



**HAL**  
open science

# Analyse globale de systèmes mécaniques non-linéaires - Application à la dynamique des rotors

Emmanuelle Sarrouy

► **To cite this version:**

Emmanuelle Sarrouy. Analyse globale de systèmes mécaniques non-linéaires - Application à la dynamique des rotors. Mécanique [physics.med-ph]. Ecole Centrale de Lyon, 2008. Français. NNT : . tel-00366857

**HAL Id: tel-00366857**

**<https://theses.hal.science/tel-00366857>**

Submitted on 9 Mar 2009

**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.

**THÈSE**

Présentée pour l'obtention du titre de

**DOCTEUR**

**DE**

**L'ÉCOLE CENTRALE DE LYON**

**SPÉCIALITÉ GÉNIE MÉCANIQUE**

**ÉCOLE DOCTORALE DE MÉCANIQUE DE LYON (UCBL/INSA/ECL)**

par

**Emmanuelle SARROUY**

**ANALYSE GLOBALE DE SYSTÈMES MÉCANIQUES NON-LINÉAIRES  
APPLICATION À LA DYNAMIQUE DES ROTORS**

Présentée et soutenue publiquement le 22 octobre 2008 devant le jury d'examen :

<b>M. B. COCHELIN</b> , Professeur, LMA, Ecole Centrale de Marseille	Rapporteur
<b>M. J-C. GOLINVAL</b> , Professeur, LTAS, Université de Liège, Belgique	Rapporteur
<b>M. J-P. LAINE</b> , Maître de Conférences, LTDS, École Centrale de Lyon	Co-encadrant
<b>M. C-H. LAMARQUE</b> , Professeur, LGM, ENTPE, Lyon	Président
<b>M. A. QUENARDEL</b> , Ingénieur de recherche et développement, Snecma	Examinateur
<b>M. F. THOUVEREZ</b> , Professeur, LTDS, École Centrale de Lyon	Directeur





**ECOLE CENTRALE DE LYON**  
**DIRECTION DE LA RECHERCHE**

**Liste des personnes Habilitées à Diriger des Recherches en poste à l'Ecole Centrale de Lyon**

Nom-Prénom	Corps grade	Laboratoire ou à défaut département ECL	Etablissement
AURIOL Philippe	professeur	AMPERE	ECL
BEROUAL Abderrahmane	professeur	AMPERE	ECL
BURET François	maître de conférences	AMPERE	ECL
JAFFREZIC-RENAULT Nicole	directeur de recherche	AMPERE	CNRS/ECL
KRÄHENBÜHL Laurent	directeur de recherche	AMPERE	CNRS/ECL
MARTELET Claude	professeur	AMPERE	ECL
NICOLAS Alain	professeur	AMPERE	ECL
NICOLAS Laurent	directeur de recherche	AMPERE	CNRS/ECL
SIMONET Pascal	chargé de recherche	AMPERE	CNRS/ECL
THOMAS Gérard	professeur	AMPERE	ECL
VOLLAIRE Christian	maître de conférences	AMPERE	ECL

**Nbre Ampère 11**

HELLOUIN Yves	maître de conférences	DER EEA	ECL
LE HELLEY Michel	professeur	DER EEA	ECL

**Nbre DER EEA 2**

GUIRALDENQ Pierre	professeur émérite	DER STMS	ECL
VINCENT Léo	professeur	DER STMS	ECL

**Nbre DER STMS 2**

LOHEAC Jean-Pierre	maître de conférences	ICJ	ECL
MAITRE Jean-François	professeur émérite	ICJ	ECL
MARION Martine	professeur	ICJ	ECL
MOUSSAOUI Mohand	professeur	ICJ	ECL
MUSY François	maître de conférences	ICJ	ECL
ROUY MIRONESCU Elisabeth	professeur	ICJ	ECL
ZINE Abdel-Malek	maître de conférences	ICJ	ECL

**Nbre ICJ 7**

DAVID Bertrand	professeur	ICTT	ECL
----------------	------------	------	-----

**Nbre ICTT 1**

CALLARD Anne-Ségolène	maître de conférences	INL	ECL
CLOAREC Jean-Pierre	maître de conférences	INL	ECL
GAFFIOT Frédéric	professeur	INL	ECL
GAGNAIRE Alain	maître de conférences	INL	ECL
GARRIGUES Michel	directeur de recherche	INL	CNRS/ECL
GENDRY Michel	directeur de recherche	INL	CNRS/ECL
GRENET Geneviève	directeur de recherche	INL	CNRS/ECL
HOLLINGER Guy	directeur de recherche	INL	CNRS/ECL
JOSEPH Jacques	professeur	INL	ECL
KRAWCZYK Stanislas	directeur de recherche	INL	CNRS/ECL
MARTIN Jean-René	professeur émérite	INL	ECL

O'CONNOR Ian	maître de conférences	INL	ECL
PHANER-GOUTORBE Magali	professeur	INL	ECL
ROBACH Yves	professeur	INL	ECL
SEASSAL Christian	chargé de recherche	INL	CNRS/ECL
SOUTEYRAND Eliane	directeur de recherche	INL	CNRS/ECL
TARDY Jacques	directeur de recherche	INL	CNRS/ECL
VIKTOROVITCH Pierre	directeur de recherche	INL	CNRS/ECL

**Nbre INL 18**

CHEN Liming	professeur	LIRIS	ECL
-------------	------------	-------	-----

**Nbre LIRIS 1**

BAILLY Christophe	professeur	LMFA	ECL
BERTOGLIO Jean-Pierre	directeur de recherche	LMFA	CNRS/ECL
BLANC-BENON Philippe	directeur de recherche	LMFA	CNRS/ECL
BOGEY Christophe	chargé de recherche	LMFA	CNRS/ECL
CAMBON Claude	directeur de recherche	LMFA	CNRS/ECL
CARRIERE Philippe	chargé de recherche	LMFA	CNRS/ECL
CHAMPOUSSIN J-Claude	professeur émérite	LMFA	ECL
COMTE-BELLOT geneviève	professeur émérite	LMFA	ECL
FERRAND Pascal	directeur de recherche	LMFA	CNRS/ECL
GALLAND Marie-Annick	maître de conférences	LMFA	ECL
GODEFERD Fabien	chargé de recherche	LMFA	CNRS/ECL
HENRY Daniel	directeur de recherche	LMFA	CNRS/ECL
JEANDEL Denis	professeur	LMFA	ECL
JUVE Daniel	professeur	LMFA	ECL
LE RIBAUT Catherine	chargée de recherche	LMFA	CNRS/ECL
LEBOEUF Francis	professeur	LMFA	ECL
PERKINS Richard	professeur	LMFA	ECL
ROGER Michel	professeur	LMFA	ECL
SCOTT Julian	professeur	LMFA	ECL
SHAO Liang	chargé de recherche	LMFA	CNRS/ECL
SIMOENS Serge	chargé de recherche	LMFA	CNRS/ECL
TREBINJAC Isabelle	maître de conférences	LMFA	ECL

**Nbre LMFA 22**

BENAYOUN Stéphane	professeur	LTDS	ECL
CAMBOU Bernard	professeur	LTDS	ECL
COQUILLET Bernard	maître de conférences	LTDS	ECL
DANESCU Alexandre	maître de conférences	LTDS	ECL
FOUVRY Siegfried	chargé de recherche	LTDS	CNRS/ECL
GEORGES Jean-Marie	professeur émérite	LTDS	ECL
GUERRET Chrystelle	chargé de recherche	LTDS	CNRS/ECL
HERTZ Dominique	past	LTDS	ECL
ICHCHOU Mohamed	maître de conférences	LTDS	ECL
JEZEQUEL Louis	professeur	LTDS	ECL
JUVE Denyse	ingénieur de recherche	LTDS	ECL
KAPSA Philippe	directeur de recherche	LTDS	CNRS/ECL
LE BOT Alain	chargé de recherche	LTDS	CNRS/ECL
LOUBET Jean-Luc	directeur de recherche	LTDS	CNRS/ECL
MARTIN Jean-Michel	professeur	LTDS	ECL
MATHIA Thomas	directeur de recherche	LTDS	CNRS/ECL
MAZUYER Denis	professeur	LTDS	ECL
PERRET-LIAUDET Joël	maître de conférences	LTDS	ECL
SALVIA Michelle	maître de conférences	LTDS	ECL
SIDOROFF François	professeur	LTDS	ECL
SINOUE Jean-Jacques	maître de conférences	LTDS	ECL
STREMSDOERFER Guy	professeur	LTDS	ECL
THOUVEREZ Fabrice	professeur	LTDS	ECL

TREHEUX Daniel	professeur	LTDS	ECL
VANNES André-Bernard	professeur émérite	LTDS	ECL

**Nbre LTDS 25**

**Total HdR ECL 89**



# Remerciements

Je tiens à remercier tout d'abord mes deux directeurs de recherche, Fabrice Thouverez et Jean-Pierre Laîné, pour m'avoir proposé ce travail, m'accordant alors plus de confiance que je ne m'en accordais moi-même, puis pour m'avoir soutenue parfois, guidée, écoutée toujours et pour avoir pris le temps - et parfois le plaisir j'espère - de réfléchir avec moi autour de deux ou trois schémas rapidement esquissés sur un tableau noir ou un coin de feuille.

Je tiens ensuite à adresser de sincères remerciements à Bruno Cochelin et Jean-Claude Golinval qui ont tous deux accepté de rapporter un travail à la limite de notre domaine usuel de réflexion. Je remercie aussi vivement Claude-Henri Lamarque qui fut un président de jury des plus agréables et dont les remarques m'auront été infiniment utiles. Je les remercie aussi pour l'enthousiasme qu'ils ont bien voulu témoigner à l'égard de mes recherches ; il est réconfortant et revigorant d'entendre de la part de personnes que l'on estime que ce travail n'est pas trop mauvais et aurait peut-être quelque intérêt à être communiqué et poursuivi.

Un chaleureux merci aussi à tous les membres de l'équipe D2S parmi lesquels j'ai trouvé des amis, des compagnons de footing, des professeurs de mécanique version cambouis et clé de douze, certains se retrouvant dans plusieurs de ces catégories, en tous les cas des personnes agréables et prêtes à aider, peu importent la nature et la taille du problème.

Enfin, j'aimerais remercier mes proches, amis et famille, et plus particulièrement mes parents qui ont su me donner durant mes jeunes années la confiance ainsi que le goût d'apprendre et de surmonter les difficultés.



# Résumé

La naissance de ce sujet de thèse a été motivée par la volonté de plus en plus importante d'introduire des non-linéarités dans les modèles afin de mieux rendre compte du comportement de structures toujours plus complexes. Ceci aboutit à l'écriture de systèmes dynamiques non-linéaires pouvant présenter des solutions de nature complexe (solutions périodiques mais aussi quasi-périodique voire chaotique) et pouvant coexister. L'enjeu d'une modélisation étant la prédiction du comportement du système pour son dimensionnement notamment, il est nécessaire de pouvoir prédire toutes les solutions vers lesquelles pourra tendre la structure en fonctionnement.

Actuellement peu d'outils ou de méthodes sont dédiés à ce type d'analyse, dite globale, permettant d'exhiber toutes les solutions d'un système dynamique non-linéaire. Le parti est généralement pris de réaliser « un grand nombre de tirages » à l'aide d'algorithmes de recherche locale. Ceci peut amener à trouver effectivement toutes les solutions mais ne permet pas de garantir qu'on en possède la totalité et est généralement très coûteux numériquement. Notre principal objectif était donc de combler cette lacune en tentant de proposer des réponses satisfaisant à la fois au but théorique mais aussi à des contraintes pratiques liées aux ressources numériques requises pour les mettre en œuvre.

Pour y parvenir, nous avons mis en œuvre quatre types de méthodes : des méthodes de cell-mapping qui consiste en une discrétisation de l'espace d'états et du temps, des méthodes basées sur des tests d'exclusion appliqués à une cellule, des méthodes utilisant l'arithmétique des intervalles de façon à réduire et exclure des cellules et enfin, des méthodes requérant une approximation polynomiale des non-linéarités et une résolution du système par homotopie. Ces différentes méthodes ont été illustrées par le cas académique de l'oscillateur de Duffing, présentant une non-linéarité simple de façon à pouvoir les comparer, puis nous les avons appliquées à différents systèmes relatifs à la modélisation de la dynamique des machines tournantes.

Parallèlement à ce but principal, nous avons étudié des techniques de continuation - permettant de suivre l'évolution d'une solution lorsqu'un des paramètres du système varie - ainsi que les méthodes d'analyse de stabilité et de bifurcation dans le cas des solutions constantes et périodiques. Ce second volet permet par exemple d'établir les fonctions de réponse en fréquence de structures non-linéaires sur lesquelles on fait apparaître les cycles limites en cas d'instabilité.

## Mots-clés

Non-linéaire, Analyse globale, Cell-mapping, Tests d'exclusion, Analyse intervalles, Homotopies, Rotors (dynamique des)



# Abstract

Birth of this PhD thesis subject was motivated by the growing will to introduce nonlinearities in models ; this is done in order to have a better description of more and more complex structures behaviour. It leads to nonlinear dynamic systems which can exhibit solutions of complex nature (periodic but also quasiperiodic or even chaotic ones) and in which multiple solutions can coexist. Prediction of the structure behaviour being at stake in modelling, in particular in view of its dimensioning, one should be able to predict all of the solutions the system can hang to.

Nowadays few tools or methods are dedicated to this kind of analysis, named global analysis, which are able to find all the solutions of a nonlinear dynamic system. One usually tests an « important » number of different initial conditions using a local search algorithm. This can lead to effectively find all the solutions but does not ensure it and is very often a costly approach according to numerical resources. Our main goal was then to fill this lack of methods and to propose answers satisfying both the theoretical request and the practical constraints bound to numerical resources they need.

In order to reach this goal, we implemented four kind of methods : cell-mapping which consists in state-space and time discretization, test exclusion algorithms which try to exclude cells using a simple test, methods based on interval analysis which aim at reducing as well as excluding cells and finally methods requiring a polynomial approximation of the nonlinearities expressions which solve the system using homotopy. These different methods were applied to a simple Duffing oscillator in order to compare them. We then applied them to many examples chosen in the field of rotordynamics.

In parallel of this main goal continuation techniques were studied - techniques which allow to follow solution evolution while one of the system parameters varies ; stability and bifurcation analysis methods were also implemented in case of constant and periodic solutions. This second part of our work allows to draw frequency response of nonlinear structures showing also limit cycles if instability occurs.

## Keywords

Nonlinear, Global analysis, Cell-mapping, Exclusion test, Interval analysis, Homotopy, Rotordynamics



# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Phénoménologie des systèmes non-linéaires</b>	<b>3</b>
1.1 Quelques définitions pour bien commencer . . . . .	4
1.2 Solutions d'un système non-linéaire : caractérisation et représentation . . . . .	4
1.2.1 Catégories de solution . . . . .	5
1.2.2 Stabilité des solutions . . . . .	5
1.2.3 Attracteurs et bassin d'attraction . . . . .	6
1.2.4 Les outils de visualisation . . . . .	7
1.2.5 Classification des solutions . . . . .	11
1.3 Notions de continuation et bifurcations . . . . .	14
1.3.1 Bifurcations . . . . .	14
1.3.2 Diagramme de bifurcation . . . . .	17
1.4 Conclusion : enjeux . . . . .	17
<b>2 Méthodes d'étude classiques (locales)</b>	<b>19</b>
2.1 Algorithmes d'intégration temporelle . . . . .	19
2.1.1 Notions essentielles pour qualifier un schéma . . . . .	20
2.1.2 Cas des systèmes linéaires . . . . .	21
2.1.3 Cas des systèmes non-linéaires . . . . .	23
2.1.4 Synthèses des différentes méthodes trouvées dans la littérature . . . . .	25
2.2 Algorithmes de résolution de systèmes non-linéaires . . . . .	25
2.2.1 Algorithmes sans gradient . . . . .	28
2.2.2 Algorithmes avec gradient . . . . .	28
2.3 Construire et étudier des solutions . . . . .	29
2.3.1 Solutions type point fixe . . . . .	29
2.3.2 Solutions périodiques . . . . .	29
2.4 Algorithmes de continuation . . . . .	34
2.4.1 Principales méthodes de prédiction et de correction . . . . .	35
2.4.2 Adaptation de la distance entre deux points successifs . . . . .	39
2.4.3 Branch switching . . . . .	40
2.5 Méthode de réduction de la taille du problème . . . . .	43
2.6 Conclusion . . . . .	44

<b>3</b>	<b>Analyse globale par cell-mapping</b>	<b>45</b>
3.1	Construction d'un mapping . . . . .	45
3.2	Discrétisation de l'espace . . . . .	47
3.3	Analyse d'un cell-mapping . . . . .	48
3.4	Interprétation d'un cell-mapping . . . . .	52
3.5	Application à un oscillateur de Duffing . . . . .	52
3.6	Conclusion . . . . .	57
<b>4</b>	<b>Analyse globale par exclusion de cellules</b>	<b>59</b>
4.1	Etapes de l'algorithme général . . . . .	60
4.2	Tests d'exclusion simples . . . . .	61
4.2.1	Fonction lipschitzienne . . . . .	62
4.2.2	Fonction monotonement décomposable . . . . .	62
4.3	Tests basés sur la domination de fonction . . . . .	63
4.3.1	Fonctions développables en série entière . . . . .	63
4.3.2	Développement de Taylor . . . . .	64
4.4	Méthode du simplexe . . . . .	66
4.4.1	Problème traité par une méthode de simplexe : forme et résolution . . . . .	66
4.4.2	Application de la méthode à un système non-linéaire . . . . .	71
4.5	Application à un oscillateur de Duffing . . . . .	72
4.5.1	Résultats . . . . .	73
4.5.2	Extrapolations . . . . .	74
4.6	Conclusion . . . . .	75
<b>5</b>	<b>Analyse globale utilisant l'arithmétique des intervalles</b>	<b>79</b>
5.1	Théorie des intervalles . . . . .	79
5.2	L'algorithme général d'Hansen et Walster . . . . .	80
5.2.1	Étape de Newton . . . . .	81
5.2.2	Méthode de Hull Consistency . . . . .	83
5.2.3	Méthode de Box Consistency . . . . .	84
5.2.4	Méthode de Newton Généralisée . . . . .	85
5.2.5	Point d'expansion . . . . .	88
5.2.6	Méthode de Box Consistency Partiel . . . . .	89
5.2.7	Division des cellules (splitting) . . . . .	89
5.2.8	Critères d'arrêt . . . . .	90
5.3	Application à un oscillateur de Duffing . . . . .	91
5.3.1	Résultats . . . . .	91
5.3.2	Extrapolations . . . . .	92
5.4	Conclusion . . . . .	95

<b>6</b>	<b>Analyse globale par homotopies polynomiales</b>	<b>99</b>
6.1	Passage d'un système algébrique non-linéaire à un système multi-polynomial . . . . .	101
6.2	Nature des solutions d'un système multi-polynomial . . . . .	102
6.3	Principe et difficultés d'une résolution par méthode homotopique . . . . .	102
6.4	Espace projectif et homogénéisation . . . . .	103
6.5	Choix d'un polynôme initial . . . . .	106
6.5.1	Méthode du degré total . . . . .	107
6.5.2	Méthode par multi-homogénéisation . . . . .	108
6.5.3	Méthode par produits d'éléments linéaires . . . . .	111
6.5.4	Comparaison des trois méthodes . . . . .	114
6.6	Application à un oscillateur de Duffing . . . . .	119
6.6.1	Résultats . . . . .	119
6.6.2	Extrapolations . . . . .	119
6.7	Conclusion . . . . .	122
<b>7</b>	<b>Application à la dynamique des rotors</b>	<b>125</b>
7.1	Comparaisons des méthodes sur un cas simple . . . . .	125
7.2	Contact permanent rotor-stator avec stator non-linéaire . . . . .	126
7.2.1	Équations . . . . .	127
7.2.2	Application des méthodes d'analyse globale . . . . .	130
7.2.3	Conclusion . . . . .	131
7.3	Rotor sur palier squeeze-film . . . . .	131
7.3.1	Généralités sur l'organe squeeze-film . . . . .	131
7.3.2	Système dynamique étudié . . . . .	133
7.3.3	Quelques résultats classiques . . . . .	135
7.3.4	Analyse globale par homotopies polynomiales . . . . .	135
7.3.5	Conclusion . . . . .	140
7.4	Contact rotor-stator rigide . . . . .	141
7.4.1	Système étudié . . . . .	141
7.4.2	Étude des points fixes . . . . .	148
7.4.3	Étude des points de bifurcation de Hopf . . . . .	156
7.4.4	Analyse globale . . . . .	163
7.4.5	Conclusion . . . . .	169
7.5	Conclusion . . . . .	169
	<b>Conclusion</b>	<b>171</b>
	<b>Bibliographie</b>	<b>173</b>
	<b>Table des figures</b>	<b>177</b>

---

<b>Liste des tableaux</b>	<b>181</b>
<b>Liste des définitions</b>	<b>183</b>
<b>Liste des lemmes et théorèmes</b>	<b>185</b>
<b>Nomenclature</b>	<b>187</b>
<b>Annexes</b>	<b>189</b>
<b>A Oscillateur de Duffing</b>	<b>191</b>
A.1 Application de la méthode des échelles multiples . . . . .	192
A.2 Application de la méthode de la balance harmonique . . . . .	195
A.3 Fonction de réponse en fréquence et stabilité . . . . .	196
<b>B Arithmétique des intervalles</b>	<b>199</b>
B.1 Définition des opérations élémentaires pour les intervalles . . . . .	199
B.2 Illustration des conséquences de la sous-distributivité . . . . .	200
B.3 Schéma détaillé de l'algorithme proposé par Hansen et Walster . . . . .	201

# Introduction

L'utilisation d'outils numériques pour modéliser et prédire le comportement dynamique de systèmes complexes apparaît aujourd'hui comme incontournable dans le cadre du développement de produits industriels ou la conception de bancs expérimentaux. L'utilisation de tels outils permet en effet de réduire les coûts et les délais en minimisant le nombre de corrections apportées à un prototype avant d'obtenir le produit final, l'optimisation, en faisant varier des jeux de paramètres pour sélectionner le meilleur, et de prédire de façon fiable les résonances, les contraintes au sein de la structure, etc. Les méthodes relatives à la modélisation linéaire d'une structure ont été largement étudiées ces cinquante dernières années et sont aujourd'hui bien maîtrisées. On dispose ainsi d'outils permettant de réaliser l'analyse du comportement dynamique de systèmes de grandes tailles dans des délais raisonnables.

Cependant, force est de constater que la nature n'a pas engendré des systèmes destinés à simplifier la tâche des mécaniciens et que l'on doit recourir parfois à une modélisation plus complexe pour lui rester fidèle. La volonté industrielle d'une part et plus généralement celle des dynamiciens d'autre part, est aujourd'hui de prendre en compte cet écart des systèmes réels aux modèles linéaires que l'on met en miroir afin de prédire mieux leur comportement. En effet, la recherche de performances toujours accrues pousse les concepteurs à aller vers des structures plus légères, donc plus flexibles, travaillant aux limites de leurs capacités mécaniques et à l'introduction d'organes complexes ayant pour but d'assurer la meilleure cohérence de l'ensemble. Deux grandes voies permettent de prendre cet écart en compte. La première consiste en l'utilisation d'un modèle linéaire dont certains paramètres comportent des incertitudes qui reflètent alors une gamme de comportements possibles que l'on ne connaît pas exactement sous forme d'une loi entrée-sortie mais sous forme d'une répartition probable d'états. La seconde est de choisir d'inclure dans le modèle une loi entrée-sortie non-linéaire dont les paramètres sont connus. Dans le premier cas, on obtient un ensemble de comportements simples (en ce sens qu'ils découlent d'une analyse linéaire) probables et leur probabilité de réalisation ; dans le second, on obtient le comportement certain de la structure en accord avec la loi implémentée, comportement qui peut se révéler complexe. Notons que rien n'empêche de croiser les deux approches, postulant que l'on connaît exactement le modèle (non-linéaire) des corps utilisés et que seuls subsistent des incertitudes sur quelques paramètres qui les caractérisent (incertitudes liées souvent à la réalisation de l'objet, comme la dispersion autour des cotes nominales, l'hétérogénéité du module d'Young d'un corps métallique qui peut contenir des impuretés, etc). Rien ne l'empêche théoriquement ; la pratique en revanche connaît actuellement certaines limites, notamment liées aux coûts de calcul, qui rendent rare cette approche.

Ce travail s'inscrit dans le cadre de la seconde approche : nous essaierons de caractériser toutes les solutions de systèmes non-linéaires déterministes. L'emploi du verbe « essayer »

découle de la volonté de trouver *toutes* les solutions des systèmes non-linéaires considérés. En effet, la difficulté majeure dans l'étude de tels systèmes provient du fait que l'on ne peut prédire la nature (périodique, chaotique, ...) ni le nombre des solutions recherchées alors que l'on sait que pour une même excitation, plusieurs solutions de natures identiques ou différentes peuvent coexister et que le système se dirige vers l'une ou l'autre en fonction des conditions initiales. On connaît aujourd'hui un certain nombre de techniques numériques ou analytiques permettant de trouver *une* solution particulière du système. L'objectif de nos recherches est de les trouver toutes de façon à garantir un dimensionnement avisé de la structure : nous souhaitons contribuer à l'élaboration d'une méthode d'analyse globale des systèmes dynamiques non-linéaires.

Le champ de l'analyse globale, bien que relativement récent, n'est pas complètement vierge. Son essor a globalement suivi celui des capacités de calcul numérique dont il est très dépendant. En effet, les seules méthodes de résolution numérique exacte (aux erreurs d'arrondis près) dont nous disposons actuellement s'appliquent uniquement à des systèmes linéaires. Le champ des systèmes non-linéaires requiert ainsi généralement l'itération d'une même étape un « grand » nombre de fois. Celui de l'analyse globale des systèmes non-linéaires multiplie encore le nombre d'itérations nécessaires. On comprend donc aisément que le principal enjeu de l'analyse globale, outre le fait de reposer sur une méthode capable théoriquement de garantir l'obtention de toutes les solutions, sera de le faire « vite », i.e. dans des temps raisonnables pour les enjeux considérés (12 à 24 heures). On trouve actuellement quelques méthodes d'analyse globale dans la littérature ; elles s'appliquent le plus souvent à des systèmes de petite taille (moins d'une dizaine d'inconnues) et sous forme algébrique. Afin d'atteindre notre objectif, il faudra faire évoluer ces méthodes de façon à pouvoir traiter des systèmes d'équations différentielles de grande taille (plus d'une centaine d'inconnues).

Nous commencerons donc par exposer dans un premier chapitre, les spécificités des systèmes non-linéaires par rapport au cas linéaire. Nous donnerons dans un second chapitre les principales méthodes d'étude locale répertoriées dans la littérature, vouées à la recherche et à la continuation d'une solution particulière. Nous explorerons ensuite dans les chapitres 3 à 6 différentes méthodes d'étude globale appliquées à l'oscillateur non-linéaire classique qu'est l'oscillateur de Duffing. Cette première application permettra de dégager les avantages et inconvénients de chaque méthode ainsi que leur comparaison dans le cadre d'un dernier chapitre. Cette dernière partie contiendra aussi l'application de certaines des méthodes à des systèmes choisis pour leur intérêt dans l'étude des structures tournantes. Ceci nous permettra d'affiner la comparaison et de dégager la méthode qui donne le plus de satisfaction.

