du deuxième cas de la proposition précédente est vraie ou fausse. Seul le troisième cas (plus faible que la réciproque du second) a pu être établie.*

Cette dernière proposition permet de construire une AE-extension minimale pour les fonctions continue $f : \mathbb{R} \rightarrow \mathbb{R}$. Il sera prouvé dans la section 13 que ces fonctions ont une unique AE-extension minimale, qui est donc celle qui est calculée dans l'exemple suivant.

Exemple IV.13. *Considérons la fonction $\exp(x)$ et un intervalle $\mathbf{x} \in \mathbb{K}\mathbb{R}$. L'intervalle \mathbf{x} est soit propre soit impropre et la Proposition IV.9 peut donc être appliquée quelque soit l'argument \mathbf{x} . Définissons la fonction d'intervalles $\exp^* : \mathbb{K}\mathbb{R} \rightarrow \mathbb{K}\mathbb{R}$ de la manière suivante :*

- si $\mathbf{x} \in \mathbb{I}\mathbb{R}$ alors $\exp^*(\mathbf{x}) = \text{range}(\exp, \mathbf{x}) = [\exp(\underline{\mathbf{x}}), \exp(\overline{\mathbf{x}})]$;
- si $\mathbf{x} \in \mathbb{I}\mathbb{R}$, alors $\exp^*(\mathbf{x}) = \text{dual range}(\exp, \text{pro } \mathbf{x}) = [\exp(\underline{\mathbf{x}}), \exp(\overline{\mathbf{x}})]$.

Une autre expression de cette fonction d'intervalles est

$$\exp^*(\mathbf{x}) = [\exp(\underline{\mathbf{x}}), \exp(\overline{\mathbf{x}})].$$

La Proposition IV.9 montre alors que \exp^ est une AE-extension minimale de \exp . De la même manière, nous pouvons construire une AE-extension minimale de $\ln x$: pour tout $\mathbf{x} \in \mathbb{K}\mathbb{R}$ tels que $\text{pro } \mathbf{x} > 0$, une expression de l'AE-extension minimale de \ln est $\ln^* \mathbf{x} = [\ln \underline{\mathbf{x}}, \ln \overline{\mathbf{x}}]$.*

Remarque. *Nous pouvons remarquer que les AE-extensions minimales \exp^* et \ln^* coïncident avec les extensions \exp^K et \ln^K définies par Kaucher dans Kaucher[21] (cf. section 3.4 page 40).*

Une question importante se pose alors : étant donné une AE-extension \mathbf{g} de f , existe-t-il toujours une AE-extension minimale qui soit plus précise que \mathbf{g} ? La proposition suivante donne une réponse positive à cette question. Avant tout, le résultat technique suivant sera utile.

Lemme IV.3. *Soit $\mathbb{E} \subseteq \mathbb{K}\mathbb{R}^n$ un ensemble non-vide, fermé et borné. Alors, \mathbb{E} contient un élément minimal au sens de l'inclusion, i.e. un élément qui n'est strictement inclus dans aucun élément de \mathbb{E} .*

Démonstration. Considérons la bijection continue $\sigma : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{R}^{2n}$ compatible avec les ordres introduit dans Shary[92][94] défini par

$$\sigma(\mathbf{x}) = (-\underline{\mathbf{x}}_1, \dots, \underline{\mathbf{x}}_n, \overline{\mathbf{x}}_1, \dots, \overline{\mathbf{x}}_n)$$

Comme σ est compatible avec les ordres, si x^* est un élément minimal de $\sigma(\mathbb{E})$ alors $\sigma^{-1}(x^*)$ est un élément minimal de \mathbb{E} . Finalement comme σ est continu, $\sigma(\mathbb{E})$ est non-vide, fermé et borné. Donc $\sigma(\mathbb{E})$ a au moins un élément minimal x^* . \square

Proposition IV.10. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et \mathbf{g} une AE-extension de f . Il existe alors une AE-extension minimale de f qui est plus précise que \mathbf{g} .*

Démonstration. Nous avons juste à prouver que pour tout $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, il existe un intervalle (f, \mathbf{x}) -interprétable minimal qui est plus précis que $\mathbf{g}(\mathbf{x})$. Considérons un intervalle $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ quelconque et définissons $\mathbb{E} = \{\mathbf{z} | \mathbf{z} \subseteq \mathbf{g}(\mathbf{x})\}$ et $\mathbb{F} = \{\mathbf{z} \in \mathbb{K}\mathbb{R}^m | \phi(\text{dual } \mathbf{x}, \mathbf{z})\}$ où $\phi(x, z) \iff z = f(x)$. Ainsi, les intervalles de $\mathbb{E} \cap \mathbb{F}$ sont à la fois (f, \mathbf{x}) -interprétable et plus précis que $\mathbf{g}(\mathbf{x})$. D'une part, \mathbb{E} est fermé (car il est l'intersection de deux fermés). D'autre part, le graphe de ϕ est fermé dans \mathbb{R}^n parce que f est continue. Ainsi \mathbb{F} est fermé comme le montre la Proposition IV.4. Donc, $\mathbb{E} \cap \mathbb{F}$ fermé. De plus, $\mathbb{E} \cap \mathbb{F}$ est borné car $\mathbf{z} \in \mathbb{E} \cap \mathbb{F}$ implique $\text{dual}(\square \text{ range}(f, \text{pro } \mathbf{x})) \subseteq \mathbf{z} \subseteq \mathbf{g}(\mathbf{x})$ (la première inclusion étant une conséquence de la Proposition IV.7). Finalement, $\mathbb{E} \cap \mathbb{F}$ non-vide car $\mathbf{g}(\mathbf{x}) \in \mathbb{E} \cap \mathbb{F}$. Nous pouvons donc appliquer le Lemme IV.3 qui prouve que $\mathbb{E} \cap \mathbb{F}$ a un élément minimal, qui est un intervalle (f, \mathbf{x}) -interprétable minimal par définition de \mathbb{E} et \mathbb{F} . \square

La proposition précédente prouve en particulier que n'importe quel boîte incluse dans $\text{range}(f, \mathbf{x})$ peut être étendue jusqu'à obtenir une approximation intérieure non-extensible. Il peut arriver qu'une fonction est de nombreuses AE-extensions minimales, en particulier lorsque les AE-extensions sont utilisées pour construire des approximations intérieures de fonctions à valeurs vectorielles. La proposition suivante donne une condition nécessaire et suffisante pour qu'une fonction continue ait une unique AE-extension minimale. Cette proposition est capitale pour les développements présentés dans le chapitre suivant.

Proposition IV.11. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. L'intervalle $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$ est l'unique intervalle (f, \mathbf{x}) -interprétable minimal si et seulement si pour tout $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$,*

$$\mathbf{z}' \text{ is } (f, \mathbf{x})\text{-interprétable} \iff \mathbf{z} \subseteq \mathbf{z}'$$

Démonstration. (\implies) Supposons que \mathbf{z} est l'unique intervalle (f, \mathbf{x}) -interprétable minimal. D'une part, par la Proposition IV.2, nous avons $\mathbf{z} \subseteq \mathbf{z}'$ implique \mathbf{z}' est (f, \mathbf{x}) -interprétable. D'autre part, supposons que \mathbf{z}' est (f, \mathbf{x}) -interprétable. Alors, par la Proposition IV.10, il existe un intervalle (f, \mathbf{x}) -interprétable minimal $\mathbf{z}'' \in \mathbb{K}\mathbb{R}^m$ qui vérifie $\mathbf{z}'' \subseteq \mathbf{z}'$. Finalement, $\mathbf{z}'' = \mathbf{z}$ parce que \mathbf{z} est l'unique intervalle (f, \mathbf{x}) -interprétable minimal, ce qui prouve que $\mathbf{z} \subseteq \mathbf{z}'$.

(\impliedby) Supposons que \mathbf{z}' est (f, \mathbf{x}) -interprétable si et seulement si $\mathbf{z} \subseteq \mathbf{z}'$. En particulier, $\mathbf{z} \subseteq \mathbf{z}$ implique que \mathbf{z} soit (f, \mathbf{x}) -interprétable. Considérons un intervalle $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$ (f, \mathbf{x}) -interprétable et qui vérifie $\mathbf{z}' \subseteq \mathbf{z}$. Par hypothèse $\mathbf{z} \subseteq \mathbf{z}'$ parce que \mathbf{z}' est (f, \mathbf{x}) -interprétable. Nous avons donc $\mathbf{z}' = \mathbf{z}$ ce qui montre que \mathbf{z} est minimal. Considérons finalement un intervalle $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$ (f, \mathbf{x}) -interprétable et minimal. Comme \mathbf{z}' est (f, \mathbf{x}) -interprétable, nous avons par hypothèse $\mathbf{z} \subseteq \mathbf{z}'$. Alors, \mathbf{z}' étant supposé minimal, nous avons $\mathbf{z} = \mathbf{z}'$. \mathbf{z} est donc l'unique intervalle (f, \mathbf{x}) -interprétable minimal. \square

En particulier, la proposition précédente sera utilisée dans le prochain chapitre pour montrer que toute fonction continue $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a une unique AE-extension minimale. La prochaine proposition présente une propriété intéressante qui est vraie si une fonction a une unique AE-extension : comme dans la contexte des extensions aux intervalles classiques (où l'unicité de l'extension minimale est toujours vérifiée) nous pouvons construire un intervalle (f, \mathbf{x}) -interprétable en intersectant deux intervalles (f, \mathbf{x}) -interprétables.

Proposition IV.12. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Supposons qu'il existe un unique intervalle (f, \mathbf{x}) -interprétable minimal \mathbf{z}^* et considérons deux intervalles (f, \mathbf{x}) -interprétables $\mathbf{z}, \mathbf{z}' \in \mathbb{K}\mathbb{R}^m$. Alors l'intervalle $\mathbf{z} \wedge \mathbf{z}'$ est aussi (f, \mathbf{x}) -interprétable.*

Démonstration. Par la Proposition IV.11 nous avons $\mathbf{z}^* \subseteq \mathbf{z}$ et $\mathbf{z}^* \subseteq \mathbf{z}'$. Donc $\mathbf{z}^* \subseteq \mathbf{z} \wedge \mathbf{z}'$, ce qui implique finalement que $\mathbf{z} \wedge \mathbf{z}'$ soit (f, \mathbf{x}) -interprétable. \square

La proposition précédente ne peut pas être généralisée à toutes les AE-extensions. En effet, un cas typique où il n'existe pas d'unique intervalle (f, \mathbf{x}) -interprétable minimal est l'utilisation des AE-extensions pour construire des approximations intérieures de l'image de fonctions à valeurs vectorielles. Dans ce cas, l'intersection des intervalles (f, \mathbf{x}) -interprétables revient à faire l'union de leur projections propres (car ces intervalles sont impropres). Et la réunion d'intervalles de deux approximations intérieures d'un sous-ensemble de \mathbb{R}^n n'est pas en générale une approximation intérieure lorsque $n > 1$.

Exemple IV.14. *Soient $f(x) = (x_1 + x_2, x_1 - x_2)^T$ et $\mathbf{x} = ([1, -1], [1, -1])^T$. Considérons les deux intervalles (f, \mathbf{x}) -interprétables $\mathbf{z} = ([1, 1], [0, 0])^T$ et $\mathbf{z}' = ([0, 0], [1, 1])^T$ ($\text{pro } \mathbf{z}$ et $\text{pro } \mathbf{z}'$ sont deux approximations intérieures de $\text{range}(f, \text{pro } \mathbf{x})$). L'intervalle $\mathbf{z}'' = \mathbf{z} \wedge \mathbf{z}' = ([1, 0], [1, 0])^T$ n'est pas (f, \mathbf{x}) -interprétable car $\text{pro } \mathbf{z}''$ n'est pas une approximation intérieure de $\text{range}(f, \text{pro } \mathbf{x})$.*

Finalement, la proposition suivante est un résultat technique qui sera utile dans la section 12.4. Avant tout, le lemme suivant doit être établi :

Lemme IV.4. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ et \mathbf{z} un intervalle (f, \mathbf{x}) -interprétable minimal. Alors, $\mathbf{z} \subseteq \mathbf{r}(\mathbf{x})$ où $\mathbf{r}(\mathbf{x}) = \square \text{range}(f, \text{pro } \mathbf{x})$.*

Démonstration. Notons la relation $z = f(x)$ par $\phi(x, z)$. Donc $\phi(\text{dual } \mathbf{x}, \mathbf{z})$ est vrai parce que \mathbf{z} est (f, \mathbf{x}) -interprétable. Définissons

$$\mathbf{y} = \square \{(x, z) \in (\text{pro } \mathbf{x}, \text{pro } \mathbf{z}) \mid \phi(x, z)\}$$

Nous avons, $\mathbf{y} \subseteq (\text{pro } \mathbf{x}, \mathbf{r}(\mathbf{x}))$ car $x \in \mathbf{x}$ implique $f(x) \in \mathbf{r}(\mathbf{x})$. Maintenant, par la Proposition IV.3, nous avons

$$\phi(\text{dual } \mathbf{x}, \mathbf{z}) \implies \phi(\mathbf{y} \wedge (\text{dual } \mathbf{x}, \mathbf{z}))$$

et par la Proposition IV.2,

$$\phi(\mathbf{y} \wedge (\text{dual } \mathbf{x}, \mathbf{z})) \implies \phi((\text{pro } \mathbf{x}, \mathbf{r}(\mathbf{x})) \wedge (\text{dual } \mathbf{x}, \mathbf{z})).$$

Maintenant, parce que $\text{dual } \mathbf{x} \subseteq \text{pro } \mathbf{x}$, nous avons $(\text{dual } \mathbf{x}) \cap (\text{pro } \mathbf{x}) = \text{dual } \mathbf{x}$ et donc

$$\phi(\text{dual } \mathbf{x}, \mathbf{z} \wedge \mathbf{r}(\mathbf{x})),$$

i.e. $\mathbf{z} \wedge \mathbf{r}(\mathbf{x})$ est (f, \mathbf{x}) -interprétable. De plus, $\mathbf{z} \wedge \mathbf{r}(\mathbf{x}) \subseteq \mathbf{z}$ et donc par la minimalité de \mathbf{z} , nous avons $\mathbf{z} \wedge \mathbf{r}(\mathbf{x}) = \mathbf{z}$. Cela implique finalement $\mathbf{z} \subseteq \mathbf{r}(\mathbf{x})$. \square

Proposition IV.13. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction localement Lipschitz continue et \mathbf{f} une AE-extension minimale de f . Alors pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n$, il existe $\gamma \in \mathbb{R}$, $\gamma > 0$, tel que pour tout $\mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}$,*

$$\|\text{wid } \mathbf{f}(\mathbf{x})\| \leq \gamma \|\text{wid } \mathbf{x}\|$$

Démonstration. Par la Proposition IV.7 et le Lemme IV.4 nous avons

$$\text{imp } \mathbf{r}(\mathbf{x}) \subseteq \mathbf{f}(\mathbf{x}) \subseteq \mathbf{r}(\mathbf{x})$$

où $\mathbf{r}(\mathbf{x}) = \square \text{range}(f, \text{pro } \mathbf{x})$. Ces deux inclusions impliquent $\|\text{rad } \mathbf{f}(\mathbf{x})\| \leq \|\text{rad } \mathbf{r}(\mathbf{x})\|$. Maintenant, il est prouvé dans Kaucher[21] que si f est localement Lipschitz continue alors $\mathbf{r}(\mathbf{x})$ l'est aussi (c'est un cas particulier du théorème 2.6 de Kaucher[21]). Donc, pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n$, il existe $\gamma \in \mathbb{R}$, $\gamma > 0$ tel que pour tout $\mathbf{x}, \mathbf{y} \in \mathbb{K}\mathbf{x}^{\text{ref}}$,

$$\text{dist}(\mathbf{r}(\mathbf{x}), \mathbf{r}(\mathbf{y})) \leq \gamma \text{dist}(\mathbf{x}, \mathbf{y})$$

Choisissons $\mathbf{y} = \text{mid } \mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}$ et remarquons que $\text{dist}(\mathbf{x}, \text{mid } \mathbf{x}) = \|\text{rad } \mathbf{x}\|$ pour obtenir $\text{dist}(\mathbf{r}(\mathbf{x}), \mathbf{f}(\text{mid } \mathbf{x})) \leq \gamma \|\text{rad } \mathbf{x}\|$. Finalement, remarquons que pour tout $z \in \mathbb{R}^n$, $\|\text{rad } \mathbf{r}(\mathbf{x})\| \leq \text{dist}(\mathbf{r}(\mathbf{x}), z)$, ce qui est donc vrai en particulier pour $z = \mathbf{f}(\text{mid } \mathbf{x})$. Nous avons donc prouvé que

$$\|\text{rad } \mathbf{f}(\mathbf{x})\| \leq \gamma \|\text{rad } \mathbf{x}\|,$$

ce qui est équivalent à l'énoncé proposé. \square

12.4 Ordre de convergence d'une AE-extension

Dans le contexte des extensions aux intervalles classiques, la qualité d'une extension est mesurée par son ordre de convergence. Cette notion est maintenant généralisée aux AE-extensions.

Définition IV.8. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{R}^m$ une AE-extension de f . L'AE-extension \mathbf{g} a un ordre de convergence $\alpha \in \mathbb{R}$, $\alpha > 0$, si et seulement s'il existe une AE-extension minimale \mathbf{f} de f plus précise que \mathbf{g} telle que pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n$, il existe $\gamma > 0$ tel que pour tout $\mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}$,*

$$\|\text{wid } \mathbf{g}(\mathbf{x}) - \text{wid } \mathbf{f}(\mathbf{x})\| \leq \gamma (\|\text{wid } \mathbf{x}\|)^\alpha$$

Remarque. *Il est clair qu'une AE-extension ayant un ordre de convergence α a aussi un ordre de convergence α' pour tout $0 < \alpha' \leq \alpha$. Les ordres de convergence habituellement considérés sont des entiers. Un ordre de convergence 1 est appelé un ordre de convergence linéaire, et un ordre de convergence 2 un ordre de convergence quadratique.*

Remarque. *Pour toute AE-extension \mathbf{g} de f , il existe au moins une AE-extension minimale de f qui soit plus précise que \mathbf{g} (voir Proposition IV.10). Donc l'énoncé de la définition précédente ne peut pas être faux à cause de la non-existence d'une telle AE-extension. S'il est montré que la fonction f a une unique AE-extension minimale, la définition coïncide avec la définition de l'ordre de convergence des extensions aux intervalles classiques.*

Dans le contexte des extensions aux intervalles classiques, l'ordre de convergence est lié à la distance à l'extension minimale. Pour étendre cette propriété aux AE-extensions, le lemme suivant est nécessaire :

Lemme IV.5. *Soit $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$ et $\mathbf{z}' \in \mathbb{K}\mathbb{R}^m$ vérifiant $\mathbf{z}' \subseteq \mathbf{z}$. Alors,*

$$\frac{1}{2} \|\text{wid } \mathbf{z} - \text{wid } \mathbf{z}'\| \leq \text{dist}(\mathbf{z}, \mathbf{z}') \leq \|\text{wid } \mathbf{z} - \text{wid } \mathbf{z}'\|$$

Démonstration. L'inclusion $\mathbf{z}' \subseteq \mathbf{z}$ implique $0 \subseteq \mathbf{z} - \text{dual } \mathbf{z}'$ et donc $\mathbf{u} := \mathbf{z} - \text{dual } \mathbf{z}'$ est propre et contient 0. Remarquons alors que $\|\text{wid } \mathbf{z} - \text{wid } \mathbf{z}'\| = \|\text{wid}(\mathbf{z} - \text{dual } \mathbf{z}')\| = \|\text{wid } \mathbf{u}\|$ et $\text{dist}(\mathbf{z}, \mathbf{z}') = \|\mathbf{z} - \text{dual } \mathbf{z}'\| = \|\mathbf{u}\|$. Finalement, l'inégalité suivante est vraie pour tout intervalle propre qui contient 0 :

$$\|\text{rad } \mathbf{u}\| \leq \|\mathbf{u}\| \leq \|\text{wid } \mathbf{u}\|$$

Cela correspond à l'énoncé du lemme. \square

Proposition IV.14. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{K}\mathbb{R}^n \rightarrow \mathbb{K}\mathbb{R}^m$ une AE-extension de f . L'AE-extension \mathbf{g} a un ordre de convergence $\alpha \in \mathbb{R}$, $\alpha > 0$, si et seulement s'il existe une AE-extension minimale \mathbf{f} de f plus précise que \mathbf{g} telle que pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n$, il existe $\gamma > 0$ tel que pour tout $\mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}$,*

$$\text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{f}(\mathbf{x})) \leq \gamma(\|\text{wid } \mathbf{x}\|)^\alpha$$

Démonstration. Supposons que \mathbf{g} ait un ordre de convergence α , c'est à dire il existe une AE-extension minimale \mathbf{f} plus précise que \mathbf{g} telle que

$$(\forall \mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n) (\exists \gamma > 0) (\forall \mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}) (\|\text{wid } \mathbf{g}(\mathbf{x}) - \text{wid } \mathbf{f}(\mathbf{x})\| \leq \gamma(\|\text{wid } \mathbf{x}\|)^\alpha)$$

Alors, par le Lemme IV.5, $\text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{f}(\mathbf{x})) \leq \|\text{wid } \mathbf{g}(\mathbf{x}) - \text{wid } \mathbf{f}(\mathbf{x})\|$ ce qui prouve que la condition proposée dans cette proposition est vraie. Supposons maintenant que cette condition soit vraie, c'est à dire il existe une AE-extension minimale \mathbf{f} plus précise que \mathbf{g} telle que

$$(\forall \mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n) (\exists \gamma > 0) (\forall \mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}) (\text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{f}(\mathbf{x})) \leq \gamma(\|\text{wid } \mathbf{x}\|)^\alpha)$$

Alors, par le Lemme IV.5, $\frac{1}{2} \|\text{wid } \mathbf{g}(\mathbf{x}) - \text{wid } \mathbf{f}(\mathbf{x})\| \leq \text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{f}(\mathbf{x}))$ ce qui prouve que \mathbf{g} a un ordre de convergence α (en utilisant la définition IV.8 avec $\gamma' = 2\gamma$). \square

Finalement, la proposition suivante prouve que toute AE-extension qui est localement Lipschitz continue a un ordre de convergence linéaire.

Proposition IV.15. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{K}\mathbb{R}^n \rightarrow \mathbb{K}\mathbb{R}^m$ une AE-extension de f . Si \mathbf{g} est localement Lipschitz continue alors \mathbf{g} a un ordre de convergence linéaire.*

Démonstration. Avant tout, il est facile de vérifier que si \mathbf{g} est localement Lipschitz continue alors f l'est aussi (parce que $f(x) = \mathbf{g}(x)$ pour tout $x \in \mathbb{R}^n$). Considérons alors une AE-extension minimale \mathbf{f} plus précise que \mathbf{g} (qui existe grâce à la Proposition IV.10). Nous utilisons la continuité locale de Lipschitz de \mathbf{g} et la Proposition IV.13 : pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{I}\mathbb{R}^n$, il existe $\gamma > 0$ et $\gamma' > 0$ tels que pour tout $\mathbf{x} \in \mathbb{K}\mathbf{x}^{\text{ref}}$ et tout $\mathbf{y} \in \mathbb{K}\mathbf{x}^{\text{ref}}$,

$$\|\text{wid } \mathbf{f}(\mathbf{x})\| \leq \gamma \|\text{wid } \mathbf{x}\| \quad \text{et} \quad \text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{g}(\mathbf{y})) \leq \gamma' \text{dist}(\mathbf{x}, \mathbf{y})$$

En choisissant $\mathbf{y} = [\text{mid } \mathbf{x}, \text{mid } \mathbf{x}] \subseteq \mathbf{x}^{\text{ref}}$, nous obtenons

$$\text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{g}(\text{mid } \mathbf{x})) \leq \gamma' \|\text{rad } \mathbf{x}\|$$

Comme $\mathbf{g}(\text{mid } \mathbf{x})$ est dégénéré, nous avons $\|\text{rad } \mathbf{g}(\mathbf{x})\| \leq \text{dist}(\mathbf{g}(\mathbf{x}), \mathbf{g}(\text{mid } \mathbf{x}))$ et donc $\|\text{rad } \mathbf{g}(\mathbf{x})\| \leq \gamma' \|\text{rad } \mathbf{x}\|$. Définissons $\gamma'' = \max\{\gamma, \gamma'\}$. Nous avons donc prouvé

$$\|\text{wid } \mathbf{f}(\mathbf{x})\| \leq \gamma'' \|\text{wid } \mathbf{x}\| \quad \text{and} \quad \|\text{wid } \mathbf{g}(\mathbf{x})\| \leq \gamma'' \|\text{wid } \mathbf{x}\|$$

Nous avons donc $\|\text{wid } \mathbf{g}(\mathbf{x}) - \text{wid } \mathbf{f}(\mathbf{x})\| \leq 2\gamma'' \|\text{wid } \mathbf{x}\|$ (le pire cas se produisant lorsque $\text{wid } \mathbf{g}(\mathbf{x}) = -\text{wid } \mathbf{f}(\mathbf{x})$). \square

12.5 Arrondi d'une AE-extension

Quand des ordinateurs sont utilisés, la précision finie des nombres et des opérations doit être prise en compte. Dans le contexte des extensions aux intervalles classiques, l'arrondi extérieur est compatible avec la sémantique des extensions. Dans le contexte plus général des AE-extensions, il est maintenant montré que l'arrondi extérieur peut être utilisé de la même manière. L'arrondi est effectué en deux étapes :

1. arrondi extérieur des arguments ;
2. arrondi extérieur de l'AE-extension.

L'arrondi extérieur d'un intervalle $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ est noté $\diamond \mathbf{x}$ et satisfait par définition $\mathbf{x} \subseteq \diamond \mathbf{x}$. D'une part, l'arrondi extérieur d'un intervalle propre agrandit l'ensemble de réels sous-jacent : par exemple $\diamond[-1, 1]$ pourrait être égale à $[-1.1, 1.1]$. D'autre part, l'arrondi extérieur d'un intervalle impropre rétrécit l'ensemble de réels sous-jacent : par exemple $\diamond[1, -1]$ pourrait être égal à $[0.9, -0.9]$. Il peut aussi arriver que l'arrondi extérieur change la qualité propre/impropre : par exemple $\diamond[0.05, -0.05]$ pourrait être égale à $[-0.05, 0.05]$. Nous allons montrer que ces différentes possibilités sont compatibles avec la modélisation des quantificateurs utilisée dans l'interprétation des AE-extensions. L'arrondi extérieur de l'AE-extension \mathbf{g} est notée \mathbf{g}^\diamond et satisfait par définition $\mathbf{g}(\mathbf{x}) \subseteq \mathbf{g}^\diamond(\mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. La prochaine proposition prouve que l'arrondi extérieur est compatible avec la sémantique des AE-extensions. Rappelons que $\mathbb{F} \subseteq \mathbb{R}$ est un sous ensemble fini de nombres réels représentables et $\mathbb{K}\mathbb{F} \subseteq \mathbb{K}\mathbb{R}$ est l'ensemble fini des intervalles généralisés dont les bornes sont des éléments de \mathbb{F} . Comme indiqué en introduction, le dépassement de capacité n'est pas explicitement pris en compte. Il suffirait d'introduire un nombre flottant ∞ qui serait absorbant pour toutes les opérations (voir Neumaier[8]).

Proposition IV.16. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{g} : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{R}^m$ une AE-extension de f . Considérons de plus une fonction $\mathbf{g}^\diamond : \mathbb{K}\mathbb{F}^n \longrightarrow \mathbb{K}\mathbb{F}^m$ qui vérifie $\mathbf{g}(\mathbf{x}) \subseteq \mathbf{g}^\diamond(\mathbf{x})$ quelque soit $\mathbf{x} \in \mathbb{K}\mathbb{F}^n$. Alors, la fonction d'intervalles $\mathbf{h} : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{F}^m$ définie par $\mathbf{h}(\mathbf{x}) = \mathbf{g}^\diamond(\diamond \mathbf{x})$ est une AE-extension faible de f .*

Démonstration. Notons la relation $z = f(x)$ par $\phi(x, z)$. Alors, \mathbf{g} étant une AE-extension de f , pour tout $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ l'intervalle $\mathbf{g}(\diamond \mathbf{x})$ est $(f, \diamond \mathbf{x})$ -interprétable, c'est à dire

$$\phi(\text{dual } \diamond \mathbf{x}, \mathbf{g}(\diamond \mathbf{x}))$$

est vraie. Comme $\text{dual } \diamond \mathbf{x} \subseteq \text{dual } \mathbf{x}$ et $\mathbf{g}(\diamond \mathbf{x}) \subseteq \mathbf{g}^\diamond(\diamond \mathbf{x})$ la Proposition IV.2 montre que

$$\phi(\text{dual } \mathbf{x}, \mathbf{g}^\diamond(\diamond \mathbf{x}))$$

est vraie (par exemple, $\forall x \in \diamond \mathbf{x}, \exists z \in \mathbf{g}(\diamond \mathbf{x}), z = f(x)$ implique $\forall x \in \mathbf{x}, \exists z \in \diamond \mathbf{g}(\diamond \mathbf{x}), z = f(x)$). Donc, $\mathbf{h}(\mathbf{x})$ est (f, \mathbf{x}) -interprétable. \square

L'exemple suivant illustre la procédure d'arrondi proposée par la proposition précédente.

Exemple IV.15. *Supposons que nous voulions calculer des approximations intérieures et extérieures de*

$$\text{range}(\ln, \mathbf{x}) = [1, 2]$$

avec $\mathbf{x} = [e, e^2]$, en utilisant une précision de 0.01. Considérons d'abord l'approximation extérieure : une approximation extérieure de de l'AE-extension $\ln^*(\mathbf{x})$ doit être calculée. Premièrement, \mathbf{x} est arrondi : $\diamond[e, e^2] = [2.71, 7.39]$. Ensuite, l'AE-extension arrondie donne

$$\diamond \ln^*(\diamond \mathbf{x}) = \diamond[0.99694863\dots, 2.0001277\dots] = [0.99, 2.01]$$

qui est bien une approximation extérieur de $[1, 2]$. Considérons maintenant l'approximation intérieure : une approximation extérieure de l'AE-extension $\ln^*(\text{dual } \mathbf{x})$ doit être calculée. Premièrement, $\text{dual } \mathbf{x} = [e^2, e]$ est arrondi en $\diamond[e^2, e] = [7.38, 2.72]$. Remarquons que l'arrondi extérieur de cet intervalle impropre rétrécit l'ensemble des réels sous-jacent. Ensuite, l'arrondi extérieur de l'AE-extension donne

$$\diamond \ln^*(\diamond(\text{dual } \mathbf{x})) = \diamond[1.9987736\dots, 1.0006319\dots] = [1.99, 1.01]$$

Encore une fois, l'arrondi extérieur de cet intervalle impropre rétrécit l'ensemble des réels sous-jacent, et la projection propre du résultat final est bien une approximation intérieure de $[1, 2]$.