# Chapitre 1

## Phénoménologie des systèmes non-linéaires

---

*Lors de la mise en équation d'un système mécanique, on peut être amené à introduire des non-linéarités dans le système si l'on souhaite traiter des grands déplacements, introduire une loi de comportement d'un matériau non-élastique, ou plus localement introduire un organe au comportement non-linéaire (roulements, amortisseurs squeeze-film, etc) ou enfin, considérer des efforts intérieurs de type contact-perte de contact. On aboutit alors à une équation de la forme (1.1), contenant une contribution linéaire (matrices de masse  $M$ , de « dissipation »  $C$  pouvant contenir outre les effets dissipatifs, une contribution anti-symétrique due aux effets gyroscopiques et de raideur  $K$  pouvant elle aussi contenir au delà des termes de raideur classiques des contributions telles que les effets d'assouplissement), une contribution non-linéaire ( $\hat{f}(q, \dot{q})$ ) et une excitation du système ( $f_e(t)$ ).*

$$M.\ddot{q} + C.\dot{q} + K.q + \hat{f}(q, \dot{q}) = f_e(t) \quad (1.1)$$

*La présence d'une contribution non-linéaire empêche l'application du strict principe de superposition (bien qu'un certain nombre de travaux tels que [SP94] et [PSP01] aient mis en place des méthodes de superposition approchée). Ainsi, même en ne considérant que des excitations simples, i.e. périodiques, on ne peut établir de cadre général permettant la résolution du système d'équations. On ne sait pas a priori la forme que la solution aura, si elle sera unique ou si plusieurs solutions coexisteront, ni prédire simplement l'évolution de la solution lorsqu'un des paramètres du système change (par exemple, la fréquence ou l'amplitude de l'excitation).*

Dans ce chapitre, nous proposons dans un premier temps de décrire les solutions possibles d'un système non-linéaire, leur caractérisation formelle et leur classification, ainsi que les outils utilisés pour les représenter après avoir fait un petit détour par quelques définitions formelles. Nous aborderons dans un second temps les notions de continuation et bifurcation, liées à l'étude de l'évolution des solutions lorsque l'on fait varier un paramètre du système. Pour une description plus détaillée de ces différents aspects, nous recommandons les ouvrages [NB95] et [Sey88].

## 1.1 Quelques définitions pour bien commencer

Lorsque l'on écrit les équations d'un système mécanique, on obtient classiquement un système différentiel d'ordre 2 de la forme (1.1). Cette écriture n'est pas la plus compacte ni la plus adaptée à l'étude mathématique du système puisqu'elle fait apparaître de façon dissociée la position  $q(t)$  et la vitesse  $\dot{q}(t)$  qui sont toutes deux nécessaires pour caractériser complètement l'état d'un système, en laissant sous-jacente leur relation de dérivée par rapport au temps. On aura donc le plus souvent recours à la forme augmentée de cette équation, aussi nommée *équation d'états* où ces deux informations sont regroupées dans un unique vecteur d'inconnues  $x(t)$  appelé vecteur d'inconnues augmenté ou *vecteur d'états* :

$$A.\dot{x} + B.x + \hat{F}(x) = F_e(t) \quad (1.2)$$

Nous renvoyons à la section Notations (p. 187) pour la définition des termes qui la constituent. On utilisera aussi par commodité de notation la forme (1.3) qui suppose, par rapport à la forme (1.2), l'inversibilité de  $A$ .

$$\dot{x} = F(x, t) \quad (1.3)$$

Soulignons que la présence du temps comme variable dans  $F$  reflète uniquement l'excitation du système. Ainsi, lorsque nous considérerons la dérivée de  $F$  par rapport à  $x$ , celle-ci ne dépendra que de  $x$  et sera notée  $D_x F(x)$ .

Outre ces différentes écritures, il est bon de connaître la définition ci-dessous, faisant la différence entre les systèmes dépendant ou non explicitement du temps :

### Définition 1.1 : Système autonome ou non-autonome

Un système est dit autonome si le temps n'apparaît pas explicitement comme variable dans l'équation (1.3). A l'inverse, on parle de système non-autonome.  $\square$

Maintenant que ces bases sont jetées, nous pouvons aborder les différents aspects d'un problème non-linéaire.

## 1.2 Solutions d'un système non-linéaire : caractérisation et représentation

Nous donnerons dans cette partie les moyens d'identifier et caractériser les différentes formes de solutions d'un système non-linéaire. Nous ne donnerons de caractérisation mathé-

matique que dans le cas des solutions dites point fixe et périodique car ce sont les deux types de solution les plus fréquemment rencontrés et sur lesquels nous focaliserons généralement notre étude.

### 1.2.1 Catégories de solution

On distingue quatre types de solutions à un problème non-linéaire : les solutions de type point fixe, ou point d'équilibre, qui sont constantes dans le temps, les solutions périodiques qui consistent en un motif se reproduisant à une fréquence  $f$  constante, les solutions quasi-périodiques qui sont une superposition de signaux périodiques de fréquences incommensurables (à rapport non rationnel) entre elles, et enfin les solutions chaotiques dont il est difficile de donner une définition claire si ce n'est par opposition aux trois autres types d'états stationnaires précédemment mentionnés ([NB95, Chap. 5]).

Formellement, ceci se traduit de la façon suivante pour les points fixes et les solutions périodiques :

- Points fixes :  $q(t)$  solution de (1.1) tel que  $\forall t > 0, \dot{q}(t) = 0$ . On note alors  $q(t) = q_0$ .
- Solutions périodiques :  $q(t)$  solution de (1.1) tel que  $\exists T > 0, \forall t > 0, q(t + T) = q(t)$ . On note alors  $f = 1/T$  la fréquence de la solution et  $\omega = 2\pi f$  sa pulsation.

Il peut sembler peu pertinent d'envisager que la solution d'un système soumis à une excitation périodique soit constante. Cela peut pourtant être le cas et ne signifie pas que la structure ne vibre pas mais que les paramètres et le référentiel choisis pour décrire son mouvement sont constants. Ceci peut être le cas notamment lors d'une formulation autonome du problème.

### 1.2.2 Stabilité des solutions

Une fois que l'on possède une solution, il est nécessaire d'en connaître la stabilité. En effet, si plusieurs solutions de natures identiques ou différentes coexistent et que l'une est instable, il y a toutes les chances pour que le système saute de cette solution instable à une solution stable. Le fait de connaître toutes les solutions et leur stabilité permet de prédire ces sauts qui peuvent être violents et amener le système hors de la zone de fonctionnement souhaitée.

Plusieurs définitions de la stabilité sont fréquemment rencontrées. Nous donnons ici la plus utilisée d'une façon générale, qui est celle de Lyapunov (1892, Russie), désignée aussi sous le terme de stabilité uniforme :

#### Définition 1.2 : Stabilité au sens de Lyapunov ou uniforme

Une solution  $x(t)$  du problème (1.3) est dite stable au sens de Lyapunov si et seulement si : quel que soit  $\epsilon > 0$ , il existe  $\delta > 0$  tel que pour toute solution  $y(t)$  satisfaisant  $\|x(t_0) - y(t_0)\| < \delta$  à l'instant  $t = t_0$ , la relation  $\|x(t) - y(t)\| < \epsilon$  est vérifiée pour tout instant ultérieur ( $\forall t > t_0$ ).  $\square$

La définition de la stabilité la plus communément utilisée en mécanique et que nous utiliserons aussi est celle de stabilité asymptotique, raffinement de la stabilité uniforme : le système légèrement perturbé à l'instant  $t_0$  revient vers la solution  $x(t)$  au bout d'un temps infini.

**Définition 1.3 : Stabilité asymptotique**

Une solution  $x(t)$  est dite asymptotiquement stable si et seulement si elle est stable (au sens de Lyapunov) et  $\lim_{t \rightarrow \infty} \|x(t) - y(t)\| = 0$ .  $\square$

Ces définitions sont relatives à la stabilité locale : elles s'intéressent à l'évolution d'une solution. On parle de stabilité globale lorsque l'on considère la stabilité de l'ensemble des solutions ; nous n'aborderons pas ce problème dans ce mémoire.

Classiquement, pour déterminer la stabilité d'une solution  $x(t)$ , on utilise une technique de perturbation en considérant l'évolution d'une solution du problème qui s'écrirait  $y(t) = x(t) + z(t)$ ,  $z(t)$  « petit ». Ainsi on peut réaliser le développement au premier ordre suivant :

$$\begin{aligned} \dot{y}(t) &= F(y(t), t) \\ \Leftrightarrow \dot{x}(t) + \dot{z}(t) &= F(x(t), t) + D_x F(x(t)).z(t) + O(\|z(t)\|^2) \end{aligned}$$

En réinjectant la relation  $\dot{x} = F(x, t)$  qui traduit le fait que  $x(t)$  est solution de l'équation, on obtient :

$$\dot{z}(t) \approx D_x F(x(t)).z(t) \tag{1.4}$$

On a donc l'équation différentielle qui gouverne l'évolution temporelle d'une petite perturbation  $z(t)$  de la solution  $x(t)$  étudiée. Si toute petite perturbation tend à s'évanouir ( $\lim_{t \rightarrow \infty} \|z(t)\| = 0$ ), la solution est stable, sinon, elle est instable. Nous verrons dans les cas particuliers des points fixes et des solutions périodiques comment conclure facilement sur la stabilité à partir de cette équation en § 1.3.1.

Il est intéressant de remarquer qu'une solution instable constitue un attracteur si l'on considère les temps indirects (sens décroissant de  $t$ ) : en effet, si une fois la solution perturbée, le système s'écarte d'une solution instable dans le sens des temps croissants (ou directs), à l'inverse, dans le sens des temps décroissants, il y revient alors qu'il s'écarterait d'une solution stable. Ceci est illustré en figure 1.1 où un point fixe instable  $A$  est représenté sous forme de portrait de phase en (a) et sous forme du déplacement fonction du temps sur la partie droite (graphiques (b) et (c)). Dans le sens des temps directs, on parcourt la spirale de  $A$  vers  $B$ , s'éloignant du point fixe, il est donc instable ; cependant, en parcourant la courbe dans le sens du temps décroissant, on parcourt la spirale de  $B$  vers  $A$ . Temporellement, au lieu d'obtenir un signal qui diverge dans le sens du temps croissant comme illustré en (b), on converge vers un point fixe dans le sens décroissant du temps (graphique (c)). Cette remarque est parfois utilisée pour arriver à retrouver des solutions instables ([MBC03]) : au lieu d'intégrer l'équation différentielle dans le sens du temps croissant comme on le fait couramment, on programme l'algorithme pour l'intégrer dans le sens des temps décroissants de façon à reconstruire des cycles limites, des points fixe, etc, difficiles voire impossibles à tracer par une intégration temporelle directe.

**1.2.3 Attracteurs et bassin d'attraction**

Les solutions que nous considérons ne sont généralement pas atteintes dès l'instant initial. En effet, le système part de conditions initiales qui correspondent rarement à celles d'une des solutions particulières puis, après un régime transitoire, va se « fixer » sur l'une des solutions ; il

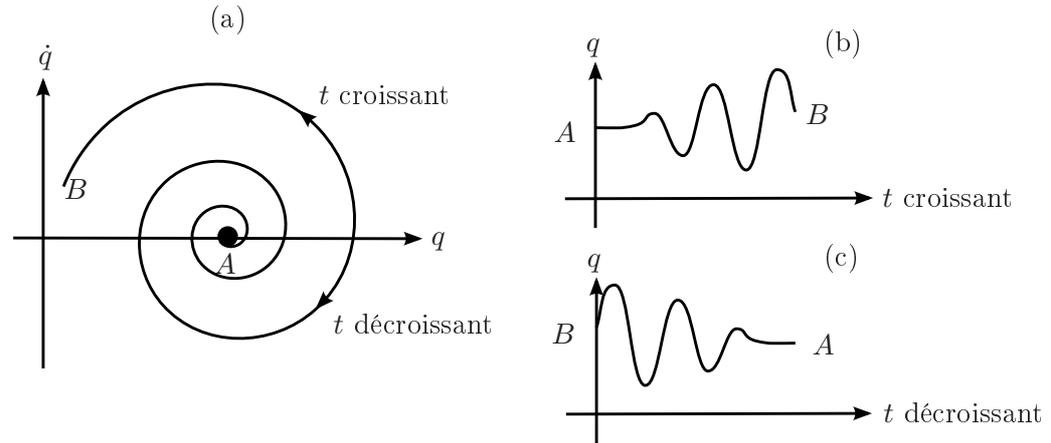


FIG. 1.1 – Illustration des temps direct et indirect

est attiré par une des solutions stables que l'on nomme aussi attracteur. Il est ainsi intéressant de définir le bassin d'attraction de chacune des solutions, c'est à dire l'ensemble des conditions initiales  $(q(0), \dot{q}(0))$  telles qu'au bout d'un certain temps, le système rejoindra cette solution. Pour le concepteur qui doit dimensionner une structure, cette information lui permet par exemple de connaître la perturbation maximale que le système, sur son régime stationnaire, peut supporter sans quitter son bassin d'attraction, c'est-à-dire à l'issue de laquelle il reviendra à son régime initial et non vers un autre.

Ce type de graphique est coûteux à obtenir dans la mesure où il nécessite d'avoir testé (donc intégré temporellement) un grand nombre de conditions initiales (idéalement une infinité). De plus, sa représentation est malaisée et oblige à procéder par plan de coupe lorsque le système est constitué de plus d'un degré de liberté (ddl).

Pour illustration, nous donnons la figure 1.2 qui représente le bassin d'attraction simplifié d'un oscillateur de Duffing obtenu grâce à une méthode de cell-mapping (cf. Sec. 3.5). Sur cette figure, on lit que trois solutions  $S_1$ ,  $S_2$  et  $S_3$  coexistent. Puis, on lit que toutes les trajectoires prenant leurs conditions initiales sur la zone blanche convergeront vers  $S_1$  : cette zone constitue le bassin d'attraction de  $S_1$ . De la même façon, toutes les trajectoires prenant leurs conditions initiales dans la zone rouge convergeront vers  $S_3$ . Enfin,  $S_2$  se trouve sur la frontière entre les bassins de  $S_1$  et de  $S_3$ , c'est-à-dire sur une séparatrice. Cette solution est instable.

#### 1.2.4 Les outils de visualisation

Afin d'identifier des solutions au comportement temporel complexe, il est judicieux d'avoir recours aux outils suivants : les portraits de phase, les sections de Poincaré et les spectres fréquentiels dans le cas des solutions périodiques ou quasi-périodiques. En effet, le tracé du signal au cours du temps ne permet bien souvent pas d'en dégager les grandeurs caractéristiques.

##### Portrait de phase

Le portrait de phase consiste en général à tracer des courbes  $(q(t), \dot{q}(t))$  paramétrées par la variable temporelle  $t$  pour différentes conditions initiales  $(q(0), \dot{q}(0))$ .  $q(t)$  est fréquemment

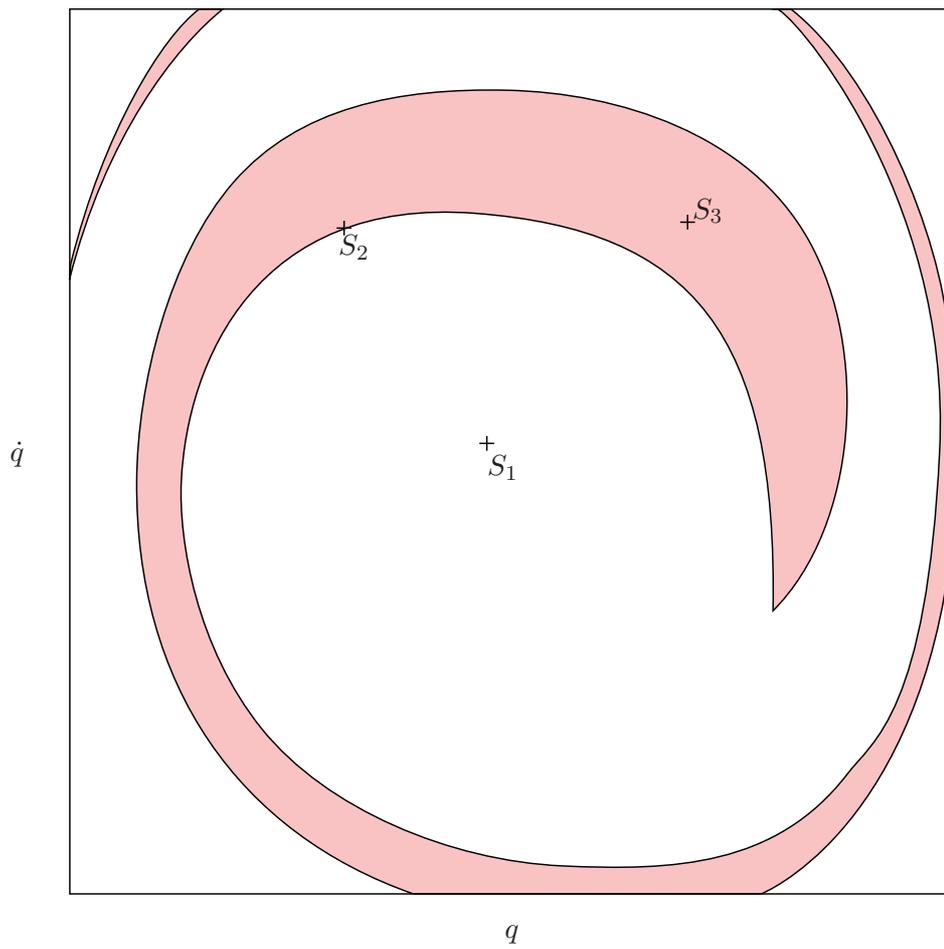


FIG. 1.2 – Bassin d'attraction schématique

un vecteur à  $N$  composantes. Il faut alors pour être rigoureux, tracer  $N$  portraits à partir de chacune des  $N$  coordonnées  $q_i(t)$ . On obtient alors un réseau de courbes fléchées dans le sens des temps croissants, qui vont converger vers des lieux que l'on dit attracteurs ou s'éloigner d'autres lieux que l'on nomme alors répulseurs ou solutions instables (si la représentation est faite pour des durées  $t$  assez longues). Ces lieux qui ressortent sont des solutions. Un exemple de portrait de phase est proposé en figure 1.3 où l'on visualise deux trajectoires issues de conditions initiales différentes (points  $B$  et  $C$ ) venant « s'enrouler » vers un point fixe  $A$  dont on déduit qu'il est stable.

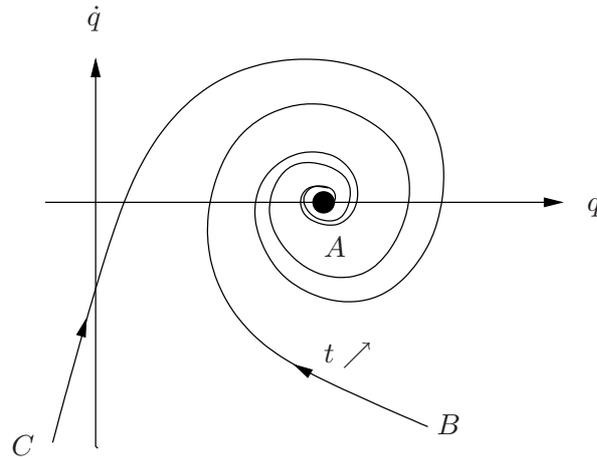


FIG. 1.3 – Portrait de phase schématisé

### Section et application de Poincaré

Une section de Poincaré consiste le plus souvent à relever  $(q(t), \dot{q}(t))$  ou  $x(t)$  à des instants  $t$  régulièrement espacés.

Ainsi, si  $q(t)$  est une solution périodique de période identique à la période d'excitation  $T_e$  et si l'échantillonnage se fait à des instants espacés de  $T_e$ , on obtient toujours le même point. Si la solution a une période égale à  $2T_e$  et que l'échantillonnage est fait tous les  $T_e$ , on obtiendra deux points distincts. Pour une solution quasi-périodique à deux fréquences maîtresses dont l'une est celle de l'excitation (comme c'est fréquemment le cas après une bifurcation de Hopf), on obtiendra une courbe fermée, et pour une solution chaotique, un amas de points. Ces sections de Poincaré types sont illustrées par la figure 1.4.

Une autre façon de construire une section de Poincaré, beaucoup moins rencontrée dans la littérature est de regarder l'intersection de  $x(t)$  avec un hyperplan  $\Sigma$  de l'espace d'états que l'on aura choisi. La démarche n'est donc plus de maîtriser l'intervalle de temps séparant deux relevés consécutifs, mais de maîtriser le lieu de l'espace d'états où ces points sont relevés. Ceci n'est pertinent que dans le cas des systèmes à formulation autonome.

Ce relevé de points effectué de l'une ou l'autre des façons peut permettre de construire une application, nommée application de Poincaré définie comme suit :

#### Définition 1.4 : Application de Poincaré

Dans le cas d'un système non-autonome, nous nommerons application de Poincaré l'applica-

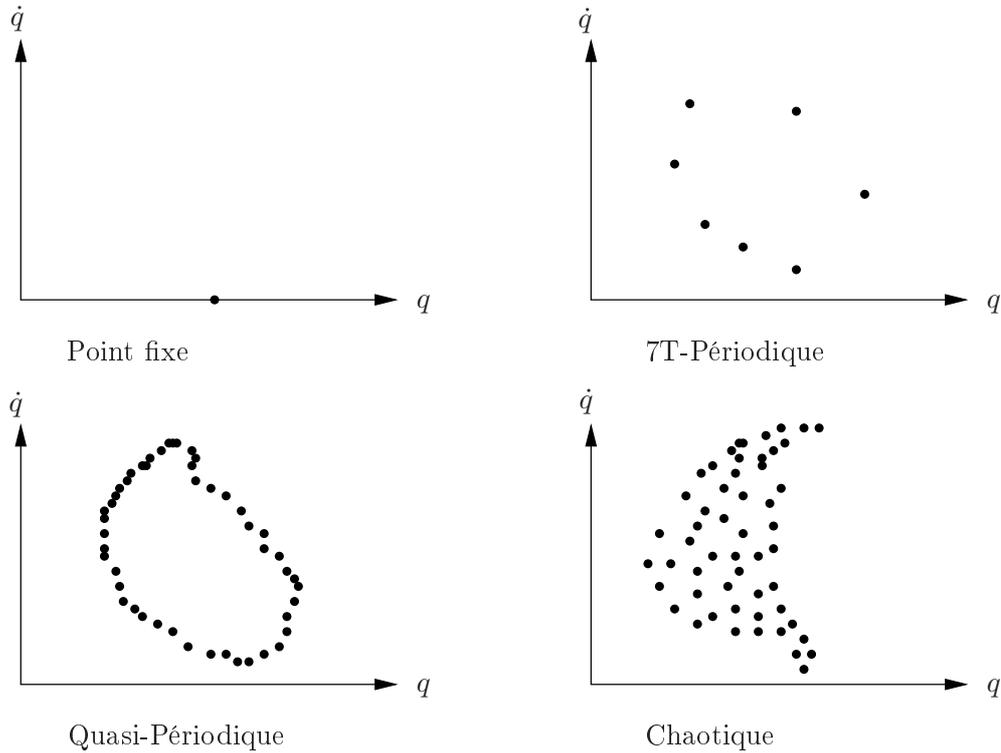


FIG. 1.4 – Sections de Poincaré schématiques

tion :

$$\begin{aligned} \mathcal{P} : \mathbb{R}^{2N} &\rightarrow \mathbb{R}^{2N} \\ x_0 &\rightarrow \mathcal{P}(x_0) = x(t_p), \quad x(t) \text{ étant solution de (1.3)} \\ &\quad \text{pour la condition initiale } x(0) = x_0 \end{aligned}$$

avec  $t_p$  un temps fixé (souvent pris égal à  $T_e$ , période d'excitation du système).

Dans le cas d'un système autonome, nous nommerons application de Poincaré l'application :

$$\begin{aligned} \mathcal{P} : \mathbb{R}^{2N-1} &\rightarrow \mathbb{R}^{2N-1} \\ x_0 &\rightarrow \mathcal{P}(x_0), \text{ première intersection avec l'hyperplan } \Sigma \text{ de } x(t) \\ &\quad \text{solution de (1.3) de condition initiale } x(0) = x_0 \end{aligned}$$

où  $\Sigma$  est un hyperplan de l'espace d'états ( $\mathbb{R}^{2N}$ ) que l'on s'est donné.

Dans ce second cas, l'application de Poincaré est aussi nommée *application de premier retour*.  $\square$

L'intérêt d'une telle application est que l'on diminue d'un la taille de l'espace dans lequel on travaille (élimination du temps  $t \in \mathbb{R}$  dans le premier cas, d'une dimension de l'espace d'états par ajout d'une équation dans le second). De plus, on peut démontrer que la stabilité des solutions de cette fonction est équivalente à la stabilité des solutions de l'équation différentielle à partir de laquelle elle a été construite ([Kuz04, Sec. 1.5]).

### Spectres fréquentiels

Un dernier outil d'analyse du signal est le tracé du spectre fréquentiel. Ceci est très pratique dans le cas de l'analyse de solutions périodiques et quasi-périodiques car il fait

ressortir nettement les contributions fréquentielles majeures. Une façon très simple d'obtenir un spectre fréquentiel à partir d'un signal expérimental ou numérique (échantillonné dans les deux cas) est le tracé de la transformée de Fourier discrète (ou DFT pour Discrete Fourier Transform) qui est définie comme suit :

**Définition 1.5 : Transformée de Fourier discrète (DFT)**

Soit  $q(t_k)$ ,  $k \in \{0, \dots, n\}$  les valeurs de la solution étudiée à des instants  $t_k$  constituant un échantillonnage de pas  $h$  de la durée  $T = n \times h$  en  $n + 1$  points. La transformée de Fourier discrète de ce signal vaut :

$$\mathcal{F}_q(\nu_j) = \frac{1}{n} \sum_{k=0}^{n-1} q(t_k) e^{-2i\pi\nu_j t_k}, \quad \nu_j = \frac{j}{T}, \quad j \in \{0, \dots, n-1\}$$

□

Le module de  $\mathcal{F}_q(\nu_j)$  donne la contribution en amplitude de la fréquence  $\nu_j$  associée au signal  $q(t)$ . Il est ainsi facile d'identifier les solutions périodiques dont le spectre ne comporte des pics qu'en des multiples d'une fréquence fondamentale et les solutions quasi-périodiques dont les spectres font apparaître des pics aux multiples des fréquences maîtresses et de leurs combinaisons linéaires.

La figure 1.5 illustre l'intérêt d'un tel outil. Le signal temporel représenté en (a) vaut  $q(t) = 0.4 + \cos(2\pi t) + 0.2 \cos(4\pi t) + \cos(\sqrt{2}2\pi t)$ . La présence de deux contributions fréquentielles incommensurables ( $1$  et  $\sqrt{2}$ ) rend le déchiffrement du signal temporel impossible. En revanche, la figure (b) représente la norme de sa DFT (pour une période d'étude  $T = 10s$  et  $n = 500$  points d'échantillonnage). On observe une symétrie des valeurs autour de  $f = 25Hz$ ; cette symétrie provient du fait que  $\nu_j = \bar{\nu}_{n-j}$ . La figure (c) réalise un zoom sur les premiers pics et permet bien de retrouver (avec la précision permise par la discrétisation,  $1/T$ ) les différentes fréquences mises en jeu et la hauteur de leur contribution :  $\nu_0 = 0$  Hz pour le terme constant,  $\nu_1 = 1$  Hz et  $\nu_2 = \sqrt{2} \approx 1.4$  Hz.

### 1.2.5 Classification des solutions

On a pour habitude de classer les solutions en points hyperboliques et non-hyperboliques. Ceci donne lieu à un certain nombre de noms qu'il est bon de connaître pour pouvoir comprendre la littérature. Nous nous limitons à la classification des points fixes et des solutions périodiques qui constitueront la majeure partie de notre étude. Nous mentionnons entre parenthèses les appellations anglophones dans la mesure où la totalité des ouvrages que nous avons considérée est ainsi rédigée.

#### Classification des points fixes

Soit  $x_0$  un point fixe de l'équation (1.3). On notera  $\lambda_i$ ,  $i = 1..2N$  les valeurs propres de la jacobienne  $D_x F(x_0)$  de  $F$  en ce point. Si toutes les valeurs propres  $\lambda_i$  ont une partie réelle non nulle, le point fixe est dit hyperbolique; sinon, il est dit non-hyperbolique. L'arborescence ci-dessous permet de raffiner les cas.