Finalement, les AE-extensions qui seront construites dans la suite seront composées de plusieurs fonctions d'intervalles. Pour obtenir une procédure d'arrondi efficace, il est nécessaire qu'il suffise d'arrondir chaque fonction mise en jeu pour arrondir l'AE-extension composée de ces fonctions. Supposons que l'AE-extension \mathbf{g} soit composée des fonctions d'intervalles \mathbf{g}_k . Supposons de plus que ces fonctions d'intervalles soit croissantes par rapport à l'inclusion. Il est alors facile de vérifier que la fonction d'intervalles obtenue en utilisant \mathbf{g}_k^\diamond au lieu des \mathbf{g}_k est un arrondi extérieur de \mathbf{g} . Maintenant, toutes les fonctions d'intervalles utilisées dans la suite ne seront pas croissantes par rapport à l'inclusion (par exemple pro et mid ne sont pas croissantes par rapport à l'inclusion). Malgré tout, un arrondi extérieur de \mathbf{g} est aussi obtenu si les fonctions d'intervalles non croissantes par rapport à l'inclusion sont mises en jeu au niveau des feuilles de l'arbre syntaxique. Par exemple, considérons $\mathbf{g}(\mathbf{x}) = (\text{pro } \mathbf{x}) + (\text{pro } \mathbf{x})$. Bien que la fonction pro ne soit pas croissante par rapport à l'inclusion, la fonction d'intervalles $\mathbf{h}(\mathbf{x}) = (\text{pro}^\diamond \mathbf{x}) +^\diamond (\text{pro}^\diamond \mathbf{x})$ est une approximation extérieure de \mathbf{g} . En effet, $(\text{pro}^\diamond \mathbf{x}) \supseteq (\text{pro } \mathbf{x})$ parce que $\text{pro}^\diamond \mathbf{x}$ est une approximation extérieure de pro . Comme $+$ est croissante par rapport à l'inclusion,

$$(\text{pro } \mathbf{x}) + (\text{pro } \mathbf{x}) \subseteq (\text{pro}^\diamond \mathbf{x}) + (\text{pro}^\diamond \mathbf{x})$$

et par ce que $+^\diamond$ est une approximation extérieure de $+$,

$$(\text{pro}^\diamond \mathbf{x}) + (\text{pro}^\diamond \mathbf{x}) \subseteq (\text{pro}^\diamond \mathbf{x}) +^\diamond (\text{pro}^\diamond \mathbf{x})$$

Toutes les AE-extensions qui seront construites par la suite vérifieront l'hypothèse que les fonctions d'intervalles dont elles sont composées sont croissantes par rapport à l'inclusion, sauf peut-être les fonctions terminales dans l'arbre syntaxique. Elle pourrons donc être arrondies en arrondissant les fonctions élémentaires dont elles sont composées.

12.6 Conclusion

Ce chapitre a posé les bases théoriques des AE-extensions des fonctions réelles. Après avoir donné la définition des AE-extensions, plusieurs définitions relatives aux AE-extensions ont été proposées : la minimalité d'une AE-extension et l'ordre de convergence d'une AE-extension. Il est à noter que ces définitions généralisent les définitions correspondantes dans la théorie des intervalles classiques. Tout en restant dans un cadre générale, des propriétés utiles ont été données (existence d'une AE-extension minimale plus précise qu'une AE-extension donnée, une condition suffisante pour qu'une AE-extension aie un ordre de convergence linéaire). Finalement, la possibilité d'utiliser l'arrondi extérieure de façon efficace pour calculer des AE-extensions en utilisant une précision finie a été montré.

Les chapitres suivants vont maintenant présenter des AE-extensions qui pourront être calculées en pratique. La même stratégie que celle utilisée dans la théorie des intervalles classique est mise en oeuvre : tout d'abord, le cas particulier des fonctions $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est étudié conduisant à une expression de l'unique AE-extension optimale de ces fonctions (Chapitre 13). Grâce à cette expression, une arithmétique des intervalles généralisés est définie. Il est montré qu'elle coïncide avec l'arithmétique de Kaucher (Chapitre 14). Ensuite, l'AE-extension naturelle (Chapitre 15) et l'AE-extension de la valeur moyenne (Chapitre 16) sont définies.

Chapitre 13

AE-extensions minimales des fonctions $f : \mathbb{R}^n \longrightarrow \mathbb{R}$

Le cas particulier et important des fonctions à valeurs réelles, i.e. $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, est maintenant étudié. Dans ce cas, nous sommes en position de prouver que la fonction f possède une unique AE-extension optimale. Une expression de cette AE-extension sera donnée au moyen des opérations d'union et d'intersection d'intervalles généralisés. Cette expression aura une grande importance car elle sera utilisée pour calculer les AE-extensions des fonctions élémentaires celle-ci étant la base de toutes les AE-extensions qui seront construites par la suite.

13.1 L'AE-extension minimale f^*

Étant donnée une fonction continue f , nous définissons maintenant la fonction d'intervalles f^* . Cette fonction va jouer un rôle central dans les développements à venir.

Définition IV.9. Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue. La fonction d'intervalles $f^* : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{R}$ est définie de la manière suivante :

$$\begin{aligned} f^*(\mathbf{x}) &= \bigvee_{x_P \in \mathbf{x}_P} \bigwedge_{x_I \in (\text{pro } \mathbf{x}_I)} f(x) \\ &= \left[\min_{x_P \in \mathbf{x}_P} \max_{x_I \in (\text{pro } \mathbf{x}_I)} f(x), \max_{x_P \in \mathbf{x}_P} \min_{x_I \in (\text{pro } \mathbf{x}_I)} f(x) \right] \end{aligned}$$

où $P = P(\mathbf{x})$ et $I = I(\mathbf{x})$.

Remarques.

1. La notation f^* a été choisie pour faire le lien avec la fonction f^* qui est définie dans la théorie des intervalles modaux. Ces deux fonctions sont égales à l'isomorphisme près entre les intervalles modaux et les intervalles généralisés.
2. Quand $I = \emptyset$ l'expression de f^* est

$$f^*(\mathbf{x}) = \left[\min_{x \in \mathbf{x}} f(x), \max_{x \in \mathbf{x}} f(x) \right]$$

Quand $P = \emptyset$ l'expression de f^* est

$$f^*(\mathbf{x}) = \left[\max_{x \in (\text{pro } \mathbf{x})} f(x), \min_{x \in (\text{pro } \mathbf{x})} f(x) \right]$$

Cependant, l'expression présentée dans la définition IV.9 sera utilisée dans tous les cas.

3. L'expression de f^* diffère de celle de f^K (cf. section 3.4 page 40) par l'ordre dans lequel les opérations d'unions et d'intersections d'intervalles sont effectuées.

Le théorème suivant joue un rôle central : d'une part il prouve qu'une fonction à valeurs réelles a une unique AE-extension minimale. D'autre part, il donne une expression pour cette AE-extension minimale. Si cette expression est trop coûteuse à évaluer pour des fonctions quelconques (il s'agit d'un problème NP-difficile) elle sera très utile pour calculer les AE-extensions des fonctions élémentaires.

Théorème IV.1. *Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue. Alors f^* est l'unique AE-extension minimale de f .*

Démonstration. En utilisant la Proposition IV.11, il suffit de montrer que quelque soit $\mathbf{z} \in \mathbb{K}\mathbb{R}$, la proposition quantifiée est vraie si et seulement si $f^*(\mathbf{x}) \subseteq \mathbf{z}$:

$$(\forall x_P \in \mathbf{x}_P)(\mathbf{Q}^{(\mathbf{z})}z \in \text{pro } \mathbf{z})(\exists x_I \in \text{pro } \mathbf{x}_I)(z = f(x)) \quad (13.1)$$

où $P = P(\mathbf{x})$, $I = I(\mathbf{x})$ et le quantificateur $\mathbf{Q}^{(\mathbf{z})}$ est défini \mathbf{z} comme dans le corollaire IV.1 (i.e. $\mathbf{Q}^{(\mathbf{z})} = \exists \iff \mathbf{z} \in \mathbb{I}\mathbb{R}$). Nous supposons que P et I sont non-vides, les autres cas étant similaires et plus simples. Pour tout $x_P \in \mathbf{x}_P$, l'ensemble de réels $\{z \in \mathbb{R} | (\exists x_I \in \mathbf{x}_I)(z = f(x))\}$ est un intervalle car f est continue. Cet intervalle sera noté $\mathbf{r}(x_P, \mathbf{x}_I)$ et nous avons

$$\mathbf{r}(x_P, \mathbf{x}_I) = \bigvee_{x_I \in (\text{pro } \mathbf{x}_I)} f(x) \quad \text{et} \quad \text{dual } \mathbf{r}(x_P, \mathbf{x}_I) = \bigwedge_{x_I \in (\text{pro } \mathbf{x}_I)} f(x)$$

Maintenant, $(\exists x_I \in \text{pro } \mathbf{x}_I)(z = f(x))$ est équivalent à $z \in \mathbf{r}(x_P, \mathbf{x}_I)$. La proposition quantifiée (13.1) est donc équivalente à

$$(\forall x_P \in \mathbf{x}_P)(\mathbf{Q}^{(\mathbf{z})}z \in \text{pro } \mathbf{z})(z \in \mathbf{r}(x_P, \mathbf{x}_I)).$$

En appliquant le corollaire IV.1, nous avons l'équivalence suivante :

$$(\mathbf{Q}^{(\mathbf{z})}z \in \text{pro } \mathbf{z})(z \in \mathbf{r}(x_P, \mathbf{x}_I)) \iff \mathbf{z} \supseteq (\text{dual } \mathbf{r}(x_P, \mathbf{x}_I)).$$

Donc, la proposition quantifiée (13.1) est équivalente à

$$(\forall x_P \in \mathbf{x}_P)(\mathbf{z} \supseteq (\text{dual } \mathbf{r}(x_P, \mathbf{x}_I))).$$

Finalement, par définition de la réunion d'intervalles (i.e. plus petit majorant) cette dernière proposition est équivalente à

$$\mathbf{z} \supseteq \bigvee_{x_P \in \mathbf{x}_P} (\text{dual } \mathbf{r}(x_P, \mathbf{x}_I)).$$

Ce qui conclut la preuve grâce à l'expression explicite de $\text{dual } \mathbf{r}(x_P, \mathbf{x}_I)$. □

Remarque. *Le Théorème IV.1 est similaire à l'équivalence (1) \iff (3) du Théorème II.3 page 55. Néanmoins, la preuve proposée ici est différente de celle proposée dans SIGLA/X[37].*

Exemple IV.16. *Considérons la fonction $f(x, y) = x^2 + y^2$ et les intervalles propres $\mathbf{x} = [-3, 3]$ et $\mathbf{y} = [4, 6]$. Alors,*

$$\begin{aligned} f^*(\mathbf{x}, \mathbf{y}) &= [\min_{x \in \mathbf{x}} \min_{y \in \mathbf{y}} f(x, y), \max_{x \in \mathbf{x}} \max_{y \in \mathbf{y}} f(x, y)] \\ &= [\min_{x \in \mathbf{x}} x^2 + \underline{\mathbf{y}}^2, \max_{x \in \mathbf{x}} x^2 + \bar{\mathbf{y}}^2] \\ &= [0 + \underline{\mathbf{y}}^2, \bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2] \\ &= [16, 9 + 36] = [16, 45] \end{aligned}$$

Comme f^ est l'unique AE-extension minimale de f , ce calcul montre que $\mathbf{z} = [16, 45]$ est le plus petit intervalle qui vérifie*

$$(\forall x \in \mathbf{x})(\forall y \in \mathbf{y})(\exists z \in \mathbf{z})(x^2 + y^2 = z).$$

C'est à dire $\text{range}(f, \mathbf{x}, \mathbf{y}) = [16, 45]$.

Exemple IV.17. *Dans la même situation que pour l'exemple précédent, calculons maintenant*

$$\begin{aligned} f^*(\text{dual } \mathbf{x}, \text{dual } \mathbf{y}) &= [\max_{x \in \mathbf{x}} \max_{y \in \mathbf{y}} f(x, y), \min_{x \in \mathbf{x}} \min_{y \in \mathbf{y}} f(x, y)] \\ &= [\max_{x \in \mathbf{x}} x^2 + \bar{\mathbf{y}}^2, \min_{x \in \mathbf{x}} x^2 + \underline{\mathbf{y}}^2] \\ &= [\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2, 0 + \underline{\mathbf{y}}^2] \\ &= [9 + 36, 16] = [45, 16] \end{aligned}$$

Comme f^ est l'unique AE-extension minimale de f , ce calcul montre que $\mathbf{z} = \text{pro}[45, 16] = [16, 45]$ est le plus grand intervalle qui vérifie*

$$(\forall z \in \mathbf{z})(\exists x \in \mathbf{x})(\exists y \in \mathbf{y})(x^2 + y^2 = z)$$

C'est à dire $\text{range}(f, \mathbf{x}, \mathbf{y}) = [16, 45]$.

Exemple IV.18. *Dans la même situation que pour les deux exemples précédent, calculons maintenant*

$$\begin{aligned} f^*(\mathbf{x}, \text{dual } \mathbf{y}) &= [\min_{x \in \mathbf{x}} \max_{y \in \mathbf{y}} f(x, y), \max_{x \in \mathbf{x}} \min_{y \in \mathbf{y}} f(x, y)] \\ &= [\min_{x \in \mathbf{x}} x^2 + \bar{\mathbf{y}}^2, \max_{x \in \mathbf{x}} x^2 + \underline{\mathbf{y}}^2] \\ &= [0 + \bar{\mathbf{y}}^2, \bar{\mathbf{x}}^2 + \underline{\mathbf{y}}^2] \\ &= [36, 9 + 16] = [36, 25] \end{aligned}$$

Comme f^ est l'unique AE-extension minimale de f , ce calcul montre que $\mathbf{z} = \text{pro}[36, 25] = [25, 36]$ est le plus grand intervalle qui vérifie*

$$(\forall z \in \mathbf{z})(\forall x \in \mathbf{x})(\exists y \in \mathbf{y})(x^2 + y^2 = z)$$

Calculer f^* est un problème NP-difficile par rapport au nombre de variables car le calcul de l'image d'un intervalle par f est NP-difficile (cf. Kreinovich[98]) et ce dernier problème est un cas particulier du calcul de f^* (les autres cas n'étant pas plus simples).

13.2 Commutation avec l'opération dual

Cette notion est introduite après avoir constaté que, dans le cas de fonctions à valeurs réelles, l'unique approximation intérieure non-extensible de l'image d'une boîte par une fonction est égale à l'unique approximation extérieure non-rétractable, toute deux étant égales à l'image exacte de la boîte. Formulée dans le langage des AE-extensions, cette coïncidence s'exprime de la manière suivante : considérons une fonction continue $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et un intervalle \mathbf{x} ,

1. $f^*(\text{dual } \mathbf{x}) = (\text{dual range}(f, \text{pro } \mathbf{x}))$ et $\text{pro } f^*(\mathbf{x})$ est l'unique approximation intérieure non-extensible de $\text{range}(f, \mathbf{x})$.
2. $f^*(\mathbf{x}) = \text{range}(f, \mathbf{x})$ et $f^*(\mathbf{x})$ est l'unique approximation extérieure non-rétractable de $\text{range}(f, \mathbf{x})$.

Et ainsi, si $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ est propre ou impropre, i.e. $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$ ou $\mathbf{x} \in \overline{\mathbb{I}\mathbb{R}^n}$, nous avons

$$f^*(\text{dual } \mathbf{x}) = \text{dual } f^*(\mathbf{x})$$

C'est à dire, l'AE-extension minimale f^* de f commute avec l'opération dual pour des arguments propres ou impropres. Cette propriété reste-t-elle vraie pour des arguments quelconques ? La réponse à cette question est non. En générale, seule une relation plus faible lie $f^*(\text{dual } \mathbf{x})$ et $\text{dual } f^*(\mathbf{x})$:

Proposition IV.17. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Alors,*

$$f^*(\text{dual } \mathbf{x}) \subseteq \text{dual } f^*(\mathbf{x})$$

Démonstration. Définissons $P = P(\mathbf{x})$ et $I = I(\mathbf{x})$. Si l'un des ensembles d'indices est vide, nous sommes dans le cas où \mathbf{x} est propre ou impropre. Ainsi, $f^*(\text{dual } \mathbf{x}) = \text{dual } f^*(\mathbf{x})$ et l'inclusion est donc vérifiée. Si les deux ensembles d'indices sont non-vides, par définition de f^* nous avons

$$f^*(\mathbf{x}) = \bigvee_{x_P \in \mathbf{x}_P} \bigwedge_{x_I \in (\text{pro } \mathbf{x}_I)} f(x).$$

Comme l'opération dual inverse les opérations du treillis des intervalles généralisés, nous obtenons

$$(\text{dual } f^*(\mathbf{x})) = \bigwedge_{x_P \in \mathbf{x}_P} \bigvee_{x_I \in (\text{pro } \mathbf{x}_I)} f(x).$$

La définition de f^* conduit à l'expression suivante (les composantes propres de \mathbf{x} deviennent les composantes impropres de $\text{dual } \mathbf{x}$ et inversement) :

$$f^*(\text{dual } \mathbf{x}) = \bigvee_{x_I \in (\text{pro } \mathbf{x}_I)} \bigwedge_{x_P \in \mathbf{x}_P} f(x)$$

Il ne reste donc plus qu'à prouver l'inclusion suivante :

$$\bigvee_{x_I \in (\text{pro } \mathbf{x}_I)} \bigwedge_{x_P \in \mathbf{x}_P} f(x) \subseteq \bigwedge_{x_P \in \mathbf{x}_P} \bigvee_{x_I \in (\text{pro } \mathbf{x}_I)} f(x)$$

Cette inclusion est une propriété usuelle des treillis : on obtient toujours un élément plus petit en effectuant d'abord les opérations minimum (ou intersection) avant les maximum (ou union). \square

La propriété abstraite donnée par la Proposition IV.17 a des implications pratiques : par exemple, étant donnés $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continue et non constante sur \mathbf{x} , i.e. $\text{range}(f, \mathbf{x})$ n'est pas un intervalle dégénéré, au moins l'un des deux intervalles $f^*(\mathbf{x})$ et $f^*(\text{dual } \mathbf{x})$ doit être propre (sans quoi l'inclusion imposée par la Proposition IV.17 serait violée car un intervalle impropre ne peut contenir un intervalle propre que si les deux sont dégénérés). En pratique, cela implique que quelque soient $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ continue, $\mathbf{x} \in \mathbb{I}\mathbb{R}$ et $\mathbf{y} \in \mathbb{I}\mathbb{R}$, au moins l'une des des propositions quantifiées suivantes est vraie :

$$(\exists z \in \mathbb{R})(\forall x \in \mathbf{x})(\exists y \in \mathbf{y})(f(x, y) = z)$$

ou

$$(\exists z \in \mathbb{R})(\forall y \in \mathbf{y})(\exists x \in \mathbf{x})(f(x, y) = z)$$

Cette conséquence de la Proposition IV.17 se semble pas être triviale. La Proposition IV.17 permet aussi d'obtenir la condition suffisante suivante pour la minimalité d'une AE-extension :

Proposition IV.18. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue et $\mathbf{g} : \mathbb{K}\mathbb{R}^n \rightarrow \mathbb{K}\mathbb{R}$ une AE-extension de f . Si f^* et \mathbf{g} commutent avec l'opération dual pour un argument \mathbf{x} alors \mathbf{g} est minimale en \mathbf{x} .*

Démonstration. Comme f^* est l'unique AE-extension minimale de f nous avons

$$f^*(\mathbf{x}) \subseteq \mathbf{g}(\mathbf{x}) \quad \text{et} \quad f^*(\text{dual } \mathbf{x}) \subseteq \mathbf{g}(\text{dual } \mathbf{x}).$$

Comme ces deux AE-extensions commutent avec l'opération dual pour l'argument \mathbf{x} , nous avons $f^*(\text{dual } \mathbf{x}) = \text{dual } f^*(\mathbf{x})$ et $\mathbf{g}(\text{dual } \mathbf{x}) = \text{dual } \mathbf{g}(\mathbf{x})$. Donc,

$$f^*(\text{dual } \mathbf{x}) \subseteq \mathbf{g}(\text{dual } \mathbf{x}) \iff (\text{dual } f^*(\mathbf{x})) \subseteq (\text{dual } \mathbf{g}(\mathbf{x})).$$

Nous avons donc $f^*(\mathbf{x}) \supseteq \mathbf{g}(\mathbf{x})$ et finalement $f^*(\mathbf{x}) = \mathbf{g}(\mathbf{x})$. □

Finalement, la prochaine proposition donne une condition suffisante pour qu'une AE-extension commute avec l'opération dual. Elle sera utilisée en conjonction avec la Proposition IV.18 pour démontrer la minimalité de l'AE-extension naturelle dans le cas des fonctions bilinéaires (cf. section 15.4 page 169).

Proposition IV.19. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue qui vérifie*

$$f(x) = \sum_{k \in [1..m]} g_k(x_{E_k})$$

où E_k sont des ensembles d'indices disjoints, $\bigcup E_k = [1..n]$, et g_k sont continues. Si chaque g_k^ commute avec l'opération dual pour l'argument \mathbf{x}_{E_k} alors f^* commute avec l'opération dual pour l'argument \mathbf{x} .*

Démonstration. En utilisant la définition de f^* , f^* commute avec dual pour \mathbf{x} si les deux égalités suivantes sont vérifiées :

$$\min_{x_P \in \mathbf{x}_P} \max_{x_I \in \mathbf{x}_I} f(x) = \max_{x_I \in \mathbf{x}_I} \min_{x_P \in \mathbf{x}_P} f(x) \quad ; \quad \max_{x_P \in \mathbf{x}_P} \min_{x_I \in \mathbf{x}_I} f(x) = \min_{x_I \in \mathbf{x}_I} \max_{x_P \in \mathbf{x}_P} f(x)$$

où $P = P(\mathbf{x})$ et $I = I(\mathbf{x})$. Considérons la première égalité et définissons $I_k = E_k \cap I$ et $P_k = E_k \cap P$. Supposons que tous les ensembles d'indices ainsi définis sont non-vides (les autres cas étant similaires et plus simples. Notons $\text{pro } \mathbf{x}_I$ par \mathbf{x}'_I). Alors,

$$\begin{aligned}
\min_{x_P \in \mathbf{x}_P} \max_{x_I \in \mathbf{x}'_I} f(x) &= \min_{x_P \in \mathbf{x}_P} \max_{x_I \in \mathbf{x}'_I} \sum_{k \in [1..m]} g_k(x_{E_k}) \\
&= \sum_{k \in [1..m]} \min_{x_{P_k} \in \mathbf{x}_{P_k}} \max_{x_{I_k} \in \mathbf{x}'_{I_k}} g_k(x_{E_k}) \\
(g_k \text{ commute avec dual}) &= \sum_{k \in [1..m]} \max_{x_{I_k} \in \mathbf{x}'_{I_k}} \min_{x_{P_k} \in \mathbf{x}_{P_k}} g_k(x_{E_k}) \\
&= \max_{x_I \in \mathbf{x}'_I} \min_{x_P \in \mathbf{x}_P} \sum_{k \in [1..m]} g_k(x_{E_k}) \\
&= \max_{x_I \in \mathbf{x}'_I} \min_{x_P \in \mathbf{x}_P} f(x)
\end{aligned}$$

La seconde égalité est prouvée de la même manière. □

Remarque. Cette proposition est une version simplifiée de la "tree-optimality" proposée dans la théorie des intervalles modaux, cf. SIGLA/X[37].

Chapitre 14

AE-extensions des fonctions élémentaires

Les fonctions élémentaires qui seront considérées sont les suivantes, leur domaines de définition étant les domaines habituels :

- fonctions à deux variables : $\Omega = \{ x + y, x - y, x \times y, x/y \}$ définies dans \mathbb{R}^2 exceptée x/y qui n'est pas définie pour $y = 0$;
- fonctions à une variable : $\Phi = \{ \exp x, \ln x, \sin x, \cos x, \tan x, \arccos x, \arcsin x, \arctan x, \text{abs } x, x^n, \sqrt[n]{x} \}$ définie dans \mathbb{R} , exceptées \tan , \arccos , \arcsin et $\sqrt[n]{x}$ qui sont définies respectivement dans $\mathbb{R} \setminus \{k\pi + \pi/2 \mid k \in \mathbb{Z}\}$, $[-1, 1]$, $[-1, 1]$ et $\{x \mid x \geq 0\}$.

Remarque. *L'utilisation d'autres fonctions élémentaires ne peut se faire que sous certaines conditions : ces fonctions doivent satisfaire les propositions IV.20 (page 152) ou IV.21 (page 154).*

Les deux sections suivantes montrent que les AE-extensions minimales f^* de ces fonctions élémentaires coïncident avec les extensions définies dans Kaucher[21]. En particulier, les AE-extensions minimales des fonctions à deux variables de Ω coïncident avec les opérations de l'arithmétique de Kaucher.

14.1 Fonctions à une variable

Dans le cas des fonctions à une variable, les définitions de f^* et f^K coïncident (voir pages 41 et 145 pour leurs définitions respectives) :

$$f^*(\mathbf{x}) = f^K(\mathbf{x}) = \bigwedge_{x \in \text{pro } \mathbf{x}}^{\mathbf{x}} f(x)$$

C'est à dire,

- si $\mathbf{x} \in \mathbb{I}\mathbb{R}$ alors $f^*(\mathbf{x}) = [\min_{x \in \mathbf{x}} f(x), \max_{x \in \mathbf{x}} f(x)] = \text{range}(f, \mathbf{x})$;
- si $\mathbf{x} \in \overline{\mathbb{I}\mathbb{R}}$ alors $f^*(\mathbf{x}) = [\max_{x \in \text{pro } \mathbf{x}} f(x), \min_{x \in \text{pro } \mathbf{x}} f(x)] = \text{dual}(\text{range}(f, \text{pro } \mathbf{x}))$.

Dans la suite, les extensions f^* et f^K des fonctions élémentaires de Φ seront notées avec le même symbole que la fonction originale.

Exemple IV.19.

- $\exp(\mathbf{x}) = [\exp(\underline{\mathbf{x}}), \exp(\overline{\mathbf{x}})]$

- $\ln(\mathbf{x}) = [\ln(\underline{\mathbf{x}}), \ln(\overline{\mathbf{x}})]$ est définie si $\underline{\mathbf{x}} > 0$ et $\overline{\mathbf{x}} > 0$
- $\sqrt{\mathbf{x}} = [\sqrt{\underline{\mathbf{x}}}, \sqrt{\overline{\mathbf{x}}}]$ est définie si $\underline{\mathbf{x}} \geq 0$ et $\overline{\mathbf{x}} \geq 0$

Maintenant que les AE-extensions sont calculées, nous avons besoin d'une propriété supplémentaire : nous devons prouver que le choix des variables quantifiées existentiellement peut être fait de manière continue par rapport aux variations des variables quantifiées universellement. Considérons la proposition quantifiée suivante :

$$(\forall z \in \mathbf{z})(\exists x \in \mathbf{x})(z = f(x))$$

Alors, x peut être choisi de manière continue par rapport à z s'il existe une fonction **continue** $s : \mathbf{z} \longrightarrow \mathbf{x}$ qui vérifie $x = s(z) \implies z = f(x)$.

Exemple IV.20. Considérons $f(x) = x^2$, $\mathbf{x} = [2, -1]$ et $\mathbf{z} = \mathbf{x}^2 = [4, 0]$ (il s'agit de l'AE-extension minimale de la fonction f). La proposition quantifiée suivante est donc vraie :

$$(\forall z \in \text{pro } \mathbf{z})(\exists x \in \text{pro } \mathbf{x})(z = f(x)),$$

c'est à dire $\text{pro } \mathbf{z} \subseteq \text{range}(f, \text{pro } \mathbf{x})$. Le choix de x peut être fait continument par rapport à z car il existe une fonction **continue** $s : \text{pro } \mathbf{z} \longrightarrow \text{pro } \mathbf{x}$ qui vérifie $x = s(z) \implies z = f(x)$. Cette fonction est $s(z) = \sqrt{x}$ et en effet $\text{range}(s, \text{pro } \mathbf{z}) \subseteq \text{pro } \mathbf{x}$ et $x = \sqrt{z} \implies x^2 = z$.

La proposition suivante généralise l'exemple précédent à toutes les fonctions de Φ et pour n'importe quel argument. Il s'agit d'un résultat technique qui sera utilisé dans la suite.

Proposition IV.20. Soit f une fonction élémentaire de Φ et $(\mathbf{x}_0, \mathbf{x}_1)^T \in \mathbb{K}\mathbb{R}^2$ satisfaisant $\mathbf{x}_0 = f^*(\mathbf{x}_1)$. Définissons les ensembles d'indices suivants : $A = P(\mathbf{x}_1) \cup I(\mathbf{x}_0)$ et $E = I(\mathbf{x}_1) \cup P(\mathbf{x}_0)$ (A contient les indices des variables quantifiées universellement et E les indices des variables quantifiées existentiellement²⁰). Alors A et E sont non-vides et il existe une fonction **continue** $s : \mathbf{x}_A \longrightarrow \mathbf{x}_E$ qui vérifie $x_E = s(x_A) \implies x_0 = f(x_1)$.

Démonstration. La démonstration de cette proposition est donnée dans la section 14.3 page 155. □

Remarque. Il existe des fonctions à une variable qui ne satisfont pas la Proposition IV.20 (voir exemple suivant). Donc, avant d'ajouter une nouvelle fonction élémentaire dans Φ , il faut vérifier qu'elle vérifie la Proposition IV.20.

Exemple IV.21. Considérons la fonction $f(x) = x^3 - 3x$ et l'intervalle $\mathbf{x} = [3, -3]$. Ainsi,

$$f^*(\mathbf{x}) = \text{dual range}(f, \text{pro } \mathbf{x}) = [18, -18] =: \mathbf{z}$$

On peut constater sur la figure 14.1 toutes les fonctions $s : \text{pro } \mathbf{z} \longrightarrow \text{pro } \mathbf{x}$ qui vérifient $x = s(z) \implies z = f(x)$ sont discontinues.

Remarque. Beaucoup de temps a été passé pour essayer de donner une caractérisation alternative pour la continuité des variations des variables existentiellement quantifiées par rapport aux variables universellement quantifiées : si la caractérisation proposée ci-dessus est suffisante pour les développements ultérieurs, une meilleur caractérisation

²⁰Dans le cas des fonctions élémentaires à une variable, soit $A = \{1\}$ et $E = \{0\}$ si $\mathbf{x} \in \mathbb{I}\mathbb{R}$, soit $A = \{0\}$ et $E = \{1\}$ si $\mathbf{x} \in \mathbb{I}\mathbb{R}$.

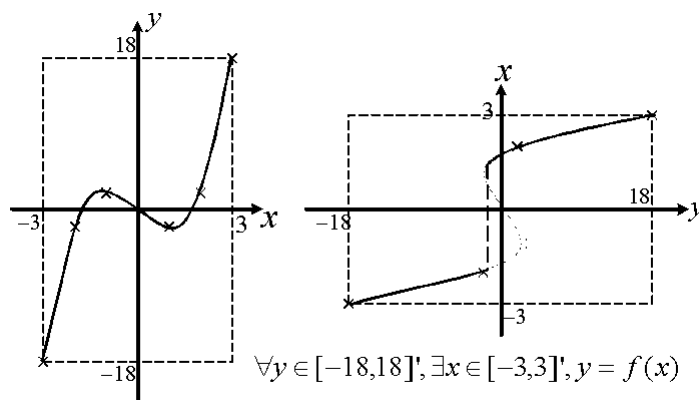


FIG. 14.1 – Une fonction simple qui ne peut pas être considérée comme fonction élémentaire.

aurait de nombreux avantages. En particulier, une voix prometteuse était la c -continuité des fonctions à valeurs ensemblistes proposée dans Neumaier[8]. Grâce à cette généralisation de la continuité, il est possible de prouver que toutes les fonctions à une variable peuvent être considérées comme fonctions élémentaires. Malheureusement, une erreur a été trouvée dans le théorème paramétrique de Miranda pour fonctions c -continues et ce théorème était une condition nécessaire pour l'introduction de la c -continuité dans les développements de cette thèse. Malgré tout, nous restons convaincus que toute fonction à une variable peut être utilisée comme fonction élémentaire, même si la preuve de cette conjecture n'est pas encore faite. Cette preuve nécessite l'introduction d'une autre modélisation pour la continuité des variations des variables existentiellement quantifiées par rapport aux variables universellement quantifiées. Des travaux ont été conduits par Peter Schodl dans cette direction (voir [71]) où l'auteur introduit la notion de fonctions à valeurs ensemblistes "implicitly-selectable continuous".

Remarquons finalement que les AE-extensions des fonctions élémentaires à une variable commutent avec l'opération dual (comme c'est illustré dans la section 13.2).

14.2 Fonctions à deux variables

Dans le cas des fonctions à deux variables, les expressions de $f^*(\mathbf{x}, \mathbf{y})$ et $f^K(\mathbf{x}, \mathbf{y})$ ne coïncident pas en général. Cependant, elles coïncident dans les cas suivants :

- $\mathbf{x} \in \mathbb{IR}$ et $\mathbf{y} \in \mathbb{IR}$
- $\mathbf{x} \in \overline{\mathbb{IR}}$ et $\mathbf{y} \in \overline{\mathbb{IR}}$
- $\mathbf{x} \in \mathbb{IR}$ et $\mathbf{y} \in \overline{\mathbb{IR}}$

Il reste à étudier le cas où $\mathbf{x} \in \overline{\mathbb{IR}}$ et $\mathbf{y} \in \mathbb{IR}$. Dans ce cas, les expressions de f^* et f^K sont les suivantes :

$$f^*(\mathbf{x}, \mathbf{y}) = \bigvee_{x \in \mathbf{x}} \bigwedge_{y \in \text{pro } \mathbf{y}} f(x, y) ; \quad f^K(\mathbf{x}, \mathbf{y}) = \bigwedge_{y \in \text{pro } \mathbf{y}} \bigvee_{x \in \mathbf{x}} f(x, y)$$

Elles conduisent à des résultats différents en général. Cependant, il est maintenant prouvé que ces deux expressions donnent le même résultat dans le cas des fonctions élémentaires

de Ω . Les opérations de l'arithmétique de Kaucher sont temporairement notées \circ^K pour marquer leur différence avec les AE-extensions minimales, notées \circ^* . Considérons tout d'abord une fonction parmi $\circ \in \{+, \times\}$. Si $\mathbf{x} \in \overline{\mathbb{R}}$ et $\mathbf{y} \in \mathbb{R}$ alors par définition de f^* et f^K , l'expression de $\mathbf{x} \circ^* \mathbf{y}$ coïncide avec l'expression de $\mathbf{y} \circ^K \mathbf{x}$. Finalement, parce que \circ^K est commutative, nous avons

$$\mathbf{x} \circ^* \mathbf{y} = \mathbf{y} \circ^K \mathbf{x} = \mathbf{x} \circ^K \mathbf{y}$$

Dans le cas de la soustraction, nous avons $\mathbf{x} -^K \mathbf{y} = (-^K \mathbf{y}) +^K \mathbf{x}$ qui est égal à $\mathbf{x} -^* \mathbf{y}$ parce que :

$$\begin{aligned} (-^K \mathbf{y}) +^K \mathbf{x} &= \bigvee_{v \in (-\mathbf{y})} \bigwedge_{x \in \text{pro } \mathbf{x}} v + x \\ &= \bigvee_{y \in \mathbf{y}} \bigwedge_{x \in \text{pro } \mathbf{x}} (-y) + x \\ &= \mathbf{x} -^* \mathbf{y} \end{aligned}$$

Finalement, dans le cas de la division nous avons $\mathbf{x}/^K \mathbf{y} = (1/^K \mathbf{y}) \times \mathbf{x}$ qui est égal à $\mathbf{x}/^* \mathbf{y}$ (démonstration similaire à la précédente). Donc, les AE-extensions minimales des fonctions élémentaires de Ω coïncident avec leur extensions définies par Kaucher. Dans la suite, les extensions \circ^* et \circ^K des fonctions de Ω seront notées avec le même symbole que les fonctions réelles qu'elles étendent, i.e. \circ . Comme dans le cas des fonctions élémentaires à une variable, nous devons maintenant prouver que le choix des valeurs des variables quantifiées existentiellement peut être fait continûment par rapport aux valeurs des variables quantifiées universellement.

Proposition IV.21. *Soit \circ une fonction élémentaire de Ω et $\mathbf{x}_{\{0,1,2\}} \in \mathbb{K}\mathbb{R}^3$ tel que $\mathbf{x}_0 = \mathbf{x}_1 \circ \mathbf{x}_2$. Définissons les ensembles d'indices suivants : $A = P(\mathbf{x}_{\{1,2\}}) \cup I(\mathbf{x}_0)$ et $E = I(\mathbf{x}_{\{1,2\}}) \cup P(\mathbf{x}_0)$ (A contient les indices des variables quantifiées universellement et E les indices des variables quantifiées existentiellement). Alors A et E sont non-vides et il existe une fonction continue $s : \mathbf{x}_A \longrightarrow \mathbf{x}_E$ qui vérifie $x_E = s(x_A) \implies x_0 = x_1 \circ x_2$.*

Démonstration. La démonstration de cette proposition est donnée dans la section 14.3 page 155. \square

Exemple IV.22. *Considérons la fonction $f(x, y) = x + y$, $\mathbf{x} = [1, -1]$, $\mathbf{y} = [0, 1]$ et $\mathbf{z} = \mathbf{x} + \mathbf{y} = [1, 0]$. La proposition quantifiée suivante est donc vraie :*

$$(\forall x \in \text{pro } \mathbf{x})(\forall z \in \text{pro } \mathbf{z})(\exists y \in \mathbf{y})(x + y = z)$$

La Proposition IV.21 énonce une propriété plus forte : elle prouve qu'il existe une fonction continue $s : (\mathbf{y}, \text{pro } \mathbf{z})^T \longrightarrow \text{pro } \mathbf{x}$ qui vérifie $x = s(y, z) \implies z = f(x, y)$. Dans ce cas, la fonction s peut-être facilement calculée : $s(y, z) = z - y$. En effet, $\text{range}(s, \mathbf{y}, \text{pro } \mathbf{z}) = \text{pro } \mathbf{z} - \mathbf{y} = \text{pro } \mathbf{x}$ (et $x = z - y \implies z = x + y$).

Remarque. *Encore une fois, il existe des fonctions continues à deux variables qui ne vérifient pas la proposition précédente (par exemple $f(x, y) = 1 - (x - y)^2$ avec $\mathbf{x} = [1, -1]$ et $\mathbf{y} = [1, -1]$). Ainsi, avant d'ajouter de nouvelles fonction élémentaires à Φ , nous devons vérifier que celles-ci vérifient la Proposition IV.21. Le cas des fonctions à deux variables est différent de celui des fonctions à une variable car il est prouvé que certaines*

fonctions à deux variables (en fait la plupart) ne peuvent pas être considérées comme fonctions élémentaires quelque soit la modélisation choisie pour la continuité des du choix des variables existentielles par rapports aux universelles (cf. Exemple IV.27 où le fait de considérer la fonction $f(x, y) = 1 - (x - y)^2$ comme une fonction élémentaire conduit à des résultats faux).

Pour toute fonction $\circ \in \{+, -, \times, /\}$ de l'arithmétique de Kaucher, il est prouvé dans Kaucher[21] que quelque soient $\mathbf{x}, \mathbf{y} \in \mathbb{K}\mathbb{R}$

$$(\text{dual } \mathbf{x}) \circ (\text{dual } \mathbf{y}) = \text{dual}(\mathbf{x} \circ \mathbf{y}).$$

Ce qui veut dire que \circ commute avec l'opération dual.

14.3 Démonstrations

Les preuves des propositions IV.20 et IV.21 sont maintenant présentées. Avant tout, le cas suivant est trivial :

Proposition IV.22. Soient $f : \mathbf{x}_{[1..n]} \longrightarrow \mathbb{R}$ une fonction continue et $\mathbf{x}_{[1..n]} \in \mathbb{I}\mathbb{R}^n$. Définissons $\mathbf{x}_0 = f^*(\mathbf{x}_{[1..n]})$ qui est propre. Il existe alors une fonction continue $s : \mathbf{x}_{[1..n]} \longrightarrow \mathbf{x}_0$ qui vérifie

$$x_0 = s(x_{[1..n]}) \implies x_0 = f(x_{[1..n]})$$

Démonstration. La fonction s n'est autre que f qui est continue. □

14.3.1 Fonctions élémentaires à une variable

Dans le contexte des fonctions à une variable, il n'y a que deux cas à étudier : soit $\mathbf{x} \in \mathbb{I}\mathbb{R}$ soit $\mathbf{x} \in \overline{\mathbb{I}\mathbb{R}}$. Le premier a déjà été traité dans la Proposition IV.22, il ne reste donc que le second cas à étudier. La prochaine proposition s'applique aux fonctions $\exp x$, $\ln x$, \tan , $\arccos x$, $\arcsin x$, $\arctan x$, x^n pour n impaire et $x^{\frac{1}{n}}$.

Proposition IV.23. Soit $f : \mathbf{x} \longrightarrow \mathbb{R}$ une fonction continue strictement croissante ou strictement décroissante. Définissons alors $\mathbf{z} = f^*(\text{dual } \mathbf{x})$, qui est impropre. Il existe alors une fonction continue $s : \text{pro } \mathbf{z} \longrightarrow \mathbf{x}$ qui vérifie $x = s(z) \implies z = f(x)$.

Démonstration. Par définition de f^* , $\text{pro } \mathbf{z} = \text{range}(f, \mathbf{x})$. Comme f est continue et strictement monotone, c'est un homéomorphisme entre \mathbf{x} et \mathbf{z} , i.e. la bijection inverse $f^{-1} : \text{pro } \mathbf{z} \longrightarrow \mathbf{x}$ est aussi continue. Finalement, $x = f^{-1}(z) \implies z = f(x)$ et $s = f^{-1}$ est la fonction recherchée. □

Finalement, la proposition suivante s'applique aux fonctions $\sin x$, $\cos x$, $\text{abs } x$, x^n pour n pair.

Proposition IV.24. Soit $f : \mathbf{x} \longrightarrow \mathbb{R}$ une fonction continue qui n'a pas de minimum local ou maximum local excepté peut-être sur les bornes de \mathbf{x} et qui est strictement monotone entre ses minimum et maximum (locaux ou globaux). Définissons alors $\mathbf{z} = f^*(\text{dual } \mathbf{x})$, qui est impropre. Il existe alors une fonction continue $s : \text{pro } \mathbf{z} \longrightarrow \mathbf{x}$ qui vérifie $x = s(z) \implies z = f(x)$.

Démonstration. Considérons un minimum global $x_1 \in \mathbf{x}$ et un maximum global $x_2 \in \mathbf{x}$ qui ne sont séparés par aucun autre minimum ou maximum. C'est bien possible grâce à l'hypothèse de la proposition. Définissons $\mathbf{x}' = x_1 \vee x_2$. La fonction $f|_{\mathbf{x}'}$ est soit strictement croissante soit strictement décroissante, donc la proposition précédente peut s'appliquer. Finalement, s est définie dans $\text{range}(f|_{\mathbf{x}'})$ qui est égale à $\text{range}(f, \mathbf{x})$ car x_1 et x_2 correspondent à des minimum et maximum globaux. De plus, $s(x) \in \mathbf{x}' \subseteq \mathbf{x}$ et $x = s(z) \implies z = f|_{\mathbf{x}'}(x) = f(x)$. \square

Remarque. La proposition précédente ne s'applique pas au cas de l'Exemple IV.21 (page 152) car la fonction définie dans cet exemple possède un minimum local et un maximum local à l'intérieur de l'intervalle.

14.3.2 Fonctions élémentaires à deux variables

Le cas où $\mathbf{x} \in \mathbb{IR}$ et $\mathbf{y} \in \mathbb{IR}$ est déjà traité par la Proposition IV.22. De plus, les cas de $-$ et $/$ s'obtiennent grâce à $+$ et \times parce que

$$\mathbf{x} - \mathbf{y} = \mathbf{x} + (-\mathbf{y}) \quad \text{et} \quad \mathbf{x}/\mathbf{y} = \mathbf{x} \times (1/\mathbf{y})$$

La proposition suivante traite le cas de toutes les fonctions élémentaire à deux variable lorsque $\mathbf{x} \in \overline{\mathbb{IR}}$ et $\mathbf{y} \in \overline{\mathbb{IR}}$. Dans la suite, nous supposons que $\text{pro } \mathbf{x}_1$ et $\text{pro } \mathbf{x}_2$ ne sont pas dégénérés, les autres cas étant similaires et plus simples.

Proposition IV.25. Soit $\mathbf{x}_1 \in \overline{\mathbb{IR}}$, $\mathbf{x}_2 \in \overline{\mathbb{IR}}$ et $\mathbf{x}_0 = \mathbf{x}_1 \circ \mathbf{x}_2$ pour $\circ \in \{+, \times\}$. Il existe alors une fonction continue $s : \mathbf{x}_0 \longrightarrow \text{pro } \mathbf{x}_{\{1,2\}}$ qui vérifie $x_{\{1,2\}} = s(x_0) \implies x_0 = f(x_1, x_2)$ où $f(x_1, x_2) = x_1 \circ x_2$.

Démonstration. Définissons $\mathbf{x}'_1 = \text{pro } \mathbf{x}_1$, $\mathbf{x}'_2 = \text{pro } \mathbf{x}_2$ et $\mathbf{x}'_0 = \text{pro } \mathbf{x}_0$. Nous avons

$$\mathbf{x}'_0 = \text{range}(f, \mathbf{x}'_1, \mathbf{x}'_2).$$

Donc $\mathbf{x}'_0 = [u_1 \circ u_2, v_1 \circ v_2]$ avec $u_{\{1,2\}} \in \mathbf{x}'_{\{1,2\}}$ et $v_{\{1,2\}} \in \mathbf{x}'_{\{1,2\}}$ ($u_1 \circ u_2 < v_1 \circ v_2$ parce que \mathbf{x}'_0 n'est pas dégénéré par hypothèse). Définissons maintenant la fonction continue $m_{\{1,2\}} : [0, 1] \longrightarrow \mathbf{x}'_{\{1,2\}}$ par

$$m_{\{1,2\}}(t) = u_{\{1,2\}}(1-t) + v_{\{1,2\}}t$$

et la fonction continue $g : [0, 1] \longrightarrow \mathbf{x}'_0$ par

$$g(t) = f(m_{\{1,2\}}(t)) = (u_1(1-t) + v_1t) \circ (u_2(1-t) + v_2t)$$

Nous avons $g(0) = \inf \mathbf{x}'_0$ et $g(1) = \sup \mathbf{x}'_0$ et donc $\text{range}(g, [0, 1]) = \mathbf{x}'_0$. Prouvons maintenant que les propositions IV.23 et IV.24 s'appliquent à g quelque soit \circ :

$\circ = +$ Nous avons $g(t) = (u_1 + u_2)(1-t) + (v_1 + v_2)t$ qui est strictement croissante parce que $u_1 + u_2 < v_1 + v_2$. Donc la Proposition IV.23 s'applique.

$\circ = \times$ Nous avons $g(t) = at^2 + bt + c$. Si $a \neq 0$ alors la Proposition IV.24 s'applique.

Si $a = 0$ et $b \neq 0$ alors la Proposition IV.23 s'applique. Nous ne pouvons pas avoir $a = 0$ et $b = 0$ car \mathbf{x}'_0 n'est pas dégénéré par hypothèse.

Il existe donc une fonction continue $\tilde{s} : \mathbf{x}'_0 \longrightarrow [0, 1]$ qui vérifie $t = \tilde{s}(x_0) \implies x_0 = g(t)$. La fonction recherchée est alors $s(x_0) = m_{\{1,2\}}(\tilde{s}(x_0))$: d'une part $s(x_0) \in \mathbf{x}'_{\{1,2\}}$ par construction de $m_{\{1,2\}}$. D'autre part, $x_{\{1,2\}} = s(x_0)$ implique $x_{\{1,2\}} = m_{\{1,2\}}(\tilde{s}(x_0))$ qui implique $f(x_{\{1,2\}}) = f(m_{\{1,2\}}(\tilde{s}(x_0))) = g(\tilde{s}(x_0)) = x_0$. \square

Il reste à étudier les cas où \mathbf{x}_1 et \mathbf{x}_2 n'ont pas la même qualité propre/impropre. La proposition suivante traite le cas de l'addition.

Proposition IV.26. *Soient $\mathbf{x}_1 \in \mathbb{K}\mathbb{R}$, $\mathbf{x}_2 \in \mathbb{K}\mathbb{R}$ et $\mathbf{x}_0 = \mathbf{x}_1 + \mathbf{x}_2$. Définissons les ensembles d'indices suivants : $A = P(\mathbf{x}_{\{1,2\}}) \cup I(\mathbf{x}_0)$ et $E = I(\mathbf{x}_{\{1,2\}}) \cup P(\mathbf{x}_0)$. Alors A et E sont non-vides et il existe une fonction continue $s : \mathbf{x}_A \longrightarrow \mathbf{x}_E$ qui vérifie $x_E = s(x_A) \implies x_0 = x_1 + x_2$.*

Démonstration. Définissons $\mathbf{x}'_1 = \text{pro } \mathbf{x}_1$, $\mathbf{x}'_2 = \text{pro } \mathbf{x}_2$ et $\mathbf{x}'_0 = \text{pro } \mathbf{x}_0$. Nous avons juste deux cas à étudier car $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_2 + \mathbf{x}_1$.

(1) $\mathbf{x}_1 \in \mathbb{I}\mathbb{R}$, $\mathbf{x}_2 \in \overline{\mathbb{I}\mathbb{R}}$ et $\mathbf{x}_0 \in \mathbb{I}\mathbb{R}$. La proposition quantifiée correspondant à l'interprétation de l'AE-extension minimale est

$$(\forall x_1 \in \mathbf{x}'_1)(\exists x_0 \in \mathbf{x}'_0)(\exists x_2 \in \mathbf{x}'_2)(x_1 + x_2 = x_0)$$

Nous devons prouver l'existence d'une fonction continue $s_{\{0,2\}} : \mathbf{x}'_1 \longrightarrow \mathbf{x}'_{\{0,2\}}$ qui vérifie $x_{\{0,2\}} = s_{\{0,2\}}(x_1) \implies x_0 = x_1 + x_2$. En appliquant les règles de l'arithmétique de Kaucher, nous obtenons $\mathbf{x}_0 = \mathbf{x}_1 + \mathbf{x}_2 \iff (\text{dual } \mathbf{x}_1) = (\text{dual } \mathbf{x}_0) - \mathbf{x}_2$. Nous pouvons alors appliquer la Proposition IV.25 dans le cas de cette soustraction, ce qui prouve l'existence d'une fonction continue $s'_{\{0,2\}} : \mathbf{x}'_1 \longrightarrow \mathbf{x}'_{\{0,2\}}$ qui vérifie $x_{\{0,2\}} = s'_{\{0,2\}}(x_1) \implies x_1 = x_0 - x_2$. Comme $x_1 = x_0 - x_2 \iff x_0 = x_1 + x_2$, $s_{\{0,2\}}$ est finalement la fonction continue recherchée.

(2) $\mathbf{x}_1 \in \mathbb{I}\mathbb{R}$, $\mathbf{x}_2 \in \overline{\mathbb{I}\mathbb{R}}$ et $\mathbf{x}_0 \in \overline{\mathbb{I}\mathbb{R}}$. La démonstration est similaire à la précédente en considérant cette fois $(\text{dual } \mathbf{x}_2) = (\text{dual } \mathbf{x}_0) - \mathbf{x}_1$, nous obtenons une fonction continue $s_2 : \mathbf{x}'_{\{0,1\}} \longrightarrow \mathbf{x}'_2$ qui vérifie les conditions demandées. \square

La proposition suivante traite le cas de la multiplication.

Proposition IV.27. *Soit $\mathbf{x}_1 \in \mathbb{K}\mathbb{R}$, $\mathbf{x}_2 \in \mathbb{K}\mathbb{R}$ et $\mathbf{x}_0 = \mathbf{x}_1 \mathbf{x}_2$. Définissons les ensembles d'indices suivants : $A = P(\mathbf{x}_{\{1,2\}}) \cup I(\mathbf{x}_0)$ et $E = I(\mathbf{x}_{\{1,2\}}) \cup P(\mathbf{x}_0)$. Alors A et E sont non-vides et il existe une fonction continue $s : \mathbf{x}_A \longrightarrow \mathbf{x}_E$ qui vérifie $x_E = s(x_A) \implies x_0 = x_1 x_2$.*

Démonstration. Définissons $\mathbf{x}'_1 = \text{pro } \mathbf{x}_1$, $\mathbf{x}'_2 = \text{pro } \mathbf{x}_2$ et $\mathbf{x}'_0 = \text{pro } \mathbf{x}_0$. Premièrement, si les intervalles mis en jeu ne contiennent pas 0, nous pouvons utiliser l'addition de Kaucher de la manière suivante :

(1) $\mathbf{x}_1 > 0$ et $\mathbf{x}_2 > 0$. Dans ce cas, la multiplication de Kaucher $\mathbf{x}_0 = \mathbf{x}_1 \mathbf{x}_2$ peut-être changée en addition : $\ln \mathbf{x}_0 = \ln \mathbf{x}_1 + \ln \mathbf{x}_2$ qui s'écrit $\tilde{\mathbf{x}}_0 = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2$ avec $\tilde{\mathbf{x}}_k = \ln \mathbf{x}_k$. Alors, on peut appliquer la Proposition IV.26 à l'addition $\tilde{\mathbf{x}}_0 = \tilde{\mathbf{x}}_1 + \tilde{\mathbf{x}}_2$: la fonction \ln ne changeant pas la qualité propre/impropre, l'application de la Proposition IV.26 montre qu'il existe une fonction continue $\tilde{s} : \text{pro } \tilde{\mathbf{x}}_A \longrightarrow \text{pro } \tilde{\mathbf{x}}_E$ qui vérifie $\tilde{x}_E = \tilde{s}(\tilde{x}_A) \implies \tilde{x}_0 = \tilde{x}_1 + \tilde{x}_2$. Définissons alors $s(x_A) = \exp \tilde{s}(\ln x_A)$ (les fonctions \ln et \exp sont appliquées composante par composante). D'une part, nous avons $\text{pro } \ln \mathbf{x}_k = \ln \text{pro } \mathbf{x}_k$, et donc s vérifie $s : \mathbf{x}'_A \longrightarrow \mathbf{x}'_E$. D'autre part, en utilisant les définitions de s ET \tilde{s} , nous obtenons $x_E = s(x_A) \implies x_E = \exp \tilde{s}(\ln x_A) \implies \ln x_E = \tilde{s}(\ln x_A) \implies \ln x_0 = \ln x_1 + \ln x_2 \implies x_0 = x_1 x_2$.

(1') $0 \notin \mathbf{x}'_1$ et $0 \notin \mathbf{x}'_2$. Nous revenons au premier cas de la manière suivante : pour $k \in \{1, 2\}$ définissons $\epsilon_k \in \{-1, 1\}$ tel que $\epsilon_k \mathbf{x}_k > 0$. Donc $\mathbf{x}_1 \mathbf{x}_2 = (\epsilon_1 \epsilon_2)(\epsilon_1 \mathbf{x}_1)(\epsilon_2 \mathbf{x}_2)$.

(2-4) Il reste seulement trois cas où $\mathbf{x}_1 \in \mathbb{I}\mathbb{R}$ et $\mathbf{x}_2 \in \overline{\mathbb{I}\mathbb{R}}$ car $\mathbf{x}_0 \mathbf{x}_1 = \mathbf{x}_1 \mathbf{x}_0$.

(2) $0 \in \mathbf{x}'_1$ et $0 \notin \mathbf{x}'_2$ qui impliquent $\mathbf{x}_0 \in \mathbb{I}\mathbb{R}$ (cf. expressions de l'arithmétique de

Kaucher). L'existence de la fonction s est prouvée de manière similaire au cas (1) de la Proposition IV.26, en considérant que $\mathbf{x}_0 = \mathbf{x}_1 \mathbf{x}_2 \iff (\text{dual } \mathbf{x}_1) = (\text{dual } \mathbf{x}_0) / \mathbf{x}_2$, qui est bien définie car $0 \notin \mathbf{x}'_2$.

(3) $0 \notin \mathbf{x}'_1$ et $0 \in \mathbf{x}'_2$ qui impliquent $\mathbf{x}_0 \in \overline{\mathbb{R}}$. Supposons que $\mathbf{x}_1 > 0$, l'autre cas étant similaire. en utilisant l'expression de la multiplication de Kaucher, nous avons

$$\mathbf{x}'_0 = [(\inf \mathbf{x}'_1)(\inf \mathbf{x}'_2), (\inf \mathbf{x}'_1)(\sup \mathbf{x}'_2)]$$

Définissons $s_2(x_0, x_1) = x_0/x_1$ (qui est bien définie dans $\mathbf{x}'_{\{0,1\}}$ car $0 \notin \mathbf{x}'_1$). D'une part, $x_2 = s_2(x_0, x_1) \implies x_0 = x_1 x_2$. D'autre part,

$$\text{range}(s_2, \mathbf{x}'_0, \mathbf{x}'_1) = \mathbf{x}'_0 / \mathbf{x}'_1 = [\inf \mathbf{x}'_0 / \inf \mathbf{x}'_1, \sup \mathbf{x}'_0 / \inf \mathbf{x}'_1] = \mathbf{x}'_2.$$

(4) $0 \in \mathbf{x}'_1$ et $0 \in \mathbf{x}'_2$. Dans ce cas, $\mathbf{x}_0 = [0, 0]$ et la fonction continue recherchée est $s_2(x_1, x_0) = 0 \in \mathbf{x}'_2$ car $x_2 = s(x_1, x_0) \implies x_1 \times x_2 = x_0$. \square

Chapitre 15

L'AE-extension naturelle

Dans le contexte de la théorie des intervalles classiques, l'extension naturelle a une forte importance car elle permet de construire un sur-ensemble de l'image d'une fonction en ne travaillant que sur les fonctions élémentaires dont la fonction est composée. Dans ce chapitre, une AE-extension naturelle est définie. Elle va permettre de construire des intervalles généralisés (f, \mathbf{x}) -interprétables en ne travaillant que sur les fonctions élémentaires. A la différence de l'extension naturelle classique, l'expression d'une fonction devra être modifiée avant de l'évaluer avec l'arithmétique de Kaucher. L'AE-extension naturelle est construite en trois étapes :

- Dans le cas des fonctions à valeurs réelles, le cas où l'expression de la fonction ne contient qu'une unique occurrence de chaque variable est tout d'abord étudié.
- Le cas général des fonctions à valeurs réelles est ensuite abordé. L'expression de l'AE-extension naturelle de ces fonctions est plus compliquée que l'expression de l'extension naturelle des intervalles classiques (l'expression de la fonction doit être modifiée avant son évaluation pour des intervalles). Néanmoins, lorsque des intervalles propres sont mis en jeux, l'AE-extension naturelle coïncide avec l'extension naturelle des intervalles classiques.
- Le cas général des fonctions à valeurs vectorielle est finalement étudié.