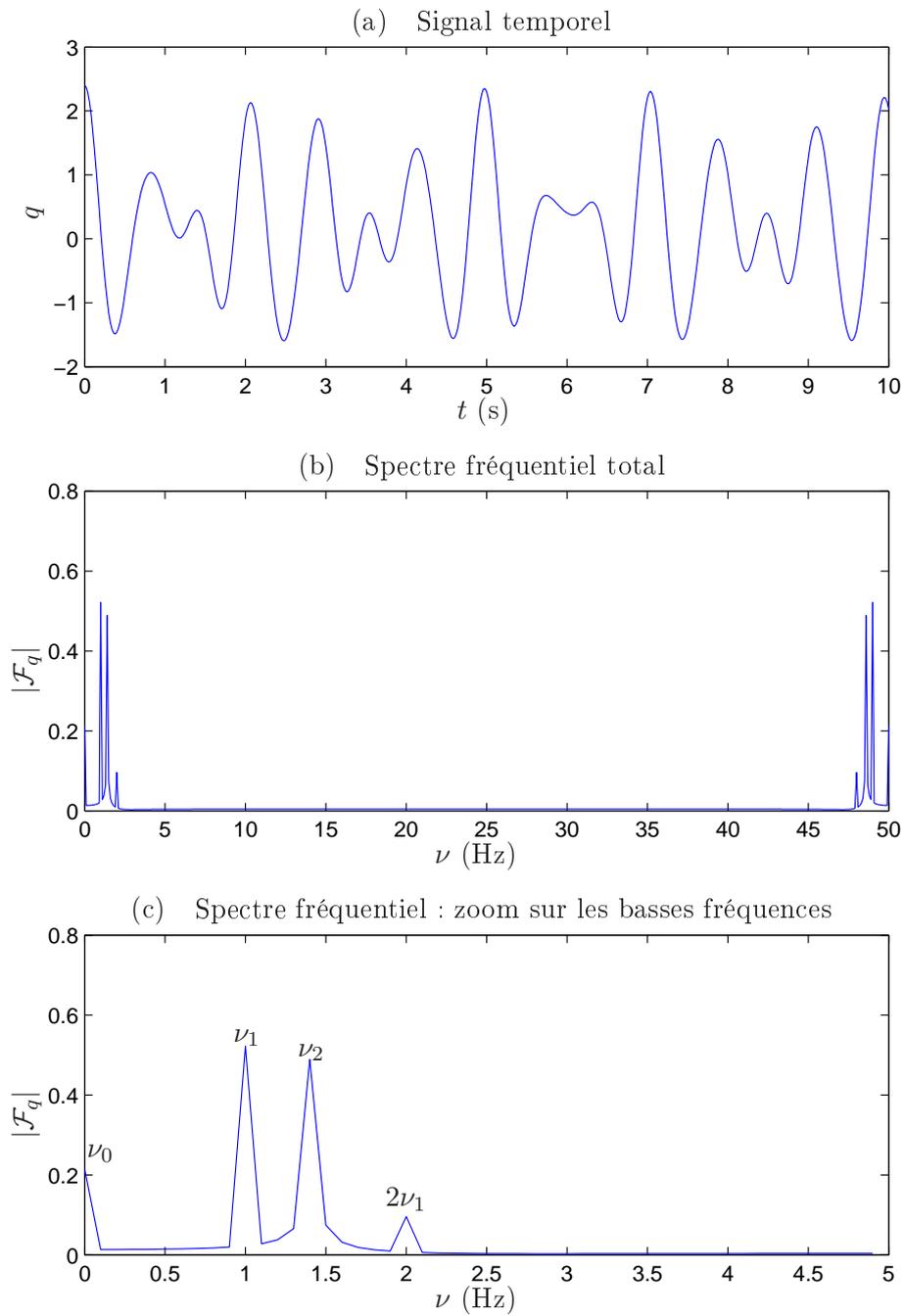


FIG. 1.5 – Intérêt du spectre fréquentiel

**Point fixe hyperbolique (hyperbolic fixed point) :**  $\forall i \in \llbracket 1, 2N \rrbracket, \operatorname{Re}(\lambda_i) \neq 0$

Puits (Sink) :  $\forall i \in \llbracket 1, 2N \rrbracket, \operatorname{Re}(\lambda_i) < 0$

– Noeud stable (Stable node) :  $\forall i \in \llbracket 1, 2N \rrbracket, \lambda_i \in \mathbb{R}$

– Foyer stable (Stable focus) :  $\exists i \in \llbracket 1, 2N \rrbracket, \operatorname{Im}(\lambda_i) \neq 0$

Source (Source) :  $\forall i \in \llbracket 1, 2N \rrbracket, \operatorname{Re}(\lambda_i) > 0$

– Noeud instable (Unstable node) :  $\forall i \in \llbracket 1, 2N \rrbracket, \lambda_i \in \mathbb{R}$

– Foyer instable (Unstable focus) :  $\exists i \in \llbracket 1, 2N \rrbracket, \operatorname{Im}(\lambda_i) \neq 0$

Noeud selle ou noeud col (Saddle) :  $\exists i, j \in \llbracket 1, 2N \rrbracket, \operatorname{Re}(\lambda_i)\operatorname{Re}(\lambda_j) < 0$

Instable à la fois dans les temps directs et indirects, on qualifie aussi en anglais un tel point de « nonstable » fixed point (et non « unstable » fixed point) pour marquer cette spécificité.

**Point fixe non-hyperbolique (non-hyperbolic fixed point) :**

$\exists i \in \llbracket 1, 2N \rrbracket, \operatorname{Re}(\lambda_i) = 0$

Point fixe non-hyperbolique instable (Unstable non-hyperbolic fixed point) :

$\exists j \in \llbracket 1, 2N \rrbracket, \operatorname{Re}(\lambda_j) > 0$

Point fixe neutralement stable (Neutrally (ou marginally) stable non-hyperbolic fixed point) :

$\forall i \in \llbracket 1, N \rrbracket, \operatorname{Re}(\lambda_i) \leq 0$

Centre (Center) :  $\forall i \in \llbracket 1, 2N \rrbracket, \operatorname{Re}(\lambda_i) = 0, \operatorname{Im}(\lambda_i) \neq 0$

Les points fixes hyperboliques sont illustrés par leurs portraits de phase en figure 1.6.

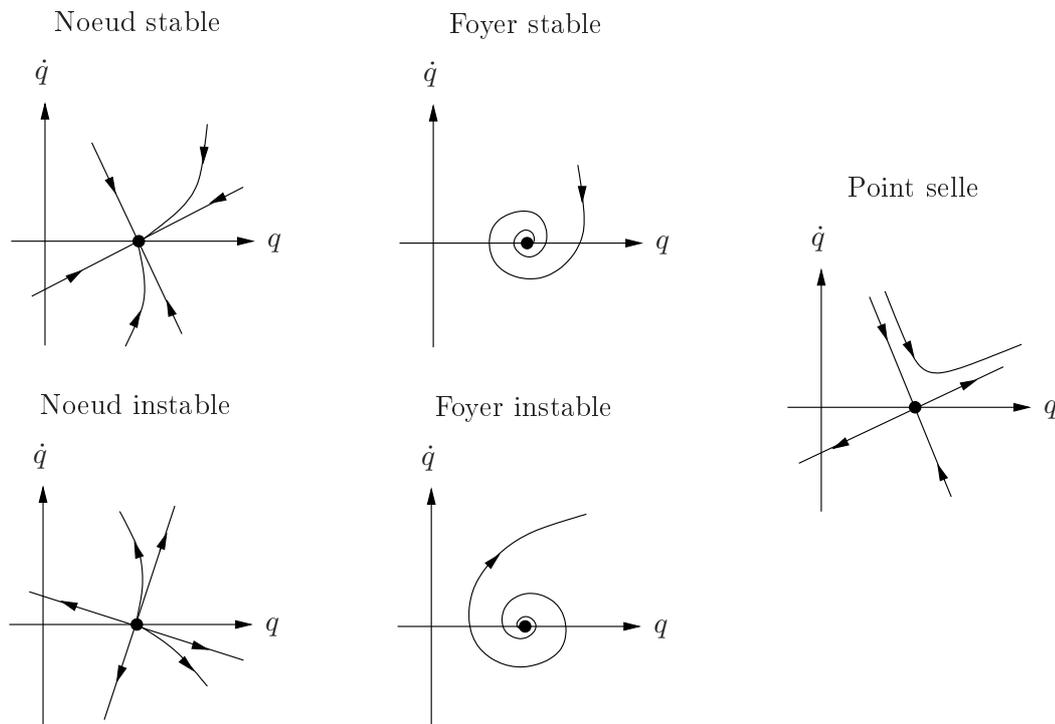


FIG. 1.6 – Portraits de phase types des différentes classes de points fixes hyperboliques

### Classification des cycles limites (ou solutions périodiques)

La classification des cycles limites calque celle du point fixe dans la mesure où l'étude d'un cycle limite peut se ramener à l'étude du point fixe de la section de Poincaré associée

(cf. § 1.2.4). Par conséquent, en s'appuyant cette fois sur la section de Poincaré associée et non plus sur le portrait de phase, l'analyse précédente peut se décliner de façon analogue.

### 1.3 Notions de continuation et bifurcations

Nous avons considéré, dans le début de ce chapitre, un système d'équations non-linéaires de la forme (1.3), c'est-à-dire dont tous les paramètres sont fixés et constants. Ce système possède des solutions dont nous avons donné un aperçu et des moyens de caractérisation dans les paragraphes précédents. Il peut maintenant sembler intéressant de chercher à savoir comment ces solutions évoluent lorsqu'un des paramètres du système, notons le  $\mu$ , varie. En formulation linéaire, on trace fréquemment des frf (fonctions de réponse en fréquence) de façon à voir l'évolution des amplitudes du mouvement d'un point de la structure en fonction de la fréquence d'excitation. Le paramètre  $\mu$  serait donc  $f_e$ , la fréquence d'excitation du système. Dans le cas de systèmes tournants où l'excitation est due à un balourd (mauvais équilibrage), l'excitation dépend, en fréquence et en amplitude, de la vitesse de rotation du système. Il est ainsi fréquent de considérer  $\Omega$ , vitesse de rotation comme le paramètre variant, aussi appelé *paramètre de continuation*. Formellement, on fait alors apparaître ce paramètre dans la fonction  $F$  régissant l'évolution temporelle du système dans une formulation analogue à (1.3) :

$$\dot{x} = F(x, t, \mu) \quad (1.5)$$

L'étude de l'évolution des solutions lorsque le paramètre  $\mu$  varie s'appelle une étude de *continuation* - partant du principe qu'il existe généralement une certaine continuité entre une solution du système pour une valeur  $\mu_0$  du paramètre et celle obtenue pour une valeur de  $\mu$  infiniment proche de  $\mu_0$ . Cependant, il arrive que les solutions changent radicalement pour certaines valeurs critiques du paramètre  $\mu$ . On appelle ces points des *points de bifurcation*. Ainsi, une étude de continuation complète d'un système consiste en le suivi et l'analyse en stabilité et bifurcation de toutes les solutions d'un système non-linéaire obtenues pour une valeur initiale de  $\mu$  lorsque ce paramètre évolue sur un domaine donné.

Nous consacrerons la seconde partie de ce chapitre à l'exposé des principales bifurcations des points fixes et des solutions périodiques.

#### 1.3.1 Bifurcations

Les points de bifurcation sont des points  $(x, \mu)$  où une solution change de nature ou de stabilité lorsque le paramètre de continuation évolue. Si l'on compare deux portraits de phase du système juste avant et juste après un point de bifurcation, il doit y avoir un changement notable dans la nature, la stabilité et le nombre des solutions observées. Les bifurcations possibles pour un système étudié sous la variation d'un seul paramètre de continuation sont bien répertoriées et caractérisées dans la littérature (on pourra par exemple se référer aux ouvrages [NB95, Kuz04] et [TS02, Chap. 13]).

Il est intéressant de faire la distinction entre les *bifurcations continues* et les bifurcations discontinues ou *catastrophiques* : dans le premier cas, la solution continuée change de nature ou disparaît mais laisse place à une autre de façon continue en amplitude. Dans le second, il

y a perte de la solution continuée (qui devient instable) mais on ne « gagne » pas de nouvelle solution stable de façon continue. On doit alors mener une analyse globale, c'est-à-dire sonder l'espace d'états, pour trouver d'autres solutions stables que le système peut accrocher.

Enfin, comme précisé en § 1.2.2, l'étude de la stabilité et celle des bifurcations reposent généralement sur les mêmes indicateurs. Nous les avons donc regroupées ici.

### Stabilité et bifurcation des points fixes

Soit  $x^*$  une solution de (1.5) pour une valeur  $\mu_0$  du paramètre de continuation. L'étude de la stabilité et des bifurcations se fait grâce aux valeurs propres de la jacobienne  $J = D_x F(x^*, \mu_0)$  qui est constante dans le temps étant donné que  $x^*$  est un point fixe. On note  $\lambda_i$ ,  $i \in \{1, \dots, 2N\}$ , ses valeurs propres.

**Stabilité :** signe de la partie réelle des valeurs propres  $\lambda_i$

- parties réelles toutes strictement négatives : stabilité asymptotique
- parties réelles nulles ou négatives, pas de valeur propre nulle : stabilité
- une partie réelle au moins strictement positive : instabilité

**Bifurcation :** annulation de la partie réelle d'une valeur propre au moins

- une valeur propre réelle s'annule : bifurcation type Turning Point (retournement), Transcritique (croisement de deux branches de points fixes) ou Pitchfork (apparition/disparition de deux branches de points fixes)

Ces bifurcations sont dites *statiques*.

- deux valeurs propres complexes conjuguées coupent l'axe imaginaire : bifurcation type Hopf (apparition/disparition d'une solution périodique).

Ces bifurcations sont dites *dynamiques*.

La figure 1.7 donne l'allure de ces bifurcations.

### Stabilité et Bifurcation des cycles limites

Soit  $x^*(t)$  une solution  $T$ -périodique de (1.5) pour une valeur  $\mu_0$  du paramètre de continuation. L'étude de sa stabilité et de ses bifurcations se fait par l'étude des valeurs propres de la *matrice de monodromie*  $\Phi$  (voir § 2.3.2 pour son calcul) qui possède la propriété suivante par rapport à une perturbation  $z(t)$  de  $x^*(t)$  :

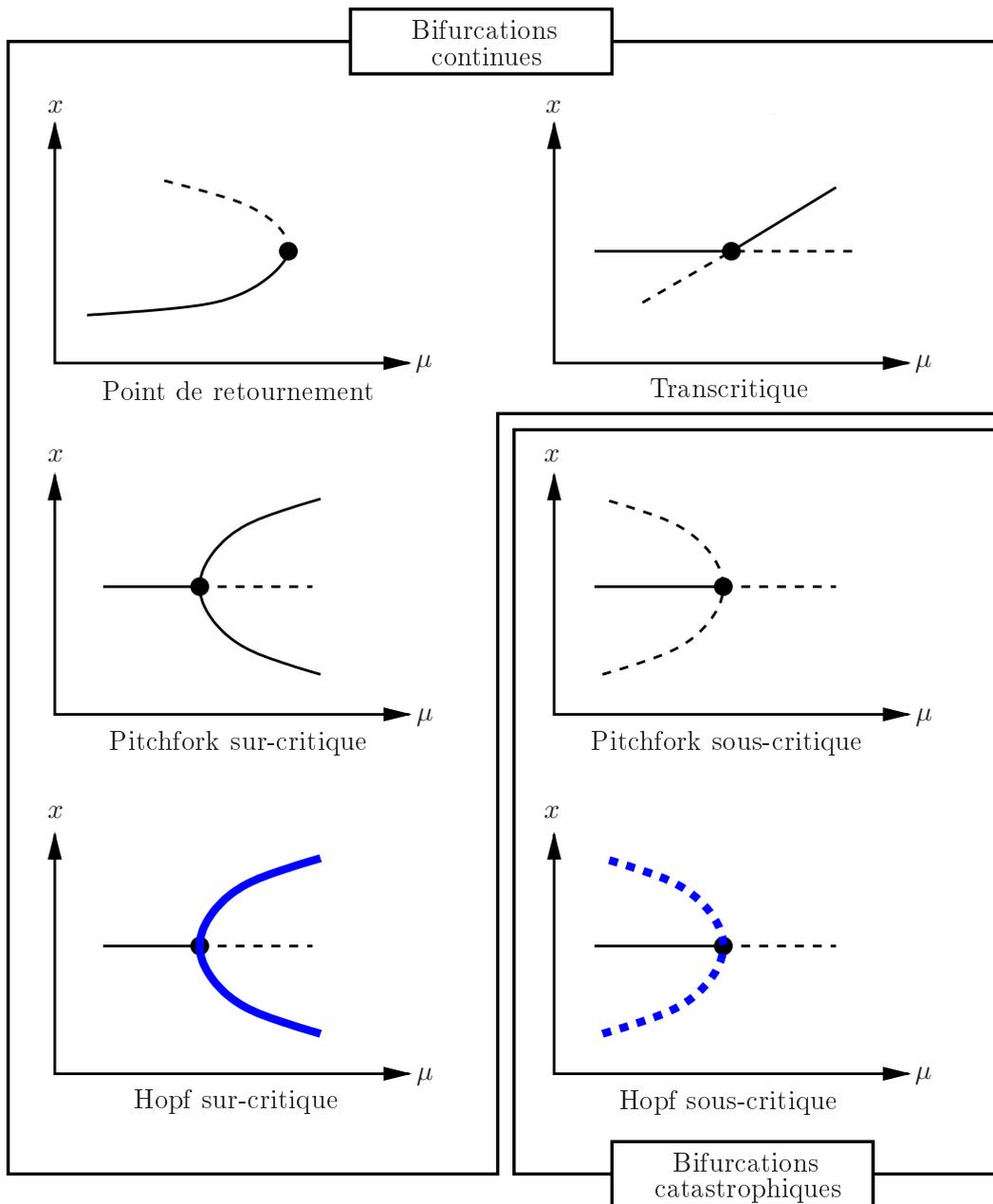
$$\forall t > 0, z(t+T) = \Phi z(t)$$

On note ses valeurs propres  $\rho_i$ ,  $i \in \{1, \dots, 2N\}$ . Elles se nomment couramment *multiplieurs de Floquet*.

Précisons que dans le cas d'un système autonome, l'une des valeurs propres vaut systématiquement 1. Il faut alors exclure cette valeur propre particulière des critères donnés ci-dessous. On pourra consulter [NB95, Sec. 3.2.1] pour la démonstration de ce résultat.

**Stabilité :** module des valeurs propres  $\rho_i$  de  $\Phi$

- module strictement inférieur à 1 : stabilité asymptotique
- module inférieur ou égal à 1 : stabilité
- un module au moins supérieur à 1 : instabilité



Stable	Instable	
—	- - - -	Point fixe
—	•••••	Cycle limite

FIG. 1.7 – Bifurcations des solutions types points fixes

- Bifurcation** : traversée du cercle unité dans le plan complexe par une des valeurs propres
- sortie du cercle par  $+1$  : transcritique, symmetry-breaking, cyclic-fold
  - sortie du cercle par  $-1$  : period doubling ou flip (doublement de la période de la solution)
  - sortie du cercle par  $\mathcal{C} \setminus \{-1, +1\}$  : Hopf (apparition d'une solution quasi-périodique)

Ces différents cas sont résumés sur la figure 1.3.1

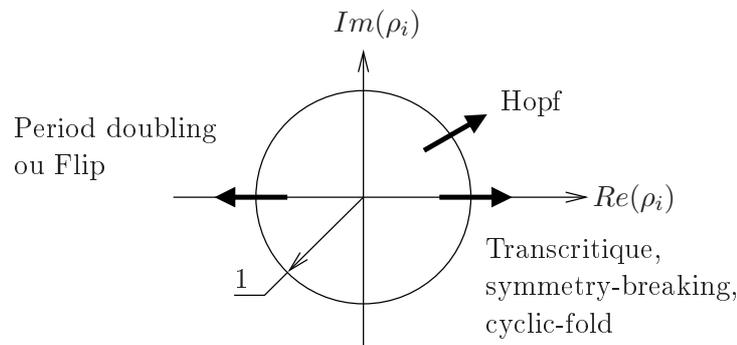


FIG. 1.8 – Bifurcations des solutions périodiques selon la position des coefficients de Floquet dans le plan complexe

### 1.3.2 Diagramme de bifurcation

Un diagramme de bifurcation représente l'évolution des solutions au cours de la variation du paramètre de continuation  $\mu$ . Il est généralement obtenu par échantillonnage du signal à la fréquence d'excitation.

On donne sur la figure 1.9 l'allure des différentes solutions vue sur un diagramme de bifurcation. On représente le paramètre de continuation en abscisse et en ordonnée, l'amplitude d'une des inconnues du système, échantillonnée à une fréquence donnée (généralement  $f_e$ ). Nous n'avons représenté ici que les solutions stables du système comme c'est généralement le cas dans la littérature.

Un tel diagramme souffre d'insuffisances dans la mesure où une perte d'informations survient par l'intermédiaire de l'échantillonnage : on ne peut par exemple pas distinguer un point fixe d'une solution périodique dont la période est multiple de celle utilisée pour l'échantillonnage. De même, on perd généralement l'information sur les amplitudes des signaux lorsqu'ils ne sont pas constants. Ceci est symptomatique de la complexité des systèmes non-linéaires : les deux dimensions (voire trois en s'appliquant) à notre disposition pour représenter visuellement toutes les informations sont généralement insuffisantes. Malgré tout, ce type de graphique permet d'avoir un aperçu rapide de l'évolution du comportement du système et se trouve très répandu dans la littérature.

## 1.4 Conclusion : enjeux

Au terme de ce premier chapitre, on aperçoit les difficultés liées à l'analyse globale d'un système non-linéaire possédant un paramètre qui varie : il faut se donner les moyens de

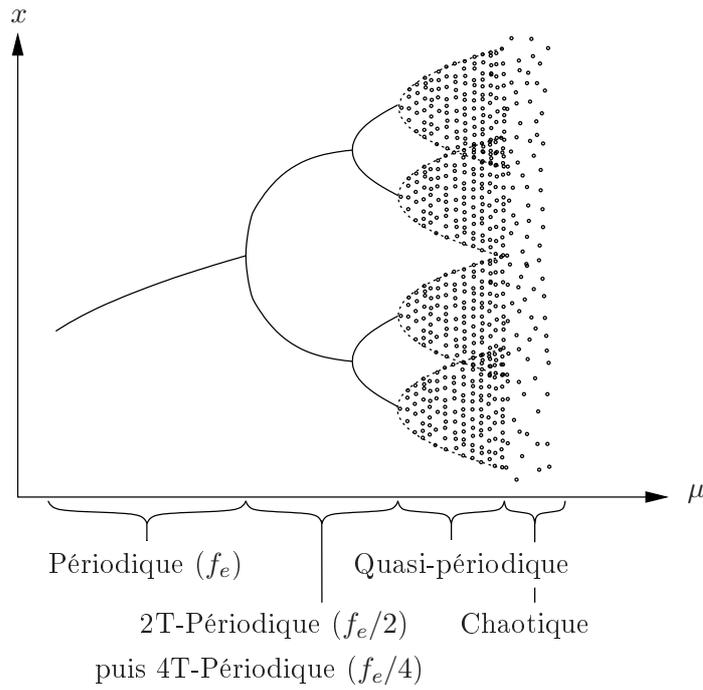


FIG. 1.9 – Diagramme de bifurcation schématique

trouver, pour une valeur du paramètre donné, toutes les solutions - qui peuvent être de natures diverses - puis il faut ensuite savoir calculer leur évolution, leur stabilité, détecter les points de bifurcation et éventuellement construire les solutions qui naissent en ces points.

Le chapitre suivant passe en revue l'essentiel des techniques numériques permettant de réaliser ces différentes tâches. Il ne traite pas l'aspect coût de calcul que nous aborderons dans des trois derniers chapitres.

## Chapitre 2

# Méthodes d'étude classiques (locales)

---

*A l'issue du premier chapitre, nous avons eu un aperçu des différents phénomènes non-linéaires qui distinguent ces systèmes des systèmes linéaires classiques. Nous dédions ce deuxième chapitre aux outils d'étude locale de tels modèles. Nous commencerons par ce qu'il y a de plus naturel, c'est-à-dire les algorithmes d'intégration temporelle. Nous donnerons ensuite les grands principes des algorithmes de résolution de systèmes non-linéaires algébriques. Munis de ces deux approches, nous exposerons les méthodes non-linéaires dédiées à la construction et à l'étude des points fixes et des cycles limites. Enfin, nous donnerons les principales clefs de la mise en œuvre d'une étude de continuation.*

### 2.1 Algorithmes d'intégration temporelle

L'utilisation d'intégration temporelle est inévitable, tant pour la construction de sections de Poincaré que pour une analyse de type Floquet (cf. § 2.3.2) et d'une façon générale elle conforte, en dernier lieu, les résultats obtenus par d'autres méthodes. Il existe différents types de schémas d'intégration ayant chacun leurs avantages et inconvénients en termes de coût numérique, de précision et de stabilité. Nous présentons dans cette section les schémas principaux et les notions qui servent à évaluer leurs performances. A titre d'introduction aux schémas d'intégration temporelle, nous recommandons la lecture de [GR93, Chap. 7].

### 2.1.1 Notions essentielles pour qualifier un schéma d'intégration temporelle

Lorsque l'on se ramène à une équation du premier ordre de la forme (1.3) et que l'on note  $x_n = x(t_n)$ , on pourra généralement mettre l'algorithme numérique sous la forme :

$$x_{n+1} = \sum_{j=1}^m \alpha_j x_{n+1-j} - h \times \sum_{j=0}^m \beta_j \dot{x}_{n+1-j}, \quad h = t_{n+1} - t_n \quad (2.1)$$

On distingue alors deux grandes catégories de schémas :

#### Définition 2.1 : Schéma d'intégration temporelle explicite ou implicite

Si  $\beta_0 = 0$ , le schéma est dit explicite :  $x_{n+1}$  peut être calculé à partir des résultats calculés aux pas de temps précédents.

Si  $\beta_0 \neq 0$ , le schéma est dit implicite :  $x_{n+1}$  dépend de sa propre dérivée.  $\square$

Les algorithmes à schéma explicite ne sont pas inconditionnellement stables mais présentent l'avantage d'un faible coût de calcul car les opérations à mener à chaque itération se résument à du calcul vectoriel (sommés et produits). Pour les algorithmes à schéma implicite, en revanche, les équations du mouvement doivent être remaniées avant d'être résolues et la solution sera obtenue de façon itérative dans le cas non-linéaire. Ils sont généralement inconditionnellement stables mais présentent un coût de calcul élevé car ils nécessitent la résolution d'un système d'équations à chaque itération.

#### Définition 2.2 : Consistance d'une méthode d'intégration

Un schéma d'intégration temporelle est consistant si et seulement si  $\lim_{h \rightarrow 0} \frac{x_{n+1} - x_n}{h} = \dot{x}(t_n)$ .  $\square$

La consistance d'un schéma d'intégration est un élément fondamental dans la mesure où elle assure que la solution exacte des équations différentielles discrétisées tend vers la solution exacte des équations continues lorsque le pas de discrétisation tend vers 0.

#### Définition 2.3 : Stabilité d'une méthode d'intégration

Un schéma d'intégration directe est dit stable si et seulement s'il existe un pas  $h_0 > 0$  tel que pour tout pas  $h \in [0, h_0]$ , pour toute perturbation finie de  $x_n$  à l'instant  $t_n$ , la modification des  $x_{n+j}$ ,  $j \geq 1$  est non croissante.  $\square$

Un schéma stable ne signifie pas qu'il converge vers la solution mais que la différence entre la solution numérique obtenue et la solution exacte est bornée. Ainsi, on utilise le terme de *précision* pour qualifier la proximité du résultat obtenu par intégration temporelle à la solution réelle. Les schémas les plus précis sont les schémas à la limite de la stabilité et de l'instabilité. En effet, un schéma instable a tendance à faire diverger les amplitudes alors qu'un schéma stable a tendance à les atténuer. L'idéal est donc un schéma à la frontière qui les reproduit exactement.

Dans le cas des systèmes non-linéaires, il n'est pas toujours possible de déterminer par avance la stabilité de l'algorithme utilisé. Il est intéressant alors de vérifier que l'on a *conservation de l'énergie*. Ceci, en effet, doit être nécessairement vérifié et constitue un critère classique de stabilité d'un algorithme.

### 2.1.2 Cas des systèmes linéaires

Nous considérerons dans cette section les problèmes différentiels linéaires de la forme (2.2).

$$\dot{x} = Jx + F_e(t) \quad (2.2)$$

Il est en effet plus aisé de définir les schémas sur des systèmes linéaires car on peut alors quantifier analytiquement les notions énoncées précédemment. Nous donnerons leurs adaptations aux systèmes non-linéaires dans la section suivante.

Nous commencerons cette section par une recherche du pas optimal en temps. Nous présentons ensuite deux méthodes particulières : la méthode de Newmark, qui est une méthode implicite à un pas ( $m=1$ ), puis la méthode des différences finies centrées qui est une méthode explicite.