15.1 Évaluation d'intervalles

Commençons par étudier un exemple. Considérons la fonction $f(u, v, w) = u(v + w)$ et trois intervalles $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{K}\mathbb{R}$. Pour illustrer le fait que l'évaluation d'intervalles $\mathbf{z} := \mathbf{u}(\mathbf{v} + \mathbf{w})$ est un intervalle $(f, \mathbf{u}, \mathbf{v}, \mathbf{w})$ -interprétable, définissons $\mathbf{t} = (\mathbf{v} + \mathbf{w})$ et supposons que $\mathbf{u} \in \overline{\mathbb{I}\mathbb{R}}$, $\mathbf{v} \in \mathbb{I}\mathbb{R}$, $\mathbf{w} \in \mathbb{I}\mathbb{R}$, $\mathbf{t} \in \mathbb{I}\mathbb{R}$ et $\mathbf{z} \in \overline{\mathbb{I}\mathbb{R}}$ (les autres cas sont similaires). Parce que les AE-extensions de $+$ et \times sont utilisées pour l'évaluation d'intervalles $\mathbf{u}(\mathbf{v} + \mathbf{w})$, les deux propositions quantifiées suivantes sont vraies :

$$(\forall v \in \mathbf{v})(\forall w \in \mathbf{w})(\exists t \in \mathbf{t})(v + w = t)$$

et

$$(\forall t \in \mathbf{t})(\forall z \in \text{pro } \mathbf{z})(\exists u \in \text{pro } \mathbf{u})(ut = z).$$

La conjonction de ces deux propositions quantifiée implique la suivante :

$$(\forall v \in \mathbf{v})(\forall w \in \mathbf{w})(\forall z \in \text{pro } \mathbf{z})(\exists u \in \text{pro } \mathbf{u})(\exists t \in \mathbf{t})(v + w = t \wedge ut = z).$$

C'est à dire,

$$(\forall v \in \mathbf{v})(\forall w \in \mathbf{w})(\forall z \in \text{pro } \mathbf{z})(\exists u \in \text{pro } \mathbf{u})(u(v+w) = z). \quad (15.1)$$

L'intervalle impropre \mathbf{z} est donc $(f, \mathbf{u}, \mathbf{v}, \mathbf{w})$ -interprétable. Ce raisonnement se généralise sans difficultés à des qualités propres/impropres quelconque. Maintenant, une propriété supplémentaire est nécessaire pour les développements à venir : le choix des valeurs des variables existentiellement quantifiées doit pouvoir être fait continûment par rapport aux valeurs des variables quantifiées universellement. Dans le cas de la proposition quantifiée 15.1 cela consiste à vérifier l'existence d'une fonction **continue** $s : (\mathbf{v}, \mathbf{w}, \text{pro } \mathbf{z})^T \longrightarrow \text{pro } \mathbf{u}$ qui vérifie $u = s(v, w, z) \implies z = f(u, v, w)$: la Proposition IV.21 prouve l'existence de deux fonction **continues**

$$\begin{aligned} s' : (\mathbf{v}, \mathbf{w})^T &\longrightarrow \mathbf{t} & ; & \quad t = s'(v, w) \implies t = v + w \\ s'' : (\mathbf{t}, \text{pro } \mathbf{z})^T &\longrightarrow \text{pro } \mathbf{u} & ; & \quad u = s''(t, z) \implies z = ut \end{aligned}$$

La fonction s est alors construite de la manière suivante :

$$s(v, w, z) = s''(s'(v, w), z)$$

Ainsi, $u = s(v, w, z)$ implique $u = s''(t, z)$ où $t = s'(v, w)$. Cela implique finalement $z = f(u, v, w)$. La proposition suivante généralise cet exemple introductif. Elle est d'une importance capitale pour la construction des AE-extension proposées ensuite.

Proposition IV.28. *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue et \mathbf{f} une expression de f mettant en jeux des fonctions élémentaires de Ω et Φ dans laquelle **chaque variable a une unique occurrence**. Pour tout $\mathbf{x}_{[1..n]} \in \mathbb{K}\mathbb{R}^n$, définissons $\mathbf{x}_0 = \mathbf{f}(\mathbf{x}_{[1..n]})$ où l'évaluation est faite en utilisant les AE-extensions minimales des fonctions mises en jeux. Définissons $A = P(\mathbf{x}_{[1..n]}) \cup I(\mathbf{x}_0)$ et $E = I(\mathbf{x}_{[1..n]}) \cup A(\mathbf{x}_0)$ (A contient les indices des variables quantifiées universellement et E les indices des variables quantifiées existentiellement). Alors A et E sont non-vides et il existe une fonction **continue** $s_E : \mathbf{x}_A \longrightarrow \mathbf{x}_E$ qui vérifie $x_E = s_E(x_A) \implies x_0 = f(x_{[1..n]})$.*

Démonstration. La démonstration est faite par récurrence sur l'expression \mathbf{f} . Premièrement, par les propositions IV.20 et IV.21, la présente proposition est vraie pour les fonctions élémentaires. Il reste donc à prouver l'étape générale de la récurrence. Considérons deux fonctions $g : \mathbb{R}^m \longrightarrow \mathbb{R}$ et $h : \mathbb{R}^{m'} \longrightarrow \mathbb{R}$ qui satisfont la proposition. Nous devons donc prouver que leur composition $f : \mathbb{R}^{m+m'-1} \longrightarrow \mathbb{R}$ définie par

$$f(x_{[1..m-1]} \cup [m+1..m+m']) = \mathbf{f}(x_{[1..m-1]} \cup [m+1..m+m'])$$

où

$$\mathbf{f}(x_{[1..m-1]} \cup [m+1..m+m']) = \mathbf{g}(x_{[1..m-1]}, \mathbf{h}(x_{[m+1..m+m']}))$$

vérifie aussi la proposition²¹.

Remarque. *Une numérotation spécifique des variables est introduite pour pouvoir considérer la variable temporaire de manière homogène. Par exemple $\mathbf{f}(x_1, x_3) = \mathbf{g}(x_1, \mathbf{h}(x_3, x_4))$ permet de définir la variable temporaire $x_2 = \mathbf{h}(x_3, x_4)$ et d'obtenir $x_0 = g(x_1, x_2)$ avec $x_2 = h(x_3, x_4)$.*

²¹L'hypothèse que l'expression \mathbf{f} ne contient qu'une occurrence de chaque variable est implicitement utilisée ici : en effet f est réellement une composition si chaque variable n'a qu'une occurrence dans son expression en fonction de g et h .

Considérons un quelconque $\mathbf{x}_{[1..m-1] \cup [m+1..m+m']} \in \mathbb{K}\mathbb{R}^{m+m'-1}$ et définissons

$$\mathbf{x}_m = \mathbf{h}(\mathbf{x}_{[m+1..m+m']}) \text{ et } \mathbf{x}_0 = \mathbf{g}(\mathbf{x}_{[1..m]}).$$

En appliquant l'hypothèse de récurrence, nous obtenons deux fonctions continues $s'_{E'}$ et $s''_{E''}$ définie de la manière suivante :

$$s'_{E'} : \text{pro } \mathbf{x}_{A'} \longrightarrow \text{pro } \mathbf{x}_{E'} \quad ; \quad x_{E'} = s'(x_{A'}) \implies x_0 = g(x_{[1..m]})$$

et

$$s''_{E''} : \text{pro } \mathbf{x}_{A''} \longrightarrow \text{pro } \mathbf{x}_{E''} \quad ; \quad x_{E''} = s''(x_{A''}) \implies x_m = h(x_{[m+1..m+m']})$$

où $A' \subseteq [0..m]$, $E' \subseteq [0..m]$, $A'' \subseteq [m..m+m']$ et $E'' \subseteq [m..m+m']$ sont définis comme dans l'énoncé de la proposition. Définissons maintenant $A = (A' \cup A'') \setminus \{m\}$, $E = (E' \cup E'') \setminus \{m\}$. Remarquons que A et E sont non-vides car A' , A'' , E' et E'' sont non-vides par hypothèse. Par définition de ces ensembles, nous avons

$$\vee \begin{array}{l} (\mathbf{x}_m \in \mathbb{I}\mathbb{R} \wedge m \in A' \wedge m \notin E' \wedge m \notin A'' \wedge m \in E'') \\ (\mathbf{x}_m \in \overline{\mathbb{I}\mathbb{R}} \wedge m \notin A' \wedge m \in E' \wedge m \in A'' \wedge m \notin E'') \end{array}$$

Remarque. Cette disjonction veut dire que si \mathbf{x}_m est propre alors x_m est quantifié universellement dans la proposition quantifiée correspondant à l'interprétation de $\mathbf{x}_0 = \mathbf{g}(\mathbf{x}_{[1..m]})$, et quantifié existentiellement dans la proposition quantifiée correspondant à l'interprétation de $\mathbf{x}_m = \mathbf{h}(\mathbf{x}_{[m+1..m+m']})$. La situation est inversée si \mathbf{x}_m est impropre.

Définissons maintenant la fonction $s_E : \text{pro } \mathbf{x}_A \longrightarrow \text{pro } \mathbf{x}_E$ de la manière suivante :

- Si $\mathbf{x}_m \in \mathbb{I}\mathbb{R}$ alors $A' \subseteq A \cup \{m\}$, $A'' \subseteq A$ et $E = E' \cup (E'' \setminus \{m\})$. Dans ce cas, définissons $s_E(x_A)$ par $s_{E'' \setminus \{m\}}(x_A) = s''_{E'' \setminus \{m\}}(x_{A''})$ et $s_{E'}(x_A) = s'_{E'}(x_{A' \setminus \{m\}}, s''_{E''}(x_{A''}))$.
- Si $\mathbf{x}_m \in \overline{\mathbb{I}\mathbb{R}}$ alors $A'' \subset A \cup \{m\}$, $A' \subseteq A$ et $E = (E' \setminus \{m\}) \cup E''$. Dans ce cas, définissons $s_E(x_A)$ par $s_{E' \setminus \{m\}}(x_A) = s'_{E' \setminus \{m\}}(x_{A'})$ et $s_{E''}(x_A) = s''_{E''}(s'_m(x_{A'}), x_{A'' \setminus \{m\}})$.

Dans les deux cas, nous avons

$$x_E = s_E(x_A) \implies (\exists x_m \in \text{pro } \mathbf{x}_m) (x_{E'} = s'_{E'}(x_{A'}) \wedge x_{E''} = s''_{E''}(x_{A''}))$$

En effet, dans le premier cas $x_m = s''_m(x_{A''})$ et dans le second cas $x_m = s'_m(x_{A'})$. Donc, grâce aux définitions de s' et s'' , nous avons

$$x_E = s_E(x_A) \implies (\exists x_m \in \text{pro } \mathbf{x}_m) (x_0 = g(x_{[1..m]}) \wedge x_m = h(x_{[m+1..m+m']}))$$

Par définition de la fonction f ,

$$(\exists x_m \text{ pro } \mathbf{x}_m) x_0 = g(x_{[1..m]}) \wedge x_m = h(x_{[m+1..m+m']}) \implies x_0 = f(x_{[1..m-1] \cup [m+1..m+m']}).$$

Finalement, il reste à renommer les indices ($i \mapsto i$ pour $i < m$ et $i \mapsto i - 1$ pour $i > m$) pour terminer la récurrence. \square

Une conséquence directe de la proposition précédente est qu'une expression de la fonction f qui ne contient qu'une unique occurrence de chaque variable peut être utilisée pour construire une AE-extension de f :

Corollaire IV.2. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue, \mathbf{f} une expression de f mettant en jeu des fonctions élémentaires de Ω et Φ dans laquelle **chaque variable a une unique occurrence** et $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Alors l'intervalle $\mathbf{f}(\mathbf{x})$ est (f, \mathbf{x}) -interprétable.

Démonstration. Définissons $\mathbf{x}_0 = \mathbf{f}(\mathbf{x})$. La Proposition IV.28 montre qu'il existe une fonction s_E qui vérifie

$$s_E : \mathbf{x}_A \longrightarrow \mathbf{x}_E \ ; \ x_E = s_E(x_A) \implies x_0 = f(x_{[1..n]})$$

Cela implique clairement que la proposition quantifiée suivante est vraie :

$$(\forall x_A \in \mathbf{x}_A)(\exists x_E \in \mathbf{x}_E)(f(x_{[1..n]}) = x_0)$$

C'est à dire l'intervalle \mathbf{x}_0 est $(f, \mathbf{x}_{[1..n]})$ -interprétable. \square

Remarque. Contrairement au cas de extensions aux intervalles classiques, l'évaluation d'intervalles généralisés d'expressions ne contenant qu'une unique occurrence de chaque variable ne conduit pas nécessairement à l'AE-extension minimale de la fonction correspondante (cf. exemple II.6.3 de SIGLA/X[37]).

15.2 L'AE-extension naturelle (cas particulier $m = 1$)

Les deux prochains exemple illustrent la manière dont les occurrences multiples sont prise en compte. Tout d'abord, les occurrences multiples des intervalles propres (correspondant aux variables quantifiées universellement) sont prises en compte.

Exemple IV.23. Considérons la fonction $f(x, y)$ définie par $f(x, y) = x(x + y)$, $\mathbf{x} \in \mathbb{IR}$ et $\mathbf{y} \in \mathbb{IR}$. Nous voulons construire un intervalle $\mathbf{z} \in \mathbb{KR}$ qui soit $(f, \mathbf{x}, \mathbf{y})$ -interprétable, c'est à dire qui vérifie la proposition quantifiée suivante :

$$(\forall x \in \mathbf{x})(\mathbb{Q}^{(\mathbf{z})} z \in \text{pro } \mathbf{z})(\exists y \in \text{pro } \mathbf{y})(f(x, y) = z) \quad (15.2)$$

Pour pouvoir utiliser la Proposition IV.28 (ou de manière équivalente son corollaire IV.2) nous considérons la fonction auxiliaire $\tilde{f}(x_1, x_2, y)$ définie par $\tilde{f}(x_1, x_2, y) = x_1(x_2 + y)$, où chaque occurrence est rendue indépendante. Définissons maintenant $\mathbf{z} = \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{x}, \mathbf{y})$, qui est aussi égal à $\mathbf{f}(\mathbf{x}, \mathbf{y})$. Maintenant, comme \tilde{f} a une unique occurrence de chaque variable, le corollaire IV.2 prouve que \mathbf{z} est $(\tilde{f}, \mathbf{x}, \mathbf{x}, \mathbf{y})$ -interprétable. La proposition quantifiée suivante est donc vraie :

$$(\forall x_1 \in \mathbf{x})(\forall x_2 \in \mathbf{x})(\mathbb{Q}^{(\mathbf{z})} z \in \text{pro } \mathbf{z})(\exists y \in \text{pro } \mathbf{y})(\tilde{f}(x_1, x_2, y) = z)$$

Finalement, en considérant la diagonale de la boite $(\mathbf{x}, \mathbf{x})^T$ et en remarquant que

$$\tilde{f}(x, x, y) = f(x, y),$$

nous montrons aisément que la proposition quantifiée précédente implique la proposition quantifiée (15.2). Donc, \mathbf{z} est (f, \mathbf{x}) -interprétable.

Comme c'est illustré par l'exemple précédent, les occurrences multiples des intervalles propres peuvent être prise en compte dans le corollaire IV.2 sans aucune adaptation (la Proposition IV.28 pourrait aussi être généralisée au cas des occurrences multiples des intervalles propres). Cependant, comme l'illustre le prochain exemple, une technique plus compliquée doit être utilisée lorsque des intervalles impropres ont des occurrences multiples.

Exemple IV.24. *Considérons la fonction $f(x, y) = x(x + y)$, $\mathbf{x} \in \overline{\mathbb{IR}}$ et $\mathbf{y} \in \mathbb{IR}$. Nous voulons construire un intervalle $\mathbf{z} \in \mathbb{KR}$ qui soit $(f, \mathbf{x}, \mathbf{y})$ -interprétable, c'est à dire qui vérifie la proposition quantifiée suivante :*

$$(\forall y \in \mathbf{y})(\mathbb{Q}^{(\mathbf{z})} z \in \mathbf{z})(\exists x \in \text{pro } \mathbf{x})(f(x, y) = z) \quad (15.3)$$

Encore une fois, nous considérons la fonction $\tilde{f}(x_1, x_2, y) = x_1(x_2 + y)$. Définissons naïvement $\mathbf{z} = \tilde{f}(\mathbf{x}, \mathbf{x}, \mathbf{y})$. Comme \tilde{f} a une unique occurrence de chaque variable, le corollaire IV.2 montre que \mathbf{z} est $(\tilde{f}, \mathbf{x}, \mathbf{x}, \mathbf{y})$ -interprétable. La proposition quantifiée suivante est donc vraie :

$$(\forall y \in \mathbf{y})(\mathbb{Q}^{(\mathbf{z})} z \in \mathbf{z})(\exists x_1 \in \text{pro } \mathbf{x})(\exists x_2 \in \text{pro } \mathbf{x})(\tilde{f}(x_1, x_2, y) = z)$$

Cependant, cette fois la proposition précédente n'implique pas en générale la proposition quantifiée (15.3) (les deux occurrences quantifiées existentiellement ne peuvent plus être mise en relation) et l'intervalle \mathbf{z} n'est donc pas $(f, \mathbf{x}, \mathbf{y})$ -interprétable en général. Pour pouvoir valider la proposition (15.3), nous considérons l'évaluation modifiée suivante : $\mathbf{z}' = (\text{pro } \mathbf{x})(\mathbf{x} + \mathbf{y})$ (l'évaluation $\mathbf{z}' = \mathbf{x}(\text{pro } \mathbf{x} + \mathbf{y})$ pourrait être considérée de manière équivalente). Maintenant, la Proposition IV.28 montre que l'intervalle \mathbf{z}' est $(\tilde{f}, \text{pro } \mathbf{x}, \mathbf{x}, \mathbf{y})$ -interprétable, c'est à dire la proposition quantifiée suivante est vraie :

$$(\forall x_1 \in \text{pro } \mathbf{x})(\forall y \in \mathbf{y})(\mathbb{Q}^{(\mathbf{z}')} z \in \mathbf{z}')(\exists x_2 \in \text{pro } \mathbf{x})(\tilde{f}(x_1, x_2, y) = z)$$

Finalemnt, la proposition quantifiée précédente implique la proposition quantifiée (15.3). Cette implication est démontrée de la manière suivante : supposons que $\mathbf{z}' \in \mathbb{IR}$, i.e. $\mathbb{Q}^{(\mathbf{z}')} = \exists$ (l'autre cas étant similaire). La Proposition IV.28 prouve qu'il existe une fonction continue $\tilde{s} = (\tilde{s}_z, \tilde{s}_2)^T : (\text{pro } \mathbf{x}, \mathbf{y})^T \rightarrow (\mathbf{z}', \text{pro } \mathbf{x})^T$ qui vérifie

$$(z, x_2)^T = \tilde{s}(x_1, y) \implies z = \tilde{f}(x_1, x_2, y)$$

Considérons la composante $\tilde{s}_2 : (\text{pro } \mathbf{x}, \mathbf{y})^T \rightarrow \text{pro } \mathbf{x}$ de \tilde{s} . Pour chaque valeur de y , la fonction $\tilde{s}_2(\cdot, y) : \text{pro } \mathbf{x} \rightarrow \text{pro } \mathbf{x}$ a un point fixe par le théorème de Brouwer. Ainsi, la proposition quantifiée suivante est vraie :

$$(\forall y \in \mathbf{y})(\exists x \in \mathbf{x})(\tilde{s}_2(x, y) = x).$$

Ce qui implique

$$(\forall y \in \mathbf{y})(\exists x \in \mathbf{x})(\exists z \in \mathbf{z}')(s_2(x, y) = x \wedge z = \tilde{s}_z(x, y)).$$

Et finalement, par construction de la fonction \tilde{s}

$$(\forall y \in \mathbf{y})(\exists x \in \mathbf{x})(\exists z \in \mathbf{z}')(z = f(x, y)).$$

Ce qui prouve que $\mathbf{z}' = (\text{pro } \mathbf{x})(\mathbf{x} + \mathbf{y})$ est $(f, \mathbf{x}, \mathbf{y})$ -interprétable. Dans le cas général, l'enchaînement des trois dernières propositions quantifiées sera fait en utilisant le corollaire VI.1 du théorème de Brouwer (page 201).

Finalemnt, les AE-extensions naturelles de la fonction $f(x, y) = x(x + y)$ sont

$$(\text{pro } \mathbf{x})(\mathbf{x} + \mathbf{y}) \text{ et } \mathbf{x}(\text{pro } \mathbf{x} + \mathbf{y})$$

Ainsi, si \mathbf{x} est propre alors $\text{pro } \mathbf{x} = \mathbf{x}$ et l'argumentation présentée dans l'Exemple IV.23 reste valable. Si \mathbf{x} est impropre, l'argumentation présentée dans l'Exemple IV.24 est aussi valable. Donc, quelque soient \mathbf{x} et \mathbf{y} , les intervalles $(\text{pro } \mathbf{x})(\mathbf{x} + \mathbf{y})$ et $\mathbf{x}(\text{pro } \mathbf{x} + \mathbf{y})$ sont $(f, \mathbf{x}, \mathbf{y})$ -interprétables. Ces idées sont formalisées dans le théorème suivant.

Théorème IV.2. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction continue et \mathbf{f} une expression de f mettant en jeux des fonctions élémentaires de Ω et Φ . Définissons l'expression \mathbf{g} à partir de \mathbf{f} en insérant une opération pro devant toutes les occurrences sauf une de chaque variable (cf. Exemple IV.25 page 166). Alors la fonction d'intervalles $\mathbf{g} : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{R}$ définie par $\mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{x})$ est une AE-extension de f , appelée une AE-extension naturelle.

Démonstration. Remarquons d'abord que $\mathbf{g}(x) = f(x)$ pour tout $x \in \mathbb{R}^n$ (car $\text{pro } x = x$). Il reste à prouver que quelque soit $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ l'intervalle $\mathbf{g}(\mathbf{x})$ est (f, \mathbf{x}) -interprétable.

Remarque. La preuve proposée reprend les deux exemples précédents, la principale difficulté étant la gestion des indices des variables dans le cas général.

Considérons un $\mathbf{x}_{[1..n]} \in \mathbb{K}\mathbb{R}^n$ quelconque, nous avons à démontrer que la proposition quantifiée suivante est vraie :

$$(\forall x_A \in \mathbf{x}_A)(\exists x_E \in \mathbf{x}_E)(f(x_{[1..n]} = x_0)), \quad (15.4)$$

où $\mathbf{x}_0 = \mathbf{g}(\mathbf{x}_{[1..n]})$, $A = P(\mathbf{x}_{[1..n]}) \cup I(\mathbf{x}_0)$ et $E = I(\mathbf{x}_{[1..n]}) \cup P(\mathbf{x}_0)$. Nous supposons que $A \neq \emptyset$, l'autre cas étant similaire et plus simple. Définissons l'expression $\tilde{\mathbf{f}}$ à partir de \mathbf{f} en rendant indépendantes toutes les occurrences de chaque variable. Notons n_i le nombre d'occurrences de la variable x_i et x_{ij} la $j^{\text{ième}}$ occurrence de cette variable. Sans perdre en généralité, nous pouvons supposer que l'unique occurrence qui n'est pas précédée de l'opération pro dans l'expression \mathbf{g} est la première, i.e. x_{i1} . Définissons

$$K = \{(i, j) \in \mathbb{N}^2 \mid i \in [1..n], j \in [1..n_i]\}$$

et $\mathbf{x}_K \in \mathbb{K}\mathbb{R}^{\sum n_i}$ par $\mathbf{x}_{ij} = \text{pro } \mathbf{x}_i$ quelque soit $(i, j) \in K$ tel que $j \neq 1$ et $\mathbf{x}_{ij} = \mathbf{x}_i$ sinon. En conséquence, nous avons l'égalité suivante :

$$\mathbf{g}(\mathbf{x}_{[1..n]}) = \tilde{\mathbf{f}}(\mathbf{x}_K).$$

Définissons $\mathbf{x}_{01} = \mathbf{x}_0$ et les ensembles d'indices suivants : $K_A := P(\mathbf{x}_K) \cup I(\mathbf{x}_{01})$, $K_E := I(\mathbf{x}_K) \cup P(\mathbf{x}_{01})$ et $A' := \{i \mid (\exists j \in \mathbb{N}) (i, j) \in K_A\}$.

Exemple. Considérons $f(x_{[1..2]}) = \mathbf{f}(x_{[1..2]}) = x_1 x_2 + x_1(x_1 + x_2)$, $\mathbf{x}_1 \in \overline{\mathbb{I}\mathbb{R}}$ et $\mathbf{x}_2 \in \mathbb{I}\mathbb{R}$. Alors,

$$\tilde{\mathbf{f}}(x_{11}, x_{12}, x_{13}, x_{21}, x_{22}) = x_{11}x_{21} + x_{12}(x_{13} + x_{22}).$$

L'AE-extension naturelle est

$$\mathbf{x}_0 = \mathbf{g}(\mathbf{x}_{[1..2]}) = \mathbf{x}_1 \mathbf{x}_2 + (\text{pro } \mathbf{x}_1)(\text{pro } \mathbf{x}_1 + \text{pro } \mathbf{x}_2)$$

Supposons que $\mathbf{x}_0 \in \overline{\mathbb{I}\mathbb{R}}$. Alors $A = \{0, 2\}$ et $E = \{1\}$. De plus,

- $K = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2)\}$
- $\mathbf{x}_K = (\mathbf{x}_{11}, \mathbf{x}_{12}, \mathbf{x}_{13}, \mathbf{x}_{21}, \mathbf{x}_{22})^T = (\mathbf{x}_1, \text{pro } \mathbf{x}_1, \mathbf{x}_2, \text{pro } \mathbf{x}_2, \text{pro } \mathbf{x}_2)^T$
- $K_A = \{(0, 1), (1, 2), (1, 3), (2, 1), (2, 2)\}$
- $K_E = \{(1, 1)\}$
- $A' = \{0, 1, 2\}$

Finalement, nous avons $\mathbf{g}(\mathbf{x}_{[1..2]}) = \tilde{\mathbf{f}}(\mathbf{x}_K)$ et nous voulons montrer que

$$(\forall x_{K_A} \in \mathbf{x}_{K_A})(\exists x_{K_E} \in \mathbf{x}_{K_E})(\tilde{\mathbf{f}}(x_K) = x_{01})$$

implique

$$(\forall x_A \in \mathbf{x}_A)(\exists x_E \in \mathbf{x}_E)(f(x_{[1..n]} = x_0)).$$

Nous avons

$$K_A \cup K_E = K \cup \{(0\ 1)\}, \quad K_A \cap K_E = \emptyset \text{ et } A' = \{i \in \mathbb{N} \mid (\exists j \in \mathbb{N})((i\ j) \in K_A)\}.$$

De plus, par définition de \mathbf{x}_K nous avons $K_E = \{(i\ 1) \mid i \in E\}$. Maintenant, comme \tilde{f} est composée de fonctions élémentaires de Ω et Φ et comme chaque variable x_{ij} , $(i, j) \in K$, a une unique occurrence dans \tilde{f} , nous pouvons appliquer la Proposition IV.28 qui prouve d'une part que K_A et K_E sont non-vides et d'autre part qu'il existe une fonction continue $\tilde{s}_{K_E} : \text{pro } \mathbf{x}_{K_A} \longrightarrow \text{pro } \mathbf{x}_{K_E}$ qui vérifie

$$x_{K_E} = \tilde{s}_{K_E}(x_{K_A}) \implies x_{01} = \tilde{f}(x_K) \quad (15.5)$$

Première étape. Les occurrences multiples des occurrences quantifiées universellement sont identifiées en utilisant le raisonnement présenté dans l'Exemple IV.23.

Définissons $\tilde{s}'_E : \text{pro } \mathbf{x}_{A'} \longrightarrow \text{pro } \mathbf{x}_E$ de la manière suivante :

$$\tilde{s}'_E(x_{A'}) = \tilde{s}_{K_E}(\sigma_{K_A}(x_{A'})) \quad (15.6)$$

où σ_{K_A} est définie par $\sigma_{K_A}(x_{A'}) = x_{K_A}$ avec x_{K_A} défini par $x_{ij} = x_i$ pour tout $(i, j) \in K_A$ (qui est bien définie car $A' = \{i \in \mathbb{N} \mid (\exists j \in \mathbb{N})((i\ j) \in K_A)\}$). La fonction $\tilde{s}'_E(x_{A'})$ est continue car elle est composée de fonctions continues.

Deuxième étape. Les occurrences multiples des variables quantifiées existentiellement sont maintenant reliées en utilisant le raisonnement présenté dans l'Exemple IV.24.

Nous avons $A' \neq \emptyset$ parce que A est supposé être non-vide et $E \neq \emptyset$ parce que K_E est non-vide. Le corollaire VI.1 du théorème de Brouwer (page 201) implique donc l'existence d'une fonction $s_E : \text{pro } \mathbf{x}_A \longrightarrow \text{pro } \mathbf{x}_E$ qui vérifie

$$x_E = s_E(x_A) \implies \tilde{s}'_E(x_{A'}) = x_E \quad (15.7)$$

Prouvons finalement que

$$s_E(x_A) = x_E \implies x_0 = f(x_{[1..n]}) \quad (15.8)$$

En effet, $s_E(x_A) = x_E$ implique $\tilde{s}'_E(x_{A'}) = x_E$ (par l'équation (15.7)) qui implique $\tilde{s}_{K_E}(\sigma_{K_A}(x_{A'})) = x_E$ (par l'équation (15.6)). Définissons $x_K \in \mathbb{R}^{\sum n_i}$ par $x_{K_A} = \sigma_{K_A}(x_{A'})$ et $x_{K_E} = x_E$. Nous avons alors $x_{K_E} = \tilde{s}_{K_E}(x_{K_A})$ et $x_{ij} = x_i$ pour tout $(ij) \in K$. Cela veut dire que toutes les occurrences de x_i ont la même valeur. Nous avons donc $\tilde{f}(x_K) = f(x_{[1..n]})$. Donc $x_{K_E} = \tilde{s}_{K_E}(x_{K_A})$ implique $x_{01} = \tilde{f}(x_K)$ (par l'équation (15.5)) ce qui implique $x_0 = f(x_{[1..n]})$. Finalement, l'existence d'une fonction s_E qui vérifie (15.8) implique que la proposition (15.4) soit vraie. \square

Remarques.

1. Comme dans le cas de l'extension naturelle classique, si l'expression \mathbf{g} n'est pas définie pour tous les intervalles $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ alors le domaine de définition de l'AE-extension naturelle doit être adapté. On peut remarquer que si l'extension naturelle classique est définie pour $\text{pro } \mathbf{x}$ alors l'AE-extension naturelle est définie pour \mathbf{x} .
2. Lorsque tous les intervalles mis en jeu sont propres, les opérations pro n'ont aucune influence et toutes les AE-extensions naturelles d'une fonction coïncident avec l'extension naturelle des intervalles classiques.

Étant donné une expression pour une fonction f , il existe de nombres AE-extensions naturelles, en fonction de la manière dont les opérations pro sont insérées dans l'expression.