#### Valeur du pas en temps pour une précision optimale

La question du pas de temps est primordiale : un pas de temps trop grand génère des imprécisions, un pas de temps trop petit rend le calcul coûteux et accumule inutilement les erreurs numériques. Il est donc important de le choisir judicieusement. Comme expliqué au paragraphe précédent, le cas idéal pour la précision du calcul est un schéma à la limite de la stabilité. Reprenons donc la définition de la stabilité pour en déduire le pas optimal et posons  $z_n$  une petite perturbation de  $x_n$  à l'instant  $t_n$ . L'équation (2.2) nous donne que la perturbation respecte l'équation  $\dot{z}_{n+1} = J.z_{n+1}$ . En réinjectant ces expressions dans l'équation (2.1), on obtient :

$$\sum_{j=0}^m (\alpha_j I - h \times \beta_j J) z_{n+1-j} = 0, \quad \alpha_0 = -1 \quad (2.3)$$

On pose ensuite  $\mu_i$ ,  $i \in \{1, \dots, 2N\}$  les valeurs propres de  $J$  (supposée de rang  $2N$ ) que l'on réunit dans la matrice diagonale  $\underline{\mu}$ , et on note  $X$  la matrice des vecteurs propres (normés) associés. On peut ainsi écrire l'équation caractéristique de (2.3) en posant  $z_{n+1-m+k} = \lambda^k X \tilde{z}$  :

$$\sum_{j=0}^m (\alpha_j I - h \times \beta_j A) \lambda^{m-j} X \tilde{z} = 0$$

En prémultipliant par  $X^t$  et en utilisant la relation d'orthogonalité des vecteurs propres  $X^t X = I$ , on obtient alors :

$$\left[ \sum_{j=0}^m (\alpha_j I - h \times \beta_j \underline{\mu}) \lambda^{m-j} \right] \tilde{z} = 0 \quad \Leftrightarrow \quad \forall k \in \llbracket 1, 2N \rrbracket, \quad \sum_{j=0}^m (\alpha_j - h \mu_k \beta_j) \lambda^{m-j} = 0$$

Comme il est facile de s'apercevoir que le schéma est stable si et seulement si  $|\lambda| \leq 1$ , on peut écrire à la frontière entre stabilité et instabilité  $\lambda = e^{i\theta}$ ,  $0 \leq \theta < 2\pi$ .

Ainsi, dans le cas général la limite de stabilité est obtenue pour

$$\mu h = \frac{\sum_{j=0}^m \alpha_j e^{i\theta(m-j)}}{\sum_{j=0}^m \beta_j e^{i\theta(m-j)}} \quad (2.4)$$

### Méthode de Newmark

La démonstration de l'élaboration de cette méthode permet de bien comprendre l'approche des schémas d'intégration temporelle et ses différents points clés; nous l'exposerons donc en détail.

Cette méthode étant initialement dédiée aux équations classiques de la dynamique, nous utiliserons le formalisme suivant, équivalent linéaire de la formulation (1.1) :

$$M.\ddot{q} + C.\dot{q} + K.q = f_e(t) \quad (2.5)$$

Cette méthode repose sur un développement au premier ordre et une approximation des restes intégraux par quadrature numérique :

- Développement au premier ordre de  $q(t_n + h)$  et  $\dot{q}(t_n + h)$  :

$$\begin{cases} q_{n+1} = q_n + h\dot{q}_n + \int_{t_n}^{t_{n+1}} \ddot{q}(\tau)(t_n + h - \tau)d\tau \\ \dot{q}_{n+1} = \dot{q}_n + \int_{t_n}^{t_{n+1}} \ddot{q}(\tau)d\tau \end{cases}$$

- Approximation des restes intégraux :

$$\begin{cases} (a) : \ddot{q}_n = \ddot{q}(\tau) + q^{(3)}(\tau)(t_n - \tau) + \dots \\ (b) : \ddot{q}_{n+1} = \ddot{q}(\tau) + q^{(3)}(\tau)(t_{n+1} - \tau) + \dots \\ (1 - \gamma)(a) + \gamma(b) : \ddot{q}(\tau) = (1 - \gamma)\ddot{q}_n + \gamma\ddot{q}_{n+1} + q^{(3)}(\tau)(\tau - h\gamma - t_n) + O(h^2q^{(4)}) \\ (1 - 2\beta(a)) + 2\beta(b) : \ddot{q}(\tau) = (1 - 2\beta)\ddot{q}_n + 2\beta\ddot{q}_{n+1} + q^{(3)}(\tau)(\tau - 2h\beta - t_n) + O(h^2q^{(4)}) \end{cases}$$

$$\Rightarrow \begin{cases} \int_{t_n}^{t_{n+1}} \ddot{q}(\tau)d\tau = [(1 - \gamma)\ddot{q}_n + \gamma\ddot{q}_{n+1}] h + r_n \\ \int_{t_n}^{t_{n+1}} \ddot{q}(\tau)(t_n + h - \tau)d\tau = [(1/2 - \beta)\ddot{q}_n + \beta\ddot{q}_{n+1}] h^2 + r'_n \end{cases},$$

$$\begin{cases} r_n = (\gamma - 1/2)h^2q^{(3)}(\tilde{\tau}) + O(h^3q^{(4)}), \quad t_n < \tilde{\tau} < t_{n+1} \\ r'_n = (\beta - 1/6)h^3q^{(3)}(\tilde{\tau}) + O(h^4q^{(4)}) \end{cases}$$

A ce stade, on voit que le choix des constantes  $\gamma$  et  $\beta$  influence grandement l'algorithme. Deux cas sont fréquemment rencontrés :

a)  $\gamma = 1/2$ ,  $\beta = 1/6$  : ce cas revient à une interpolation linéaire des accélérations sur  $[t_n \ t_{n+1}]$ ,  $\ddot{q}(\tau) = \ddot{q}_n + (\tau - t_n)(\ddot{q}_{n+1} - \ddot{q}_n)/h$ .

b)  $\gamma = 1/2$ ,  $\beta = 1/4$  : ce cas revient à utiliser une valeur moyenne de l'accélération sur  $[t_n \ t_{n+1}]$ ,  $\ddot{q}(\tau) = (\ddot{q}_{n+1} + \ddot{q}_n)/2$ .

On a donc dans le cas général le schéma implicite suivant :

$$\begin{Bmatrix} q_{n+1} \\ \dot{q}_{n+1} \end{Bmatrix} = \begin{Bmatrix} q_n \\ \dot{q}_n \end{Bmatrix} + \begin{Bmatrix} h\dot{q}_n + h^2(1/2 - \beta)\ddot{q}_n + h^2\beta\ddot{q}_{n+1} \\ (1 - \gamma)h\ddot{q}_n + \gamma h\ddot{q}_{n+1} \end{Bmatrix} \quad (2.6)$$

Comme nous l'avons déjà dit en introduisant les méthodes implicites, il est nécessaire de reformuler les équations pour arriver à résoudre le système. De façon à pouvoir calculer une valeur approchée de  $\ddot{q}_{n+1}$ , on réinjecte le schéma (2.6) dans l'équation (2.5) :

$$[M + \gamma hC + \beta h^2 K] \cdot \ddot{q}_{n+1} = f_e(t_{n+1}) - [C \cdot \tilde{q}_{n+1} + K \cdot \tilde{q}_{n+1}] \quad (2.7)$$

avec  $\tilde{q}_{n+1} = q_n + h\dot{q}_n + (1/2 - \beta)h^2\ddot{q}_n$  et  $\tilde{\dot{q}}_{n+1} = \dot{q}_n + (1 - \gamma)h\ddot{q}_n$

Une fois  $\ddot{q}_{n+1}$  évalué, il ne reste plus qu'à effectuer les corrections sur  $q_{n+1}$  et  $\dot{q}_{n+1}$ .

L'algorithme complet permettant de mettre en œuvre la méthode est présenté en figure 2.1.

### Méthode des différences finies centrées

La méthode des différences finies centrées est équivalente à une méthode de Newmark avec  $\gamma = 1/2$  et  $\beta = 0$  :

$$\begin{Bmatrix} q_{n+1} \\ \dot{q}_{n+1} \end{Bmatrix} = \begin{Bmatrix} q_n \\ \dot{q}_n \end{Bmatrix} + \begin{Bmatrix} h\dot{q}_n + (h^2/2)\ddot{q}_n \\ (h/2)(\ddot{q}_n + \ddot{q}_{n+1}) \end{Bmatrix} \quad (2.8)$$

On peut alors mettre ce système sous une forme à trois pas en exprimant  $q_n$  :

$$q_n = q_{n-1} + h\dot{q}_{n-1} + (h^2/2)\ddot{q}_{n-1}$$

En réinjectant dans la deuxième équation de (2.8) et en divisant le tout par  $h$ , on obtient :

$$\frac{q_{n+1} - q_n}{h} - \frac{q_n - q_{n-1}}{h} = \dot{q}_n - \dot{q}_{n-1} + (h/2)(\ddot{q}_n - \ddot{q}_{n-1})$$

Puis, en remplaçant  $\dot{q}_n$  par  $\dot{q}_{n-1} + (h/2)(\ddot{q}_n + \ddot{q}_{n-1})$  (première équation de (2.8) écrite à l'instant précédent), on obtient :

$$\ddot{q}_n = \frac{q_{n+1} - 2q_n + q_{n-1}}{h^2} \quad (2.9)$$

#### 2.1.3 Cas des systèmes non-linéaires

Nous présentons maintenant l'adaptation de la méthode de Newmark à des systèmes non-linéaires. On utilisera la formulation (1.1) des équations. Le principe est le même que dans le cas des systèmes linéaires : on effectue une prédiction  $\tilde{q}_{n+1}$  et  $\tilde{\dot{q}}_{n+1}$  en fonction des termes  $q_n, \dot{q}_n$  et  $\ddot{q}_n$ , puis on évalue  $\ddot{q}_{n+1}$  que l'on réinjecte pour obtenir  $q_{n+1}$  et  $\dot{q}_{n+1}$ . La différence principale avec le traitement des systèmes linéaires est l'évaluation de  $\ddot{q}_{n+1}$  et la correction des termes  $q_{n+1}$  et  $\dot{q}_{n+1}$  qui en résulte. En effet, lorsque l'on écrit l'équivalent de l'équation (2.7), on obtient le système non-linéaire suivant (d'inconnue  $\ddot{q}_{n+1}$ ) :

$$[M + \gamma hC + \beta h^2 K] \cdot \ddot{q}_{n+1} = f_e(t_{n+1}) - [C \cdot \tilde{q}_{n+1} + K \cdot \tilde{q}_{n+1} + \hat{f}(\tilde{q}_{n+1} + \beta h^2 \ddot{q}_{n+1}, \tilde{\dot{q}}_{n+1} + \gamma h \ddot{q}_{n+1})]$$

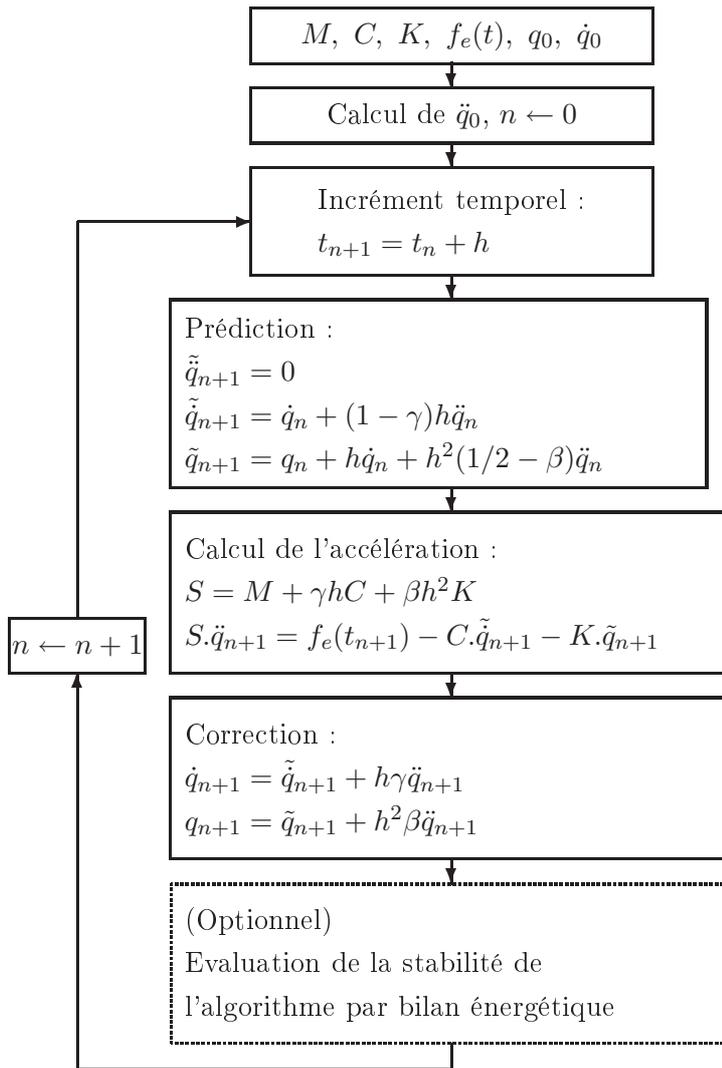


FIG. 2.1 – Algorithme d'intégration temporelle par une méthode de Newmark - Système linéaire

Il est alors classique d'utiliser un algorithme de type Newton-Raphson pour le résoudre.

On pose donc :

$$\ddot{q}_{n+1}^{(0)} = \tilde{\ddot{q}}_{n+1}, \dot{q}_{n+1}^{(0)} = \tilde{\dot{q}}_{n+1}, q_{n+1}^{(0)} = \tilde{q}_{n+1}$$

Puis on cherche les corrections successives  $\Delta\ddot{q}^{(i)}$  telles que :

$$S_{n+1}^{(i)} \cdot \Delta\ddot{q}^{(i)} = -r_{n+1}^{(i)}$$

avec  $r_{n+1}^{(i)} = M \cdot \ddot{q}_{n+1}^{(i)} + C \cdot \dot{q}_{n+1}^{(i)} + K \cdot q_{n+1}^{(i)} + \hat{f}(q_{n+1}^{(i)}, \dot{q}_{n+1}^{(i)}) - f_e(t_{n+1})$ ,

$$S_{n+1}^{(i)} = M + \gamma h(G + D + D_q \hat{f}(q_{n+1}^{(i)}, \dot{q}_{n+1}^{(i)})) + \beta h^2(K + D_q \hat{f}(q_{n+1}^{(i)}, \dot{q}_{n+1}^{(i)}))$$

Enfin on applique les corrections :

$$\ddot{q}_{n+1}^{(i+1)} = \ddot{q}_{n+1}^{(i)} + \Delta\ddot{q}^{(i)}, \dot{q}_{n+1}^{(i+1)} = \dot{q}_{n+1}^{(i)} + \gamma h \Delta\ddot{q}^{(i)}, q_{n+1}^{(i+1)} = q_{n+1}^{(i)} + \beta h^2 \Delta\ddot{q}^{(i)}$$

Et on itère tant que :

$$\|r_{n+1}^{(i+1)}\| > \epsilon$$

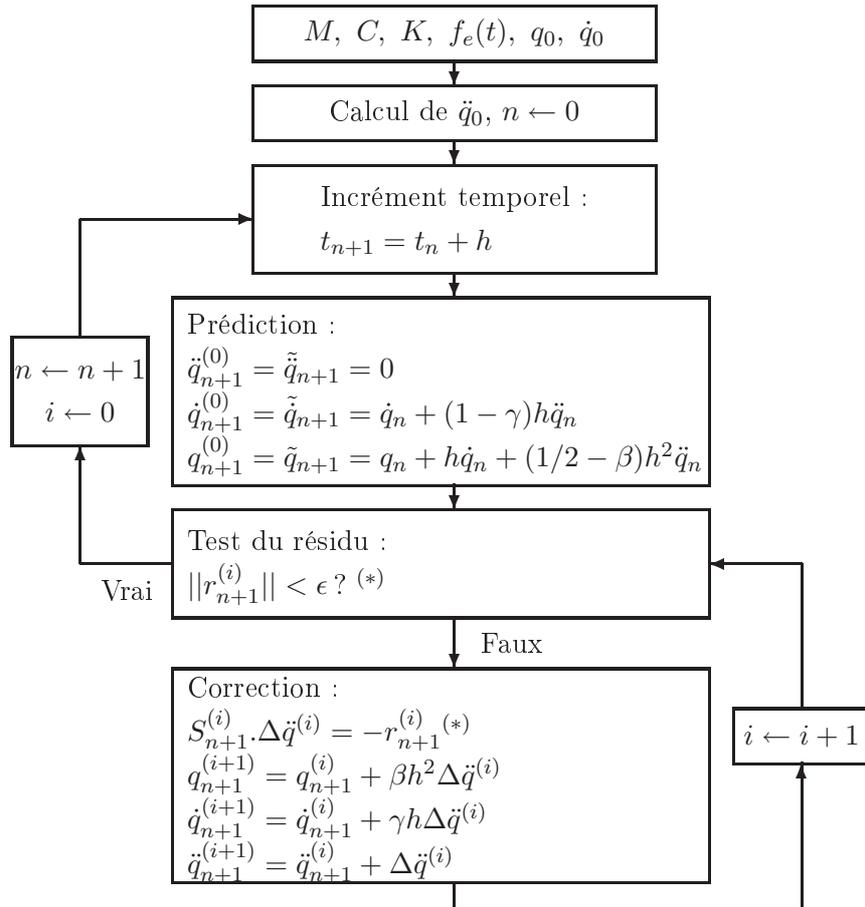
L'algorithme du schéma entier est donné en figure 2.2.

#### 2.1.4 Synthèses des différentes méthodes trouvées dans la littérature

Nous avons détaillé dans les deux paragraphes précédents des techniques fondamentales s'appuyant tant dans le cas non-linéaire que dans le cas linéaire sur une approche de type Newmark. Certains raffinements permettent d'introduire de l'amortissement numérique de façon à stabiliser le schéma (HHT- $\alpha$ ), et à pouvoir optimiser la répartition fréquentielle de cet amortissement (Generalized- $\alpha$ ). Nous donnons dans le tableau 2.1 les éléments principaux de ces méthodes. Appliquées au cas des systèmes linéaires, elles sont performantes et l'on peut en assurer la stabilité et le degré d'approximation (cf. [BBBE06]). En revanche, elles perdent fréquemment leur stabilité une fois appliquées aux systèmes non-linéaires (cf. [EBB02]). Un certain nombre de méthodes forçant la conservation de l'énergie totale et/ou des moments énergétiques dans le cas de non-linéarités structurales sont alors apparues. On en trouvera une présentation générale dans [KC99] et une mise en œuvre dans [Gon00]. On pourra consulter les références [ST92, NSP04] pour l'exposé de la méthode fondatrice dite EMCA (Energy Momentum Conserving Algorithm).

## 2.2 Algorithmes de résolution de systèmes non-linéaires

Les équations du dynamiqueien sont initialement formulées sous la forme (1.1), c'est-à-dire sous forme d'équations différentielles. Une première façon de résoudre ces équations est, comme nous venons de le voir, d'utiliser un schéma d'intégration temporelle qui donne à partir de conditions initiales l'évolution des différentes coordonnées du système dans le temps. Cette approche peut être relativement coûteuse en temps de calcul car l'obtention d'un état stationnaire - ce qui nous intéresse le plus fréquemment - peut demander d'intégrer le système sur des durées longues. Comme nous le verrons par la suite (§ 2.3), il existe donc des méthodes qui, partant d'une hypothèse sur la forme de l'état stationnaire recherché permettent de transformer les équations différentielles en équations algébriques (non-linéaires) et permettent



(\*) Définis par :

$$r_{n+1}^{(i)} = M \cdot \ddot{q}_{n+1}^{(i)} + C \cdot \dot{q}_{n+1}^{(i)} + K \cdot q_{n+1}^{(i)} + \hat{f}(q_{n+1}^{(i)}, \dot{q}_{n+1}^{(i)}) - f_e(t_{n+1})$$

$$S_{n+1}^{(i)} = M + \gamma h (C + D_{\dot{q}} \hat{f}(q_{n+1}^{(i)}, \dot{q}_{n+1}^{(i)})) + \beta h^2 (K + D_q \hat{f}(q_{n+1}^{(i)}, \dot{q}_{n+1}^{(i)}))$$

FIG. 2.2 – Algorithme d'intégration temporelle par une méthode de Newmark - Système non-linéaire

Nom - Paramètres Équations Référence	Caractéristiques
<b>Newmark</b> - $\gamma, \beta$ $M.\ddot{q} + C.\dot{q} + K.q = f_e$ $q_{n+1} = q_n + h\dot{q}_n + \frac{h^2}{2}(1 - 2\beta)\ddot{q}_n + \frac{h^2}{2}(2\beta)\ddot{q}_{n+1}$ $\dot{q}_{n+1} = \dot{q}_n + (1 - \gamma)h\ddot{q}_n + \gamma\ddot{q}_{n+1}$ c.f. [GR93]	<ul style="list-style-type: none"> <li>- Implicite</li> <li>- Pas d'amortissement numérique</li> <li>- Pour <math>\gamma = \frac{1}{2}, \beta = \frac{1}{4}</math> :               <ul style="list-style-type: none"> <li>- Stabilité inconditionnelle</li> <li>- Pas d'erreur d'amplification</li> </ul> </li> </ul>
<b>Différences finies centrées</b> $M.\ddot{q} + C.\dot{q} + K.q = f_e$ $q_{n+1} = q_n + h\dot{q}_{n+\frac{1}{2}}$ $\dot{q}_{n+1+\frac{1}{2}} = \dot{q}_{n+\frac{1}{2}} + h\ddot{q}_{n+1}$ c.f. [GR93]	<ul style="list-style-type: none"> <li>- Explicite</li> <li>- Pas d'amortissement numérique</li> <li>- Stabilité conditionnelle : <math>\omega h \leq 2</math></li> <li>- Pas d'erreur d'amplification</li> </ul>
<b>Hilber-Hughes-Taylor : HHT-<math>\alpha</math></b> - $\gamma, \beta, \alpha$ $M.\ddot{q}_{n+1} + (1 - \alpha)f_{n+1}^{int} + \alpha f_n^{int} = (1 - \alpha)f_{e\ n+1} + \alpha f_{e\ n}$ $q_{n+1} = q_n + h\dot{q}_n + \frac{h^2}{2}(1 - 2\beta)\ddot{q}_n + \frac{h^2}{2}(2\beta)\ddot{q}_{n+1}$ $\dot{q}_{n+1} = \dot{q}_n + (1 - \gamma)h\ddot{q}_n + \gamma\ddot{q}_{n+1}$ c.f. [GR93]	<ul style="list-style-type: none"> <li>- Implicite</li> <li>- Amortissement numérique</li> <li>- Pour <math>0 \leq \alpha \leq \frac{1}{3}, \gamma = \frac{1}{2} + \alpha, \beta = \frac{1}{4}(1 + \alpha)^2</math> :               <ul style="list-style-type: none"> <li>- Stabilité inconditionnelle</li> <li>- Erreur d'amplification du second ordre</li> </ul> </li> </ul>
<b>Generalized-<math>\alpha</math></b> - $(\gamma, \beta, \alpha_m, \alpha_f)$ ou $\rho^\infty$ $M.((1 - \alpha_m)\ddot{q}_{n+1} + \alpha_m\ddot{q}_n) + (1 - \alpha_f)f_{n+1}^{int} + \alpha_f f_n^{int} = (1 - \alpha_f)f_{e\ n+1} + \alpha_f f_{e\ n}$ $q_{n+1} = q_n + h\dot{q}_n + \frac{h^2}{2}(1 - 2\beta)\ddot{q}_n + \frac{h^2}{2}(2\beta)\ddot{q}_{n+1}$ $\dot{q}_{n+1} = \dot{q}_n + (1 - \gamma)h\ddot{q}_n + \gamma\ddot{q}_{n+1}$ c.f. [CH93]	<ul style="list-style-type: none"> <li>- Implicite</li> <li>- Amortissement numérique optimal :               <ul style="list-style-type: none"> <li>- Dissipation hautes freq max si <math>\beta = \frac{1}{4}(1 - \alpha_m + \alpha_f)^2</math></li> <li>- Dissipation basses freq min si <math>\alpha_f = \frac{\alpha_m + 1}{3}</math></li> </ul> </li> <li>- Pour <math>\alpha_m \leq \alpha_f \leq \frac{1}{2}</math> et <math>\beta \geq \frac{1}{4} + \frac{1}{2}(\alpha_f - \alpha_m)</math> : stab. incond.</li> <li>- Pour <math>\gamma = \frac{1}{2} - \alpha_m + \alpha_f</math> : erreur d'amplification du <math>2^{nd}</math> ordre</li> <li>- En respectant tous les paramètres optimum : il suffit de choisir <math>\rho^\infty</math>, rayon spectral des hautes fréquences puis :               <ul style="list-style-type: none"> <li><math>\alpha_m = \frac{2\rho^\infty - 1}{\rho^\infty + 1}</math> <math>\alpha_f = \frac{\rho^\infty}{\rho^\infty + 1}</math></li> </ul> </li> </ul>

TAB. 2.1 – Synthèse des différents algorithmes d'intégration temporelle en linéaire

de trouver plus rapidement les états stationnaires. Nous présentons donc ici de façon très simple les idées maîtresses des algorithmes pouvant être mis en œuvre de façon à résoudre le problème non-linéaire algébrique suivant :

$$H(\tilde{x}) = 0, \quad H(\tilde{x}) = \bar{H}.\tilde{x} + \hat{H}(\tilde{x}) - H_c \quad (2.10)$$

où la fonction  $H$  se décompose en une contribution linéaire,  $\bar{H}$ , une contribution non-linéaire  $\hat{H}$  et enfin, une contribution constante  $H_c$ , souvent imputable aux efforts extérieurs. Pour une présentation illustrée des principes et des algorithmes, nous recommandons la consultation de [PTVF92, Chap. 9].

### 2.2.1 Algorithmes sans gradient

L'idée de cette catégorie de méthodes est de construire la surface  $\|H(\tilde{x})\|$  à partir des valeurs en quelques points  $\tilde{x}^{(i)}$ ,  $i = 1, \dots, p$ . Classiquement, on utilise une interpolation polynomiale de degré  $p - 1$ . Une fois réalisée l'approximation de  $\|H\|$  par un polynôme  $\mathbf{P}$  (c.f. [DJKP07, Chap. 3]), il ne reste plus qu'à trouver les zéros du polynôme.

### 2.2.2 Algorithmes avec gradient

Ces algorithmes sont des méthodes itératives reposant sur le développement de l'équation (2.10) au premier ordre (voire au second). On se donne un point de départ  $\tilde{x}^{(0)}$  qui ne vérifie généralement pas (2.10). On cherche donc une correction au premier ordre  $\Delta\tilde{x}$  telle que  $H(\tilde{x}^{(0)} + \Delta\tilde{x}) = 0$ . On écrit pour cela le développement :

$$H(\tilde{x}_0 + \Delta\tilde{x}) = H(\tilde{x}^{(0)}) + D_{\tilde{x}}H(\tilde{x}^{(0)})\Delta\tilde{x} + O(\|\Delta\tilde{x}\|^2) \approx H(\tilde{x}^{(0)}) + D_{\tilde{x}}H(\tilde{x}^{(0)})\Delta\tilde{x}$$

Il faut donc résoudre le système linéaire suivant pour trouver la correction :

$$D_{\tilde{x}}H(\tilde{x}^{(0)})\Delta\tilde{x} = -H(\tilde{x}^{(0)})$$

On note alors  $\tilde{x}^{(1)} = \tilde{x}^{(0)} + \Delta\tilde{x}$  et on réitère le processus jusqu'à avoir  $\|H(\tilde{x}^{(i)})\| < \epsilon$ ,  $\epsilon$  étant une quantité positive que l'on s'est donnée comme permettant d'approximer numériquement la nullité. Cette méthode est couramment nommée méthode de *Newton* ou de *Newton-Raphson*.

On peut aussi chercher du côté des méthodes de minimisation en ayant pour but de trouver un minimum nul de la norme de la fonction à annuler. On tente donc de minimiser :

$$\tilde{H}(\tilde{x}) = \|H(\tilde{x})\| \text{ ou } \tilde{H}(\tilde{x}) = \|H(\tilde{x})\|^2$$

De nombreuses méthodes raffinent le principe de base de Newton (qui s'appelle alors méthode de la plus profonde descente) en réalisant un développement non pas au premier mais au second ordre sans toutefois requérir l'évaluation de la hessienne (méthode du gradient conjugué), etc.

Cependant, elles partagent toutes des travers tels que la dépendance forte de la convergence au choix du point initial ou la possibilité de tomber dans un puits (un minimum local non nul) et de ne plus en sortir. Elles ne sont donc à utiliser que lorsque l'on est capable d'initialiser l'algorithme pour une valeur « proche » d'un zéro.