Exemple IV.25. *Considérons la fonction $f(x, y) = xy + x(x + y)$. Les fonctions suivantes sont les différentes AE-extensions naturelles de f :*

$$\begin{aligned} \mathbf{xy} + (\text{pro } \mathbf{x})(\text{pro } \mathbf{x} + \text{pro } \mathbf{y}) & ; \quad \mathbf{x}(\text{pro } \mathbf{y}) + (\text{pro } \mathbf{x})(\text{pro } \mathbf{x} + \mathbf{y}) \\ (\text{pro } \mathbf{x})\mathbf{y} + \mathbf{x}(\text{pro } \mathbf{x} + \text{pro } \mathbf{y}) & ; \quad (\text{pro } \mathbf{x})(\text{pro } \mathbf{y}) + \mathbf{x}(\text{pro } \mathbf{x} + \mathbf{y}) \\ (\text{pro } \mathbf{x})\mathbf{y} + (\text{pro } \mathbf{x})(\mathbf{x} + \text{pro } \mathbf{y}) & ; \quad (\text{pro } \mathbf{x})(\text{pro } \mathbf{y}) + (\text{pro } \mathbf{x})(\mathbf{x} + \mathbf{y}) \end{aligned}$$

Les différentes AE-extensions naturelles ne donnent pas le même résultat en général. Dans le cas particulier des fonctions à valeurs réelles (qui ont une unique AE-extension minimale) la Proposition IV.12 montre que l'on peut intersecter les différents résultats obtenus pour chaque AE-extension naturelle pour en obtenir un plus précis que chacun des originaux. Néanmoins, le nombre d'AE-extensions naturelles possible croît exponentiellement avec le nombre d'occurrences total des variables et un algorithme permettant de décider laquelle serait la plus précise est encore à trouver. Quoiqu'il en soit, comme dans le cas de l'extension naturelle classique, un nombre important d'occurrences des variables rend en général le résultat de l'AE-extension naturelle trop pessimiste pour être utile.

15.3 L'AE-extension naturelle (cas général)

Nous construisons maintenant une AE-extension naturelle d'une fonction à valeur vectorielle $f = (f_1, \dots, f_n)$ en utilisant l'évaluation d'intervalles des expressions des composantes f_i de f . Commençons par étudier un exemple. Considérons une fonction à valeurs vectorielles $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ et trois intervalles $\mathbf{u} \in \mathbb{IR}$, $\mathbf{v} \in \mathbb{IR}$ et $\mathbf{w} \in \mathbb{IR}$. Nous voulons maintenant construire un intervalle $\mathbf{z} \in \mathbb{KR}^2$ ($f, \mathbf{u}, \mathbf{v}, \mathbf{w}$)-interprétable en utilisant des AE-extensions naturelles \mathbf{g}_1 et \mathbf{g}_2 des fonctions à valeurs réelles f_1 and f_2 . Pour simplifier cet exemple, considérons que $\mathbf{z} \in \mathbb{IR}^2$. Ainsi, la proposition quantifiée suivante doit être validée :

$$(\forall u \in \mathbf{u})(\forall z_1 \in \text{pro } \mathbf{z}_1)(\forall z_2 \in \text{pro } \mathbf{z}_2)(\exists v \in \text{pro } \mathbf{v})(\exists w \in \text{pro } \mathbf{w})(f(u, v, w) = z) \quad (15.9)$$

Considérons l'intervalle $\mathbf{z} \in \mathbb{KR}^2$ défini naïvement par

$$\mathbf{z}_1 = \mathbf{g}_1(\mathbf{u}, \mathbf{v}, \mathbf{w}) ; \quad \mathbf{z}_2 = \mathbf{g}_2(\mathbf{u}, \mathbf{v}, \mathbf{w}).$$

Supposons que $\mathbf{z} \in \mathbb{IR}^2$. Comme \mathbf{g}_1 et \mathbf{g}_2 sont des AE-extensions de f_1 et f_2 , les deux propositions quantifiées suivantes sont vraies :

$$(\forall u \in \mathbf{u})(\forall z_1 \in \text{pro } \mathbf{z}_1)(\exists v \in \text{pro } \mathbf{v})(\exists w \in \text{pro } \mathbf{w})(f_1(u, v, w) = z_1)$$

et

$$(\forall u \in \mathbf{u})(\forall z_2 \in \text{pro } \mathbf{z}_2)(\exists v \in \text{pro } \mathbf{v})(\exists w \in \text{pro } \mathbf{w})(f_2(u, v, w) = z_2).$$

Cependant, la conjonction de ces deux propositions quantifiées n'implique pas en général la proposition quantifiée (15.9) (comme dans l'Exemple IV.24 les occurrences quantifiées existentiellement ne peuvent pas être en relation). De manière similaire à l'AE-extension

pour fonctions à valeurs réelles, l'utilisation des AE-extensions \mathbf{g}_1 et \mathbf{g}_2 est modifiée de la manière suivante :

$$\mathbf{z}'_1 = \mathbf{g}_1(\mathbf{u}, \text{pro } \mathbf{v}, \mathbf{w}) \quad ; \quad \mathbf{z}'_2 = \mathbf{g}_1(\text{pro } \mathbf{u}, \mathbf{v}, \text{pro } \mathbf{w})$$

Comme \mathbf{u} est propre, l'opération pro n'a pas d'influence sur lui. Supposons que $\mathbf{z}' \in \overline{\mathbb{R}^2}$. Alors les deux propositions quantifiées suivantes sont vraies :

$$(\forall u \in \mathbf{u})(\forall z_1 \in \text{pro } \mathbf{z}'_1)(\forall v \in \text{pro } \mathbf{v})(\exists w \in \text{pro } \mathbf{w})(f_1(u, v, w) = z_1)$$

et

$$(\forall u \in \mathbf{u})(\forall z_2 \in \text{pro } \mathbf{z}'_2)(\forall w \in \text{pro } \mathbf{w})(\exists v \in \text{pro } \mathbf{v})(f_2(u, v, w) = z_2)$$

Avec une hypothèse supplémentaire sur \mathbf{g}_1 et \mathbf{g}_2 la conjonction de ces deux propositions quantifiées implique la proposition quantifiée (15.9). Cette hypothèse supplémentaire est satisfaite si les AE-extensions \mathbf{g}_1 et \mathbf{g}_2 sont des AE-extensions naturelles. **Cependant, toute les AE-extensions ne peuvent pas être utilisées dans ce raisonnement** : des contres exemples peuvent être obtenus en utilisant par exemple les AE-extensions minimales f_1^* et f_2^* (cf. Exemple IV.27 page 168). Le théorème IV.3 précise de quelle manière ce raisonnement peut être utiliser pour construire une AE-extension naturelle pour les fonctions à valeurs vectorielles.

Remarque. Cette technique pour construire des AE-extensions de fonctions à valeurs vectorielles a été initialement proposée dans la théorie des intervalles modaux (SIGLA/X[37]) et a été reprise dans Ratschan[76]. Dans les deux cas, les preuves proposées ne sont pas correctes (Goldsztejn[3] et Schodl[71]).

Le théorème suivant est une généralisation du Théorème IV.2 qui prend en compte les fonctions à valeurs vectorielles. Il généralise l'argumentation qui vient d'être présentée.

Théorème IV.3. Soient $f_{[1..m]} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continue et $\mathbf{f}_{[1..m]}$ une expression de $f_{[1..m]}$, i.e. \mathbf{f}_i est une expression de f_i , mettant en jeux des fonctions élémentaires de Ω et Φ . Définissons l'expression $\mathbf{g}_{[1..m]}$ à partir de $\mathbf{f}_{[1..m]}$ en insérant une opération pro devant toutes les occurrences sauf une (cf. Exemple IV.26). Alors la fonction d'intervalles $\mathbf{g} : \mathbb{K}\mathbb{R}^n \longrightarrow \mathbb{K}\mathbb{R}^m$ définie par $\mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{x})$ est une AE-extension de f , appelée une AE-extension naturelle de f .

Démonstration. Premièrement, $\mathbf{g}(x) = f(x)$ pour tout $x \in \mathbb{R}^n$. Il reste à prouver que $\mathbf{g}(\mathbf{x})$ est (f, \mathbf{x}) -interprétable pour tout $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$. Considérons $\mathbf{x}_{[1..n]} \in \mathbb{K}\mathbb{R}^n$ quelconque, nous devons prouver que la proposition quantifiée suivante est vraie :

$$(\forall x_A \in \mathbf{x}_A)(\exists x_E \in \mathbf{x}_E)(f(x_{[1..n]}) = x_{[n+1..n+m]}) \quad (15.10)$$

où $\mathbf{x}_{[n+1..n+m]} = \mathbf{g}(\mathbf{x}_{[1..n]})$, $A = P(\mathbf{x}_{[1..n]}) \cup I(\mathbf{x}_{[n+1..n+m]})$ et $E = I(\mathbf{x}_{[1..n]}) \cup P(\mathbf{x}_{[n+1..n+m]})$. Nous supposons que $A \neq \emptyset$, l'autre cas étant similaire et plus simple. Définissons la fonction $\pi : E \longrightarrow [1..n]$ qui associe à chaque variable existentiellement quantifiée l'indice de la composante de f où l'une de ses occurrence n'est pas précédée par l'opération pro (il s'agit bien d'une fonction par construction de l'expression \mathbf{g}). Définissons $E_k = \pi^{-1}(k)$, qui forment une partition de E , et $A_k = A \cup E'_k$ où $E'_k \subseteq E$ est l'ensemble des indices des variables x_i , $i \in E$, qui ont plus d'une occurrence dans l'expression \mathbf{f}_k . Maintenant,

en suivant la preuve du Théorème IV.2 appliquée pour chaque $k \in [1..m]$, nous obtenons m fonctions continues $\tilde{s}'_{E_k}(x_{A_k})$ qui satisfont

$$x_{E_k} = \tilde{s}'_{E_k}(x_{A_k}) \implies x_{n+k} = f_k(x_{[1..n]}) \quad (15.11)$$

De plus, pour chaque $k \in [1..m]$, A_k et E_k sont non-vides (donc E contient au moins m éléments). Maintenant, les E_k formant une partition de E , la fonction $\tilde{s}'_E : \mathbf{x}_{A'} \longrightarrow \mathbf{x}_E$, où $A' = \bigcup_{k \in [1..m]} A_k$, est bien définie. Par construction, nous avons $A = A' \setminus E$, qui est non-vide par supposition. Nous pouvons donc appliquer le corollaire VI.1 du théorème de Brouwer (page 201) qui prouve l'existence d'une fonction $s_E : \mathbf{x}_A \longrightarrow \mathbf{x}_E$ satisfaisant

$$x_E = s_E(x_A) \implies x_E = \tilde{s}'_E(x_{A'})$$

Finalement, $x_E = \tilde{s}'_E(x_{A'})$ implique $\forall k \in [1..m], x_{E_k} = \tilde{s}'_{E_k}(x_{A_k})$. Par l'équation (15.11), cela implique $\forall k \in [1..m], x_{n+k} = f_k(x_{[1..n]})$, c'est à dire $x_{[n+1..n+m]} = f(x_{[1..n]})$. Finalement, $x_E = s_E(x_A)$ implique $x_{[n+1..n+m]} = f(x_{[1..n]})$ et l'existence d'une telle fonction s_E implique la proposition quantifiée (15.10). \square

Comme dans le cas des fonctions à valeurs réelles, il existe autant d'AE-extensions naturelles qu'il y a manières d'insérer les opérations pro dans l'expression.

Exemple IV.26. *Considérons la fonction $f(x, y) = (xy, x(x+y))^T$. Ses AE-extensions naturelles sont*

$$\begin{aligned} & (\mathbf{xy}, (\text{pro } \mathbf{x})(\text{pro } \mathbf{x} + \text{pro } \mathbf{y}))^T ; & (\mathbf{x}(\text{pro } \mathbf{y}), (\text{pro } \mathbf{x})(\text{pro } \mathbf{x} + \mathbf{y}))^T \\ & ((\text{pro } \mathbf{x})\mathbf{y}, \mathbf{x}(\text{pro } \mathbf{x} + \text{pro } \mathbf{y}))^T ; & ((\text{pro } \mathbf{x})(\text{pro } \mathbf{y}), \mathbf{x}(\text{pro } \mathbf{x} + \mathbf{y}))^T \\ & ((\text{pro } \mathbf{x})\mathbf{y}, (\text{pro } \mathbf{x})(\mathbf{x} + \text{pro } \mathbf{y}))^T ; & ((\text{pro } \mathbf{x})(\text{pro } \mathbf{y}), (\text{pro } \mathbf{x})(\mathbf{x} + \mathbf{y}))^T \end{aligned}$$

L'exemple suivant montre que pour construire une AE-extension d'une fonction à valeurs vectorielles, on ne peut pas utiliser n'importe quelle AE-extension pour les composantes de la fonction à valeur vectorielle.

Exemple IV.27. *Considérons la fonction*

$$f(x, y) = \begin{pmatrix} 1 - (x - y)^2 \\ x - y \end{pmatrix}$$

et les intervalles $\mathbf{x} = [1, -1]$ et $\mathbf{y} = [1, -1]$. Nous calculons maintenant l'intervalle

$$\mathbf{z} = \begin{pmatrix} f_1^*(\text{pro } \mathbf{x}, \mathbf{y}) \\ f_2^*(\mathbf{x}, \text{pro } \mathbf{y}) \end{pmatrix}$$

Comme f_2 est linéaire, nous avons $\mathbf{z}_2 = f_2^*(\mathbf{x}, \text{pro } \mathbf{y}) = \mathbf{x} - \text{pro } \mathbf{y} = [0, 0]$ (voir Théorème IV.5 page 169) et la proposition quantifiée suivante est vraie :

$$(\forall y \in \text{pro } \mathbf{y})(\exists x \in \text{pro } \mathbf{x})(f_2(x, y) = 0)$$

Maintenant, calculons $\mathbf{z}_1 = f_1^*(\text{pro } \mathbf{x}, \mathbf{y})$ en utilisant l'expression de f_1^* :

$$\begin{aligned} f_1^*(\text{pro } \mathbf{x}, \mathbf{y}) &= \bigvee_{x \in \text{pro } \mathbf{x}} \text{dual range}(f_1, x, \text{pro } \mathbf{y}) \\ &= \bigvee_{x \in \text{pro } \mathbf{x}} [1, 1 - (1 + |x|)^2] \\ &= [1, \max_{x \in \text{pro } \mathbf{x}} 1 - (1 + |x|)^2] \\ &= [1, 0] \end{aligned}$$

Donc la proposition quantifiée suivante est vraie :

$$(\forall z_1 \in [0, 1]) (\forall x \in \text{pro } \mathbf{x}) (\exists y \in \text{pro } \mathbf{y}) (f_1(x, y) = z_1)$$

En suivant l'argumentation proposée au début de la section, nous concluons que les deux propositions quantifiées précédentes impliquent la suivante :

$$(\forall z_1 \in [0, 1]) (\forall z_2 \in [0, 0]) (\exists y \in \text{pro } \mathbf{y}) (\exists x \in \text{pro } \mathbf{x}) (f(x, y) = z).$$

Cependant, cette proposition quantifiée est fautive : elle implique en effet que l'équation $f(x, y) = 0$ a une solution dans $(\mathbf{x}, \mathbf{y})^T$ alors que $f_2(x, y) = 0$ implique $x = y$ ce qui est incompatible avec $f_1(x, y) = 0$.

Remarque. Les articles décrivant la théorie des intervalles modaux n'insistent pas sur l'existence de situations comme celle présentée dans l'exemple précédent. En particulier, beaucoup de précautions doivent être prises pour utiliser les résultats de la section III.4 de SIGLA/X[37].

15.4 Qualité de l'AE-extension naturelle

Deux résultats théoriques sur la qualité de l'AE-extension naturelle sont maintenant présentés. Premièrement, son ordre de convergence est précisé par le prochain théorème.

Lemme IV.6. *L'opération $\text{pro} : \mathbb{K}\mathbb{R} \longrightarrow \mathbb{K}\mathbb{R}$ est Lipschitz continue, et donc localement Lipschitz continue.*

Démonstration. La preuve est triviale. □

Théorème IV.4. *Lorsque les fonctions $\sqrt[n]{x}$ sont restreinte au réels strictement positifs, l'AE-extension naturelle a un ordre de convergence linéaire.*

Démonstration. Dans ce cas, l'AE-extension naturelle est composée d'une part des AE-extensions minimales des fonctions élémentaires qui sont localement Lipschitz continues (Kaucher[21]) et d'autre part d'opérations pro qui sont aussi localement Lipschitz continues. L'AE-extension naturelle est donc localement Lipschitz continue. Cela implique finalement un ordre de convergence linéaire par la Proposition IV.15. □

Remarque. Le cas où les fonctions $\sqrt[n]{x}$ sont utilisées en 0 pose le même problème que celui décrit dans le cadre des extensions aux intervalles classiques (cf. section 2.6 page 34).

Finalement, la proposition suivante montre que l'AE-extension naturelle des fonctions bilinéaire est minimale.

Théorème IV.5. *Considérons la fonction bilinéaire*

$$f : \mathbb{R}^{2n} \longrightarrow \mathbb{R} ; (x, y) \longmapsto \sum_{i \in [1..n]} x_i y_i$$

L'AE-extension naturelle de f est minimale.

Démonstration. Comme \times commute avec l'opération dual, la Proposition IV.19 montre que f^* commute aussi avec dual. De plus, l'AE-extension naturelle de f ne contient d'opération pro (parce que chaque variable a une unique occurrence dans l'expression de f) et commute donc avec l'opération dual (parce que $+$ et \times commutent avec dual). Finalement, la Proposition IV.18 montre que cette AE-extension naturelle est minimale. \square

Chapitre 16

L'AE-extension de la valeur moyenne

Ce chapitre présente l'AE-extension de la valeur moyenne. Cette AE-extension n'a pas de correspondante dans le théorie des intervalles modaux : elle est obtenue grâce au cadre théorique de la nouvelle formulation proposé ici. Son expression est similaire à l'extension de la valeur moyenne classique et elle présente des avantages similaires comparée à l'AE-extension naturelle.

16.1 Généralisation du théorème de la valeur moyenne

Cette section présente une généralisation du théorème de la valeur moyenne. Le théorème de la valeur moyenne permet de prouver qu'il existe un vecteur $c \in \text{range}(f', \mathbf{x})$ tel que

$$f(x) = f(\tilde{x}) + c(x - \tilde{x})$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est continûment dérivable et $x, \tilde{x} \in \mathbf{x} \in \mathbb{I}\mathbb{R}^n$. Cela s'exprime de manière équivalente en disant qu'il existe une fonction $s : \mathbf{x} \rightarrow \Delta \supseteq \text{range}(f', \mathbf{x})$ qui vérifie

$$f(x) = f(\tilde{x}) + s(x)(x - \tilde{x})$$

La généralisation proposée dans cette section indique qu'il existe une fonction s continue vérifiant la propriété précédente.

Proposition IV.29. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continûment dérivable, $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$, $\tilde{x} \in \mathbf{x}$ et $\Delta \in \mathbb{I}\mathbb{R}^n$ tel que pour tout $k \in [1..n]$,*

$$\Delta_k \supseteq \text{range} \left(\frac{\partial f}{\partial x_k}, \mathbf{x}_1, \dots, \mathbf{x}_k, \tilde{x}_{k+1}, \dots, \tilde{x}_n \right)$$

Alors il existe une fonction continue $s : \mathbf{x} \rightarrow \Delta$ qui vérifie

$$f(x) = f(\tilde{x}) + s(x)^T \times (x - \tilde{x}) \tag{16.1}$$

Démonstration. Écrivons $f(x)$ de la manière suivante :

$$f(x) = f(\tilde{x}) + \sum_{k \in [1..n]} g_k(x) \text{ avec } g_k(x) = f(x_{[1..k]}, \tilde{x}_{[k+1..n]}) - f(x_{[1..k-1]}, \tilde{x}_{[k..n]}).$$

Remarques.

1. La fonction g_k ne dépend que des variable x_i pour $i \in [1..k]$.
2. Pour obtenir des notations lisibles et compactes, nous autorisons à noter l'expression $g_n(x) = f(x_{[1..n]}) - f(x_{[1..n-1]}, \tilde{x}_n)$ par $g_n(x) = f(x_{[1..n]}, \tilde{x}_{[n+1..n]}) - f(x_{[1..n-1]}, \tilde{x}_{[n..n]})$. Le vecteur $\tilde{x}_{[n+1..n]}$ n'est pas pris en compte.

Exemple. Pour $n = 2$, l'expression précédente s'écrit

$$f(x) = f(\tilde{x}) + (f(x_1, \tilde{x}_2) - f(\tilde{x})) + (f(x) - f(x_1, \tilde{x}_2)).$$

Définissons maintenant $s_k(x)$ de la manière suivante :

$$s_k(x) = \frac{g_k(x)}{x_k - \tilde{x}_k} \text{ si } x_k \neq \tilde{x}_k \text{ et } s_k(x) = \frac{\partial f}{\partial x_k}(x_{[1..k-1]}, \tilde{x}_{[k..n]}) \text{ sinon}$$

Ainsi $g_k(x) = s_k(x)(x_k - \tilde{x}_k)$ quelque soit $x \in \mathbf{x}$. En effet, si $x_k \neq \tilde{x}_k$ alors

$$s_k(x)(x_k - \tilde{x}_k) = \frac{g_k(x)}{x_k - \tilde{x}_k}(x_k - \tilde{x}_k) = g_k(x).$$

Si $x_k = \tilde{x}_k$ alors $g_k(x) = 0 = s_k(x) \times 0$. Ainsi, (16.1) est vérifiée. Prouvons maintenant que s_k est continue dans \mathbf{x} . D'une part, si $x_k \neq \tilde{x}_k$ alors s_k est la composition de fonctions continues. Il reste à montrer que s_k est continue en x si $x_k = \tilde{x}_k$. Considérons une suite $x^{(i)}$ qui tendent vers x . Nous avons alors $s_k(x) = \frac{\partial f}{\partial x_k}(x_{[1..k-1]}, \tilde{x}_{[k..n]})$ et

$$s_k(x^{(i)}) = \frac{f(x_{[1..k]}^{(i)}, \tilde{x}_{[k+1..n]}) - f(x_{[1..k-1]}^{(i)}, \tilde{x}_{[k..n]})}{x_k^{(i)} - \tilde{x}_k} = \frac{\partial f}{\partial x_k}(x_{[1..k-1]}^{(i)}, \xi_k^{(i)}, \tilde{x}_{[k+1..n]})$$

avec $x_k^{(i)} \in x_k^{(i)} \vee \tilde{x}_k$ (par le théorème de la valeur moyenne). Comme $x_k^{(i)}$ tend vers \tilde{x}_k , la suite $\xi_k^{(i)}$ tend aussi vers \tilde{x}_k . Ainsi, $s_k(x^{(i)})$ tend vers $s_k(x)$ car $\frac{\partial f}{\partial x_k}$ est continue. Il reste finalement à vérifier que s_k est à valeurs dans Δ_k . D'une part, si $x_k \neq \tilde{x}_k$ le théorème de la valeur moyenne montre que quelque soit $x_{[1..k]} \in \mathbf{x}_{[1..k]}$,

$$\begin{aligned} \frac{f(x_{[1..k]}, \tilde{x}_{[k+1..n]}) - f(x_{[1..k-1]}, \tilde{x}_{[k..n]})}{x_k - \tilde{x}_k} &\in \text{range} \left(\frac{\partial f}{\partial x_k}, x_1, \dots, x_{k-1}, x_k \vee \tilde{x}_k, \tilde{x}_{k+1}, \dots, \tilde{x}_n \right) \\ &\subseteq \Delta_k \end{aligned}$$

D'autre part, si $x_k = \tilde{x}_k$ alors $s_k(x) \in \Delta_k$ par définition de s_k . □

Remarques.

1. L'utilisation de l'expression

$$\Delta_k \supseteq \text{range} \left(\frac{\partial f}{\partial x_k}, \mathbf{x}_1, \dots, \mathbf{x}_k, \tilde{x}_{k+1}, \dots, \tilde{x}_n \right)$$

à la place de l'expression

$$\Delta_k \supseteq \text{range} \left(\frac{\partial f}{\partial x_k}, \mathbf{x} \right)$$

a été proposée initialement dans Hansen[17] dans le contexte des extensions aux intervalles classiques.

2. L'hypothèse $\tilde{x} \in \mathbf{x}$ peut être remplacée par $\tilde{x} \in \mathbb{R}^n$ dans la proposition précédente à condition que

$$\Delta_k \supseteq \text{range} \left(\frac{\partial f}{\partial x_k}, \mathbf{x}_1, \dots, \mathbf{x}_k, \tilde{x}_{k+1}, \dots, \tilde{x}_n \right)$$

soit changée en

$$\Delta_k \supseteq \text{range} \left(\frac{\partial f}{\partial x_k}, \mathbf{x}_1 \vee \tilde{x}_1, \dots, \mathbf{x}_k \vee \tilde{x}_k, \tilde{x}_{k+1}, \dots, \tilde{x}_n \right).$$

16.2 L'AE-extension de la valeur moyenne (cas particulier $m = 1$)

L'AE-extension de la valeur moyenne est tout d'abord illustrée par un exemple. Considérons une fonction continûment dérivable $f : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbf{x} \in \overline{\mathbb{R}}$, $\tilde{x} \in \text{pro } \mathbf{x}$ et $\Delta \supseteq \text{range}(f', \text{pro } \mathbf{x})$. Définissons l'intervalle \mathbf{z} par

$$\mathbf{z} = f(\tilde{x}) + \Delta(\mathbf{x} - \tilde{x}) \quad (16.2)$$

qui est impropre (la multiplication par un intervalle impropre dont la projection propre contient 0 conduit à un intervalle impropre dont la projection contient 0). Donc, pour prouver que \mathbf{z} est (f, \mathbf{x}) -interprétable, il faut montrer que la proposition suivante est vraie :

$$(\forall z \in \text{pro } \mathbf{z})(\exists x \in \text{pro } \mathbf{x})(z = f(x)), \quad (16.3)$$

et donc que $\text{pro } \mathbf{z} \subseteq \text{range}(f, \text{pro } \mathbf{x})$. D'une part, l'expression (16.2) correspond à l'AE-extension naturelle de la fonction $m(x, \delta) = f(\tilde{x}) + \delta(\mathbf{x} - \tilde{x})$ évaluée en $\mathbf{x} \in \overline{\mathbb{R}}$ et $\Delta \in \mathbb{R}$. Ainsi la proposition quantifiée suivante est vraie :

$$(\forall \delta \in \Delta)(\forall z \in \text{pro } \mathbf{z})(\exists x \in \mathbf{x})(z = m(x, \delta)) \quad (16.4)$$

D'autre part, le théorème de la valeur moyenne implique la proposition quantifiée suivante :

$$(\forall x \in \text{pro } \mathbf{x})(\exists \delta \in \Delta)(f(x) = m(x, \delta)) \quad (16.5)$$

Il reste à prouver que (16.4) et (16.5) implique (16.3). La preuve de cette implication est faite de la manière suivante : remarquons tout d'abord que

1. comme l'expression (16.2) ne contient qu'une occurrence de chaque variable, la **Proposition IV.28** montre qu'il existe une fonction **continue** $s' : (\Delta, \text{pro } \mathbf{z})^T \rightarrow \text{pro } \mathbf{x}$ qui vérifie $x = s'(\delta, z) \implies z = m(x, \delta)$ (ce qui est plus fort que (16.4)).
2. la **Proposition IV.29** montre qu'il existe une fonction **continue** $s'' : \text{pro } \mathbf{x} \rightarrow \Delta$ qui vérifie $\delta = s''(x) \implies f(x) = m(x, \delta)$ (ce qui est plus fort que (16.5)).

Construisons la fonction continue $s : (\text{pro } \mathbf{x}, \text{pro } \mathbf{z})^T \rightarrow \text{pro } \mathbf{x}$ de la manière suivante : $s(x, z) = s'(s''(x), z)$. On peut alors vérifier que

$$x = s(x, z) \implies z = f(x) \quad (16.6)$$

Maintenant, pour chaque $z \in \text{pro } \mathbf{z}$, la fonction continue $s(\cdot, z)$ est définie dans $\text{pro } \mathbf{x}$ à valeurs dans $\text{pro } \mathbf{x}$. Ainsi, le théorème de Brouwer montre que $x \in \text{pro } \mathbf{x}$. La proposition quantifiée suivante est donc vraie :

$$(\forall z \in \text{pro } \mathbf{z})(\exists x \in \text{pro } \mathbf{x})(x = s(x, z))$$

Finalement, grâce à (16.6), la proposition quantifiée précédente implique (16.3), et \mathbf{z} est ainsi (f, \mathbf{x}) -interprétable. Le théorème suivante généralise ce raisonnement à une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continûment dérivable quelconque et à tout intervalle $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$.

Théorème IV.6. *Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continûment dérivable, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, $c : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tel que $c(\mathbf{x}) \in \mathbf{x}$ et $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{I}\mathbb{R}^{1 \times n}$ une extension de f' aux intervalles classiques²². Alors, la fonction d'intervalles \mathbf{h} définie par*

$$\mathbf{h}(\mathbf{x}) = f(c(\mathbf{x})) + \mathbf{g}(\text{pro } \mathbf{x}) \times (\mathbf{x} - c(\mathbf{x}))$$

est une AE-extension de f , appelée l'AE-extension de la valeur moyenne de f .

Remarque. *Rappelons l'expression de l'extension classique de la valeur moyenne :*

$$f(c(\mathbf{x})) + \mathbf{g}(\mathbf{x}) \times (\mathbf{x} - c(\mathbf{x})).$$

Ainsi les deux extensions sont très similaires.

Démonstration. Premièrement, pour tout $x \in \mathbb{R}^n$ nous avons $c(x) = x$ et donc $\mathbf{h}(x) = f(x) + \mathbf{g}(\text{pro } \mathbf{x}) \times (x - x) = f(x)$. Ensuite, pour tout $\mathbf{x}_{[1..n]} \in \mathbb{K}\mathbb{R}^n$, définissons $\Delta_{[1..n]} = \mathbf{g}(\text{pro } \mathbf{x}_{[1..n]})^T \in \mathbb{I}\mathbb{R}^n$, $\tilde{x}_{[1..n]} = c(\mathbf{x}_{[1..n]})$ et $\mathbf{x}_0 := \mathbf{h}(\mathbf{x}_{[1..n]})$. Nous avons donc

$$\mathbf{x}_0 = f(\tilde{x}_{[1..n]}) + \Delta^T \times (\mathbf{x} - \tilde{x}) \quad (16.7)$$

Définissons les ensembles d'indices $A = P(\mathbf{x}_{[1..n]}) \cup I(\mathbf{x}_0)$ et $E = I(\mathbf{x}_{[1..n]}) \cup P(\mathbf{x}_0)$. Nous devons donc prouver que la proposition quantifiée suivante est vraie :

$$(\forall x_A \in \text{pro } \mathbf{x}_A) (\exists x_E \in \text{pro } \mathbf{x}_E) (x_0 = f(x_{[1..n]})) \quad (16.8)$$

D'une part, par l'expression (16.7), \mathbf{x}_0 est l'AE-extension naturelle de la fonction

$$m(x_{[1..n]}, \delta_{[1..n]}) := f(\tilde{x}_{[1..n]}) + \delta^T \times (x - \tilde{x})$$

évaluée en $\mathbf{x}_{[1..n]}$ et $\Delta_{[1..n]}$. Comme l'expression de m ne contient qu'une occurrence de chaque variable, nous pouvons appliquer la Proposition IV.28 (page 160) qui prouve qu'il existe une fonction continue

$$s'_E : (\text{pro } \mathbf{x}_A, \Delta_{[1..n]})^T \rightarrow \text{pro } \mathbf{x}_E$$

qui vérifie

$$x_E = s'_E(x_A, \delta_{[1..n]}) \implies x_0 = m(x_{[1..n]}, \delta_{[1..n]}) \quad (16.9)$$

D'autre part, comme \mathbf{g} est une extension de f' nous avons $\Delta^T \supseteq \text{range}(f', \text{pro } \mathbf{x})$ et nous pouvons appliquer la Proposition IV.29 qui prouve qu'il existe une fonction continue

$$s''_{[1..n]} : \text{pro } \mathbf{x}_{[1..n]} \rightarrow \Delta_{[1..n]}$$

qui vérifie

$$\delta_{[1..n]} = s''_{[1..n]}(x_{[1..n]}) \implies f(x_{[1..n]}) = g(x_{[1..n]}, \delta_{[1..n]}) \quad (16.10)$$

²²Ou une AE-extension, car elle ne sera évaluée que sur des intervalles propres.

Maintenant, nous construisons la fonction $s_E : \text{pro } \mathbf{x}_A \cup E \longrightarrow \text{pro } \mathbf{x}_E$ en composant s' et s'' de la manière suivante :

$$s_E(x_{A \cup E}) = s'_E(x_A, s''_{[1..n]}(x_{[1..n]})). \quad (16.11)$$

Prouvons maintenant que

$$x_E = s_E(x_{A \cup E}) \implies x_0 = f(x_{[1..n]}). \quad (16.12)$$

Par (16.11), $s_E(x_{A \cup E}) = x_E$ implique

$$\delta_{[1..n]} = s''_{[1..n]}(x_{[1..n]}) \wedge x_E = s'_E(x_A, \delta_{[1..n]})$$

Maintenant, par (16.9) et (16.10), nous avons

$$f(x_{[1..n]}) = g(x_{[1..n]}, \delta_{[1..n]}) \wedge x_0 = g(x_{[1..n]}, \delta_{[1..n]})$$

C'est à dire $x_0 = f(x_{[1..n]})$. Finalement, grâce au théorème de Brouwer, pour chaque valeur de $x_A \in \text{pro } \mathbf{x}_A$, la fonction $s_E(x_A, \cdot) : \text{pro } \mathbf{x}_E \longrightarrow \text{pro } \mathbf{x}_E$ a un point fixe. C'est à dire,

$$(\forall x_A \in \text{pro } \mathbf{x}_A) (\exists x_E \in \text{pro } \mathbf{x}_E) (s_E(x_A \cup E) = x_E)$$

ce qui implique (16.8) grâce à (16.12). \square

Remarques.

1. Conformément à l'énoncé de la Proposition IV.29, l'extension \mathbf{g} peut être remplacée par la fonction d'intervalles propres $\tilde{\mathbf{g}}$ définie par

$$\tilde{\mathbf{g}}_{1k}(\mathbf{x}) = \mathbf{g}_{1k}(\mathbf{x}_1, \dots, \mathbf{x}_k, \tilde{x}_{k+1}, \dots, \tilde{x}_n)$$

dans l'énoncé du théorème précédent. Comme indiqué dans Hansen[18], cette modification peut conduire à une amélioration sensible de la précision de l'AE-extension de la valeur moyenne.

2. L'hypothèse $c(\mathbf{x}) \in \mathbf{x}$ peut être remplacée par $c(\mathbf{x}) \in \mathbb{R}^n$ à condition que $\mathbf{g}(\text{pro } \mathbf{x})$ soit remplacée par $\mathbf{g}(c(\mathbf{x}) \vee \text{pro } \mathbf{x})$. Cependant dans ce cas \mathbf{h} n'est plus qu'une AE-extension faible de f .
3. Comme dans le cas des extensions aux intervalles classiques, si la fonction d'intervalles \mathbf{g} n'est pas définie pour tout $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$ alors le domaine de définition de l'AE-extension de la valeur moyenne doit être adapté.
4. Lorsque tous les intervalles mis en jeu sont propres, les opérations pro n'ont aucune influence et l'AE-extension de la valeur moyenne coïncide avec l'extension de la valeur moyenne classique.

L'exemple suivant illustre la manière dont l'AE-extension de la valeur moyenne permet de construire des approximations intérieures et extérieures de l'image d'un intervalle par un fonction.

Exemple IV.28. Considérons la fonction $f(x) = x^2$ et l'intervalle propre $\mathbf{x} = [1, 1.2]$. Nous avons alors :

$$\text{range}(f, \mathbf{x}) = [1, 1.44]$$

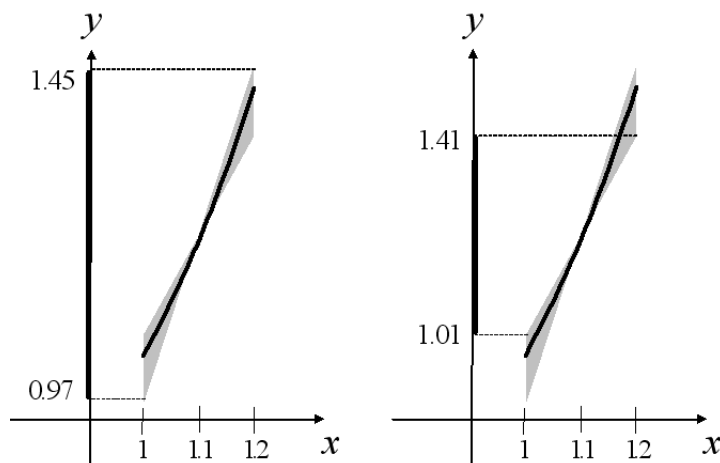


FIG. 16.1 – Graphe de la fonction $f(x) = x^2$ avec le cône des dérivées $f(\tilde{\mathbf{x}}) + f'(\mathbf{x})(x - \tilde{\mathbf{x}})$.

Considérons l'intervalle $\Delta = [2, 2.4]$ qui vérifie $\Delta \supseteq \text{range}(f', \mathbf{x})$. Quand l'AE-extension de la valeur moyenne est évaluée en \mathbf{x} , nous obtenons

$$\text{range}(f, \mathbf{x}) \subseteq f(\tilde{\mathbf{x}}) + \Delta(\mathbf{x} - \text{mid } \mathbf{x}) = [0.97, 1.45]$$

qui est bien une approximation extérieure de $\text{range}(f, \mathbf{x})$. Quand elle est évaluée en dual \mathbf{x} , nous obtenons

$$f(\tilde{\mathbf{x}}) + \Delta(\text{dual } \mathbf{x} - \text{mid } \mathbf{x}) = [1.41, 1.01]$$

L'intervalle $\text{pro}[1.41, 1.01] = [1.01, 1.41]$ est une approximation intérieure de $\text{range}(f, \mathbf{x})$. La figure 16.1 illustre ces approximations obtenues grâce à la linéarisation de f : le graphique de gauche (respectivement de droite) illustre l'approximation extérieure (respectivement intérieure). Ces deux approximations sont obtenues avec le même cône des dérivées.

Le Théorème IV.6 a prouvé que l'intervalle

$$\mathbf{z} = f(\tilde{x}) + \mathbf{g}(\text{pro } \mathbf{x}) \times (\mathbf{x} - \tilde{x})$$

est (f, \mathbf{x}) -interprétable. Comme f^* est l'unique AE-extension minimale de f , cela s'exprime de manière équivalente par $f^*(\mathbf{x}) \subseteq \mathbf{z}$. Il sera utile de construire un intervalle contenu dans $f^*(\mathbf{x})$ (en particulier l'ordre de convergence quadratique de l'AE-extension de la valeur moyenne sera obtenu grâce à cette construction). La proposition suivante présente une telle construction.

Proposition IV.30. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continûment dérivable, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, $\tilde{x} \in \mathbf{x}$ et $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^{1 \times n}$ une extension de f' aux intervalles classiques. Définissons l'intervalle

$$\mathbf{z} = f(\tilde{x}) + (\text{dual } \mathbf{g}(\text{pro } \mathbf{x})) \times (\mathbf{x} - \tilde{x}).$$

Alors, $\mathbf{z} \subseteq f^*(\mathbf{x})$.

Démonstration. Définissons la fonction

$$m(x, \delta) = f(\tilde{x}) + \delta^T \times (x - \tilde{x}).$$

L'intervalle \mathbf{z} est alors l'AE-extension naturelle de m évaluée en \mathbf{x} et dual Δ , où $\Delta = \mathbf{g}(\text{pro } \mathbf{x})^T \in \mathbb{I}\mathbb{R}^n$. Par la Proposition IV.5, cette AE-extension naturelle est minimale. Nous avons donc $\mathbf{z} = m^*(\mathbf{x}, \text{dual } \Delta)$. Maintenant, l'intervalle $\mathbf{z}' := f^*(\mathbf{x})$ est (f, \mathbf{x}) -interprétable et la proposition quantifiée suivante est donc vraie :

$$(\forall x_P \in \mathbf{x}_P)(\mathbf{Q}^{(\mathbf{z}')} z \in \text{pro } \mathbf{z}')(\exists x_I \in \text{pro } \mathbf{x}_I)(z = f(x))$$

Par la Proposition IV.29 (page 171) la proposition quantifiée suivante est vraie :

$$(\forall x \in \text{pro } \mathbf{x})(\exists \delta \in \Delta)(f(x) = m(x, \delta))$$

Les deux propositions quantifiées précédente impliquent la suivante :

$$(\forall x_P \in \mathbf{x}_P)(\mathbf{Q}^{(\mathbf{z}')} z \in \mathbf{z}')(\exists x_I \in \text{pro } \mathbf{x}_I)(\exists \delta \in \mathbf{g}(\text{pro } \mathbf{x}))(z = m(x, \delta))$$

C'est à dire, l'intervalle $\mathbf{z}' = f^*(\mathbf{x})$ est $(m, \mathbf{x}, \text{dual } \Delta)$ -interprétable, ce qui implique

$$m^*(\mathbf{x}, \text{dual } \mathbf{g}(\mathbf{y})) \subseteq f^*(\mathbf{x})$$

C'est à dire finalement $\mathbf{z} \subseteq f^*(\mathbf{x})$. □

Remarque. *Encore une fois, la proposition précédente reste valable avec $\tilde{x} \notin \mathbf{x}$ si $\mathbf{g}(\text{pro } \mathbf{x})$ est remplacée par $\mathbf{g}(\tilde{x} \vee \text{pro } \mathbf{x})$.*

16.3 L'AE-extension de la valeur moyenne (cas général)

Étant donné une fonction continûment dérivable $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et un intervalle $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, les AE-extensions de la valeur moyenne des composantes f_i sont maintenant utilisées pour construire un intervalle (f, \mathbf{x}) -interprétable $\mathbf{z} \in \mathbb{K}\mathbb{R}^m$, i.e. qui vérifie

$$(\forall x_A \in \mathbf{x}_A)(\forall z_{A'} \in \text{pro } \mathbf{z}_{A'})(\exists z_{E'} \in \mathbf{z}_{E'})(\exists x_E \in \text{pro } \mathbf{x}_E)(z = f(x)) \quad (16.13)$$

où $A = P(\mathbf{x})$, $E = I(\mathbf{x})$, $E' = P(\mathbf{z})$ et $A' = I(\mathbf{z})$. Cette construction est similaire à la construction de l'AE-extension naturelle pour fonctions à valeurs vectorielles présentées dans le chapitre précédent. Pour $i \in [1..m]$ considérons une AE-extension de la valeur moyenne \mathbf{h}_i de la fonction f_i . L'intervalle \mathbf{z} défini naïvement par $\mathbf{z}_i = \mathbf{h}_i(\mathbf{x})$ n'est pas (f, \mathbf{x}) -interprétable en général parce que la conjonction

$$\bigwedge_{i \in [1..m]} (\forall x_A \in \mathbf{x}_A)(\mathbf{Q}^{(\mathbf{z}_i)} z_i \in \text{pro } \mathbf{z}_i)(\exists x_E \in \text{pro } \mathbf{x}_E)(z_i = f_i(x)) \quad (16.14)$$

où $\mathbf{Q}^{(\mathbf{z}_i)} = \exists$ si $i \in P'$ et $\mathbf{Q}^{(\mathbf{z}_i)} = \forall$ sinon, n'implique pas (16.13) in general.

Exemple IV.29. *Considérons la fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, $\mathbf{x}_1 \in \mathbb{I}\mathbb{R}$ et $\mathbf{x}_2, \mathbf{x}_3 \in \overline{\mathbb{I}\mathbb{R}}$. Supposons que $\mathbf{z}_1 = \mathbf{h}_1(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ et $\mathbf{z}_2 = \mathbf{h}_2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ sont impropres. Alors les deux propositions quantifiées suivantes sont vraies :*

$$(\forall x_1 \in \mathbf{x}_1)(\forall z_1 \in \text{pro } \mathbf{z}_1)(\exists x_2 \in \text{pro } \mathbf{x}_2)(\exists x_3 \in \text{pro } \mathbf{x}_3)(z_1 = f_1(x))$$

et

$$(\forall x_1 \in \mathbf{x}_1)(\forall z_2 \in \text{pro } \mathbf{z}_2)(\exists x_2 \in \text{pro } \mathbf{x}_2)(\exists x_3 \in \text{pro } \mathbf{x}_3)(z_2 = f_2(x))$$

Cependant, la conjonction de ces deux dernières propositions quantifiées n'implique pas la suivante en général

$$(\forall x_1 \in \mathbf{x}_1)(\forall z \in \text{pro } \mathbf{z})(\exists x_2 \in \text{pro } \mathbf{x}_2)(\exists x_3 \in \text{pro } \mathbf{x}_3)(z = f(x)).$$

L'intervalle $\mathbf{z} \in \overline{\mathbb{R}}^2$ n'est donc par (f, \mathbf{x}) -interprétable en générale.

Pour pouvoir valider la proposition quantifiée (16.13), les calculs précédents doivent être modifiés de la même manière que dans le cas de l'AE-extension naturelle pour fonctions à valeur vectorielles. Cela peut se formaliser de la manière suivante : introduisons une fonction $\pi : [1..n] \longrightarrow [1..m]$ qui associe à une variable x_j l'indice de la composante \mathbf{h}_i dans laquelle cette variable restera quantifiée existentiellement. Alors, les AE-extensions \mathbf{h}_i sont utilisées de la manière suivante :

$$\mathbf{z}_i = \mathbf{h}_i(\mathbf{y})$$

où $\mathbf{y}_j = \mathbf{x}_j$ si $i = \pi(j)$ et $\mathbf{y}_j = \text{pro } \mathbf{x}_j$ sinon. Les composantes propres de \mathbf{x} n'étant pas sensibles à l'opération pro, l'intervalle \mathbf{z} vérifie

$$\bigwedge_{i \in [1..m]} (\forall x_{A \cup A_i} \in \mathbf{x}_{A \cup A_i})(\mathbf{Q}^{(\mathbf{z}_i)} z \in \text{pro } \mathbf{z}_i)(\exists x_{E_i} \in \text{pro } \mathbf{x}_{E_i})(z = f(x)) \quad (16.15)$$

où $E_i = E \cap \pi^{-1}(i)$ et $A_i = E \setminus E_i$. Maintenant, comme chaque variable quantifiée existentiellement n'apparaît que dans une des propositions quantifiées de la conjonction précédente, cette conjonction implique la proposition quantifiée (16.13).

Exemple IV.30. *Considérons la fonction $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$, $\mathbf{x}_1 \in \mathbb{R}$, $\mathbf{x}_2, \mathbf{x}_3 \in \overline{\mathbb{R}}$ et la fonction*

$$\pi : (1 \rightarrow 1 ; 2 \rightarrow 2 ; 3 \rightarrow 1)$$

Les calculs suivants sont donc effectués :

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{h}_1(\mathbf{x}_1, \text{pro } \mathbf{x}_2, \mathbf{x}_3) \\ \mathbf{z}_2 &= \mathbf{h}_2(\text{pro } \mathbf{x}_1, \mathbf{x}_2, \text{pro } \mathbf{x}_3) \end{aligned}$$

Supposons que les deux intervalles ci-dessus sont impropres. Les deux propositions quantifiées suivantes sont donc vraies (remarquons que $(\text{pro } \mathbf{x}_1) = \mathbf{x}_1$ car \mathbf{x}_1 étant propre) :

$$(\forall x_1 \in \mathbf{x}_1)(\forall x_2 \in \text{pro } \mathbf{x}_2)(\forall z_1 \in \text{pro } \mathbf{z}_1)(\exists x_3 \in \text{pro } \mathbf{x}_3)(z_1 = f_1(x))$$

et

$$(\forall x_1 \in \mathbf{x}_1)(\forall x_3 \in \text{pro } \mathbf{x}_3)(\forall z_2 \in \text{pro } \mathbf{z}_2)(\exists x_2 \in \text{pro } \mathbf{x}_2)(z_2 = f_2(x))$$

Nous démontrerons que la conjonction de ces deux dernières propositions quantifiées implique

$$(\forall x_1 \in \mathbf{x}_1)(\forall z \in \text{pro } \mathbf{z})(\exists x_2 \in \text{pro } \mathbf{x}_2)(\exists x_3 \in \text{pro } \mathbf{x}_3)(z = f(x))$$

Ainsi, $\mathbf{z} \in \overline{\mathbb{R}}^2$ est (f, \mathbf{x}) -interprétable.

L'expression de l'AE-extension de la valeur moyenne sera facilitée par l'introduction du produit matrice/vecteur spécifique suivant.

Définition IV.10. Soient $\mathbf{A} \in \mathbb{K}\mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$ et $\pi : [1..n] \longrightarrow [1..m]$. Le produit $\mathbf{A} *_{\pi} \mathbf{x} \in \mathbb{K}\mathbb{R}^m$ est alors défini de la manière suivante : pour $i \in [1..m]$,

$$(\mathbf{A} *_{\pi} \mathbf{x})_i = \sum_{j \in \pi^{-1}(i)} \mathbf{A}_{ij} \mathbf{x}_j + \sum_{j \in [1..n] \setminus \pi^{-1}(i)} \mathbf{A}_{ij} (\text{pro } \mathbf{x})$$

Le cas particulier suivant sera souvent rencontré : si $m = n$ et $\pi = \text{id}$ alors le produit $\mathbf{A} *_{\pi} \mathbf{x}$ est noté $\mathbf{A} * \mathbf{x}$. Dans ce cas, la définition précédente s'écrit

$$(\mathbf{A} * \mathbf{x})_i = \mathbf{A}_{ii} \mathbf{x}_i + \sum_{j \in [1..n] \setminus \{i\}} \mathbf{A}_{ij} (\text{pro } \mathbf{x}_j)$$

Maintenant, l'AE-extension de la valeur moyenne pour fonctions à valeurs vectorielles peut-être définie.

Théorème IV.7. Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ une fonction continûment dérivable, $\mathbf{x} \in \mathbb{K}\mathbb{R}^n$, $c : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ tel que $c(\mathbf{x}) \in \mathbf{x}$ et $\pi : [1..n] \longrightarrow [1..m]$. Considérons une extension $\mathbf{g} : \mathbb{R}^n \longrightarrow \mathbb{R}^{m \times n}$ de f' aux intervalles classiques. Alors la fonction d'intervalles \mathbf{h} définie par

$$\mathbf{h}(\mathbf{x}) = f(c(\mathbf{x})) + \mathbf{g}(\text{pro } \mathbf{x}) *_{\pi} (\mathbf{x} - c(\mathbf{x}))$$

est une AE-extension de f , appelée une AE-extension de la valeur moyenne f .

Remarque. Rappelons l'expression de l'extension classique de la valeur moyenne dans le cas des fonctions à valeurs vectorielles :

$$f(c(\mathbf{x})) + \mathbf{g}(\mathbf{x}) \times (\mathbf{x} - c(\mathbf{x})).$$

Ainsi les deux extensions sont très similaires.

Démonstration. Premièrement, quelque soit $x \in \mathbb{R}^n$, nous avons $\mathbf{h}(x) = f(x) + \mathbf{g}(\text{pro } \mathbf{x}) \times (x - x) = f(x)$ parce que $c(x) = x$. Considérons maintenant $\mathbf{x}_{[1..n]} \in \mathbb{K}\mathbb{R}^n$ et définissons $\Delta = \mathbf{g}(\text{pro } \mathbf{x}_{[1..n]})$ et $\tilde{x}_{[1..n]} = c(\mathbf{x}_{[1..n]})$. Pour obtenir des notations homogènes, l'évaluation de \mathbf{h} est faite de la manière suivante :

$$\mathbf{x}_{[n+1..n+m]} = f(\tilde{x}_{[1..n]}) + \Delta *_{\pi} (\mathbf{x}_{[1..n]} - \tilde{x}_{[1..n]}) \quad (16.16)$$

Nous devons démontrer que $\mathbf{x}_{[n+1..n+m]}$ est $(f, \mathbf{x}_{[1..n]})$ -interprétable. Définissons les ensembles d'indices $A = P(\mathbf{x}_{[1..n]}) \cup I(\mathbf{x}_{[n+1..n+m]})$ et $E = I(\mathbf{x}_{[1..n]}) \cup P(\mathbf{x}_{[n+1..n+m]})$. Nous devons donc prouver que la proposition quantifiée suivante est vraie :

$$(\forall x_A \in \mathbf{x}_A) (\exists x_E \in \mathbf{x}_E) (f(x_{[1..n]}) = x_{[n+1..n+m]}) \quad (16.17)$$

Considérons $i \in [1..m]$. La i^{ieme} de l'égalité (16.16) est

$$\mathbf{x}_{n+i} = f_i(\tilde{x}) + \sum_{j \in [1..n]} \Delta_{ij} (\mathbf{y}_j^{(i)} - \tilde{x})$$

où $\mathbf{y}_j^{(i)} = \mathbf{x}_j$ si $\pi(j) = i$ et $\mathbf{y}_j^{(i)} = \text{pro } \mathbf{x}_j$ sinon. Cela correspond à l'AE-extension de la valeur moyenne de f_i évaluée en $\mathbf{y}^{(i)}$. Définissons $E_i = E \cap \pi^{-1}(i)$ et $A_i = E \setminus E_i$. Ainsi, $\mathbf{y}_j^{(i)}$ est propre si et seulement si \mathbf{x}_j est propre ou $j \in A_i$. De manière équivalente, $\mathbf{y}_j^{(i)}$

est impropre si et seulement si $j \in E_i$. En utilisant le même raisonnement que dans la preuve du Théorème IV.6, nous obtenons pour chaque $i \in [1..m]$ une fonction continue

$$s_{E_i} : \mathbf{x}_{A \cup A_i} \longrightarrow \mathbf{x}_{E_i}$$

qui vérifie

$$x_{E_i} = s_{E_i}(x_{A \cup A_i}) \implies x_{n+i} = f_i(x_{[1..n]})$$

Grâce à la définition de π , nous pouvons vérifier que pour tout $i \in [1..m]$ et $i' \in [1..m]$ tels que $i \neq i'$, nous avons $E_i \cap E_{i'} = \emptyset$ et que $\bigcup \{E_i | i \in [1..m]\} = E$. Nous avons aussi $\bigcup \{A \cup A_i | i \in [1..m]\} = A \cup E$. La fonction $s_E : \mathbf{x}_{A \cup E} \longrightarrow \mathbf{x}_E$ est donc bien définie, continue et satisfait de plus

$$\begin{aligned} s_E(x_{A \cup E}) = x_E &\implies \bigwedge_{i \in [1..m]} x_{E_i} = s_{E_i}(x_{A \cup A_i}) \\ &\implies \bigwedge_{i \in [1..m]} x_{n+i} = f_i(x_{[1..n]}) \\ &\implies x_{[n+1..n+m]} = f(x_{[1..n]}) \end{aligned}$$

Finalement, par le théorème de Brouwer, pour tout $x_A \in \mathbf{x}_A$ la fonction continue $s_E(x_A, \cdot) : \mathbf{x}_E \longrightarrow \mathbf{x}_E$ a un point fixe. C'est à dire,

$$(\forall x_A \in \mathbf{x}_A) (\exists x_E \in \mathbf{x}_E) (s_E(x_A \cup E) = x_E)$$

ce qui conclue la démonstration. □

Remarques.

1. Dans le cas $m = 1$, le Théorème IV.7 coïncide avec le Théorème IV.6 : il n'y a alors qu'une seule fonction π possible qui est $\pi(i) = 1$ pour tout $i \in [1..n]$. Alors,

$$\mathbf{g}(\text{pro } \mathbf{x}) *_{\pi} (\mathbf{x} - c(\mathbf{x})) = \mathbf{g}(\text{pro } \mathbf{x}) \times (\mathbf{x} - c(\mathbf{x}))$$

Le Théorème IV.7 peut donc être utilisé pour définir l'AE-extension de la valeur moyenne en général.

2. Comme dans le cas du Théorème IV.6, l'extension \mathbf{g} peut être remplacée par $\tilde{\mathbf{g}}$ définie par

$$\tilde{\mathbf{g}}_{ij}(\mathbf{x}) = \mathbf{g}_{ij}(\mathbf{x}_1, \dots, \mathbf{x}_j, \tilde{x}_{j+1}, \dots, \tilde{x}_n).$$

3. Si $\mathbf{x} \in \mathbb{IR}^n$ alors

$$\mathbf{g}(\text{pro } \mathbf{x}) *_{\pi} (\mathbf{x} - c(\mathbf{x})) = \mathbf{g}(\mathbf{x}) \times (\mathbf{x} - c(\mathbf{x}))$$

et les AE-extensions de la valeur moyenne coïncident avec l'extension classique de la valeur moyenne.

Exemple IV.31. Considérons la fonction

$$f(x) = \begin{pmatrix} 81x_1^2 + x_2^2 + 18x_1x_2 - 100 \\ x_1^2 + 81x_2^2 + 18x_1x_2 - 100 \end{pmatrix}$$

et les expressions suivantes de ses dérivées partielles

$$\mathbf{A} = \begin{pmatrix} 162(\text{pro } \mathbf{x}_1) + 18(\text{pro } \mathbf{x}_2) & 2(\text{pro } \mathbf{x}_2) + 18(\text{pro } \mathbf{x}_1) \\ 2(\text{pro } \mathbf{x}_1) + 18(\text{pro } \mathbf{x}_2) & 162(\text{pro } \mathbf{x}_2) + 18(\text{pro } \mathbf{x}_1) \end{pmatrix}.$$

Alors, nous pouvons construire l'AE-extension suivante de f :

$$\mathbf{h}(\mathbf{x}) = f(\text{mid } \mathbf{x}) + \mathbf{A} * (\mathbf{x} - \text{mid } \mathbf{x}).$$

C'est à dire

$$\begin{pmatrix} f_1(\text{mid } \mathbf{x}) + \mathbf{A}_{11}(\mathbf{x}_1 - \text{mid } \mathbf{x}_1) + \mathbf{A}_{12}(\text{pro } \mathbf{x}_2 - \text{mid } \mathbf{x}_2) \\ f_2(\text{mid } \mathbf{x}) + \mathbf{A}_{21}(\text{pro } \mathbf{x}_1 - \text{mid } \mathbf{x}_1) + \mathbf{A}_{12}(\mathbf{x}_2 - \text{mid } \mathbf{x}_2) \end{pmatrix}$$

Si $\mathbf{x} = ([1.1, 0.9], [1.1, 0.9])^T$, l'AE-extension donne

$$\mathbf{h}(\mathbf{x}) = ([14, -14] , [14, -14])^T$$

et prouve donc que $\text{pro } \mathbf{h}(\mathbf{x}, \mathbf{y}) \subseteq \text{range}(f, \text{pro } \mathbf{x}, \text{pro } \mathbf{y})$. Par conséquent, il existe $x \in \text{pro } \mathbf{x}$ tel que $f(x) = 0$. Et en effet, $f(1, 1) = 0$.

16.4 Qualité de l'AE-extension de la valeur moyenne

Dans le cas particulier des fonctions à valeurs réelles, l'AE-extension de la valeur moyenne a un ordre de convergence quadratique. Cependant, son ordre de convergence est linéaire en général dans le cas des fonctions à valeurs vectorielles.

Théorème IV.8. Avec les notations du Théorème IV.7, supposons que \mathbf{g} est localement Lipschitz continue. Alors,

- l'AE-extension de la valeur moyenne a un ordre de convergence linéaire ;
- si $m = 1$, i.e. dans le cas particulier du Théorème IV.6, l'AE-extension de la valeur moyenne a un ordre de convergence quadratique.