## 2.3 Construire et étudier des solutions

Nous avons présenté jusqu'ici des algorithmes généraux. Détaillons maintenant les techniques relatives à l'étude des points fixes et des solutions périodiques répondant au problème (1.3). Nous entendons par « étude » la construction de telles solutions et l'obtention de grandeurs permettant d'en caractériser la stabilité.

### 2.3.1 Solutions type point fixe

#### Construction des points fixes

La recherche d'un point fixe revient à la résolution du problème :

$$\text{Trouver } x^* \in \mathbb{R}^{2N} \text{ solution de } F(x^*, t) = 0 \quad (2.11)$$

Le plus souvent, le système est autonome (absence de la variable  $t$  dans l'expression de  $F$ ).  $F$  étant un système d'équations algébriques non-linéaires, on se trouve dans le cas évoqué en section 2.2.

#### Etude de stabilité des points fixes

Lorsque l'on perturbe une solution  $x^*$  de type point fixe par  $z(t)$ , on obtient l'équation (2.12) (cf. § 1.2.2 pour la démonstration).

$$\dot{z} = D_x F(x^*) \cdot z \quad (2.12)$$

$x^*$  étant constant, la matrice jacobienne gouvernant l'évolution de la perturbation l'est aussi.  $z$  est donc solution d'une équation différentielle du premier ordre à coefficients constants réels dont on sait que les solutions s'écrivent :

$$z(t) = \sum_{i=1}^{2N} z_i e^{\lambda_i t}, \quad \forall i \in \llbracket 1, 2N \rrbracket, \quad z_i \in \mathcal{C}^{2N}$$

avec  $\lambda_i$  les  $2N$  valeurs propres complexes de  $D_x F(x^*)$ .

La décroissance en norme de  $z$  dépend donc du signe des parties réelles des  $\lambda_i$ . Les différents cas ont déjà été traités en section 1.3.1.

### 2.3.2 Solutions périodiques

Une solution périodique est une solution telle qu'il existe une période  $T$  non nulle telle que  $\forall t > 0, x^*(t+T) = x^*(t)$ . Cette caractérisation n'est pas évidente à retracer numériquement. Le point de vue adopté est donc le suivant : on cherche un point du cycle  $x_c \in \mathbb{R}^{2N}$  et une période  $T^* > 0$  tels que la résolution numérique de (1.3) avec pour condition initiale  $x^*(0) = x_c$  au bout d'un temps  $t = T^*$  redonne  $x^*(T^*) = x_c$ . On n'est pas sûr alors d'avoir la plus petite période et il est recommandé de tester la périodicité en  $T^*/k, k \in \mathbb{N}$ .

Il est fréquent que l'on connaisse la période de la solution cherchée dans la mesure où l'on cherche souvent une réponse à une sollicitation connue et sinusoïdale (la réponse est alors

souvent de période égale à celle de la sollicitation). Cependant, nous donnerons les méthodes de construction en faisant l'hypothèse que cette période est inconnue, ce qui se transpose très aisément au cas de période connue. Nous considérerons des systèmes dans leur formulation autonome qui est le cas le plus fréquent pour lequel la période de réponse est inconnue. Il est alors intéressant de considérer les deux problèmes équivalents suivants :

$$\begin{aligned} &\text{Trouver } (x_c, T^*) \in \mathbb{R}^{2N} \times \mathbb{R}^{+*} \text{ tel qu'il existe } x^*(t) \text{ solution de :} \\ &\quad \begin{cases} \dot{x}(t) = F(x(t)) \\ x(0) = x_c, x(T^*) = x_c \end{cases} \end{aligned} \quad (2.13)$$

En posant  $\tau = t/T^*$ , on obtient le problème aux limites où  $x^*(\tau)$  est périodique de période 1 :

$$\begin{aligned} &\text{Trouver } (x_c, T^*) \in \mathbb{R}^{2N} \times \mathbb{R}^{+*} \text{ tel qu'il existe } x^*(t) \text{ solution de :} \\ &\quad \begin{cases} x'(\tau) = T^*F(x(\tau)) \\ x(0) = x_c, x(1) = x_c \end{cases} \end{aligned} \quad (2.14)$$

avec la convention d'écriture  $x' = \frac{dx}{d\tau}$

### Construire une solution périodique

Il existe quatre grandes familles de méthodes pouvant servir à construire une solution périodique. Trois d'entre elles sont des méthodes dites temporelles et permettent de déterminer un point  $x_c$  du cycle limite recherché : différences finies, tir (shooting) et collocation orthogonale. La quatrième, très répandue lorsque la période de la solution recherchée est connue, est la balance harmonique ; c'est une méthode fréquentielle. Nous donnons pour chacune d'elles le principe de la méthode.

**Différences finies** On utilise ici la formulation (2.14) du problème. Le principe de cette méthode repose sur l'échantillonnage de la période avec un pas  $h = 1/m$ . En notant  $\tau_k = kh$  et  $x^k = x(\tau_k)$ , on obtient le système suivant :

$$\begin{cases} \dot{x}^0 = T^*F(x^0) \\ \vdots \\ \dot{x}^k = T^*F(x^k) \\ \vdots \\ \dot{x}^{m-1} = T^*F(x^{m-1}) \end{cases}$$

On a ainsi  $m \times 2N + 1$  inconnues (les  $m$  vecteurs  $x^k$  à  $2N$  composantes et  $T^*$ ) pour  $2mN$  équations. Il y a en effet un problème de phase (« phase shift ») à imposer : si  $x^*(t)$  est solution du problème,  $x^*(t + t')$  l'est aussi, quel que soit  $t'$ . Il faut donc fixer une des coordonnées des  $x^k$ ,  $x_{j_0}^k$ , ce qui permet de se ramener à  $mN$  inconnues. L'inconvénient est que si les solutions sont totalement inconnues, rien ne garantit que l'espace de solutions périodiques a une intersection non nulle avec l'hyperplan ainsi défini.

Il faut ensuite choisir un schéma d'intégration temporelle. La méthode tire son nom du schéma classiquement utilisé : les différences finies que l'on peut choisir centrées ( $x^{k+1} - x^k =$

$(hT^*/2)(F(x^{k+1})+F(x^k))$ ), avant ou arrière en fonction des problèmes de stabilité numérique que l'on peut rencontrer.

On obtient finalement le système non-linéaire suivant à résoudre dont la dernière ligne intègre la condition de périodicité :

$$\begin{cases} x^1 - x^0 &= \frac{hT^*}{2}(F(x^1) + F(x^0)) \\ &\vdots \\ x^{k+1} - x^k &= \frac{hT^*}{2}(F(x^{k+1}) + F(x^k)) \\ &\vdots \\ x^0 - x^{m-1} &= \frac{hT^*}{2}(F(x^0) + F(x^{m-1})) \end{cases} \quad (2.15)$$

La résolution de ce système peut se faire par une méthode classique de Newton-Raphson.

**Tir (shooting)** On utilise ici plutôt la formulation (2.13) du problème. Cette méthode est basée sur des corrections successives du couple supposé solution  $(x_c, T^*)$ . On donne donc initialement un premier couple, puis on le corrige d'après la valeur de  $x$  obtenue après un temps  $T^*$  en partant de  $x_c$ . On note  $x(T^*, x_c)$  cette valeur pour faire apparaître la dépendance du résultat considéré avec le point initialement choisi. On cherche donc  $\Delta T^*$  et  $\Delta x_c$  tels que

$$x(T^* + \Delta T^*, x_c + \Delta x_c) = x_c$$

Le développement au premier ordre autour de  $(T^*, x_c)$  donne

$$x(T^*, x_c) + \frac{\partial x}{\partial T^*}(T^*, x_c) \cdot \Delta T^* + \frac{\partial x}{\partial x_c}(T^*, x_c) \cdot \Delta x_c - x_c = 0$$

Les expressions des dérivées partielles s'obtiennent comme suit :

$$\text{a) } \frac{\partial x}{\partial T^*}(T^*, x_c) = \frac{\partial x}{\partial t} \times \frac{\partial t}{\partial T^*} = \dot{x} \times 1 = F(x(T^*, x_c))$$

$$\text{b) } \frac{\partial}{\partial t} \frac{\partial x}{\partial x_c}(T^*, x_c) = \frac{\partial}{\partial x_c} \frac{\partial x}{\partial t}(T^*, x_c) = \frac{\partial F(x(T^*, x_c))}{\partial x_c} = \frac{\partial F}{\partial x} \frac{\partial x}{\partial x_c}$$

Puis on remarque que  $\frac{\partial x}{\partial x_c}(t=0) = I$ , matrice identité.

Ainsi,  $\frac{\partial x}{\partial x_c}$  est solution de l'équation différentielle :

$$\dot{Y} = D_x F(x(t, x_c))Y, \quad Y(0) = I$$

Étant donné que l'on réalise une intégration temporelle depuis  $x(0, x_c) = x_c$  jusqu'à l'instant  $t = T^*$ , on peut obtenir aisément  $\frac{\partial x}{\partial x_c}(T^*, x_c)$  en intégrant temporellement l'équation différentielle précédente. Il est intéressant de remarquer que l'on effectue ainsi sans le vouloir le calcul de la matrice de monodromie.

Une fois obtenues les deux dérivées partielles de  $x$ , il ne reste plus qu'à résoudre :

$$\frac{\partial x}{\partial T^*}(T^*, x_c) \cdot \Delta T^* + \frac{\partial x}{\partial x_c}(T^*, x_c) \cdot \Delta x_c = x_c - x(T^*, x_c) \quad (2.16)$$

Il nous manque donc une équation : on retrouve en effet le problème du décalage de phase. On peut donc choisir de fixer une des coordonnées de la correction ce qui revient à supprimer une

des inconnues, ou de rajouter une équation, par exemple en imposant l'orthogonalité entre  $\Delta x_0$  et  $F(x_c)$  ou  $F(x(T^*, x_c))$ .<sup>1</sup>

Une mise en œuvre très complète et appliquée aux rotors est proposée dans [SN97].

**Collocation orthogonale** Cette méthode utilise comme celle des différences finies un découpage de la période 1 (formulation (2.14) du problème) en  $m+1$  points  $0 = \tau_0 < \tau_1 < \dots < \tau_m = 1$ . Sur chacun des segments  $[\tau_i, \tau_{i+1}]$ , on approche la solution par un polynôme  $x^{(i)}(\tau)$  de degré maximal  $p$ . Le terme collocation signifie que l'on requiert que notre approximation satisfasse l'équation exacte en  $p$  points sur chaque intervalle  $[\tau_i, \tau_{i+1}]$  :  $\tau_i < \xi_{i,1} < \xi_{i,2} < \dots < \xi_{i,p} < \tau_{i+1}$ .

On veut donc obtenir :

$$\forall 0 \leq i \leq m-1, \forall 1 \leq j \leq p, \frac{dx^{(i)}}{d\tau}(\xi_{i,j}) = T_0 F(x^{(i)}(\xi_{i,j})) \quad (2.17)$$

Il est pratique de définir les polynômes  $x^{(i)}$  par leurs valeurs  $x^{(i,j)}$  aux  $m+1$  points équidistants  $\tau_i = \tau_{i,0} < \tau_{i,1} < \dots < \tau_{i,j} < \dots < \tau_{i,p} = \tau_{i+1}$ . Puis de les écrire en les décomposant sur la base des polynômes de Lagrange :

$$x^{(i)}(\tau) = \sum_{j=0}^p x^{(i,j)} L_{i,j}(\tau), \quad L_{i,j}(\tau) = \prod_{k=0, k \neq j}^p \frac{\tau - \tau_{i,k}}{\tau_{i,j} - \tau_{i,k}}$$

On génère ainsi  $2N(p+1)m$  inconnues que l'on ramène à  $2N[(p+1)m - (m-1)] = 2N(pm+1)$  en remarquant que par continuité  $\forall i \in \llbracket 1, m-1 \rrbracket$ ,  $x^{(i-1,p)} = x^{(i,0)}$ . En prenant en compte que la période est elle aussi inconnue, on a au total  $2Npm + 2N + 1$  inconnues.

En réinjectant l'expression des  $x^{(i)}(\tau)$  dans (2.17), on obtient  $2Nmp$  équations. Reste à exprimer la condition de périodicité et le blocage de la phase.

La périodicité ramène  $2N$  équations par l'écriture de

$$x^{(m-1,p)} = x^{(0,0)}$$

Enfin, le blocage de la phase redonne une équation. [Kuz04, Sec. 10.1.4] propose une condition sur la phase moins contraignante que le blocage d'une des coordonnées à une valeur donnée :

$$\int_0^1 \langle x(\tau), \dot{y}(\tau) \rangle d\tau = 0$$

où  $y(\tau)$  est une solution périodique connue. Par exemple, la solution du pas précédent lorsque l'on réalise une continuation.

On a donc réussi à écrire les  $2Nmp + 2N + 1$  équations nécessaires à la résolution du système.

Il faut en dernier lieu choisir les points de collocation  $\xi_{ij}$ . [Kuz04] donne que ceux qui permettent de minimiser l'erreur sont les points de Gauss dont le calcul est répertorié dans la littérature.

Comme on peut le voir, cette méthode génère un grand système. Il est donc nécessaire pour le résoudre de prendre en compte la structure particulière de la matrice jacobienne qui en découle.

<sup>1</sup>D'autres techniques sont envisageables comme cela est précisé dans [NB95] p. 451-452

**Balance harmonique** C'est la seule méthode qui nécessite de connaître la période de la solution recherchée par avance. Elle fait partie des méthodes fréquentielles (en opposition aux trois précédentes qui sont des méthodes dites temporelles). Elle repose sur la décomposition de la solution en sa série de Fourier que l'on tronque à un ordre  $p$  donné. On utilise plutôt l'équation différentielle d'ordre 2 (1.1) dans la mesure où la forme de  $q$  étant imposée,  $\dot{q}$  en découle directement et ne constitue pas une inconnue véritable. Si la période de la solution est  $T^* = 2\pi/w^*$ , on écrit donc :

$$q^*(t) = a_0 + \sum_{k=1}^p (a_k \cos(kw^*t) + b_k \sin(kw^*t)) \quad (2.18)$$

où  $a_0 \in \mathbb{R}^N$ ,  $\forall k \in \llbracket 1, p \rrbracket$ ,  $a_k \in \mathbb{R}^N$ ,  $b_k \in \mathbb{R}^N$

La condition de périodicité est donc automatiquement vérifiée et il n'y a plus qu'à réinjecter cette forme dans l'équation différentielle puis à la projeter sur la base  $(1, \cos(kw^*t), \sin(kw^*t))$  à l'aide du produit scalaire  $(\cdot, \cdot)_{w^*}$  défini par (2.19) pour obtenir le système (2.20) de  $(2p+1)N$  équations algébriques non-linéaires régissant les  $a_k$  et les  $b_k$ .

$$(f, g)_\omega = \frac{2}{T} \int_0^T f(t)g(t)dt, \quad \omega = \frac{2\pi}{T} \quad (2.19)$$

$$\underbrace{\begin{bmatrix} K & & & \\ & \ddots & & \\ & & \Lambda_k & \\ & & & \ddots \end{bmatrix}}_{\hat{H}} \underbrace{\begin{bmatrix} a_0 \\ \vdots \\ a_k \\ b_k \\ \vdots \end{bmatrix}}_{\tilde{x}} + \underbrace{\begin{bmatrix} \frac{1}{2}(\hat{f}(q, \dot{q}), 1)_{w^*} \\ \vdots \\ (\hat{f}(q, \dot{q}), \cos(kw^*t))_{w^*} \\ (\hat{f}(q, \dot{q}), \sin(kw^*t))_{w^*} \\ \vdots \end{bmatrix}}_{\hat{H}(\tilde{x})} = \underbrace{\begin{bmatrix} \frac{1}{2}(f_e(t), 1)_{w^*} \\ \vdots \\ (f_e(t), \cos(kw^*t))_{w^*} \\ (f_e(t), \sin(kw^*t))_{w^*} \\ \vdots \end{bmatrix}}_{H_e} \quad (2.20)$$

$$\text{avec } \Lambda_k = \begin{bmatrix} K - (kw^*)^2 M & kw^* C \\ -kw^* C & K - (kw^*)^2 M \end{bmatrix} \text{ et } 1 \leq k \leq p.$$

On trouvera une mise en œuvre de cette méthode appliquée au cas d'un contact rotor/stator dans [vGE01].

### Etudier une solution périodique

L'étude d'une solution périodique se fait grâce aux *coefficients de Floquet*, valeurs propres de la *matrice de monodromie*  $\Phi$ . L'intérêt et la signification de cette matrice viennent de l'étude de l'évolution de la perturbation  $z(t)$  d'une solution  $T^*$ -périodique  $x^*(t)$  : si l'on considère une solution  $y(t) = x^*(t) + z(t)$ , c'est à dire une perturbation de la solution périodique considérée, on peut écrire (cf. § 1.2.2) :

$$\dot{z} = D_x F(x^*(t)).z \quad (2.21)$$

On peut alors considérer une base de solutions  $z_1(t), \dots, z_{2N}(t)$ . Puis, on peut montrer que si  $z(t)$  est solution  $\tilde{z}(t) = z(t+T^*)$  l'est aussi. Ainsi, on peut construire une seconde base de solutions  $\tilde{z}_1(t), \dots, \tilde{z}_{2N}(t)$ . Il existe alors une matrice  $\Phi$  telle que  $[\tilde{z}_1(t) \dots \tilde{z}_{2N}(t)] = [z_1(t) \dots z_{2N}(t)]\Phi$ .

Cette matrice est appelée matrice de monodromie. Pour l'obtenir, il suffit de remarquer que si on choisit pour conditions initiales l'ensemble :  $[z_1(0)\dots z_{2N}(0)] = I$ , matrice identité (conditions initiales choisies indépendantes pour chaque solution  $z_i$ ), alors  $[\tilde{z}_1(0)\dots \tilde{z}_{2N}(0)] = [z_1(T^*)\dots z_{2N}(T^*)] = I\Phi$ . Il suffit donc d'obtenir  $z_i(T^*)$  par intégration temporelle de (2.21) ayant pour condition initiale  $z_i(0) = \{0\ 0\dots 1\dots 0\}^t$ , le 1 étant placé au  $i$ -ième rang.

On a alors de façon simple toute solution  $z(t)$  de (2.21) qui s'écrit  $z(t) = \sum_{i=1}^{2N} \alpha_i z_i(t)$  et ainsi :

$$\begin{aligned} z(kT) &= \sum_{i=1}^{2N} \alpha_i z_i(kT) = [z_1(kT)\dots z_{2N}(kT)]\{\alpha_1\dots\alpha_{2N}\}^t = [z_1(0)\dots z_{2N}(0)]\Phi^k\{\alpha_1\dots\alpha_{2N}\}^t \\ &= \Phi^k\{\alpha_1\dots\alpha_{2N}\}^t = \Phi^k z(0) \end{aligned}$$

On voit donc que la stabilité d'une solution périodique dépend du module des valeurs propres  $\rho_i$  de  $\Phi$  (c.f. section 2.3.2 pour l'analyse de ses valeurs).

Ce résultat succinctement démontré repose sur un théorème démontré par Gaston Floquet (1847-1920) et qui porte son nom :

### **Théorème 2.1 : Théorème de Floquet**

Si  $A(t)$  est une matrice périodique de période minimale  $T$  et  $X(t)$  le système fondamental de solutions associé à l'équation  $\dot{X} = A(t)X$ , alors il existe une matrice périodique inversible  $P(t)$  et une matrice constante  $R$  telles que  $X(t) = P(t)e^{Rt}$  pour tout  $t$ .  $\square$

La matrice périodique mentionnée dans ce théorème n'est autre pour nous que la matrice  $D_x F(x^*(t))$  à laquelle la  $T^*$ -périodicité de  $x^*$  se transmet. Le système fondamental, quant à lui est donné par  $Z(t) = [z_1(t)\dots z_{2N}(t)]$ .

En prenant pour base de départ  $B$  la base canonique,  $B = I_{2N}$ , on peut écrire :

$$Z(0) = P(0) = I \text{ et } Z(T^*) = \Phi = P(T^*)e^{RT^*} = P(0)e^{RT^*} \Rightarrow \Phi = e^{RT^*}$$

Ainsi, aux valeurs propres de  $\Phi$  - les  $\rho_i$  - on peut aisément faire correspondre les valeurs propres de  $R$ , notées  $\lambda_i$  :

$$\rho_i = e^{\lambda_i T^*}$$

On retrouve alors de façon évidente le résultat énoncé précédemment sur le lien entre le module des valeurs propres  $\rho_i$  et la stabilité de la solution  $x^*$ .

## **2.4 Algorithmes de continuation**

Nous consacrons cette dernière partie aux méthodes de suivi des solutions, appelées méthodes de *continuation* lorsque l'on fait varier un des paramètres du système (noté  $\mu$ , formulation (1.5) du problème). On se place dans le contexte où l'on possède déjà une ou plusieurs solutions pour une valeur initiale  $\mu_0$  du paramètre.

Pour chaque solution, nous cherchons donc à construire la courbe  $H(x(s), \mu(s)) = 0$ , paramétrée par l'abscisse curviligne  $s$ . Un paramétrage par  $\mu$  n'est en effet pas un bon choix dans la mesure où dans le cas de bifurcation type turning point, la courbe se « replie ».  $\mu$  est

donc une inconnue du problème. Lorsque l'on possède un point  $(x_n(s_n), \mu_n(s_n))$  de la courbe, le principe général de la continuation est ainsi de faire une prédiction sur le point suivant  $(x_{n+1}, \mu_{n+1})$  pour une abscisse curviligne  $s_{n+1} = s_n + \Delta s$ , puis de corriger ce point prédit jusqu'à retomber sur la courbe étudiée.

Nous allons voir dans un premier temps les principaux moyens de prédiction et de correction. Dans un second temps, nous exposons brièvement les principes permettant d'adapter la finesse de description de la courbe (i.e. la distance entre deux points successifs, liée à  $\Delta s$ ) et quelques méthodes utiles lorsque plusieurs solutions se rencontrent en un même point.

### 2.4.1 Principales méthodes de prédiction et de correction

Comme on a pu le voir en § 2.3, une solution non constante de (1.5) à  $\mu$  donné, n'est pas décrite par son signal temporel continu mais par certaines variables (les  $a_k$  et  $b_k$  dans le cas de la balance harmonique, le couple  $(x_c, T^*)$  dans le cas d'une méthode de tir, etc). Ainsi, lorsque nous parlons de la continuation d'une solution, c'est en fait la continuation de ses descripteurs (dont nous noterons  $P$  le nombre). Ces variables de description ne sont alors pas solution d'une équation différentielle mais d'une équation algébrique de la forme (2.10). C'est donc plus souvent sur cette équation que nous nous appuyerons, équation dans laquelle on fait apparaître le paramètre de continuation explicitement :

$$H(\tilde{x}, \mu) = 0, \text{ avec } H : \mathbb{R}^P \times \mathbb{R} \rightarrow \mathbb{R}^P \quad (2.22)$$

Nous avons donc pour objectif de construire la courbe  $H(\tilde{x}(s), \mu(s)) = 0$  paramétrée par son abscisse curviligne  $s$ . L'algorithme général d'une méthode de continuation est donné en figure 2.3 et les deux sous-sections suivantes proposent les méthodes de prédiction et de correction les plus fréquemment rencontrées. Cet algorithme général est à appliquer à chacune des solutions obtenues par une analyse globale pour la valeur  $\mu_0$  du paramètre de continuation. En cas de bifurcation détectée, il faut, selon le type de bifurcation, ajouter une solution à la liste des solutions à continuer (bifurcation continue) ou procéder à une analyse globale pour détecter la ou les nouvelles solutions stables apparaissant (bifurcation catastrophique).

Nous utiliserons dans cette section la notation suivante qui regroupe la valeur des  $P$  descripteurs et celle du paramètre de continuation :

$$y = \begin{Bmatrix} \tilde{x} \\ \mu \end{Bmatrix} \in \mathbb{R}^{P+1}$$

#### Méthodes de prédiction

Nous considérons comme acquis les points  $y_0, \dots, y_n$  d'abscisses curvilignes respectives  $s_0, \dots, s_n$ , solution de (2.22). Les méthodes ci-dessous servent à donner une approximation  $y_{n+1}^{(0)}$  du prochain point  $y_{n+1}$ . La prédiction est faite pour une abscisse curviligne  $s_n + \Delta s$ ,  $\Delta s$  étant imposé. Le réglage de sa valeur requiert une certaine attention et est traité en § 2.4.2. Ces méthodes sont illustrées en figure 2.4.

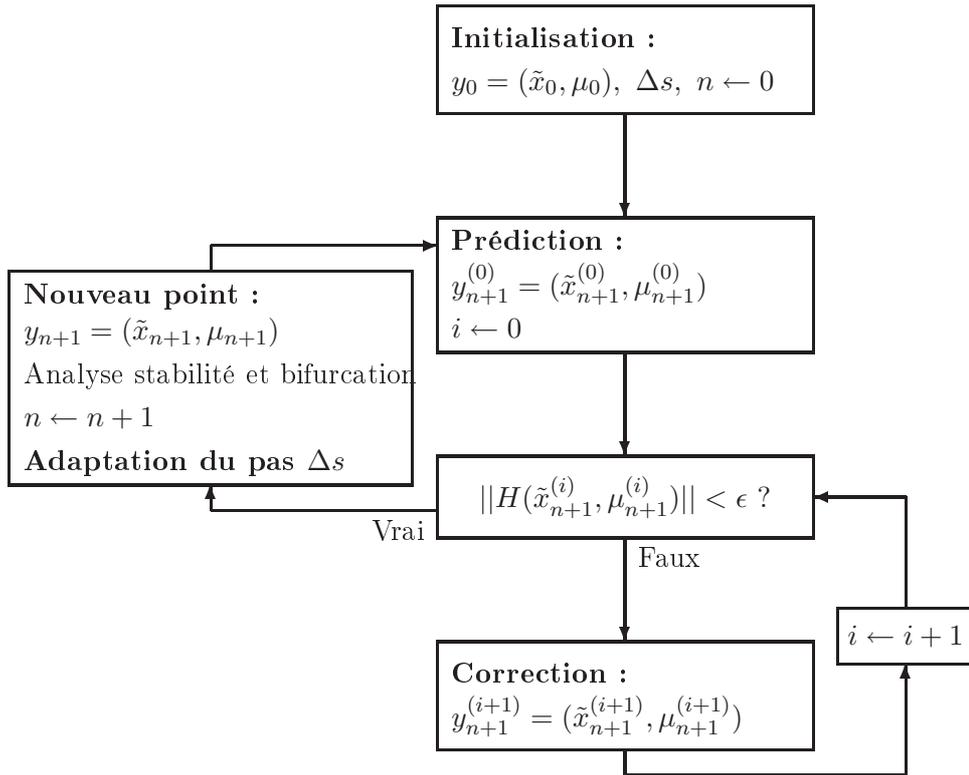


FIG. 2.3 – Algorithme général d'une méthode de continuation

**Méthode de la sécante** On considère les deux dernier points  $y_n = \{x_n, \mu_n\}^t$ ,  $y_{n-1}$ . La prédiction est faite selon le vecteur défini par ces deux points :

$$y_{n+1}^{(0)} = y_n + \Delta s \cdot \frac{y_n - y_{n-1}}{\|y_n - y_{n-1}\|}$$

**Méthode de la tangente** La prédiction est réalisée selon la tangente  $\vec{t}$  à la courbe orientée  $H(x(s), \mu(s)) = 0$  au dernier point connu  $y_n$ . On peut aussi considérer ce vecteur comme la tangente induite par la matrice jacobienne du problème  $D_y H(y_n)$ . Elle est alors définie de la façon suivante :

**Définition 2.4 : Tangente induite par une matrice**

La tangente induite par une matrice  $A$  est l'unique vecteur  $t$  tel que :

$$\begin{aligned} At &= 0 \\ \|t\| &= 1 \\ \det \begin{pmatrix} A \\ t^t \end{pmatrix} &> 0 \end{aligned}$$

□

Le calcul de  $\vec{t}$  est fait en écrivant :

$$D_{\tilde{x}} H(y_n) \cdot \Delta \tilde{x} + D_{\mu} H(y_n) \cdot \Delta \mu = 0, \Leftrightarrow \Delta \tilde{x} = \Delta \mu \vec{z}, \vec{z} = (-D_{\tilde{x}} H(y_n)^{-1} \cdot D_{\mu} H(y_n))$$

ce qui permet d'obtenir une droite porteuse. Puis, la longueur du vecteur  $\vec{t}$  étant normée à 1, on écrit :

$$\Delta\tilde{x}^t \cdot \Delta\tilde{x} + \Delta\mu^2 = (\vec{z}^t \cdot \vec{z} + 1)\Delta\mu^2 = 1 \Rightarrow \Delta\mu = \pm 1/\sqrt{\vec{z}^t \cdot \vec{z} + 1}$$

La direction enfin est donnée par le signe du déterminant :

$$\det \begin{pmatrix} D_{\tilde{x}}H & D_{\mu}H \\ \vec{t}^t & 1 \end{pmatrix} = \det \begin{pmatrix} D_{\tilde{x}}H & D_{\mu}H \\ \vec{z}^t & 1 \end{pmatrix} \cdot \Delta\mu > 0$$

Ainsi, on obtient le point prédit par :

$$y_{n+1}^{(0)} = y_n + \Delta s \cdot \vec{t}$$

**Méthodes d'ordres supérieurs** La méthode de la sécante revient en terme de polynôme à un polynôme d'ordre 1 en  $s$  où  $s$  est l'abscisse curviligne (où plutôt l'abscisse curviligne d'une approximation polygonale de la courbe suivie). La méthode de la tangente, quand à elle, est issue d'un développement limité au premier ordre de  $H(y_n + \Delta y)$ . On en parle donc comme des prédicteurs d'ordre 1. Si l'on utilise plus de points déjà connus, on peut construire des prédicteurs d'ordre supérieurs. Nous présentons ici une catégorie de prédicteurs basés sur une interpolation par les polynômes de Lagrange.