Démonstration. L'AE-extension de la valeur moyenne est la composition de fonctions d'intervalles qui sont localement Lipschitz continues. Elle est donc localement Lipschitz continue. Par conséquent, l'AE-extension de la valeur moyenne a un ordre de convergence linéaire (Proposition IV.15). Étudions en détail le cas $m = 1$. Notons $\mathbf{g}(\text{pro } \mathbf{x})^T$ par $\Delta \in \mathbb{I}\mathbb{R}^n$ et définissons

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) = f(\tilde{x}) + \Delta^T \times (\mathbf{x} - \tilde{x}) \quad \text{et} \quad \mathbf{z}' = f(\tilde{x}) + (\text{dual } \Delta)^T \times (\mathbf{x} - \tilde{x})$$

Par le Théorème IV.6 et la Proposition IV.30, nous avons $\mathbf{z}' \subseteq f^*(\mathbf{x}) \subseteq \mathbf{z}$. Grâce au lemme 2.5 de Kaucher[21], cela implique $\text{dist}(f^*(\mathbf{x}), \mathbf{z}) \leq \text{dist}(\mathbf{z}, \mathbf{z}')$. Il reste donc à majorer la distance entre \mathbf{z} et \mathbf{z}' . En utilisant les relations entre l'arithmétique de Kaucher et la distance, nous obtenons

$$\begin{aligned} \text{dist}(\mathbf{z}, \mathbf{z}') &= \text{dist}(\Delta^T \times (\mathbf{x} - \tilde{x}), (\text{dual } \Delta)^T \times (\mathbf{x} - \tilde{x})) \\ &\leq \sum_{i \in [1..n]} \text{dist}(\Delta_i \times (\mathbf{x}_i - \tilde{x}_i), (\text{dual } \Delta_i) \times (\mathbf{x}_i - \tilde{x}_i)) \\ &\leq \sum_{i \in [1..n]} |\mathbf{x}_i - \tilde{x}_i| \text{dist}(\Delta_i, \text{dual } \Delta_i) \end{aligned}$$

Maintenant, comme $\tilde{x} \in \mathbf{x}$, nous avons $|\mathbf{x}_i - \tilde{x}_i| \leq |\text{wid } \mathbf{x}_i| \leq \|\text{wid } \mathbf{x}\|$. De plus, il est clair que $\text{dist}(\Delta_i, \text{dual } \Delta_i) = \text{wid } \Delta_i \leq \|\text{wid } \Delta\|$. Nous avons donc

$$\text{dist}(\mathbf{z}, \mathbf{z}') \leq n \|\text{wid } \mathbf{x}\| \|\text{wid } \Delta\|$$

Maintenant, comme $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^{1 \times n}$ est localement Lipschitz continue, $\mathbf{g}' : \mathbb{R} \rightarrow \mathbb{R}^n$ définie par $\mathbf{g}'(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T$ l'est aussi. Et parce que pro est localement Lipschitz continue, la composition $\mathbf{g}' \circ \text{pro} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est aussi localement Lipschitz continue (avec ces notations, nous avons $\Delta = \mathbf{g}' \circ \text{pro}(\mathbf{x})$). Donc, pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{R}^n$, il existe $\gamma > 0$ tel que pour tout $\mathbf{y} \in \mathbb{K}\mathbf{x}^{\text{ref}}$,

$$\text{dist}(\Delta, \mathbf{g}' \circ \text{pro}(\mathbf{y})) \leq \gamma \text{dist}(\mathbf{x}, \mathbf{y})$$

Choisissons $\mathbf{y} = \text{mid } \mathbf{x} \in \mathbf{x}^{\text{ref}}$. D'une part,

$$\|\text{rad } \Delta\| \leq \text{dist}(\Delta, \mathbf{g}' \circ \text{pro}(\text{mid } \mathbf{x}))$$

parce que $\mathbf{g}' \circ \text{pro}(\text{mid } \mathbf{x}) \in \mathbb{R}$. D'autre part, nous avons $\text{dist}(\mathbf{x}, \text{mid } \mathbf{x}) = \|\text{rad } \mathbf{x}\|$. Donc $\|\text{rad } \Delta\| \leq \gamma \|\text{rad } \mathbf{x}\|$ ce qui est équivalent à $\|\text{wid } \Delta\| \leq \gamma \|\text{wid } \mathbf{x}\|$. Donc, $\text{dist}(\mathbf{z}, \mathbf{z}') \leq n\gamma \|\text{wid } \mathbf{x}\|^2$. Finalement, nous avons prouvé que pour tout $\mathbf{x}^{\text{ref}} \in \mathbb{R}^n$, il existe $\gamma' = n\gamma$ tel que

$$\text{dist}(\mathbf{h}(\mathbf{x}), f^*(\mathbf{x})) \leq \gamma' \|\text{wid } \mathbf{x}\|^2$$

où γ' ne dépend pas du choix de \mathbf{x}^{ref} (elle dépend seulement de la dimension du problème) ce qui correspond à un ordre de convergence quadratique. \square

L'exemple suivant illustre que l'AE-extension de la valeur moyenne n'a pas un ordre de convergence quadratique en général pour $m > 1$.

Exemple IV.32. *Considérons la fonction*

$$f(x) = \begin{pmatrix} x_1 + 0.1x_2 \\ 0.1x_1 + x_2 \end{pmatrix}$$

dont les dérivées sont

$$f'(x) = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$$

Notons cette matrice Δ . Considérons la fonction $\pi : [1..2] \rightarrow [1..2]$ définie par $\pi(1) = 2$ et $\pi(2) = 1$. L'AE-extension de la valeur moyenne correspondante est

$$\begin{aligned} \mathbf{g}(\mathbf{x}) &= f(\text{mid } \mathbf{x}) + \Delta *_{\pi} (\mathbf{x} - \text{mid } \mathbf{x}) \\ &= f(\text{mid } \mathbf{x}) + \begin{pmatrix} (\text{pro } \mathbf{x}_1 - \text{mid } \mathbf{x}_1) + 0.1(\mathbf{x}_2 - \text{mid } \mathbf{x}_2) \\ 0.1(\mathbf{x}_1 - \text{mid } \mathbf{x}_1) + (\text{pro } \mathbf{x}_2 - \text{mid } \mathbf{x}_2) \end{pmatrix} \end{aligned}$$

Considérons les intervalles $\mathbf{x}_{\epsilon} = ([\epsilon, -\epsilon], [\epsilon, -\epsilon])^T$, donc

$$\mathbf{g}(\mathbf{x}_{\epsilon}) = ([-0.9\epsilon, 0.9\epsilon], [-0.9\epsilon, 0.9\epsilon])^T$$

Notons cet intervalle \mathbf{z}_{ϵ} . La proposition quantifiée suivante est donc vraie.

$$(\exists z \in \mathbf{z}_{\epsilon})(\exists x \in \text{pro } \mathbf{x}_{\epsilon})(z = f(x))$$

Étudions maintenant l'ordre de convergence de cette AE-extension. Considérons un intervalle (f, \mathbf{x}_ϵ) -interprétable minimal \mathbf{z}_ϵ^* qui est plus précis que \mathbf{z}_ϵ . Comme \mathbf{x} est impropre, au moins une composante de \mathbf{z}_ϵ^* est impropre (parce qu'un intervalle propre ne peut pas être minimal si \mathbf{x} est impropre, cf. Proposition IV.9). Nous avons donc

$$\text{dist}(\mathbf{z}_\epsilon, \mathbf{z}_\epsilon^*) \geq \|\text{rad } \mathbf{z}_\epsilon\| = 0.9 \|\text{wid } \mathbf{x}_\epsilon\|$$

Et finalement

$$\frac{\text{dist}(\mathbf{z}_\epsilon, \mathbf{z}_\epsilon^*)}{\|\text{wid } \mathbf{x}_\epsilon\|^2} \geq \frac{0.9 \|\text{wid } \mathbf{x}_\epsilon\|}{\|\text{wid } \mathbf{x}_\epsilon\|^2} = \frac{0.9}{\|\text{wid } \mathbf{x}_\epsilon\|}$$

Ce rapport n'est pas majoré quand $\epsilon \rightarrow 0$, c'est à dire quand $\text{wid } \mathbf{x}_\epsilon \rightarrow 0$. Cette AE-extension de la valeur moyenne n'a donc pas un ordre de convergence quadratique.

Cet exemple illustre aussi que le choix de la fonction π est important pour la qualité de l'AE-extension : le choix $\pi(1) = 1$ et $\pi(2) = 2$ aurait donné une AE-extension de meilleur qualité. Cependant, un choix efficace de la fonction π n'est pas toujours possible. Dans certains cas, un préconditionnement est nécessaire. L'utilisation d'un préconditionnement avec l'AE-extension de la valeur moyenne est illustré dans la partie suivante.

Cinquième partie

Applications

Chapitre 17

Approximation intérieure de l'image d'un intervalle par une fonction à valeurs vectorielles

Dans ce chapitre, l'AE-extension de la valeur moyenne est utilisée avec un procédé de préconditionnement pour construire une boîte penchée incluse dans l'image d'une boîte par une fonction continûment dérivable $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Ce problème n'est pas bien traité aujourd'hui par l'analyse d'intervalles classiques. Avant tout, le lemme suivant est nécessaire.

Lemme V.1. Soient $\mathbf{d} \in \mathbb{IR}$ et $\mathbf{x} \in \mathbb{IR}$ tel que $\text{mid } \mathbf{x} = 0$. Alors,

1. $\text{wid}(\mathbf{d} \mathbf{x}) = |\mathbf{d}|(\text{wid } \mathbf{x})$.
2. si $0 \notin \mathbf{d}$ alors $\text{wid}(\mathbf{d}(\text{dual } \mathbf{x})) = -\langle \mathbf{d} \rangle (\text{wid } \mathbf{x})$, où $\langle \mathbf{d} \rangle$ est la mignitude de \mathbf{x} , i.e. $\langle \mathbf{d} \rangle = \min\{|\underline{\mathbf{d}}|, |\overline{\mathbf{d}}|\}$ dans le cas où $0 \notin \mathbf{d}$;

Démonstration. il suffit d'appliquer les propriétés de l'arithmétique de Kaucher de la manière suivante :

1. $\mathbf{d} \mathbf{x} = [-|\mathbf{d}|\overline{\mathbf{x}}, |\mathbf{d}|\underline{\mathbf{x}}]$. Donc, $\text{wid}(\mathbf{d} \mathbf{x}) = |\mathbf{d}|(\text{wid } \mathbf{x})$.
2. D'une part, si $\mathbf{d} > 0$ alors $\mathbf{d}(\text{dual } \mathbf{x}) = [\underline{\mathbf{d}}\overline{\mathbf{x}}, \overline{\mathbf{d}}\underline{\mathbf{x}}] = \langle \mathbf{d} \rangle (\text{dual } \mathbf{x})$. D'autre part, si $\mathbf{d} < 0$ alors $\mathbf{d}(\text{dual } \mathbf{x}) = [\overline{\mathbf{d}}\underline{\mathbf{x}}, \underline{\mathbf{d}}\overline{\mathbf{x}}] = \langle \mathbf{d} \rangle (\text{dual } \mathbf{x})$ (la dernière égalité étant une conséquence de $\underline{\mathbf{x}} = -\overline{\mathbf{x}}$ et $\overline{\mathbf{d}} = -\langle \mathbf{d} \rangle$). Finalement, $\text{wid}(\langle \mathbf{d} \rangle (\text{dual } \mathbf{x})) = -\langle \mathbf{d} \rangle (\text{wid } \mathbf{x})$. □

Le théorème suivant permet la construction d'une boîte penchée incluse dans l'image d'une fonction à valeur vectorielle. Cette partie du théorème est une simple utilisation de l'AE-extension de la valeur moyenne.

Théorème V.1. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction continûment dérivable, $\mathbf{x} \in \mathbb{IR}^n$, $\tilde{x} \in \mathbf{x}$ et $C \in \mathbb{R}^{n \times n}$ une matrice régulière. Considérons de plus une matrice d'intervalles $\mathbf{A} \in \mathbb{IR}^{n \times n}$ qui vérifie $\forall x \in \mathbf{x}, f'(x) \in \mathbf{A}$. Définissons l'intervalle

$$\mathbf{u} = Cf(\tilde{x}) + (C\mathbf{A}) * \text{dual}(\mathbf{x} - \tilde{x}). \quad (17.1)$$

Si \mathbf{u} est impropre alors

$$\{C^{-1}u | u \in \text{pro } \mathbf{u}\} \subseteq \text{range}(f, \mathbf{x}).$$

De plus, si \mathbf{u} est strictement impropre alors \mathbf{A} est fortement régulière.

Démonstration. Comme C est régulière, $f(x) = z$ est équivalent à $Cf(x) = Cz$. L'équation (17.1) est l'AE-extension de la valeur moyenne de la fonction Cf évaluée en dual \mathbf{x} (la jacobienne de Cf est $Cf'(x)$). Donc si \mathbf{u} est impropre alors la proposition quantifiée suivante est vraie :

$$(\forall u \in \text{pro } \mathbf{u})(\exists x \in \mathbf{x})(u = Cf(x)).$$

Remarquons que $u = Cf(x)$ est équivalent à $C^{-1}u = f(x)$. La proposition quantifiée précédente est donc équivalente à la suivante :

$$(\forall z \in \{C^{-1}u | u \in \text{pro } \mathbf{u}\})(\exists x \in \mathbf{x})(z = f(x)),$$

c'est à dire $\{C^{-1}u | u \in \mathbf{u}\} \subseteq \text{range}(f, \mathbf{x})$. Il reste à étudier la régularité de la matrice d'intervalles \mathbf{A} . Supposons que \mathbf{u} soit strictement impropre est notons $C\mathbf{A}$ par \mathbf{A}' . La i^{ieme} ligne de l'équation (17.1) est

$$\mathbf{u}_i = (Cf(\tilde{x}))_i + \mathbf{A}'_{ii}(\text{dual } \mathbf{x}_i - \tilde{x}_i) + \sum_{j \neq i} \mathbf{A}'_{ij}(\mathbf{x}_j - \tilde{x}_j). \quad (17.2)$$

Supposons que pour $i \in [1..n]$ nous ayons $0 \in \mathbf{A}'_{ii}$. Par définition de la multiplication de Kaucher, un intervalle propre qui contient 0 multiplié par un intervalle impropre dont la projection propre contient 0 donne $[0, 0]$. Nous avons donc $\mathbf{A}'_{ii}(\text{dual } \mathbf{x}_i - \tilde{x}_i) = 0$ et donc $\mathbf{u}_i \in \mathbb{I}\mathbb{R}$. Cela contredit l'hypothèse que \mathbf{u} est strictement impropre, et par conséquent $0 \notin \mathbf{A}'_{ii}$ pour tout $i \in [1..n]$. De la même manière, nous avons $\text{wid } \mathbf{x} > 0$. Maintenant, en appliquant le Lemme V.1 à (17.2), nous obtenons

$$\text{wid } \mathbf{u}_i = 0 - \langle \mathbf{A}'_{ii} \rangle (\text{wid } \mathbf{x}_i) + \sum_{j \neq i} |\mathbf{A}'_{ij}| (\text{wid } \mathbf{x}_j).$$

En regroupant ces équations, nous obtenons

$$\text{wid } \mathbf{u} = - \langle \mathbf{A}' \rangle (\text{wid } \mathbf{x})$$

où $\langle \mathbf{A}' \rangle$ est la matrice de comparaison de \mathbf{A}' . Comme $-\text{wid } \mathbf{u} > 0$, il existe un vecteur positif $v = (\text{wid } \mathbf{x})$ qui vérifie $\langle \mathbf{A}' \rangle v \geq 0$, ce qui correspond à la définition d'une M-matrice. Comme $\langle \mathbf{A}' \rangle$ est une M-matrice, $\mathbf{A}' = C\mathbf{A}$ est une H-matrice et \mathbf{A} est finalement fortement régulière (Neumaier[8] théorème 4.1.2 et son corollaire). \square

Exemple V.1. *Considérons la fonction*

$$f(x) = \begin{pmatrix} x_1^6 + x_2^6 + x_1x_2 \\ x_1^6 - x_2^6 - x_1x_2 \end{pmatrix}$$

et l'intervalle $\mathbf{x} = ([0.99, 1.01], [0.99, 1.01])^T$. Définissons la matrice suivante :

$$\mathbf{A} = \begin{pmatrix} 6\mathbf{x}_1^5 + \mathbf{x}_2 & 6\mathbf{x}_2^5 + \mathbf{x}_1 \\ 6\mathbf{x}_1^5 - \mathbf{x}_2 & -6\mathbf{x}_2^5 - \mathbf{x}_1 \end{pmatrix} \supseteq \{f'(x) \mid x \in \mathbf{x}\}$$

Choisissons $C = (\text{mid } \mathbf{A})^{-1}$ et $\tilde{x} = \text{mid } \mathbf{x}$. L'équation (17.1) donne alors

$$\mathbf{u} \approx ([0.18, 0.16], [0.27, 0.25])^T$$

Ce qui permet de construire la boîte penchée suivante

$$\{(\text{mid } \mathbf{A})u \mid u \in \text{pro } \mathbf{u}\}$$

L'image exacte (trait continu) et la boîte penchée qui vient d'être construite (trait pointillé) sont tracées sur la Figure 17.1.

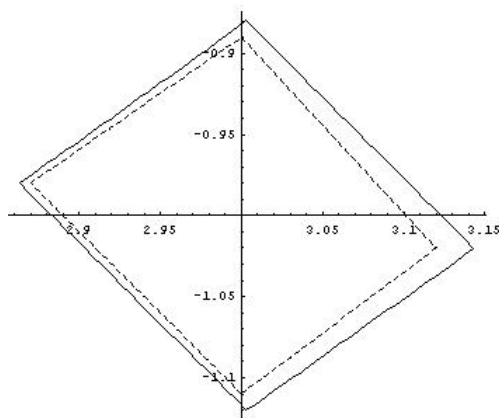


FIG. 17.1 – Approximation intérieure de l'image d'un intervalle par une fonction à valeur vectorielle.

Un test d'existence et d'unicité est facilement dérivé du Théorème V.1.

Corollaire V.1. Soient $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction continûment dérivable, $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$ un intervalle non dégénéré, $\tilde{x} \in \mathbf{x}$ et $C \in \mathbb{R}^{n \times n}$ une matrice régulière. Considérons une matrice d'intervalles $\mathbf{A} \in \mathbb{I}\mathbb{R}^{n \times n}$ qui vérifie $\forall x \in \mathbf{x}, f'(x) \subseteq \mathbf{A}$. Définissons l'intervalle

$$\mathbf{u} = Cf(\tilde{x}) + (C\mathbf{A}) * \text{dual}(\mathbf{x} - \tilde{x})$$

Si $\mathbf{u} \subseteq 0^{23}$ alors il existe $x \in \mathbf{x}$ tel que $f(x) = 0$. Si de plus \mathbf{u} est strictement impropre alors cette solution est unique.

Démonstration. L'inclusion $\mathbf{u} \subseteq 0$ est équivalente à \mathbf{u} impropre et $0 \in \text{pro } \mathbf{u}$. Le Théorème V.1 implique donc

$$\{C^{-1}u | u \in \text{pro } \mathbf{u}\} \subseteq \text{range}(f, \mathbf{x}) \quad (17.3)$$

Maintenant, $0 \in \text{pro } \mathbf{u}$ implique $0 \in \{C^{-1}u | u \in \text{pro } \mathbf{u}\}$ qui implique finalement $0 \in \text{range}(f, \mathbf{x})$. Il existe donc $x \in \mathbf{x}$ tel que $f(x) = 0$. L'unicité est prouvée en remarquant que \mathbf{u} strictement impropre implique \mathbf{A} fortement régulière, par le Théorème V.1. Alors, la forte régularité de la matrice d'intervalle \mathbf{A} implique l'unicité de la solution (cf. Neumaier[8] théorème 5.1.6). \square

Ce test d'existence est en fait équivalent au test d'existence associé à l'opérateur de Hansen-Sengupta. En effet, l'expression du test proposé dans le corollaire V.1 est

$$(\forall i \in [1..n]) \quad (Cf(\tilde{x}))_i + (C\mathbf{A})_{ii}(\text{dual } \mathbf{x}_i - \tilde{x}_i) + \sum_{j \neq i} (C\mathbf{A})_{ij}(\mathbf{x}_j - \tilde{x}_j) \subseteq 0.$$

En utilisant les règles de l'arithmétique de Kaucher, nous obtenons l'expression équivalente suivante :

$$(\forall i \in [1..n]) \quad - (Cf(\tilde{x}))_i - \sum_{j \neq i} (C\mathbf{A})_{ij}(\mathbf{x}_j - \tilde{x}_j) \subseteq \text{dual}((C\mathbf{A})_{ii})(\mathbf{x}_i - \tilde{x}_i),$$

²³C'est à dire si \mathbf{u} est impropre et si $0 \in \text{pro } \mathbf{u}$.

qui est finalement équivalente à

$$(\forall i \in [1..n]) \quad \tilde{x}_i - \frac{1}{(C\mathbf{A})_{ii}} \left((Cf(\tilde{x}))_i + \sum_{j \neq i} (C\mathbf{A})_{ij} (\mathbf{x}_j - \tilde{x}_j) \right) \subseteq \mathbf{x}_i$$

parce que $0 \notin \text{pro}(C\mathbf{A})_{ii}$. Cette dernière expression correspond au test d'existence associé à l'opérateur de Hansen Sengupta.

Chapitre 18

L'opérateur généralisé de Hansen-Sengupta

L'opérateur classique de Hansen-Sengupta permet de contracter une boîte \mathbf{x} sans perdre de solution d'une équation donnée $f(x) = 0$. Comme nous l'avons vu dans la section 5.1 (page 63), cette propriété peut s'interpréter comme l'approximation extérieure d'un "united solution set" dans lequel les intervalles associés aux paramètres sont dégénérés :

$$\Sigma(f, a, 0) = \{x \in \mathbf{x} \mid f(a, x) = 0\}$$

Ce chapitre présente une généralisation de l'opérateur de Hansen-Sengupta qui permet de prendre en comptes des intervalles non dégénérés pour les paramètres et de plus de considérer des quantifications différentes pour chaque paramètre. Ainsi, il permet de contacter une boîte \mathbf{x} sans perdre d'élément de l'"AE-solution set" $\Sigma(f, \mathbf{a}, \mathbf{b})$.

18.1 Définition de l'opérateur généralisé de Hansen-Sengupta

Nous considérons les "AE-solution sets"

$$\Sigma(f, \mathbf{a}, \mathbf{b}) = \{x \in \mathbb{R}^n \mid (\forall a_P \in \mathbf{a}_P) (\forall b_{P'} \in \mathbf{b}_{P'}) (\exists b_{I'} \in \mathbf{b}_{I'}) (\exists a_I \in \mathbf{a}_I), f(a, x) = b\}$$

où $f : \mathbb{R}^p \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ (autant d'équations que de variables) est une fonction continûment dérivable, $\mathbf{a} \in \mathbb{K}\mathbb{R}^p$ et $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$. Plus particulièrement, nous nous intéressons à $\Sigma(f, \mathbf{a}, \mathbf{b}) \cap \mathbf{x}$ où $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$ est un intervalle donné. Le lemme suivant permet de construire un "AE-solution set" linéaire qui contient le "AE-solution set" non-linéaire original. Il est une conséquence de la Proposition IV.30 page 176.

Lemme V.2. Soient $f : \mathbb{R}^p \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ une fonction continûment dérivable, $\mathbf{a} \in \mathbb{K}\mathbb{R}^p$, $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$, $\tilde{x} \in \mathbf{x}$ et $\tilde{a} \in \text{pro } \mathbf{a}$. Considérons deux matrices d'intervalles $\mathbf{J}^x \in \mathbb{I}\mathbb{R}^{n \times n}$ et $\mathbf{J}^a \in \mathbb{I}\mathbb{R}^{n \times p}$ telles que $f'_x(a, x) \in \mathbf{J}^x$ et $f'_a(a, x) \in \mathbf{J}^a$ pour tout $a \in \text{pro } \mathbf{a}$ et $x \in \mathbf{x}$. Alors,

$$\Sigma(f, \mathbf{a}, \mathbf{b}) \cap \mathbf{x} \subseteq \tilde{x} + \Sigma(\mathbf{J}^x, \mathbf{b} - f(\tilde{a}, \tilde{x}) - \mathbf{J}^a(\mathbf{a} - \tilde{a}))$$

Démonstration. Considérons $x \in \mathbf{x} \cap \Sigma(f, \mathbf{a}, \mathbf{b})$. La proposition quantifiée suivante est donc vraie :

$$(\forall a_I \in \text{pro } \mathbf{a}_I) (\forall b_I \in \text{pro } \mathbf{b}_I) (\exists b_P \in \mathbf{b}_P) (\exists a_P \in \mathbf{a}_P) (f(a, x) = b)$$

Cette proposition quantifiée implique clairement chacune des propositions suivantes, pour $k \in [1..n]$:

$$(\forall a_I \in \text{pro } \mathbf{a}_I) (\mathbf{Q}^{(\mathbf{b}_k)} b_k \in \text{pro } \mathbf{b}_k) (\exists a_P \in \mathbf{a}_P) (f_k(a, x) = b_k)$$

Ainsi, l'intervalle \mathbf{b}_k is $(f_k, \mathbf{a}, [x, x])$ -interprétable. Ou de manière équivalente, l'inclusion suivante est vraie :

$$f_k^*(\text{dual } \mathbf{a}, x) \subseteq \mathbf{b}_k.$$

L'application de la Proposition IV.30 prouve l'inclusion suivante (voir la remarque suivant la Proposition IV.30 qui permet d'utiliser $(\tilde{a}, \tilde{x})^T \notin (\text{pro } \mathbf{a}, x)^T$ parce que $\tilde{x} \vee x \subseteq \mathbf{x}$, les dérivées étant évaluées sur $(\text{pro } \mathbf{a}, \mathbf{x})^T$).

$$f_k(\tilde{a}, \tilde{x}) + \sum_{l \in [1..n]} (\text{dual } \mathbf{J}^x)_{kl} (x_l - \tilde{x}_l) + \sum_{l \in [1..p]} (\text{dual } \mathbf{J}^a)_{kl} (\text{dual } \mathbf{a}_l - \tilde{a}_l) \subseteq f_k^*(\text{dual } \mathbf{a}, x) \subseteq \mathbf{b}_k.$$

En utilisant les règles de l'arithmétique de Kaucher, nous obtenons

$$\sum_{l \in [1..n]} (\text{dual } \mathbf{J}^x)_{kl} (x_l - \tilde{x}_l) \subseteq \mathbf{b}_k - f_k(\tilde{a}, \tilde{x}) - \sum_{l \in [1..p]} (\mathbf{J}^a)_{kl} (\mathbf{a}_l - \tilde{a}_l).$$

En regroupant les inclusions précédentes, nous obtenons

$$(\text{dual } \mathbf{J}^x)(x - \tilde{x}) \subseteq \mathbf{b} - f(\tilde{a}, \tilde{x}) - \mathbf{J}^a(\mathbf{a} - \tilde{a}).$$

Finalement, en utilisant le Théorème II.10 page 69, nous obtenons

$$(x - \tilde{x}) \in \Sigma(\mathbf{J}^x, \mathbf{b} - f(\tilde{a}, \tilde{x}) - \mathbf{J}^a(\mathbf{a} - \tilde{a}))$$

ce qui conclue la démonstration. □

Remarque. Le lemme V.2 s'interprète comme une généralisation de l'opérateur de Newton aux systèmes d'équations paramétrés avec prise en compte de quantifications pour les paramètres. En effet, si \mathbf{a} est dégénéré et si $\mathbf{b} = \mathbf{0}$ alors nous retrouvons l'expression de l'opérateur classique de Newton par intervalles.

Le théorème suivant définit l'opérateur généralisé de Hansen-Sengupta. Cette généralisation est obtenue en appliquant l'opérateur généralisé de Gauss-Seidel au "AE-solution set" linéaire obtenu dans le lemme précédent.

Théorème V.2. Soient $f : \mathbb{R}^p \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ une fonction continûment dérivable, $\mathbf{a} \in \mathbb{K}\mathbb{R}^p$, $\mathbf{x} \in \mathbb{I}\mathbb{R}^n$, $\mathbf{b} \in \mathbb{K}\mathbb{R}^n$, $\tilde{x} \in \mathbf{x}$ et $\tilde{a} \in \text{pro } \mathbf{a}$. Considérons deux matrices d'intervalles $\mathbf{J}^x \in \mathbb{I}\mathbb{R}^{n \times n}$ et $\mathbf{J}^a \in \mathbb{I}\mathbb{R}^{n \times p}$ telles que $f'_x(a, x) \in \mathbf{J}^x$ et $f'_a(a, x) \in \mathbf{J}^a$ pour tout $a \in \text{pro } \mathbf{a}$ et $x \in \mathbf{x}$. Définissons l'opérateur suivant :

$$H^{AE}(\mathbf{x}, \mathbf{u}, \mathbf{v}, \tilde{x}, \tilde{u}) = \tilde{x} + \Gamma\left(C\mathbf{J}^x, C(\mathbf{b} - f(\tilde{a}, \tilde{x}) - \mathbf{J}^a(\mathbf{a} - \tilde{a}))\right), \mathbf{x} - \tilde{x}$$

En notant $H^{AE}(\mathbf{x}, \mathbf{u}, \mathbf{v}, \tilde{x}, \tilde{u})$ par \mathbf{z} ,

- si \mathbf{z} est propre alors $\Xi(f, \mathbf{u}, \mathbf{v}) \cap \mathbf{x} \subseteq \mathbf{z}$
- si \mathbf{z} n'est pas propre alors $\Xi(f, \mathbf{u}, \mathbf{v}) \cap \mathbf{x} = \emptyset$.

Démonstration. Notons l'intervalle $\mathbf{b} - f(\tilde{a}, \tilde{x}) - \mathbf{J}^a(\mathbf{a} - \tilde{a})$ par \mathbf{b}' . Alors,

$$\mathbf{z} = \Gamma(C\mathbf{J}^x, C\mathbf{b}', \mathbf{x} - \tilde{x}).$$

Grâce aux propriétés de l'opérateur généralisé de Gauss-Seidel et en utilisant le préconditionnement proposé dans Shary[96] (cf. section 5.2 page 67) si \mathbf{z} est propre alors

$$\Sigma(\mathbf{J}^x, \mathbf{b}') \cap (\mathbf{x} - \tilde{x}) \subseteq \Sigma(C\mathbf{J}^x, C\mathbf{b}') \cap (\mathbf{x} - \tilde{x}) \subseteq \mathbf{z}. \quad (18.1)$$

Maintenant, par le Lemme V.2 nous avons

$$\Sigma(f, \mathbf{a}, \mathbf{b}) \cap \mathbf{x} \subseteq \tilde{x} + \Sigma(\mathbf{J}^x, \mathbf{b}'),$$

ce qui implique

$$\begin{aligned} \Sigma(f, \mathbf{a}, \mathbf{b}) \cap \mathbf{x} &\subseteq (\tilde{x} + \Sigma(\mathbf{J}^x, \mathbf{b}')) \cap \mathbf{x} \\ &= \tilde{x} + \Sigma(\mathbf{J}^x, \mathbf{b}') \cap (\mathbf{x} - \tilde{x}). \end{aligned}$$

En utilisant l'équation (18.1), nous obtenons finalement

$$\Xi(f, \mathbf{u}, \mathbf{v}) \cap \mathbf{x} \subseteq \tilde{x} + \mathbf{z}.$$

Supposons maintenant que \mathbf{z} est impropre. Alors, grâce aux propriétés de l'opérateur généralisé de Gauss-Seidel, nous avons

$$\Sigma(\mathbf{J}^x, \mathbf{b}') \cap (\mathbf{x} - \tilde{x}) = \emptyset.$$

Par le Lemme V.2, cela implique $\Sigma(f, \mathbf{a}, \mathbf{b}) \cap \mathbf{x} = \emptyset$. □

L'opérateur $H^{AE}(\mathbf{x}, \mathbf{u}, \mathbf{v}, \tilde{x}, \tilde{u})$ est une généralisation de l'opérateur classique de Hansen-Sengupta : en effet, ce dernier est obtenu en considérant un intervalle dégénéré pour les paramètres, i.e. $\mathbf{a} \in \mathbb{R}^p$, et $\mathbf{b} = 0$. Dans ce cas,

$$\Sigma(f, a, 0) = \{x \in \mathbb{R}^n \mid f(a, x) = 0\},$$

et l'opérateur généralisé a la même interprétation que l'opérateur classique. Cependant, l'opérateur classique de Hansen-Sengupta peut être utilisé pour prouver l'existence d'une solution $x \in \mathbf{x}$ à l'équation $f(u, x) = 0$. Cette preuve d'existence n'est pas transmise à l'opérateur généralisé de Hansen-Sengupta.