Si l'on veut approcher localement  $H$  par un polynôme  $P$  d'ordre  $m$ , il faut se donner  $m + 1$  points  $y_n, \dots, y_{n-m}$  qui sont d'abscisses curvilignes respectives  $s_n, \dots, s_{n-m}$ . Puis, on détermine le polynôme  $P$  grâce à  $\forall i \in \{n-m, \dots, n\}$ ,  $P(s_i) = y_i$ . Les polynômes qui permettent de résoudre rapidement ce système sont les polynômes de Lagrange.  $P$  s'écrira donc :

$$P(s) = \sum_{i=n-m}^n \left( y_i \prod_{\substack{k=n-m \\ k \neq i}}^n \frac{s - s_k}{s_i - s_k} \right)$$

Il ne reste plus qu'à effectuer la prédiction en calculant  $P(s_n + \Delta s)$  :

$$y_{n+1}^{(0)} = P(s_n + \Delta s)$$

Pratiquement, il est simple de se donner pour origine de l'abscisse curviligne  $s_n = 0$  (i.e.  $P(0) = y_n$ ), puis d'approcher le calcul des  $s_{n-1}, \dots, s_{n-m}$  qui découle normalement d'une intégration le long de la courbe étudiée (et que l'on ne possède pas ici de façon continue) par la somme des longueurs des segments qui relient  $y_n$  à  $y_{n-i}$  :

$$\begin{cases} s_n = 0 \\ \forall i \in \{1, \dots, m\}, s_{n-i} = s_{n-i+1} - \|y_{n-i+1} - y_{n-i}\| < 0 \end{cases}$$

### Méthodes de correction

Nous considérons ici comme acquis les points  $y_0, \dots, y_n$  d'abscisses curvilignes respectives  $s_0, \dots, s_n$  et une prédiction au rang suivant  $y_{n+1}^{(0)}$  que nous corrigerons par une des méthodes ci-dessous pour obtenir le point  $y_{n+1}$ . Dans la mesure où l'on doit corriger  $P + 1$  inconnues à l'aide de  $P$  équations ( $H(\tilde{x}, \mu) \in \mathbb{R}^P$ ), il est nécessaire de fixer arbitrairement une des inconnues (ce que fait la méthode de Newton) ou d'ajouter une équation, ce qui est souvent un choix plus judicieux et le cas des méthodes Pseudo-arclength et de Moore-Penrose. Ces méthodes sont illustrées en figure 2.5.

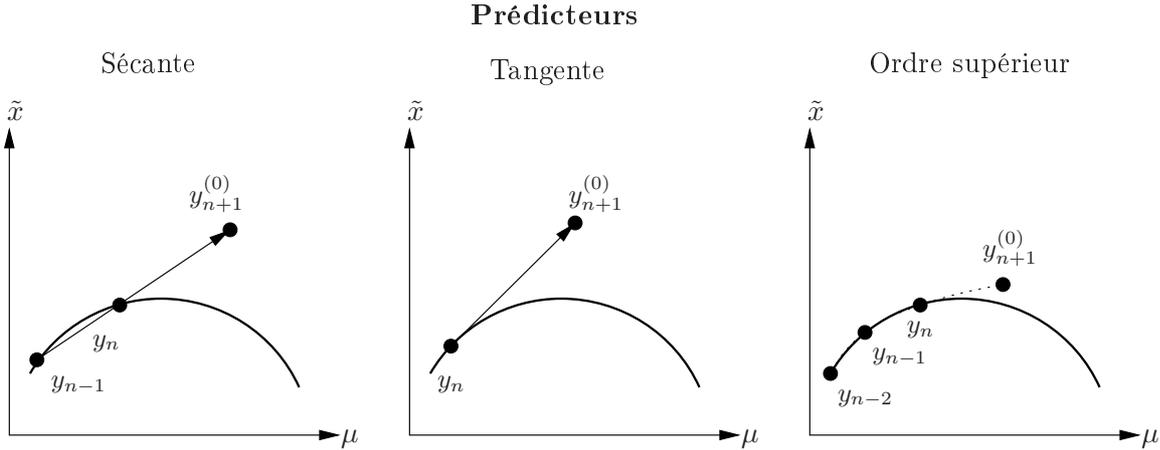


FIG. 2.4 – Continuation : Principaux prédicteurs

**Méthode naturelle ou de Newton** Cette méthode a pour particularité de fixer l'inconnue  $\mu_{n+1}$  à sa valeur de prédiction :

$$\mu_{n+1} = \mu_{n+1}^{(0)}$$

Ce type de correction est assez peu efficace lorsque l'on effectue la continuation de courbes présentant des retournements. On écrit ensuite simplement un schéma de type Newton-Raphson :

$$H(\tilde{x}_{n+1}^{(i+1)}, \mu_{n+1}) \approx H(\tilde{x}_{n+1}^{(i)}, \mu_{n+1}) + D_{\tilde{x}}H(\tilde{x}_{n+1}^{(i)}, \mu_{n+1}) \cdot \Delta\tilde{x}$$

Ce qui permet d'écrire le point corrigé :

$$y_{n+1}^{(i+1)} = \left\{ \begin{array}{c} \tilde{x}_{n+1}^{(i)} + \Delta\tilde{x} \\ \mu_{n+1} \end{array} \right\}, \quad \Delta\tilde{x} = -D_{\tilde{x}}H(\tilde{x}_{n+1}^{(i)}, \mu_{n+1})^{-1}H(\tilde{x}_{n+1}^{(i)}, \mu_{n+1})$$

**Méthode Pseudo-arclength** Cette méthode consiste en la correction simultanée de  $\tilde{x}$  et  $\mu$  en imposant comme équation manquante que le point corrigé se trouve sur l'hyperplan de  $\mathbb{R}^{P+1}$  normal au vecteur  $y_{n+1}^{(0)} - y_n$ , ce que l'on peut écrire :

$$\begin{aligned} & \langle y_{n+1}^{(i+1)} - y_{n+1}^{(0)}, y_{n+1}^{(0)} - y_n \rangle = 0 \\ \Leftrightarrow & \langle y_{n+1}^{(i)} - y_{n+1}^{(0)}, y_{n+1}^{(0)} - y_n \rangle + \langle \Delta y, y_{n+1}^{(0)} - y_n \rangle = 0 \\ \Leftrightarrow & \langle \Delta y, y_{n+1}^{(0)} - y_n \rangle = 0 \end{aligned}$$

Les équations s'écrivent donc :

$$y_{n+1}^{(i+1)} = y_{n+1}^{(i)} + \begin{Bmatrix} \Delta\tilde{x} \\ \Delta\mu \end{Bmatrix}, \quad \begin{bmatrix} D_{\tilde{x}}H(y_{n+1}^{(i)}) & D_{\mu}H(y_{n+1}^{(i)}) \\ \tilde{x}_{n+1}^{(0)} - x_n & \mu_{n+1}^{(0)} - \mu_n \end{bmatrix} \begin{Bmatrix} \Delta\tilde{x} \\ \Delta\mu \end{Bmatrix} = \begin{Bmatrix} -H(y_{n+1}^{(i)}) \\ 0 \end{Bmatrix}$$

**Méthode de Moore-Penrose** Cette méthode effectue aussi des corrections à la fois sur  $x$  et  $\mu$  en utilisant l'inverse (ou pseudo-inverse) de Moore-Penrose. En effet, l'équation manquante se traduit par le caractère non-carré de la matrice jacobienne du problème (2.22). On ne peut donc pas inverser la jacobienne pour trouver la correction à appliquer. Il est donc intéressant d'utiliser alors non pas l'inverse mais la pseudo-inverse de Moore-Penrose dont la définition est la suivante :

**Définition 2.5 : Matrice Pseudo-Inverse de Moore-Penrose**

Si  $A$  est une matrice de taille  $P \times M$ , son inverse (ou pseudo-inverse) de Moore-Penrose est déterminée par :

$$A^+ = A^t(AA^t)^{-1}$$

Elle possède les propriétés suivantes :

- $\forall u \in \mathbb{R}^P, A.(A^+.u) = u$
- $\forall (u, v) \in \mathbb{R}^P \times \text{Ker}(A), \langle A^+.u, v \rangle = 0$

□

En notant  $D^+H(y_{n+1}^{(i)})$  la pseudo-inverse de la jacobienne  $D_y H(y_{n+1}^{(i)}) = [D_{\tilde{x}} H(y_{n+1}^{(i)}) \ D_{\mu} H(y_{n+1}^{(i)})]$ , et effectuant la correction suivante :

$$y_{n+1}^{(i+1)} = y_{n+1}^{(i)} + \begin{Bmatrix} \Delta \tilde{x} \\ \Delta \mu \end{Bmatrix}, \quad \begin{Bmatrix} \Delta \tilde{x} \\ \Delta \mu \end{Bmatrix} = -D^+H(y_{n+1}^{(i)}) \cdot H(y_{n+1}^{(i)})$$

on voit d'après la seconde propriété énoncée que la correction est effectuée orthogonalement au noyau de  $D_y H(y_{n+1}^{(i)})$  (direction  $(\Delta_i)$  en figure 2.5), ce qui est intéressant étant donné que toute contribution appartenant à son noyau (symbolisé par les vecteurs  $\vec{v}_i$  en figure 2.5) ne permet pas de diminuer la norme de  $H(y_{n+1}^{(i)})$ .

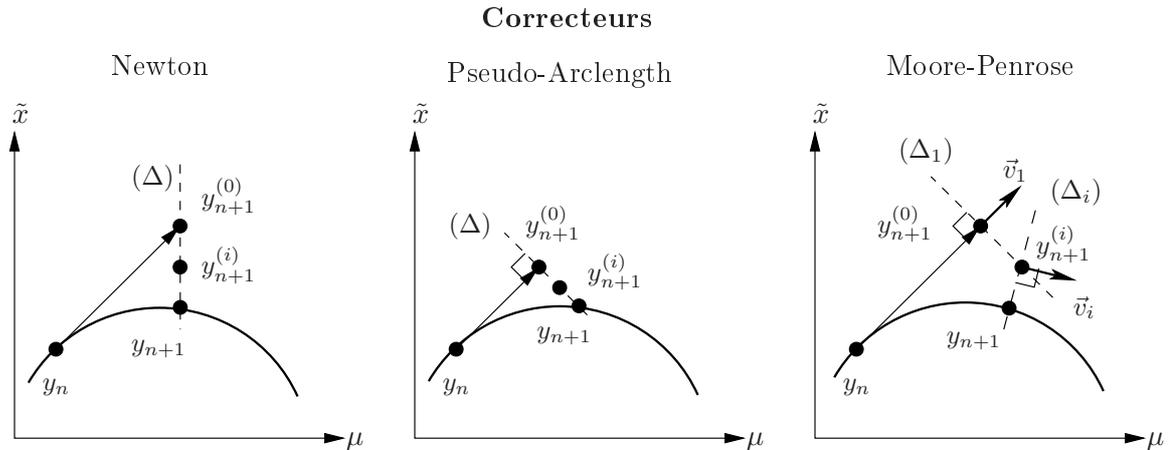


FIG. 2.5 – Continuation : Principaux correcteurs

**2.4.2 Adaptation de la distance entre deux points successifs**

Lorsque l'on effectue la prédiction d'un nouveau point, nous avons vu qu'il était nécessaire de se donner  $\Delta s$ . Cet incrément d'abscisse curviligne est un paramètre important à régler. En effet, plus il est grand, moins on utilise de points pour décrire la courbe, ce qui est a priori avantageux en terme de coût de calcul. Inversement, plus il est grand, plus le point prédit est éloigné du précédent et plus mauvaise est la prédiction, donc plus il faudra d'itérations de correction pour retrouver notre courbe. On voit donc qu'il va falloir faire un compromis entre ces deux aspects. La figure 2.6 illustre deux cas de figure classiques. La première partie de la courbe présente une courbure faible ; on peut donc espacer les points la décrivant : les prédictions resteront bonnes et le nombre d'itérations de correction nécessaires sera faible. En

revanche, la seconde partie de la courbe présente un point de retournement, donc une forte courbure ; il faut donc un grand nombre de points pour le décrire correctement. Il est logique de ne pas utiliser le même pas  $\Delta s$  dans les deux cas mais de l'adapter (on parle de *steplength adaptation* en anglais).

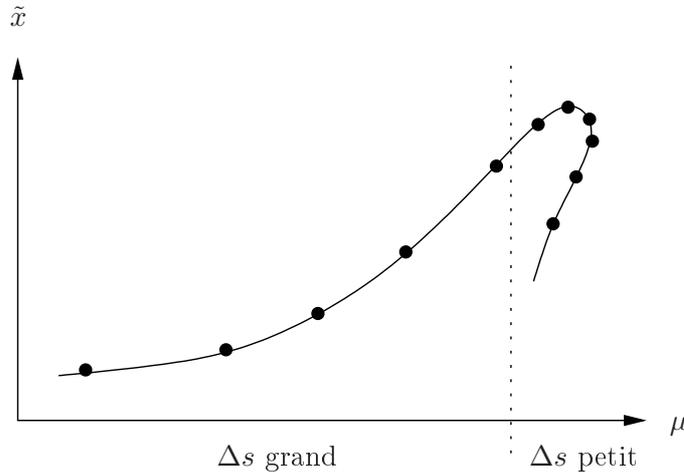


FIG. 2.6 – Intérêt de l'adaptation du pas de continuation

Il existe plusieurs tactiques pour adapter ce pas. On trouvera l'essentiel résumé dans [AG03, Chap. 6], [Sey88, Sec. 4.4] et [Kuz04, Sec. 10.2.1], qui propose une solution très simple et relativement efficace à mettre en œuvre : on se donne un pas initial et un nombre d'itérations de correction maximal  $N_{cor}^{max}$  et minimal  $N_{cor}^{min}$ . Si après la prédiction  $n + 1$  utilisant un pas  $\Delta s$ , la correction nécessite plus de  $N_{cor}^{max}$  itérations, la correction est arrêtée et l'on refait une prédiction avec un pas plus petit (par exemple,  $\Delta s/2$ ). Si la correction converge après moins de  $N_{cor}^{min}$  itérations, on utilisera pour la prédiction suivante un pas plus important (par exemple  $2\Delta s$ ). Si le nombre d'itérations nécessaires à la correction est compris entre les deux bornes, le même pas est conservé pour la prédiction suivante.

En ce qui concerne l'initialisation de la valeur du pas, nous préconisons d'utiliser un ordre de grandeur identique à la discrétisation en  $\mu$  attendue. En effet, on peut s'attendre à un démarrage « plat de la courbe ». L'essentiel de la variation d'abscisse curviligne revient alors à celle du paramètre de continuation :

$$\Delta s = \sqrt{\Delta \tilde{x}^2 + \Delta \mu^2} \approx |\Delta \mu| \quad (2.23)$$

### 2.4.3 Branch switching

Lorsque plusieurs solutions *de type point fixe* se rejoignent ou partent d'un même point, on ne poursuit avec une technique de continuation usuelle que l'une de ces branches de solutions. Or on désire généralement les tracer toutes. Il faut donc pouvoir passer d'une branche de solution à l'autre. On appelle le point de rencontre de ces solutions un *point d'embranchement* (branching point) dont nous donnons la définition issue de [Kuz04, Sec. 10.2.4] :

**Définition 2.6 : Point d'embranchement (Branching Point)**

Un point  $x_0$  est appelé point d'embranchement pour le problème de continuation :

$$F(x, \mu) = 0, \quad F : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^N \quad (2.24)$$

si  $F(x_0, \mu_0) = 0$  et qu'il existe au moins deux courbes régulières satisfaisant  $F(x, \mu) = 0$  et passant par  $(x_0, \mu_0)$ .  $\square$

Ainsi, lorsque l'on réalise une continuation d'un problème continu en temps et qu'une bifurcation de type Hopf se produit, on ne se trouve pas dans ce cas de figure. En revanche, l'étude menée dans le cas discret à l'aide d'une section de Poincaré permet de se ramener à l'étude de points fixes, et donc de correspondre à la définition donnée.

Deux grandes catégories d'approches permettent alors de suivre l'une ou l'autre des branches : soit par construction des tangentes, soit par perturbation du système, ce qui tend généralement à « casser » l'embranchement en deux branches qui ne se coupent plus (voir figure 2.7).

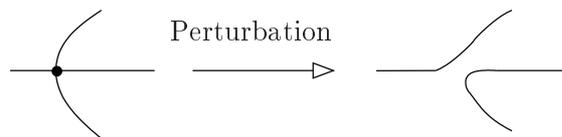


FIG. 2.7 – Destruction du point d'embranchement par perturbation

**Points d'embranchement**

Nous proposons de détailler les méthodes d'embranchement par recherche des tangentes. L'idée générale est d'avoir recours à la hessienne de  $F$  étant donné que les tangentes recherchées annulent la Jacobienne. On se ramènera au cas le plus fréquent où seules deux branches se rencontrent, et on posera une condition sur la hessienne pour qu'il existe deux tangentes distinctes.

Notons  $y_0 = \{x_0, \mu_0\}^t \in \mathbb{R}^{N+1}$  un point d'embranchement.

Localement, on peut écrire :

$$F(y_0 + \Delta y) = F(y_0) + D_y F(y_0) \cdot \Delta y + D_{yy} F(y_0) \cdot (\Delta y, \Delta y) + O(\|\Delta y\|^3)$$

$$\text{avec } D_{yy} F(y_0) \cdot (\Delta x, \Delta y) = \left\{ \sum_{j,k=1}^{N+1} \frac{\partial^2 F_i(y_0)}{\partial y_j \partial y_k} x_j y_k \right\}_{i=1..N}, \text{ la Hessienne de } F \text{ en } y_0.$$

Paramétrons selon l'abscisse curviligne une des branches de solution :  $c_1(s)$  telle que  $c_1(0) = y_0$  et pour  $s$  au voisinage de 0,  $F(c_1(s)) = 0$ . (L'existence de  $c_1$  est fournie par le théorème des fonctions implicites). En différenciant la relation précédente par rapport à  $s$ , on obtient :  $D_y F(c_1(s)) \cdot \dot{c}_1(s) = 0$ ; les tangentes recherchées  $\dot{c}_i(s)$  annulent donc la jacobienne  $D_y F(y_0)$ . On note  $K = \ker(D_y F(y_0))$  le noyau de la jacobienne en  $y_0$ .

Le cas le plus fréquent est la rencontre de deux branches distinctes de solution. On est donc en recherche de deux tangentes  $t_1$  et  $t_2$  :  $\dim(K) = 2$ . Notons  $q_1, q_2$  deux vecteurs indépendants formant une base de  $K$ . Le théorème du rang nous permet de déduire de  $\dim(K) = 2$

que  $\dim(\ker(D_y F(y_0)^t)) = 1$ . Il existe donc un unique vecteur  $\phi$  non nul (à une multiplication par un scalaire près) tel que  $D_y F(y_0)^t \cdot \phi = 0$ . Nous allons utiliser ce vecteur pour pouvoir construire une preuve d'existence de deux tangentes distinctes.

Différencions deux fois  $F(c_1(s)) = 0$  par rapport à  $s$  et prenons l'équation obtenue en  $s = 0$  :

$$D_y F(y_0) \cdot \dot{c}_1(0) + D_{yy} F(y_0) \cdot (\dot{c}_1(0), \dot{c}_1(0)) = 0$$

Si l'on fait le produit scalaire à gauche avec  $\phi$  de cette relation, on obtient :

$$\langle \phi, D_y F(y_0) \cdot \dot{c}_1(0) \rangle + \langle \phi, D_{yy} F(y_0) \cdot (\dot{c}_1(0), \dot{c}_1(s)) \rangle = 0$$

En remarquant que  $\langle \phi, D_y F(y_0) \cdot \dot{c}_1(0) \rangle = \langle D_y F^t(y_0) \cdot \phi, \dot{c}_1(0) \rangle = \langle 0, \dot{c}_1(0) \rangle = 0$ , on obtient enfin :

$$\langle \phi, D_{yy} F(y_0) \cdot (\dot{c}_1(0), \dot{c}_1(0)) \rangle = 0$$

On peut alors considérer le membre de gauche comme une fonction quadratique définie sur  $K$ . En écrivant que tout vecteur  $v$  de  $K$  se décompose sur  $q_1$  et  $q_2$  selon  $v = v_1 q_1 + v_2 q_2$ , on a :

$$\langle \phi, D_{yy} F(y_0) \cdot (v, v) \rangle = v_1^2 b_{11} + v_2^2 b_{22} + 2v_1 v_2 b_{12}, \quad b_{ij} = \langle \phi, D_{yy} F(y_0) \cdot (q_i, q_j) \rangle$$

En posant que l'un des vecteurs de la base de  $K$  vaut  $q_1 = t_1 = \dot{c}_1(0)$  (connu d'après la continuation), on obtient facilement un vecteur  $q_2$  en le cherchant orthogonal à  $q_1$  et dans  $K$  (tel que  $D_y F(y_0) \cdot q_2 = 0$ ). Il ne reste donc plus qu'à trouver les coefficients  $\beta_1$  et  $\beta_2$  tels que  $t_2 = \beta_1 q_1 + \beta_2 q_2$ .

Quels que soient ces coefficients,  $t_2 \in K$  par construction. La dernière équation à satisfaire pour  $t_2 = \dot{c}_2(0)$  est donc :  $\langle \phi, D_{yy} F(y_0) \cdot (\dot{c}_2(0), \dot{c}_2(0)) \rangle = 0$ . Ce que l'on peut écrire :

$$\begin{Bmatrix} \beta_1 \\ \beta_2 \end{Bmatrix}^t \begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix} \begin{Bmatrix} \beta_1 \\ \beta_2 \end{Bmatrix} = 0$$

Une solution évidente est  $(\beta_1, \beta_2) = (1, 0)$  qui redonne  $t_1$ . Le problème ci-dessus possède deux solutions distinctes (donc deux tangentes distinctes) non triviales si le déterminant est négatif :  $b_{11} b_{22} - b_{12}^2 < 0$ . En réinjectant que  $t_1 = 1 \cdot q_1 + 0 \cdot q_2$ , on obtient  $b_{11} = 0$  et par suite  $\beta_1 = -b_{22}/(2b_{12})\beta_2$ . Il ne reste plus qu'à fixer  $\beta_2$  en imposant par exemple une condition de normalisation du type  $\|t_2\| = 1$ .

Obtenir numériquement  $\phi$ ,  $q_1$  et  $q_2$  ne pose pas de problème. Pour approcher les quantités  $b_{ij}$ , on peut utiliser les différences finies en écrivant :

$$D_{yy} F(y_0) \cdot (q_i, q_i) \approx (F(y_0 + h q_i) + F(y_0 - h q_i)) / h^2 \text{ et}$$

$$D_{yy} F(y_0) \cdot (q_i, q_j) = (1/4) (D_{yy} F(y_0) \cdot (q_i + q_j, q_i + q_j) - D_{yy} F(y_0) \cdot (q_i - q_j, q_i - q_j))$$

Ce qui précède justifie la définition et le lemme suivants :

### Définition 2.7 : Point d'embranchement simple (Simple branching point)

Un point d'embranchement  $y_0$  est dit simple si :

1.  $\dim(\ker(D_y F(y_0))) = 2$
2.  $b_{11} b_{22} - b_{12}^2 < 0$

□

**Lemme 2.2 : Lemme relatif aux points d'embranchement**

Soit  $y_0$  un point d'embranchement simple. Alors il existe deux courbes régulières passant par  $y_0$  et satisfaisant (2.24). De plus, leurs tangentes  $t_i$  en  $y_0$  sont linéairement indépendantes et satisfont toutes deux  $\langle \phi, D_{yy}F(y_0) \cdot (t_i, t_i) \rangle = 0$ .  $\square$

**Points de bifurcation de Hopf**

Comme on a pris le soin de le préciser en introduction de cette section, la naissance d'une branche de solutions périodiques ne rentre pas dans le cadre de l'étude précédente. On souhaite néanmoins pouvoir « attraper » ces branches de solutions périodiques. On peut alors utiliser deux approches : la première, plus radicale, est de transformer le problème périodique en un problème de point fixe par le biais d'une section de Poincaré. On peut alors appliquer la théorie de la section 2.4.3. La seconde méthode est plus pragmatique : si une bifurcation de type Hopf a été détectée en  $(x_0, \mu_0)$ , on possède les deux valeurs propres conjuguées de partie réelle nulle :  $\lambda_i = +j\omega_h$  et  $\bar{\lambda}_i = -j\omega_h$ . On sait alors que la solution naissante vaut la solution type point fixe initiale perturbée d'une composante de période  $T_h$ , avec  $T_h = 2\pi/\omega_h$ . De plus, on peut facilement se procurer les deux vecteurs propres associés  $v_i$  et  $\bar{v}_i$ . Il suffit donc de chercher une solution périodique à proximité du point de bifurcation, avec pour point de départ de la recherche  $x_0 + \text{Re}(v_i) \cos(\omega_h t) + \text{Im}(v_i) \sin(\omega_h t)$ .

Cependant, il existe deux types principaux de bifurcation de Hopf (c.f. figure 1.7) lorsque l'on écarte le cas dégénéré. Pour savoir dans quel cas, sur ou sous-critique, on se trouve et donc de quel côté de  $\mu_0$  chercher, il est possible de calculer  $l_1$ , premier nombre de Lyapunov. Les étapes de calcul étant quelque peu fastidieuses, nous ne les détaillons pas ici. Elles sont très clairement résumées dans [Kuz04, Sec. 10.2.3]. Lorsque l'on a effectué ce calcul, le signe de  $l_1$  donne le cas dans lequel on se trouve :  $l_1 < 0$  correspond au cas sur-critique et  $l_1 > 0$  correspond au cas sous-critique.

**2.5 Méthode de réduction de la taille du problème**

La recherche des cycles limites est fréquemment réalisée par une méthode de balance harmonique (cf. § 2.3.2). Il est possible, lors de l'utilisation de cette méthode, de réaliser la majeure partie des calculs sur un faible nombre de ddl non-linéaires en procédant comme suit : on partitionne les variables  $\tilde{x}_i$ ,  $i = 1 \dots P$  en  $V_{nl}$  variables non-linéaires dont dépend la fonction non-linéaire  $\hat{H}(\tilde{x})$  en  $V_l = P - V_{nl}$  variables dites linéaires n'intervenant pas dans l'expression de cette fonction. On peut alors réordonner le vecteur d'inconnues en plaçant les inconnues linéaires  $\tilde{x}_l$  en tête et les inconnues non-linéaires  $\tilde{x}_{nl}$  en pied de vecteur, et réécrire synthétiquement les équations du système (2.20) comme suit :

$$\begin{bmatrix} \bar{H}_{ll} & \bar{H}_{lnl} \\ \bar{H}_{nll} & \bar{H}_{nlnl} \end{bmatrix} \begin{Bmatrix} \tilde{x}_l \\ \tilde{x}_{nl} \end{Bmatrix} + \begin{Bmatrix} \hat{H}_l(\tilde{x}_{nl}) \\ \hat{H}_{nl}(\tilde{x}_{nl}) \end{Bmatrix} = \begin{Bmatrix} H_{cl} \\ H_{cnl} \end{Bmatrix} \quad (2.25)$$

En séparant les deux groupes d'équations, on peut ainsi effectuer la résolution sur les variables non-linéaires uniquement en traitant (2.26b) puis obtenir les variables linéaires par la relation (2.26a).

$$\tilde{x}_l = \bar{H}_{ll}^{-1} \left[ H_{cl} - \bar{H}_{lnl} \tilde{x}_{nl} - \hat{H}_l(\tilde{x}_{nl}) \right] \quad (2.26a)$$

$$[\bar{H}_{nlnl} - \bar{H}_{nll}\bar{H}_{ll}^{-1}\bar{H}_{lnl}] \tilde{x}_{nl} + [\hat{H}_{nl}(\tilde{x}_{nl}) - \bar{H}_{nll}\bar{H}_{ll}^{-1}\hat{H}_l(\tilde{x}_{nl})] = H_{cnl} - \bar{H}_{nll}\bar{H}_{ll}^{-1}H_{cl} \quad (2.26b)$$

Ainsi, si l'on avait  $N$  ddl initialement dont  $N_{nl}$  ddl non-linéaires (impliqués dans l'expression des non-linéarités) et que l'on utilise une méthode de balance harmonique à  $p$  harmoniques, on résout un système de taille au plus  $V_{nl} = (2p + 1)N_{nl}$  au lieu de considérer un système de taille  $P = (2p + 1)N$ . Pour des structures dont les non-linéarités sont issues d'organes spécifiques ou de contact, le gain peut être très important.

## 2.6 Conclusion

Les outils présentés dans ce chapitre permettent d'étudier localement les solutions de type point fixe ou périodique de systèmes non-linéaires : ils permettent de construire des solutions, d'étudier leur stabilité, leur évolution en fonction de la variation d'un paramètre du système et dans ce cadre, la détection et l'étude des bifurcations.

Il nous reste maintenant à aborder les techniques permettant de trouver non pas une seule mais toutes les solutions d'un système dynamique non-linéaire sur lesquelles nous pourrions appliquer ces techniques locales. C'est à cela que sont consacrés les quatre chapitres suivants de ce mémoire.

## Chapitre 3

# Analyse globale par cell-mapping

---

*Nous abordons avec ce chapitre l'étude des méthodes d'analyse globale adaptées aux cas de systèmes mécaniques dynamiques.*

*Premier volet des quatre familles qui seront explorées, le cell-mapping ou cell-to-cell mapping consiste d'une part en la suppression de l'aspect temporel du problème et d'autre part en une discrétisation de l'espace d'états en cellules. L'objectif est de transformer un système continu en espace et en temps en un système discret. Pour bien comprendre le cell-mapping, il est intéressant de découpler les deux aspects : suppression ou transformation de l'aspect temporel d'une part, ce qui est relatif au terme « mapping » et discrétisation de l'espace d'autre part, ce qui est relatif à la notion de cellule (cell). L'ouvrage faisant référence à ce sujet est le livre de C. S. Hsu, [Hsu87].*

### 3.1 Construction d'un mapping

Partant d'une équation différentielle (1.3) décrivant le comportement d'un système et d'un ensemble de conditions initiales  $\mathcal{E}_{IC}$ , on souhaite généralement connaître le comportement du système pour des temps longs : points fixes, solutions périodiques, quasi-périodiques et solutions chaotiques. Une méthode directe bien qu'impossible à réaliser matériellement est d'intégrer le système (1.3) en tout point  $x_0 \in \mathcal{E}_{IC}$  sur une durée infinie.

Afin d'obtenir ces solutions, nous réalisons un mapping  $\mathcal{G}$  qui est une application de  $\mathbb{R}^Q$  dans  $\mathbb{R}^Q$  ( $Q = 2N$  dans notre cas) qui discrétise l'aspect temporel. Deux types de mapping peuvent être construits afin de représenter un système de la forme (1.3) :

- en considérant un point  $x$  de  $\mathbb{R}^Q$ , on pose  $x' = \mathcal{G}(x)$  le point obtenu après l'intégration de (1.3) pendant un temps  $\tau$  fixé avec  $x$  pour condition initiale
- en considérant un hyperplan  $\Sigma$  de  $\mathbb{R}^Q$  et en attribuant pour image à tout point  $x$  de  $\Sigma$  le premier point de la trajectoire issue de l'intégration temporelle de (1.3) ayant  $x$  pour condition initiale qui coupe  $\Sigma$  de façon transverse et directe

Il va de soi que dans le second cas, si l'hyperplan  $\Sigma$  n'a pas été choisi de façon judicieuse (par exemple, les trajectoires issues de  $\Sigma$  restent dans  $\Sigma$ ), le mapping n'est pas représentatif du comportement du système. Ces mappings correspondent à la définition donnée en § 1.2.4 d'une application de Poincaré.

La figure 3.1 illustre les deux méthodes de construction. Dans le cas (a),  $Q = 2$  et dans le cas (b) l'étude d'un système autonome de dimension  $Q = 3$  est ramenée à celle d'un système de dimension 2 par intersection avec l'hyperplan  $\Sigma$ .

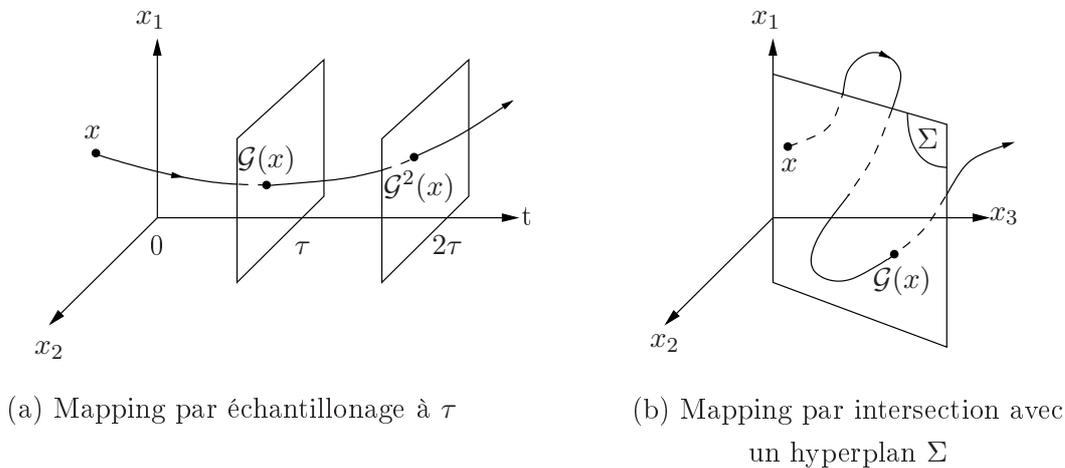


FIG. 3.1 – Modes de construction d'un mapping

Le choix entre l'une ou l'autre des méthodes de construction dépend essentiellement de la nature autonome ou non du problème. En effet, lorsque qu'un système est non-autonome, c'est fréquemment parce qu'il est soumis à une excitation périodique. Il semble alors assez naturel d'utiliser la première méthode en prenant  $\tau$  égal à la période d'excitation. Dans le cas d'un système autonome en revanche, puisqu'aucun temps particulier ne se dégage, il est préférable d'utiliser une section de Poincaré, dans la mesure où cela a l'avantage de ramener l'étude d'un système de dimension  $Q$  à celle d'un système de taille  $Q - 1$  par restriction à un hyperplan. Si l'on souhaite néanmoins utiliser une application de premier retour avec un système non-autonome  $\dot{x} = F(x, t)$ , on peut augmenter ce système en utilisant le changement de variable  $\theta = 2\pi t/\tau$  et en rajoutant l'équation  $\dot{\theta} = 2\pi/\tau$ ,  $\tau$  ayant été choisi arbitrairement.

L'intérêt d'utiliser de tels mappings est qu'il a été démontré que les propriétés des systèmes étaient conservées en termes de nombre, natures et stabilité de solutions ainsi qu'en termes de bifurcation. Bien que [Hsu87] soit l'ouvrage de référence en matière de cell-mapping, on trouvera des explications satisfaisantes sur les mapping ou point-mapping décrits ci-dessus, dans [NB95].

## 3.2 Discrétisation de l'espace

Il va de soi que s'être affranchi du problème temporel en utilisant un mapping ne suffit pas : en effet, il faudrait construire le mapping pour un espace continu inclus dans  $\mathbb{R}^Q$  ou  $\mathbb{R}^{Q-1}$  ce qui relève de l'impossible à moins de pouvoir obtenir une expression analytique du mapping. Comme ce dernier cas est peu fréquent, on a recours à une discrétisation de l'espace en cellules. On borne donc le domaine d'étude puis on le découpe en un nombre fini de cellules ; on obtient alors une collection de cellules à laquelle on ajoute une cellule puits (sink cell) qui sert à représenter tous les états en-dehors de la zone d'intérêt  $\mathcal{E}_{IC}$  que l'on s'est fixée.

La façon la plus courante de définir ces cellules est de se donner un pavé  $\mathcal{E}_{IC}$  de  $\mathbb{R}^Q$  (resp.  $\mathbb{R}^{Q-1}$  selon le mapping adopté) et de choisir un nombre de morceaux dans chaque direction de l'espace  $N_1 \times N_2 \times \dots \times N_Q$  (resp.  $N_1 \times N_2 \times \dots \times N_{Q-1}$ ) afin d'obtenir  $\prod N_i$  cellules partitionnant  $\mathcal{E}_{IC}$  auxquelles on ajoute la cellule puits pour obtenir une collection  $\mathcal{C}$  de  $N_C = \prod N_i + 1$  cellules.

Si l'on note :

- $\mathcal{E}_{IC} = [E_1^- \ E_1^+] \times \dots \times [E_Q^- \ E_Q^+]$  : l'espace d'intérêt,
- $h_i = (E_i^+ - E_i^-)/N_i$  : le pas de discrétisation dans chaque direction,

chaque cellule couvre donc les  $x$  tels que :

$$\forall i \in \llbracket 1, Q \rrbracket, E_i^- + (z_i - 1)h_i \leq x_i < E_i^- + z_i h_i$$

les  $z_i \in \{1, \dots, N_i\}$  dépendant de la cellule considérée. Il est donc assez naturel de désigner une cellule, notée  $\underline{z}$ , par ses coordonnées  $\{z_1, z_2, \dots, z_Q\}$ . On peut choisir d'attribuer à la cellule puits la coordonnée  $\{0, 0, \dots, 0\}$ .

Une fois cette discrétisation de l'espace effectuée, il faut attribuer à chaque cellule une image. Trois méthodes sont alors possibles :

- Cell-mapping simple (Simple cell-mapping) : seul un point par cellule est retenu et fait office de conditions initiales. Généralement, c'est le centre de la cellule qui est utilisé
- Cell-mapping interpolé (Interpolated cell-mapping) : plusieurs points sont utilisés sur la cellule et une interpolation est réalisée sur les différentes images
- Cell-mapping généralisé (Generalized cell-mapping) : plusieurs points sont utilisés sur la cellule et une probabilité est attribuée aux différentes images obtenues

Dans l'ordre d'énumération, le coût numérique augmente car un nombre croissant de points est utilisé pour décrire l'évolution de la cellule dans le temps. En même temps que ce coût, augmente bien évidemment la qualité de la description du système. C'est donc un compromis temps/qualité qu'il faut réaliser.

La figure 3.2 illustre un simple cell-mapping pour  $Q = 2$ . Selon cette figure, l'image de la cellule  $\underline{z}_1 = \{z_1, z_2\} = (1, 1)$  est la cellule  $\underline{z}_2 = (2, 2)$  et celle de la cellule  $\underline{z}_2$  est la cellule  $\underline{z}_3 = (3, 6)$ . Il apparaît clairement qu'il y a une approximation réalisée en amont et en aval : d'une part, le centre de la cellule vaut pour représentant de la cellule tout entière or cette approximation n'est valable que si la cellule est suffisamment petite. D'autre part, l'image du centre de  $\underline{z}_2$  n'est pas exactement le centre de  $\underline{z}_3$ . L'analyse de ce mapping, comme on le verra dans la section suivante consiste à dire que l'image de  $\underline{z}_1$  par  $\mathcal{G}^2$  est  $\underline{z}_3$ . Si l'on redonne

son sens physique à ceci, c'est que l'on considère que les points de la cellule  $\underline{z}_1$  au bout d'une durée  $2\tau$  se trouvent dans la cellule  $\underline{z}_3$ . On voit donc qu'il y a un cumul d'erreurs qui peut devenir gênant en termes d'interprétation si les cellules sont trop grossières.

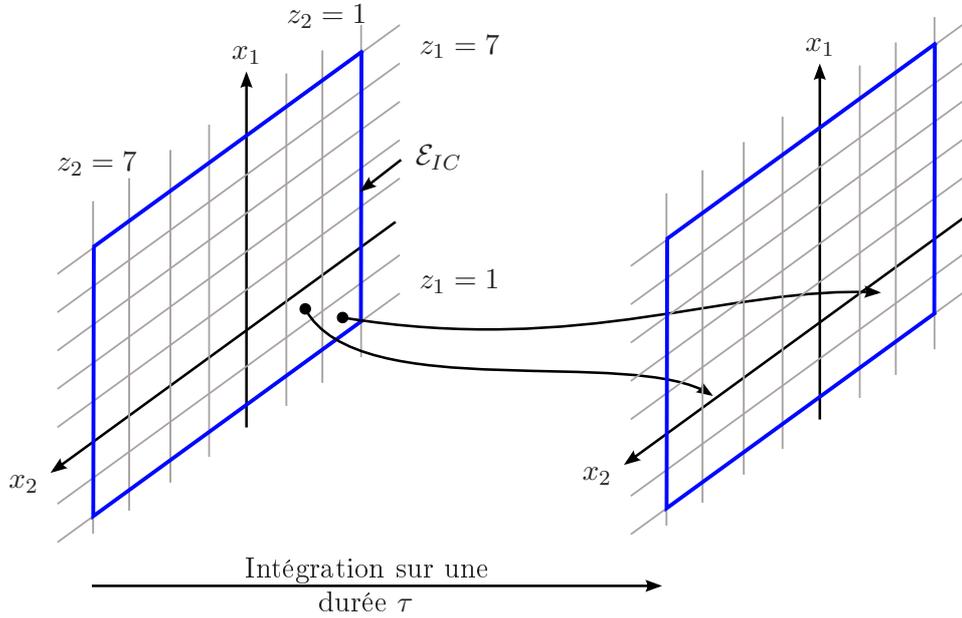


FIG. 3.2 – Illustration d'un simple cell-mapping

### 3.3 Analyse d'un cell-mapping

Une fois construit le cell-mapping  $\mathcal{G}$  qui à chaque cellule associe une cellule image selon l'une ou l'autre des techniques de mapping et de description de l'image d'une cellule, il faut analyser le résultat, c'est à dire trouver les chemins qui se referment sur eux-mêmes. Nous décrivons ici brièvement l'algorithme d'analyse proposé par [Hsu87] dans le cas d'un cell-mapping simple.

Il est proposé d'attribuer trois nombres à chaque cellule :

- $Gr$ , le numéro du groupe de cellules : toutes les cellules attirées par une même séquence périodique sont du même groupe
- $St$ , le nombre d'itérations de  $\mathcal{G}$  nécessaires pour qu'une cellule rejoigne le groupe périodique vers lequel elle se dirige ( $St = 0$  pour les cellules constituant la séquence périodique)
- $P$ , la période de la séquence périodique à laquelle appartient la cellule ou vers laquelle elle se dirige

L'analyse du cell-mapping se fait alors selon l'algorithme donné en figure 3.3 dont nous détaillons les étapes ci-dessous :

1. On initialise ces trois nombres à 0 pour chacune des  $N_C$  cellules.
2. On attribue à la cellule puits  $\underline{z}_s$  les nombres suivants :  $Gr(\underline{z}_s) = 1$ ,  $St(\underline{z}_s) = 0$  et  $P(\underline{z}_s) = 1$ .
3. On construit des séquences de cellules  $L_z = [\underline{z}_0, \underline{z}_1 = \mathcal{G}(\underline{z}), \dots, \underline{z}_n = \mathcal{G}^n(\underline{z})]$  jusqu'à tomber sur une séquence périodique déjà repérée ou refermer la séquence sur elle-même,

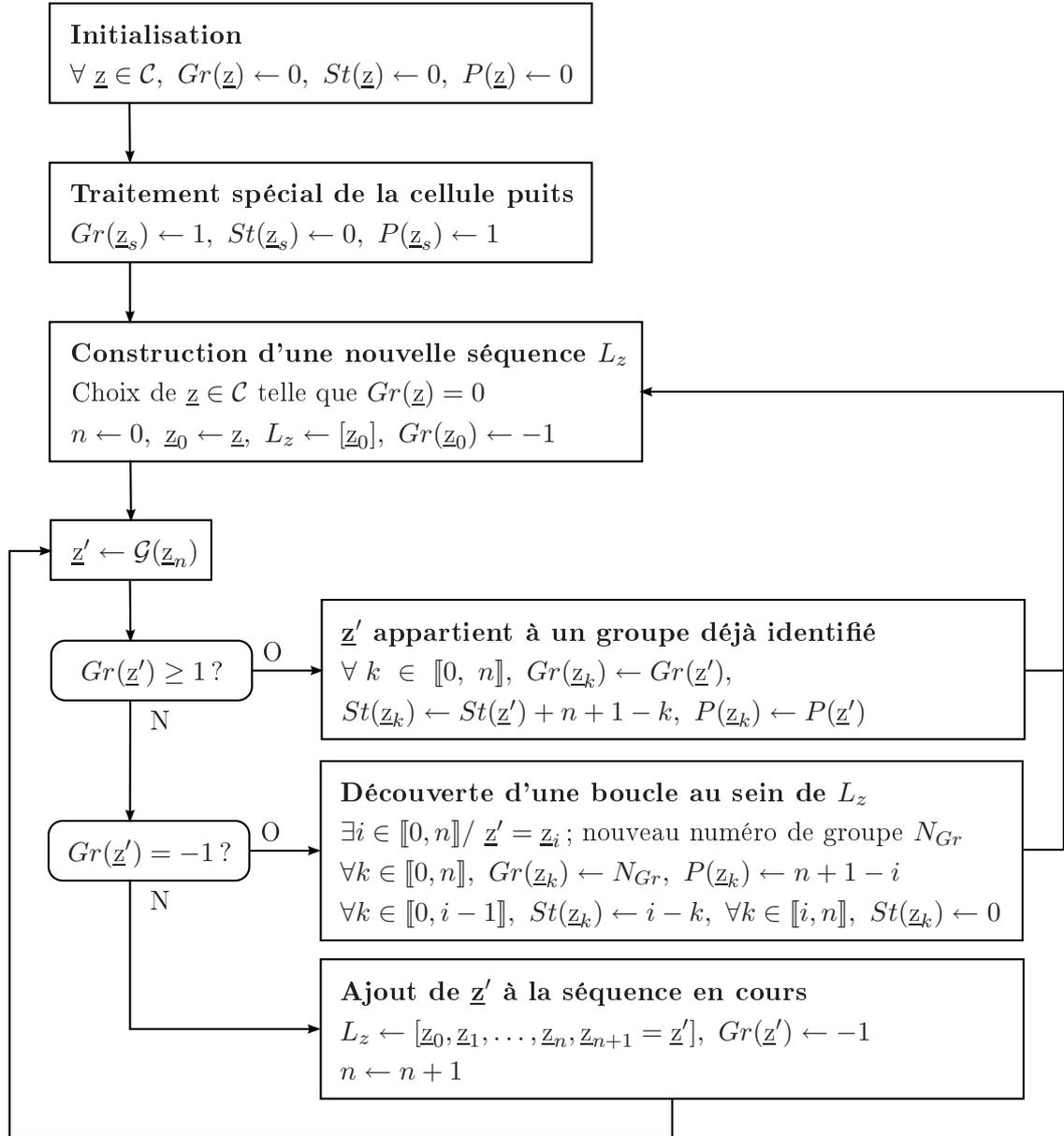


FIG. 3.3 – Algorithme d'analyse d'un cell-mapping

ce qui revient à découvrir une nouvelle séquence périodique, en partant d'une cellule  $\underline{z}_0$  dont le nombre  $Gr(\underline{z}_0)$  est nul (ce qui signifie qu'elle n'a pas encore été traitée). Un nombre  $Gr(\underline{z}_k) = -1$  est attribué temporairement à toute cellule de la séquence en cours de traitement.

Considérons donc la séquence  $L_z = [\underline{z}_0, \underline{z}_1 = \mathcal{G}(\underline{z}), \dots, \underline{z}_n = \mathcal{G}^n(\underline{z})]$ .

$\forall k \in \llbracket 0, n \rrbracket, Gr(\underline{z}_k) = -1, St(\underline{z}_k) = 0, P(\underline{z}_k) = 0$

Considérons les trois cas suivants pour l'image de  $\underline{z}_n, \underline{z}' = \mathcal{G}(\underline{z}_n)$  :

(a) Si  $Gr(\underline{z}') \geq 1, \underline{z}'$  appartient à un groupe déjà identifié : on répercute les informations sur les cellules de la séquence :

- $\forall k \in \llbracket 0, n \rrbracket, Gr(\underline{z}_k) = Gr(\underline{z}')$
- $\forall k \in \llbracket 0, n \rrbracket, St(\underline{z}_k) = St(\underline{z}') + n + 1 - k$
- $\forall k \in \llbracket 0, n \rrbracket, P(\underline{z}_k) = P(\underline{z}')$

On choisit une nouvelle cellule  $\underline{z}_0$  telle que  $Gr(\underline{z}_0) = 0$  pour démarrer une nouvelle séquence.

(b) Si  $Gr(\underline{z}') = -1$ , une nouvelle séquence périodique est découverte au sein de  $L_z$  :  $\exists i \in \llbracket 0, n \rrbracket$  tel que  $\underline{z}' = \underline{z}_i$ . Un nouveau numéro de groupe  $N_{Gr}$  est attribué à chacune des cellules.

- $\forall k \in \llbracket 0, n \rrbracket, Gr(\underline{z}_k) = N_{Gr}$
- $\forall k \in \llbracket 0, i-1 \rrbracket, St(\underline{z}_k) = i - k, \forall k \in \llbracket i, n \rrbracket, St(\underline{z}_k) = 0$
- $\forall k \in \llbracket 0, n \rrbracket, P(\underline{z}_k) = n + 1 - i$

On choisit une nouvelle cellule  $\underline{z}_0$  telle que  $Gr(\underline{z}_0) = 0$  pour démarrer une nouvelle séquence.

(c) Si  $Gr(\underline{z}') = 0$ , la séquence s'agrandit d'une nouvelle cellule :

- $Gr(\underline{z}') = -1$
- $L_z = [\underline{z}_0, \underline{z}_1, \dots, \underline{z}_n, \underline{z}_{n+1} = \underline{z}']$

On poursuit le raisonnement avec  $\underline{z}' = \mathcal{G}(\underline{z}_{n+1})$ .

Enfin, nous donnons l'exemple suivant : soit le cell-mapping trivial représenté en figure 3.4 ; les cellules ont été numérotées de 1 à 8, le numéro 9 étant attribué à la cellule puits et la cellule image de chaque cellule est désignée par une flèche. Cette figure résume en somme les informations que l'on possède après discrétisation de l'espace  $\mathcal{E}_{IC}$  puis construction du cell-mapping simple par intégration temporelle du centre de chaque cellule.

Ces informations sont reportées dans les trois premières colonnes du tableau 3.1. La première colonne reporte les coordonnées discrètes  $z_1$  et  $z_2$  selon les axes  $x$  et  $y$ , la seconde indique le numéro de la cellule correspondante et la troisième, le numéro de la cellule image. Les colonnes suivantes reflètent l'état des trois nombres  $Gr, St$  et  $P$  lors de la construction d'une séquence commençant par la cellule 1. On voit ainsi que l'on ajoute successivement les cellules 1 ( $n=0$ ), 2 ( $n=1$ ), 7 ( $n=2$ ) puis 6 ( $n=3$ ). On se trouve alors avec  $\underline{z}'$ , image de la cellule 6, valant la cellule 2 pour laquelle le nombre  $Gr$  vaut -1. Conformément à l'étape 3.b de l'algorithme, on identifie une nouvelle séquence et on cherche l'indice  $i$  tel que  $\underline{z}' = \underline{z}_i$  :  $i = 1$  pour notre exemple. On en conclut, en dernière colonne, que l'on a un cycle de période 3 constitué des cellules 2, 7 et 6 ayant la cellule 1 dans son bassin d'attraction.

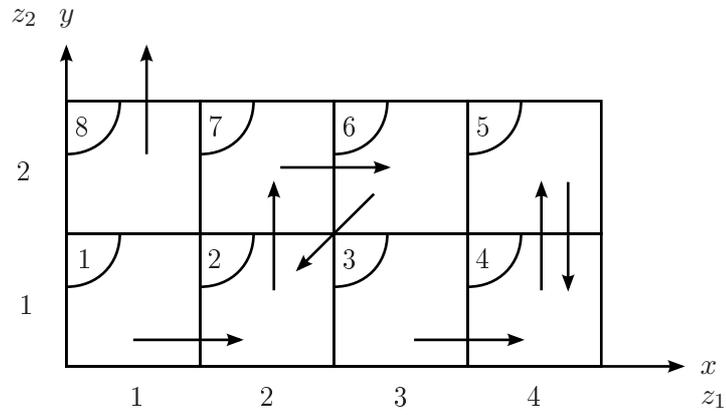


FIG. 3.4 – Cell-mapping - Illustration de la méthode d'analyse

				n=0			n=1			n=2			n=3			Fin		
$z_1$	$z_2$	Nb	Im	Gr	St	P												
1	1	1	2	-1	0	0	-1	0	0	-1	0	0	-1	0	0	2	1	3
2	1	2	7	0	0	0	-1	0	0	-1	0	0	-1	0	0	2	0	3
3	1	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	6	2	0	0	0	0	0	0	0	0	0	-1	0	0	2	0	3
2	2	7	6	0	0	0	0	0	0	-1	0	0	-1	0	0	2	0	3
1	2	8	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x	x	9	x	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1

TAB. 3.1 – Cell-mapping - Illustration de la méthode d'analyse

### 3.4 Interprétation d'un cell-mapping

Après l'étape d'analyse, reste à interpréter les résultats. En effet, selon la finesse de la discrétisation adoptée, les séquences périodiques ou cycles trouvés, reflètent plus ou moins fidèlement le comportement du système. En poussant à l'extrême, si l'on a fait le choix de prendre une seule cellule pour discrétiser  $\mathcal{E}_{IC}$ , deux groupes au plus peuvent être dégagés : le groupe 1, relatif à la cellule puits et le groupe 2 qui donne que l'image de la cellule est elle-même, ce qui ne veut rien dire de plus qu'il existe au moins une trajectoire revenant dans  $\mathcal{E}_{IC}$ . Ceci ne donne que peu d'informations relativement à la dynamique du système. De la même façon, si une solution périodique réelle existe dans l'espace continu, il se peut qu'elle soit traduite par plusieurs cycles du cell-mapping, ceci provenant du fait qu'une cellule sera toujours plus large qu'un point. On trouvera des exemples d'application de méthodes de cell-mapping et donc d'analyse, dans les articles [LSW97] et [DCL05].

Afin de parer à ces problèmes, il est intéressant d'adopter une technique de raffinement : partant d'une discrétisation grossière de l'espace  $\mathcal{E}_{IC}$ , on la raffine autour des zones où des cycles ont été détectés. Ainsi, on travaille avec un maillage lâche aux endroits de peu d'intérêt et fin sur les zones contenant des solutions remarquables. Une telle approche mêlant cell-mapping simple en première approximation et cell-mapping généralisé sur les zones de fort intérêt est proposée par [Hsu87] et dénommée « From large to small » ; [GK99] propose aussi une méthode de raffinement adaptatif dans le cas du cell-mapping généralisé.

### 3.5 Application à un oscillateur de Duffing

Nous proposons de tester chacune des méthodes d'analyse globale sur le système simple que constitue l'oscillateur de Duffing dont une étude est donnée en annexe A. Nous appliquons donc un cell-mapping simple pour ce système en étudiant la zone de condition initiales  $\mathcal{E}_{IC} = [-2, 2] \times [-3, 3]$ . Le tableau 3.2 résume les caractéristiques numériques des cell-mapping utilisés pour créer les figures 3.5 à 3.9. Les calculs ont été réalisés sur un serveur Intel Core 2 Duo E6600 (4Go de Ram). La ligne intitulée « Cycles » donne la période des cycles trouvés en notation abrégée : n-P signifie que le cycle est constitué de n cellules ; cette indication est suivie par le nombre de cycles de cette période.