18.2 Exemple

Considérons le robot parallèle à deux bras illustré par la figure 18.1. Les longueurs r_1 et r_2 des bras du robot sont les commandes. Nous supposons qu'elles sont entachées d'erreurs, et donc que les longueurs réelles sont $r_k + e_k$. Les équations du robot sont $f(x, e, r) = 0$ avec

$$f(x, e, r) = \begin{pmatrix} x_1^2 + x_2^2 - (r_1 + e_1)^2 \\ (x_1 - 10)^2 + x_2^2 - (r_2 + e_2)^2 \end{pmatrix}.$$

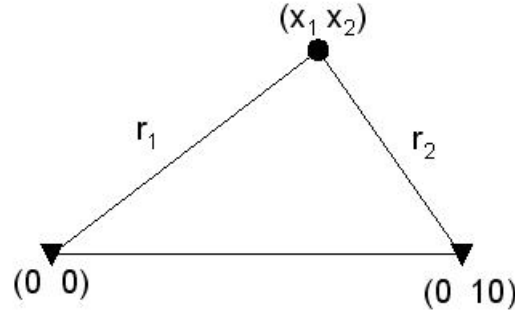


FIG. 18.1 – Le robot parallèle à deux bras.

Considérons les intervalles suivants :

$$\mathbf{x} = ([4.9, 5.1], [4.9, 5.1])^T ; \quad \mathbf{e} = ([-\epsilon, \epsilon], [-\epsilon, \epsilon])^T ; \quad \mathbf{r} = ([1, 10], [1, 10])^T,$$

où $\epsilon = 0.15$. Nous sommes intéressés par l'ensemble suivant :

$$\Sigma = \{r \in \mathbf{r} \mid \forall e \in \mathbf{e}, \exists x \in \mathbf{x}, f(x, r, e) = 0\},$$

c'est à dire, l'ensemble des commandes r pour lesquelles il existe un point de fonctionnement x dans l'intervalle \mathbf{x} , ceci quelque soient les erreurs de commande $e \in \mathbf{e}$. Il est clair que la restriction $x \in \mathbf{x}$ impose que pour r et e fixés, il n'existe qu'un unique $x \in \mathbf{x}$ vérifiant $f(x, e, r) = 0$. Et ce x correspond au point de fonctionnement du robot. Ainsi, Σ est l'ensemble des commandes r pour lesquelles le point de fonctionnement x est dans l'intervalle \mathbf{x} , ceci quelque soient les erreurs de commande $e \in \mathbf{e}$.

Pour pouvoir utiliser l'opérateur généralisé de Hansen-Sengupta, nous devons déterminer un sur-ensemble des dérivées de f pour des valeurs des variables dans leurs domaines. Nous utiliserons les matrices suivantes : la dérivée de f par rapport à r

$$\mathbf{A} = \begin{pmatrix} -(\mathbf{r}_1 + \mathbf{e}_1) & 0 \\ 0 & -(\mathbf{r}_2 + \mathbf{e}_2) \end{pmatrix} \quad (18.2)$$

et la dérivée de f par rapport à (x, e)

$$\mathbf{B} = \begin{pmatrix} 2\mathbf{x}_1 & 2\mathbf{x}_2 & -(\mathbf{r}_1 + \mathbf{e}_1) & 0 \\ 2(\mathbf{x}_1 - 10) & 2\mathbf{x}_2 & 0 & -(\mathbf{r}_2 + \mathbf{e}_2) \end{pmatrix}.$$

Maintenant, la suite $\mathbf{r}^{(k)}$ est construite avec $\mathbf{r}^{(0)} = \mathbf{r}$ et

$$\mathbf{r}^{(k+1)} = H^{AE}(\mathbf{r}^{(k)}, \mathbf{u}, 0, \text{mid } \mathbf{r}^{(k)}, \text{mid } \mathbf{u})$$

où $\mathbf{u} = (\text{dual } \mathbf{x}, \mathbf{e})^T$ et

$$H^{AE}(\mathbf{r}, \mathbf{u}, 0, \tilde{r}, \tilde{u}) = \tilde{r} + \Gamma(\mathbf{A}, -f(\text{mid } \mathbf{u}, \tilde{r}) - \mathbf{B}(\mathbf{u} - \tilde{u}), \mathbf{r} - \tilde{r}).$$

Les matrices \mathbf{A} et \mathbf{B} sont mises à jour à chaque itération en utilisant l'intervalle $\mathbf{r}^{(k-1)}$. Finalement, la matrice de préconditionnement utilisée est la matrice identité. Le quatrième élément de la suite est

$$\mathbf{r}^{(4)} = ([7.073, 7.068], [7.073, 7.068])^T$$

qui n'est pas propre. Ainsi, $\Sigma \cap \mathbf{r} = \emptyset$, c'est à dire que la proposition quantifiée suivante est vraie :

$$(\forall r \in \mathbf{r})(\exists e \in \mathbf{e})(\forall x \in \mathbf{x})(f(x, e, r) \neq 0).$$

L'erreur sur la commande est trop grande pour obtenir la précision de positionnement voulue. A notre connaissance, seules deux autres techniques peuvent montrer que cet "AE-solution set" est vide :

- l'élimination de quantificateur (cf. Davenport[46] ou Collins[15]) est restreinte à de petits systèmes qui ne mettent en jeux que des polynômes. Dans le cas de l'exemple précédent, l'implémentation de l'élimination de quantificateurs de Mathematica5.1 (cf. Weisstein[26]) ne parvient pas à prouver que le "AE-solution set" est vide : les calculs ont été stoppé après une heure sur un Pentium IV cadencé à 1.5GHz.
- l'algorithme de "continuous first-order constraint satisfaction" (voir Ratschan[83, 84]). La comparaison de ces deux techniques est difficile car elles sont très différentes. D'une part, la "continuous first-order constraint satisfaction" est plus précise car elle repose sur un algorithme de bisection, et plus générale car l'ordre des quantificateurs n'est pas contraint. D'autre part, la preuve que l'"AE-solution set" de l'exemple précédent est vide est beaucoup plus rapide en utilisant l'opérateur généralisé de Hansen-Sengupta qui ne nécessite pas de point de choix. La précision de l'opérateur généralisé de Hansen-Sengupta pourra être améliorée en l'incluant dans un algorithme de bisection.

Sixième partie

Conclusion

Depuis les années 80, deux approches ont permis d'améliorer les interprétations offertes par la théorie des intervalles classiques : d'une part, la théorie des intervalles modaux (couples intervalle classique/quantificateur) propose des interprétations enrichies pour les fonctions linéaires et non-linéaires. D'autre part, les intervalles généralisés (intervalles dont les bornes ne sont pas contraintes à être ordonnées) avec leur arithmétique, appelée l'arithmétique de Kaucher, ont permis de construire des algorithmes prometteurs dans le cadre des systèmes linéaires avec une quantification des paramètres. Si la théorie des intervalles modaux semble être prometteuse, elle reste malheureusement confidentielle à cause de sa construction originale et compliquée. Cette thèse propose d'unifier ces deux approches dans un nouveau cadre théorique, tout en les rapprochant de la théorie des intervalles classiques. Ce nouveau cadre théorique a aussi permis de proposer de nouveaux résultats intéressants.

Dans le cadre des systèmes linéaires, trois contributions ont été proposées : premièrement, nous avons proposé une nouvelle convention pour l'écriture des "AE-solution sets". Cette nouvelle convention a permis d'unifier l'écriture des "AE-solution sets" et des algorithmes qui sont dédiés à leur approximation avec la théorie des intervalles classiques. Deuxièmement, une nouvelle procédure de préconditionnement a été proposée : elle permet de construire des approximations intérieures et extérieures sous la forme de boîtes penchées, i.e. de l'image d'un intervalle par une application linéaire. Nous avons illustré les avantages apportés par ce procédé de préconditionnement compatible avec l'approximation intérieure. En particulier, le préconditionnement proposé permet de construire des approximations intérieures de "AE-solution sets" linéaires (et donc en particulier de "united solution sets" linéaires) dans des situations où aucune des techniques précédemment proposées étaient en mesure de la faire (cf. Exemple III.5 page 102). Enfin, l'opérateur généralisé de Gauss-Seidel, dédié à l'approximation extérieure des "AE-solution sets" linéaires, a été amélioré pour permettre d'être appliqué à des matrices qui ne sont pas régulières.

La contribution principale de cette thèse est la proposition d'une nouvelle formulation de la théorie des intervalles modaux. Cette nouvelle formulation est proposée dans le cadre des intervalles généralisés. Grâce à sa construction originale, la nouvelle formulation est une généralisation de la théorie des intervalles classiques. Elle conserve les aspects positifs de la théorie des intervalles modaux mais améliore ses points négatifs. Nous espérons que la nouvelle formulation de la théorie des intervalles modaux d'une part permettra son utilisation par les chercheurs de la communauté des intervalles et d'autre part facilitera son évolution future. Ainsi, nous avons déjà pu proposer une nouvelle extension de la valeur moyenne aux intervalles généralisés. Celle-ci permet de linéariser un système de manière compatible avec les interprétations enrichies proposées dans la théorie des intervalles modaux. Grâce à cette nouvelle extension, nous avons pu aborder un problème mal résolu dans le cadre de la théorie des intervalles classiques, qui est la construction d'un sous-ensemble de l'image d'un intervalle par une fonction à valeurs vectorielles. De plus, ce nouveau cadre théorique nous a permis de proposer un opérateur généralisé de Hansen-Sengupta dédié à l'approximation extérieure des "AE-solution sets" non-linéaires.

Travaux futurs

Les nouveaux résultats qui ont été proposés dans cette thèse ont été illustrés et testés sur des exemples académiques. Si leur champ d'application est prometteur il reste à les appliquer dans des situations plus réalistes. En particulier, une bibliothèque fiable de calcul pour les intervalles généralisés permettrait de promouvoir ces nouveaux outils. De plus, des essais ont été effectués qui montrent que les nouveaux résultats proposés dans cette thèse (en particulier l'opérateur généralisé de Hansen-Sengupta) permettent de résoudre dans certains cas les problèmes ouverts d'approximation intérieure des "AE-solution sets" non-linéaires lorsque la fonction est à valeurs vectorielles. Il reste néanmoins du travail à effectuer avant de pouvoir présenter ces résultats.

Du point de vue théorique, plusieurs directions peuvent être explorées : certains théorèmes de la théorie des intervalles modaux (en particulier les théorèmes de "coercion to optimality") n'ont pas été intégrés dans la nouvelle formulation de la théorie. Bien que la nouvelle extension de la valeur moyenne qui a été proposée puisse pallier ce manque, il serait certainement intéressant de comparer ces deux approches. D'autre part, l'extension de la valeur moyenne et l'opérateur généralisé de Hansen-Sengupta sont basés sur les dérivées des fonctions mises en jeu. Il serait certainement avantageux d'utiliser les pentes à la place des dérivées, mais certains développements théoriques sont nécessaires pour cette étape.

Annexe A

Un corollaire du théorème de Brouwer

Cette section présente le corollaire suivant du théorème de Brouwer. Il est d'abord illustré par l'exemple suivant.

Exemple VI.1. *Considérons $\mathbf{x}_{[1..4]} \in \mathbb{I}\mathbb{R}^4$ et la fonction continue $f_{[2..4]} : \mathbf{x}_{[1..3]} \longrightarrow \mathbf{x}_{[2..4]}$. Le corollaire suivant montre qu'il existe une fonction (pas nécessairement continue) $s_{[2..4]} : \mathbf{x}_1 \longrightarrow \mathbf{x}_{[2..4]}$ qui vérifie*

$$x_{[2..4]} = s_{[2..4]}(x_1) \implies f_{[2..4]}(x_{[1..3]}) = x_{[2..4]}$$

Cette fonction est construite de la manière suivante : le théorème de Brouwer est appliqué à la fonction $f_{\{2,3\}}(x_1, \dots) : \mathbf{x}_{\{2,3\}} \longrightarrow \mathbf{x}_{\{2,3\}}$ pour chaque valeur de $x_1 \in \mathbf{x}_1$ pour obtenir la fonction $s_{[2..3]} : \mathbf{x}_1 \longrightarrow \mathbf{x}_{[2..3]}$. Ensuite s_4 est obtenue en considérant $s_4(x_1) = f_4(x_1, s_2(x_1), s_3(x_1))$.

Corollaire VI.1. *Soit I et J des ensembles d'indices non-vides qui vérifient $I \setminus J \neq \emptyset$, et un intervalle $\mathbf{x}_{I \cup J} \in \mathbb{I}\mathbb{R}^n$ ($n = \text{card}(I \cup J)$). Considérons une fonction continue $f_J : \mathbf{x}_I \longrightarrow \mathbf{x}_J$. Alors, il existe une fonction $s_J : \mathbf{x}_{I \setminus J} \longrightarrow \mathbf{x}_J$ qui vérifie*

$$x_J = s_J(x_{I \setminus J}) \implies x_J = f_J(x_I)$$

Cette implication peut être écrite $x_J = s_J(x_{I \setminus J}) \implies y_J = f_J(x_I) \wedge x_J = y_J$. De manière équivalente, la proposition quantifiée suivante est vraie :

$$(\forall x_{I \setminus J} \in \mathbf{x}_{I \setminus J})(\exists x_J \in \mathbf{x}_J)(f(x_I) = x_J)$$

Démonstration. Définissons $A = I \setminus J \neq \emptyset$, $B = I \cap J$ et $C = J \setminus I$. Premièrement, si $B = \emptyset$ alors $A = I$ et donc $s_J = f_J$ est la fonction recherchée. Supposons maintenant que $B \neq \emptyset$. Pour tout $x_A \in \mathbf{x}_A$, le fonction continue $g_B : \mathbf{x}_B \longrightarrow \mathbf{x}_B$ définie par $g_B(x_B) = f_B(x_I)$ a un point fixe par le théorème de Brouwer. Ce point fixe dépendant de x_A , il est appelé $s_B(x_A) \in \mathbf{x}_B$. Donc $x_B = s_B(x_A) \implies f_B(x_I) = x_B$. Si $C = \emptyset$ alors $B = J$ et la démonstration est terminée. Autrement, il reste à construire $s_C : s_C(x_A) = f_C(x_A, s_B(x_A))$ qui est bien définie car $A \cup B = I$. Donc, $x_J = s_J(x_A)$ implique $x_B = s_B(x_A)$ et $x_C = s_C(x_A)$ (qui est bien définie car $C \cup B = J$ et $C \cap B = \emptyset$). Cela implique finalement $f_B(x_I) = x_B$ et $x_C = f_C(x_A, s_B(x_A)) = f_C(x_A, x_B)$, c'est à dire $x_J = f_J(x_I)$ parce que $A \cup B = I$ et $B \cup C = J$. □

Bibliographie

- [1] Goldsztejn A. A mean-value extension to generalized intervals. *Submitted for publication*.
- [2] Goldsztejn A. Modal intervals revisited. *Submitted for publication*.
- [3] Goldsztejn A. Correction of coercion to interpretability theorem in modal intervals theory. Research report, 2002.
- [4] Goldsztejn A. Modal intervals revisited : a mean-value extension to generalized intervals. In *Proceedings of QCP-2005 (Quantification in Constraint Programming), Barcelona, Spain, 2005*, 2005.
- [5] Goldsztejn A. A right-preconditioning process for the formal-algebraic approach to inner and outer estimation of AE-solution sets. *Reliable Computing*, 11(6) :443–478, 2005.
- [6] Neumaier A. Tolerance Analysis with Interval Arithmetic. *Freiburger Intervall-Berichte*, 86(9) :5–19, 1986.
- [7] Neumaier A. Overestimation in linear interval equations. *SIAM J. Numer. Anal.*, 24(1) :207–214, 1987.
- [8] Neumaier A. *Interval Methods for Systems of Equations*. Cambridge Univ. Press, Cambridge, 1990.
- [9] Neumaier A. The wrapping effect, ellipsoid arithmetic, stability and confidence regions. *Computing Supplementum*, 9 :175–190, 1993.
- [10] Neumaier A. Taylor forms - use and limits. *Reliable Computing*, 9 :43–79, 2002.
- [11] Hayes B. A Lucid Interval. *American Scientist*, 91(6) :484–488, 2003.
- [12] Edited by Hansen E. *Topics in Interval Analysis*. Oxford University Press, 1969.
- [13] Bliiek C. *Computer Methods for Design Automation*. PhD thesis, Massachusetts Institute of Technology, 1992.
- [14] Goldberg D. What every computer scientist should know about floating-point arithmetic. *Computing Surveys*, 1991.
- [15] Collins G. E. Quantifier elimination by cylindrical algebraic decomposition—twenty years of progress. In *Quantifier Elimination and Cylindrical Algebraic Decomposition*, pages 8–23, 1998.
- [16] Gardenyes E., Mielgo H., and Trepap A. Modal Intervals : Reason and Ground Semantics. *Interval Mathematics, Lecture Notes in Computer Science*, 212 :27–35, 1985.
- [17] Hansen E. On solving systems of equations using interval arithmetic. *Math. Compt.*, 22 :374–384, 1968.

- [18] Hansen E. *Global Optimization Using Interval Analysis, Second Edition, Revised and Expanded*. Marcel Dekker, NY, 1992.
- [19] Hansen E. and Sengupta S. Bounding solutions of systems of equations using interval analysis. *BIT*, 21 :203–211, 1981.
- [20] Kaucher E. *Über metrische und algebraische Eigenschaften einiger beim numerischen Rechnen auftretender Räume*. PhD thesis, Karlsruhe, 1973.
- [21] Kaucher E. Interval Analysis in the Extended Interval Space \mathbb{IR} . *Computing*, Suppl. 2 :33–49, 1980.
- [22] Loh E. and Walster G.W. Rump’s example revisited. *Reliable Computing*, 8(3) :245–248, 2002.
- [23] Popova E.D. Multiplication distributivity of proper and improper intervals. *Reliable computing*, 7(2) :129–140, 2001.
- [24] Hansen E.R. Bounding the solution of interval linear equations. *SIAM journal on numerical analysis*, 29 :1493–1503, 1992.
- [25] Kearfott R.B. et al. Standardized notation in interval analysis. 2002.
- [26] Weisstein E.W. Quantifier elimination. From MathWorld – A Wolfram Web Resource. <http://mathworld.wolfram.com/QuantifierElimination.html>.
- [27] Benhamou F. and Goualard F. Universally quantified interval constraints. In *Principles and Practice of Constraint Programming - CP 2000*, Lecture Notes in Computer Science. Springer, 2000.
- [28] Messine F. *Méthodes d’Optimisation Globale basées sur l’Analyse d’Intervalle pour la Résolution de Problèmes avec Contraintes*. PhD thesis, Institut National Polytechnique de Toulouse, 1997.
- [29] Ris F.N. *Interval analysis and applications to linear algebra*. PhD thesis, Oxford, 1972.
- [30] Preparata F.P. and Shamos M.I. *Computational geometry, an introduction*. Springer-Verlag, NY, 1985.
- [31] Alefeld G. and Herzberger J. *Introduction to Interval Computations*. Computer Science and Applied Mathematics, 1974.
- [32] Birkhoff G. and MacLane S. *A Survey of Modern Algebra. rev. edn.* Macmillan, 1953.
- [33] Mayer G. and Warnke I. On the Fixed Points of the Interval Function $f([x]) = [a][x] + [b]$. *Linear Algebra and its Applications*, 363 :201–216, 2003.
- [34] Jürgen Garloff. Convergent bounds for the range of multivariate polynomials. In *Interval Mathematics*, Lecture Notes in Computer Science, pages 37–56, 1985.
- [35] Walster G.H. The extended real interval system. Available on the internet, <http://www.mscs.mu.edu/globsol/readings.html>., 1998.
- [36] Walster G.H. Impty intervali. Available on the internet, <http://www.mscs.mu.edu/globsol/readings.html>., 1998.
- [37] SIGLA/X group. Modal intervals (basic tutorial). *Applications of Interval Analysis to Systems and Control (Proceedings of MISC’99)*, pages 157–227, 1999.
- [38] SIGLA/X group. Modal intervals. *Reliable Computing*, 7 :77–111, 2001.

- [39] Collavizza H., Delobel F., and Rueher M. Comparing partial consistencies. *Reliable Computing*, 1 :1–16, 1999.
- [40] Collavizza H., Delobel F., and Rueher M. Extending consistent domains of numeric csp. In *Proceedings of IJCAI-99, Stockholm, Sweden, 31 July - 6 August, 1999*.
- [41] Ratschek H. Teilbarkeitskriterien der Intervallarithmetik. *Journal für die reine und angewandte Mathematik*, Bd. 252 :128–138, 1972.
- [42] Ortoft H.-J. Eine Verallgemeinerung der Intervallarithmetik. *Gesellschaft fuer Mathematik und Datenverarbeitung, Bonn*, 11 :1–71, 1969.
- [43] T. Hickey, Q. Ju, and M. H. Van Emden. Interval arithmetic : From principles to implementation. *Journal of the ACM*, 48(5) :1038–1068, 2001.
- [44] H. Hong and V. Stahl. Bernstein form is inclusion monotone. *Computing*, Vol. 55(1) :43–53, 1995.
- [45] Armengol J., Vehí J., Travé-Massuyès L., and Sainz M.A. Application of modal intervals to the generation of error-bounded envelopes. *Reliable Computing*, 7(2) :171–185, 2001.
- [46] Davenport J. and Heintz J. Real quantifier elimination is doubly exponential. *J. Symb. Comput.*, 5 :29–35, 1988.
- [47] Dixmier J. *Cours de mathématiques du premier cycle - 1re année*. Dunod, 1994.
- [48] Dixmier J. *Cours de mathématiques du premier cycle - 2e année*. Dunod, 1994.
- [49] Vehí J., Sainz M.A., Armengol J., and Ferrer I. Applications of modal interval analysis to systems and control. In *Current trends in Qualitative Reasoning and Applications*, pages 49–64. Universidad de Sevilla, 2000.
- [50] Wolff von Gudenberg J. Determination of minimum sets of the set of zeros of a function. *computing*, 24(2–3) :203–212, 1980.
- [51] Muller J.M. Ordinateurs en quête d’arithmétique. *La Recherche*, 278 :772–777, 1995.
- [52] Muller J.M. Ordinateurs en quête d’arithmétique. *La Recherche*, hors-série Août 1999 :90–96, 1999.
- [53] W. Kühn. Rigorously Computed Orbits of Dynamical Systems Without the Wrapping Effect. *Computing*, 61 :47–67, 1998.
- [54] L. Kupriyanova. Inner estimation of the united solution set of interval linear algebraic system. *Reliable Computing*, 1(1) :15–31, 1995.
- [55] Jaulin L., M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis with Examples in Parameter and State Estimation, Robust Control and Robotics*. Springer-Verlag, 2001.
- [56] Christie M. *Spécification de trajectoires de caméra sous contraintes*. PhD thesis, Université de Nantes, 2003.
- [57] Jirstand M. Nonlinear control system design by quantifier elimination. *Journal of Symbolic Computation*, 24(2) :137–152, 1997.
- [58] Spivak M. *Calculus*. Addison-Wesley, 1967.
- [59] Warmus M. Calculus of Approximations. *Bulletin de l’Académie Polonaise des Sciences*, 4(5) :253–257, 1956.

- [60] Sainz M.A., Gardenyes E., and Jorba L. Formal Solution to Systems of Interval Linear or Non-Linear Equations. *Reliable computing*, 8(3) :189–211, 2002.
- [61] Sainz M.A., Gardenyes E., and Jorba L. Interval Estimations of Solutions Sets to Real-Valued Systems of Linear or Non-Linear Equations. *Reliable computing*, 8(4) :283–305, 2002.
- [62] Sainz M.A., Armengol J., and Vehí J. Fault Detection and Isolation of the Three-Tank System Using Modal Interval Analysis. *Journal of Process Control*, 12(2) :325–338, 2002.
- [63] Svetoslav Markov. On the algebra of intervals and convex bodies. *J. UCS*, 4(1) :34–47, 1998.
- [64] K Nickel. How to fight the wrapping effect. In *Proceedings of the International Symposium on interval mathematics on Interval mathematics 1985*, pages 121–132. Springer-Verlag, 1985.
- [65] Dimitrova N.S., Markov S.M., and Popova E.D. Extended interval arithmetics : New results and applications. *Computer Arithmetics and Enclosure Methods*, pages 225–232, 1992.
- [66] Beaumont O. *Algorithmique pour les Intervalles*. PhD thesis, Université de Rennes 1, 1999.
- [67] Beaumont O. Solving interval linear systems with oblique boxes. *Publication interne IRISA*, 2000. (<http://www.irisa.fr/sage/publications/Category/RR.html>).
- [68] Knueppel O. PROFIL/BIAS - A Fast Interval Library. *Computing*, 53(3-4) :277–287, 1994.
- [69] Herrero P., M.A. Sainz, Vehí J., and Jaulin L. Quantified set inversion with applications to control. In *IEEE International Symposium on Computer Aided Control Systems Design*, 2004.
- [70] Herrero P., M.A. Sainz, Vehí J., and Jaulin L. Quantified set inversion algorithm with applications to control. In *Proceedings of Interval Mathematics and Constrained Propagation methods, Novosibirsk, 2004*, volume 11(5) of *Reliable Computing*, 2005.
- [71] Schodl P. Multi-valued mappings and their zeros. Diploma thesis, Faculty of Mathematics, University of Vienna, Vienna, Austria, 2005.
- [72] Krawczyk R. Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehler-schranken. *Computing*, 4 :187–201, 1969.
- [73] Krawczyk R. and Neumaier A. Interval slopes for rational functions and associated centered forms. *SIAM J. Numer. Anal.*, 22 :604–616, 1985.
- [74] Moore R. *Interval analysis*. Prentice-Hall, 1966.
- [75] Stefan Ratschan. Approximate quantified constraint solving by cylindrical box decomposition. *Reliable Computing*, 8(1) :21–42, 2002.
- [76] Stefan Ratschan. Continuous first-order constraint satisfaction with equality and disequality constraints. In Pascal van Hentenryck, editor, *Proc. 8th International Conference on Principles and Practice of Constraint Programming*, number 2470 in Lecture Notes in Computer Science, pages 680–685. Springer, 2002.

- [77] Stefan Ratschan. Solving existentially quantified constraints with one equality and arbitrarily many inequalities. In Francesca Rossi, editor, *Proceedings of the Ninth International Conference on Principles and Practice of Constraint Programming*, number 2833 in Lecture Notes in Computer Science, pages 615–633. Springer, 2003.
- [78] Kearfott R.B. Interval computation : Introduction, uses, and resources. *Euromath, Bulletin* 2(1) :95–112, 1996.
- [79] Kearfott R.B. Interval extensions of non-smooth functions for global optimization and nonlinear systems solvers. *Computing*, 57(2) :149–162, 1996.
- [80] Kearfott R.B. *Rigorous Global Search : Continuous Problems*. Kluwer, 1996.
- [81] Kearfott R.B., M. Novoa, and C. Hu. A Review of Preconditioners for the Interval Gauss-Seidel Method. *Interval Computations*, 1(1) :59–85, 1991.
- [82] Gelbaum R.G. and Olmsted J.M.H. *Counterexamples in Analysis*. Dover Publications, 2003.
- [83] Ratschan S. Continuous first-order constraint satisfaction. In *Proceedings of Artificial Intelligence and Symbolic Computation*, Lecture Notes in Computer Science. Springer, 2002.
- [84] Ratschan S. et al. Rsolver. <http://www.mpi-sb.mpg.de/~ratschan/rsolver>, 2004. Software Package.
- [85] Zuhe S. and Wolfe M. A. On interval enclosures using slope arithmetic. *Appl. Math. Comput.*, 39(1) :89–105, 1990.
- [86] S. P. Shary. On Controlled Solution set of Interval Algebraic Systems. *Interval Computations*, 4(6) :66–75, 1992.
- [87] Markov S.M. Extended interval arithmetic and some applications. *Freiburger Intervall-Berichte, Univ. Freiburg*, 78/4 :1–12, 1978.
- [88] Markov S.M. Isomorphic embeddings of abstract interval systems. *Reliable Computing*, 3 :199–207, 1997.
- [89] Markov S.M. On the algebraic properties of convex bodies. *Pliska Stud. Math. Bulgar.*, 12 :119–132, 1998.
- [90] Markov S.M., Popova E.D., and Ullrich Ch. On the solution of linear algebraic equations involving interval coefficients. *Iterative Methods in Linear Algebra, IMACS Series on Computational and Applied Mathematics*, 3 :216–225, 1996.
- [91] Rump S.M. Algorithms for verified inclusions - theory and practice. *Perspectives in Computing*, 19 :109–126, 1988.
- [92] Shary S.P. Algebraic approach to the interval linear static identification, tolerance and control problems, or one more application of Kaucher arithmetic. *Reliable computing*, 2 :3–33, 1996.
- [93] Shary S.P. Algebraic Solutions to Interval Linear Equations and their Application. In *Proceedings of the IMACS - GAMM International Symposium on Numerical Methods and Error Bounds, Oldenburg, 1995*. Editors : G. Alefeld and J. Herzberger, 1996.
- [94] Shary S.P. Algebraic approach in the "outer problem" for interval linear equations. *Reliable computing*, 3 :103–135, 1997.
- [95] Shary S.P. Interval Gauss-Seidel method for generalized solution sets to interval linear systems. *Reliable computing*, 7 :141–155, 2001.

- [96] Shary S.P. A new technique in systems analysis under interval uncertainty and ambiguity. *Reliable computing*, 8 :321–418, 2002.
- [97] Sunaga T. Theory of interval algebra and its application to numerical analysis. In *RAAG Memoirs, Ggujutsu Bunken Fukuy-kai*, 2 :29–46, 1958.
- [98] Kreinovich V., Lakeyev A., Rohn J., and Kahl P. *Computational Complexity and Feasibility of Data Processing and Interval Computations*. Kluwer Academic Publishers, 1998.
- [99] Ménessier-Morain V. Arbitrary precision real arithmetic : design and algorithms. Submitted to the Journal of Symbolic Computation, September 1996. (http://www-calfor.lip6.fr/vmm/documents/submission_JSC.ps.gz).
- [100] Stahl V. *Interval Methods for Bounding the Range of Polynomials and Solving Systems of Nonlinear Equations*. PhD thesis, University of Linz, Austria, 1995.
- [101] Istratescu V.I. *Fixed point theory : an introduction*. Mathematics and its applications. D. Reidel, 1981.