Les figures 3.5 et 3.6 représentent les bassins d'attraction obtenus pour les valeurs respectives de  $\omega = 1.2 \text{ rad.s}^{-1}$  et  $\omega = 1.5 \text{ rad.s}^{-1}$ . Sur ces deux figures, on distingue trois solutions, marquées par des symboles + et leurs bassins d'attraction représentés par des symboles • de même couleur. Les solutions  $S_1$ , en bleu et  $S_3$ , en rouge sont stables ; elles possèdent des bassins d'aire non nulle. La solution  $S_2$ , en vert, en revanche, se situe sur la séparatrice entre les deux bassins de  $S_1$  et de  $S_3$  et ne possède pas son propre bassin : elle est instable.

Nous avons raffiné les zones encadrées en figure 3.5 ( $\omega = 1.2 \text{ rad.s}^{-1}$ ) qui contiennent les îlots de solution  $S_1$ ,  $S_2$  et  $S_3$ . Nous appliquons sur ces zones un nouveau cell-mapping simple afin d'affiner la localisation des solutions. On obtient alors respectivement les figures 3.7, 3.8 et 3.9. Nous avons aussi tracé sur ces trois dernières figures le résultat de l'intégration temporelle issue du centre de chaque cellule. Partant du centre matérialisé par un point bleu, nous menons un trait jusqu'au point obtenu à la fin de l'intégration temporelle (donc

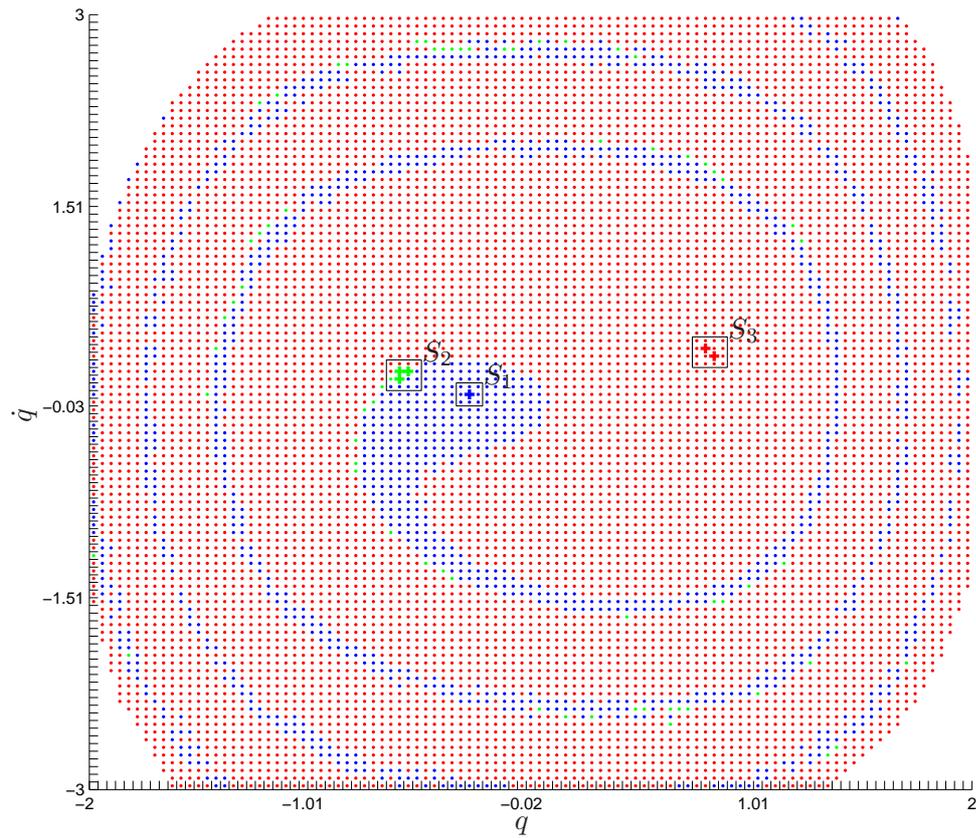
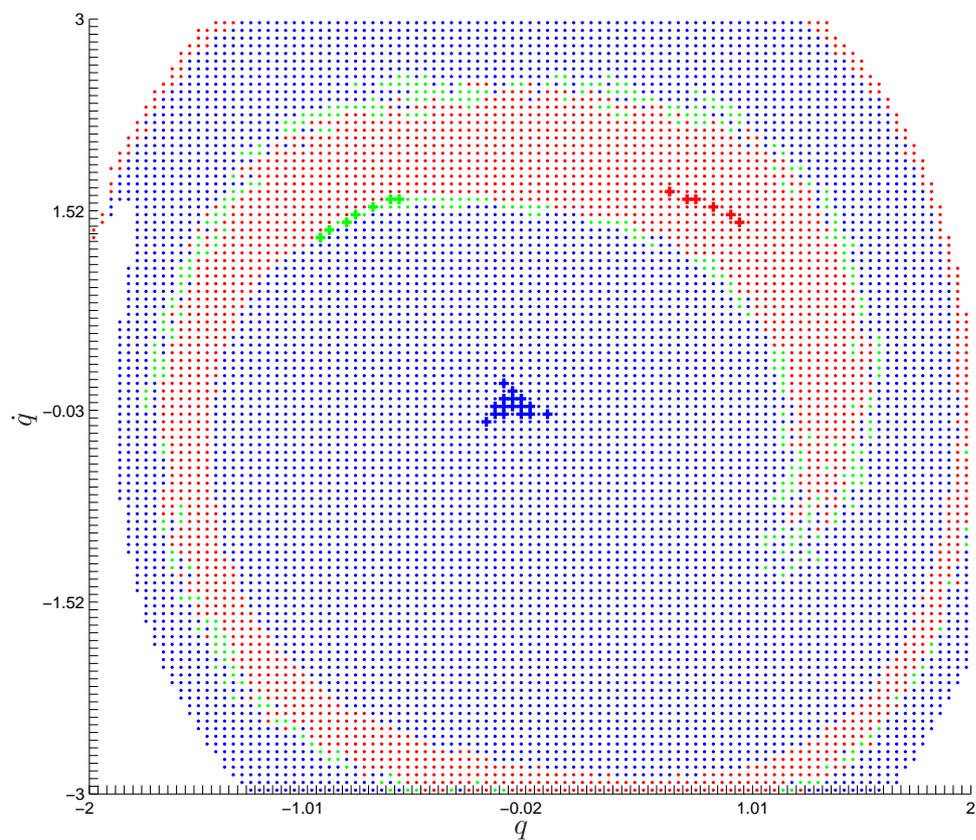
FIG. 3.5 – Cell-mapping - Oscillateur de Duffing : Bassin d'attraction pour  $\omega = 1.2 \text{ rad.s}^{-1}$ FIG. 3.6 – Cell-mapping - Oscillateur de Duffing : Bassin d'attraction pour  $\omega = 1.5 \text{ rad.s}^{-1}$

	Fig. 3.5	Fig. 3.6	Fig. 3.7	Fig. 3.8	Fig. 3.9
$\mathcal{E}_{IC}$	$[-2, 2]$ $[-3, 3]$	$[-2, 2]$ $[-3, 3]$	$[-0.34, -0.22]$ $[-0.03, 0.15]$	$[-0.65, -0.49]$ $[0.09, 0.33]$	$[0.73, 0.89]$ $[0.27, 0.51]$
Divisions ( $N_1 \times N_2$ )	$101 \times 101$	$101 \times 101$	$51 \times 51$	$51 \times 51$	$51 \times 51$
Temps cpu (s)	112	95	17	20	21
Cycles : période et nombre	1-P : 6	1-P : 14 3-P : 5	1-P : 2	1-P : 5	1-P : 1 8-P : 1

TAB. 3.2 – Cell-mapping - Oscillateur de Duffing :  
tableau récapitulatif des caractéristiques numériques

au bout d'un temps  $\tau = 2\pi/\omega$ ). On voit ainsi se dessiner la dynamique du système aux abords de ces trois solutions. Nous avons souligné le mouvement dominant au voisinage des solutions à l'aide d'un trait noir fléché. On peut lire ainsi que, conformément à la classification établie en § 1.2.5, les solutions  $S_1$  et  $S_3$  sont des foyers stables tandis que la solution  $S_2$  consiste en un point selle dont nous avons représenté les directions attractrices et répulsives. La stabilité ainsi lue des solutions se retrouve en figure 3.10 qui donne les portraits de phase obtenus pour des intégrations temporelles dont les conditions initiales correspondent aux zones identifiées comme étant des solutions : en bleu est représentée la trajectoire durant les 100 périodes d'excitation pour lesquelles l'intégration temporelle a été menée ; en vert, seule la première période d'excitation est représentée et en rouge, la dernière. On voit donc que pour les solutions stables  $S_1$  et  $S_3$  la trajectoire forme une ellipse inchangée au cours des 100 périodes d'intégration : les premiers et derniers cycles se superposent. En revanche, le cycle instable  $S_2$  représenté en vert est vite quitté par le système qui regagne ici le cycle  $S_3$ .

Nous venons de voir que les figures produites par un cell-mapping sont très séduisantes et que la quantité d'informations quant à la dynamique du système est importante : solutions, bassins d'attraction, stabilité. Cependant, cette étude a été menée sur un système très simple à un seul degré de liberté. Qu'en est-il lorsque la taille du système augmente ? En effet, si la quantité d'informations résultant d'un cell-mapping est importante c'est entre autres parce que la quantité d'informations traitée l'est aussi : cette méthode génère un grand nombre de cellules donc un grand nombre d'informations à stocker et demande une intégration temporelle par cellule, donc un temps de calcul que l'on peut supposer conséquent même si cela n'est pas perceptible sur notre exemple. Essayons donc d'extrapoler ces résultats pour des systèmes de taille plus importante. D'après les temps de calcul recensés pour nos cinq essais, on en déduit un temps moyen de calcul  $T_{moy} = 0.0085$  s par cellule. Pour un système dynamique à  $N$  ddl - donc un espace d'états de dimension  $2N$  - que l'on subdiviserait en  $N_d$  morceaux dans chaque direction ( $N_1 = \dots = N_{2N} = N_d$ ), on obtient les grandeurs clé suivantes :

$$\text{Nombre de cellules à traiter} : N'_c = N_d^{2N} \quad (3.1a)$$

$$\text{Temps total de calcul} : T_{tot} = T_{moy} N'_c \quad (3.1b)$$

$$\text{Espace de stockage} : S = s_z N'_c \quad (3.1c)$$

où  $s_z$  désigne l'espace dédié au stockage des informations pour une cellule.

En prenant  $s_z = 1$  octet, c'est à dire l'espace permettant juste de créer un vecteur dont

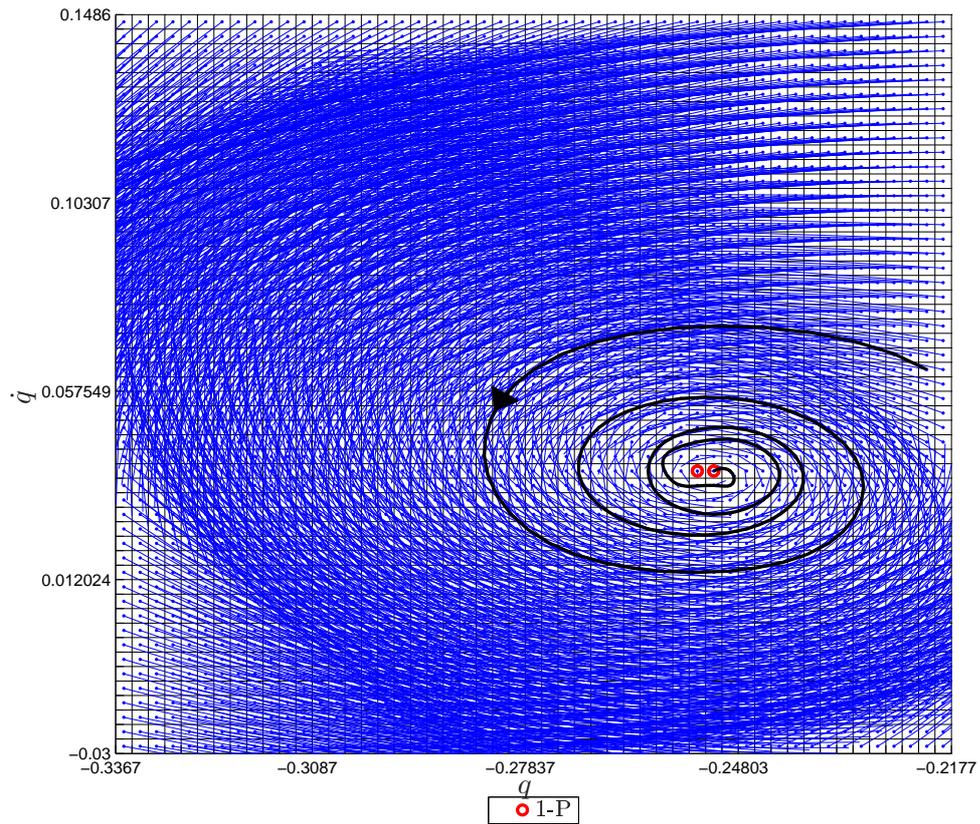


FIG. 3.7 – Cell-mapping - Oscillateur de Duffing : raffinement autour de la première solution

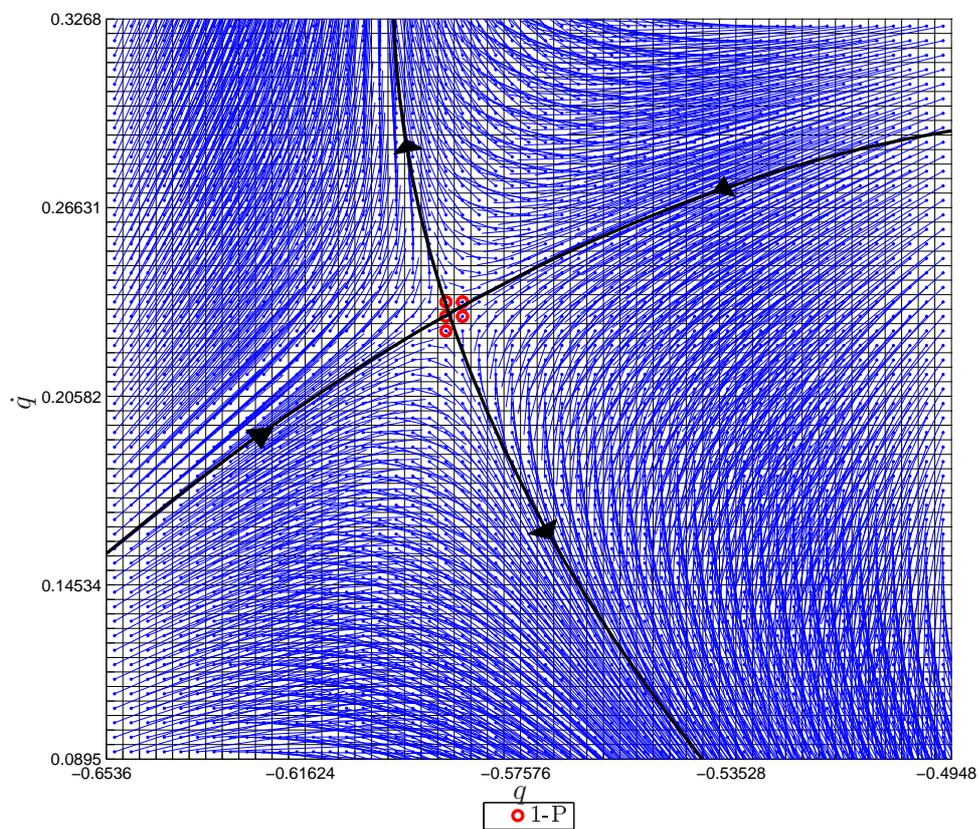


FIG. 3.8 – Cell-mapping - Oscillateur de Duffing : raffinement autour de la deuxième solution

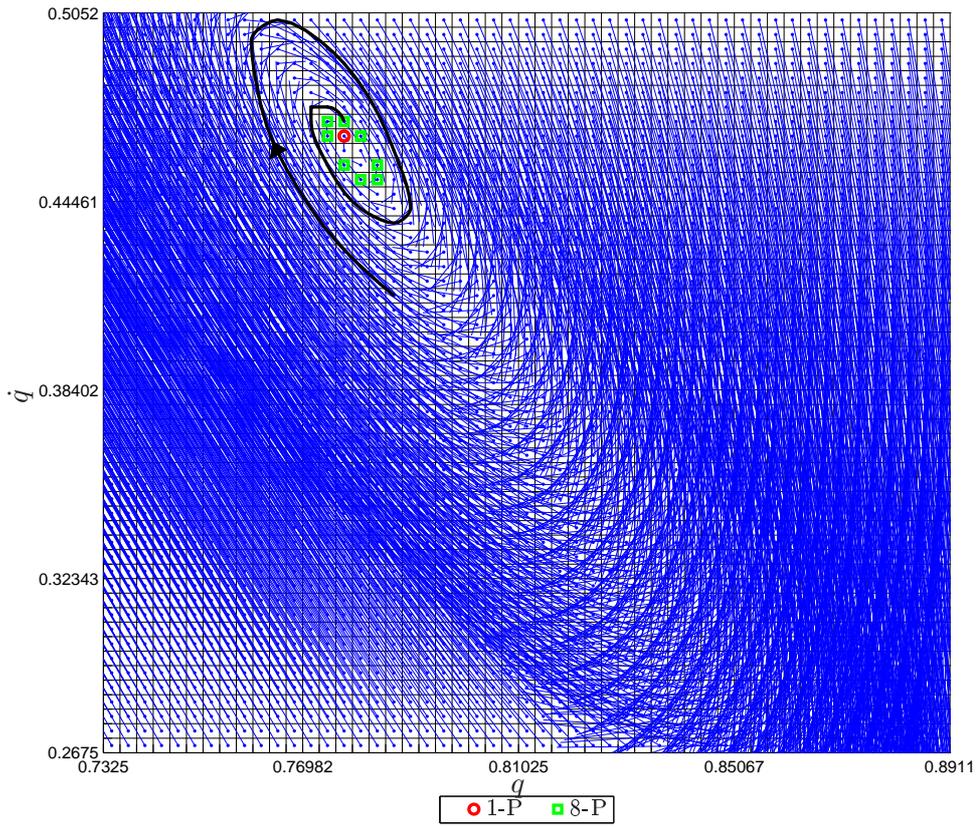


FIG. 3.9 – Cell-mapping - Oscillateur de Duffing : raffinement autour de la troisième solution

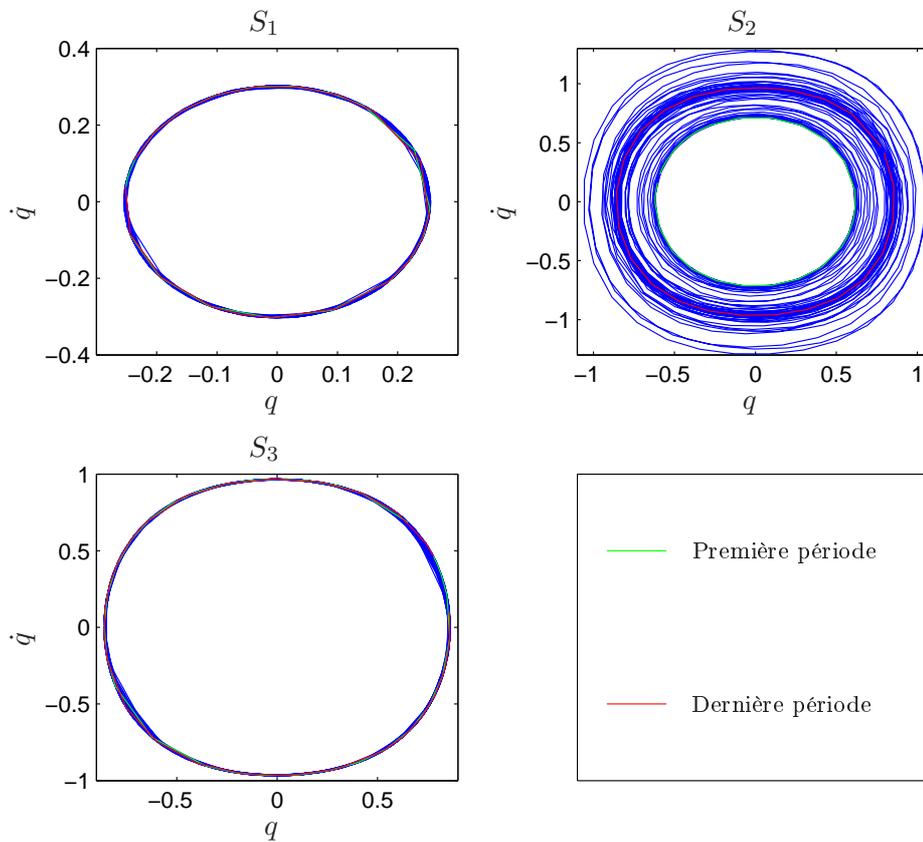


FIG. 3.10 – Cell-mapping - Oscillateur de Duffing : portraits de phase des trois solutions

l'élément  $i$  contient l'indice  $j$  de la cellule image lors d'un cell-mapping simple, et en appliquant ces formules pour des systèmes de 1 à 5 ddl et des valeurs de  $N_d$  variant entre 10 et 100, on obtient les diagrammes représentés en figure 3.11. On voit ainsi que pour un système de 4 ddl pour lequel on réaliserait 20 divisions dans chaque direction, il faudrait stocker au minimum 500 Go de données obtenues après 10 ans de calcul... quand bien même on diviserait par 10 le temps de calcul par cellule, l'utilisation d'une telle méthode n'est pas envisageable. Cette extrapolation nous amène donc à nous tourner vers d'autres méthodes, certainement moins riches dans leur résultats mais permettant de traiter des systèmes de taille plus importante.

### 3.6 Conclusion

Basé sur l'intégration temporelle de points de l'espace d'états, donc donnant une image globale de l'évolution du système, le cell-mapping s'avère être une méthode séduisante en termes de richesse des résultats : solutions de différentes natures, bassins d'attraction et stabilité. Cependant elle s'est révélée trop coûteuse numériquement pour pouvoir satisfaire à notre besoin. En effet, elle ne permet de traiter dans des délais raisonnables (quelques jours au maximum) que des systèmes à un ou deux degrés de liberté qui ne suffisent pas pour décrire des structures industrielles. Nous allons donc nous tourner dans les prochains chapitres vers des techniques donnant une vision moins complète de la dynamique du système - en faisant par exemple des hypothèses sur la forme des solutions recherchées donc en restreignant le champ d'investigation - mais plus rapides en termes de calcul.

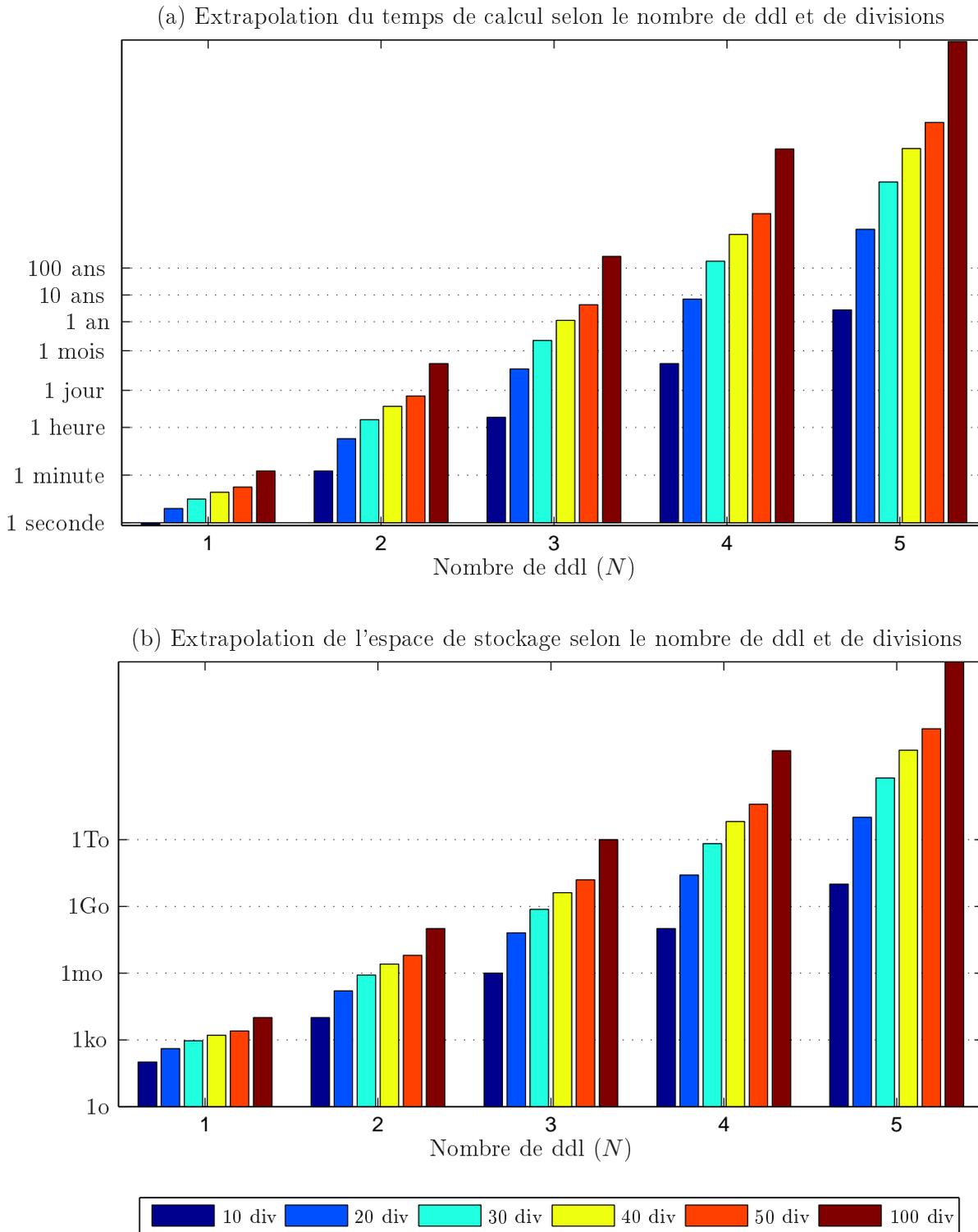


FIG. 3.11 – Cell-mapping - Extrapolation des temps de calcul et des espaces de stockage

# Chapitre 4

## Analyse globale par exclusion de cellules

---

Nous abordons avec ce chapitre l'utilisation de méthodes permettant la résolution de systèmes algébriques et non de systèmes dynamiques. Ces méthodes nécessitent donc la transformation de l'équation différentielle (1.1) en un système algébrique (2.10) voué à représenter un type de solution particulier (point fixe, solution périodique, etc). Il faut donc appliquer ces méthodes sur des systèmes de la forme (2.10) issus de transformations de (1.1) différentes, pour obtenir toutes les solutions du problème considéré. Leur avantage sur les méthodes de cell-mapping est en revanche leur capacité à traiter de plus gros systèmes.

Le principe des méthodes par exclusion de cellule est de traiter l'ensemble  $\Omega$  des valeurs possibles pour les variables  $\tilde{x}_i$  sous forme de cellules  $\tilde{X}$  réunies dans une partition  $\Gamma_k$  de l'espace inclus dans  $\Omega$  dont on va chercher à montrer qu'elle ne contiennent pas de solution. Si l'on y arrive, la cellule est éliminée, sinon, elle est subdivisée en sous-cellules. Le moyen utilisé pour montrer qu'aucune solution n'est contenue dans la cellule est un test numériquement peu coûteux. Ce test est nommé test d'exclusion ou cell discarded condition que nous noterons sous sa forme abrégée CDC dans la suite de ce chapitre. Ce processus est décrit par l'algorithme 4.1.

L'efficacité de ces méthodes dépend essentiellement de deux choses. Premièrement et majoritairement, elle dépend du test d'exclusion : s'il n'élimine pas facilement de larges zones vides de solutions, ces zones sont subdivisées en zones plus petites qui sont testées à leur tour. Si le test n'est capable d'éliminer que des cellules de très petite taille, il faudra l'exécuter sur un nombre de cellules très important ce qui ralentira l'exécution de l'algorithme. Plusieurs

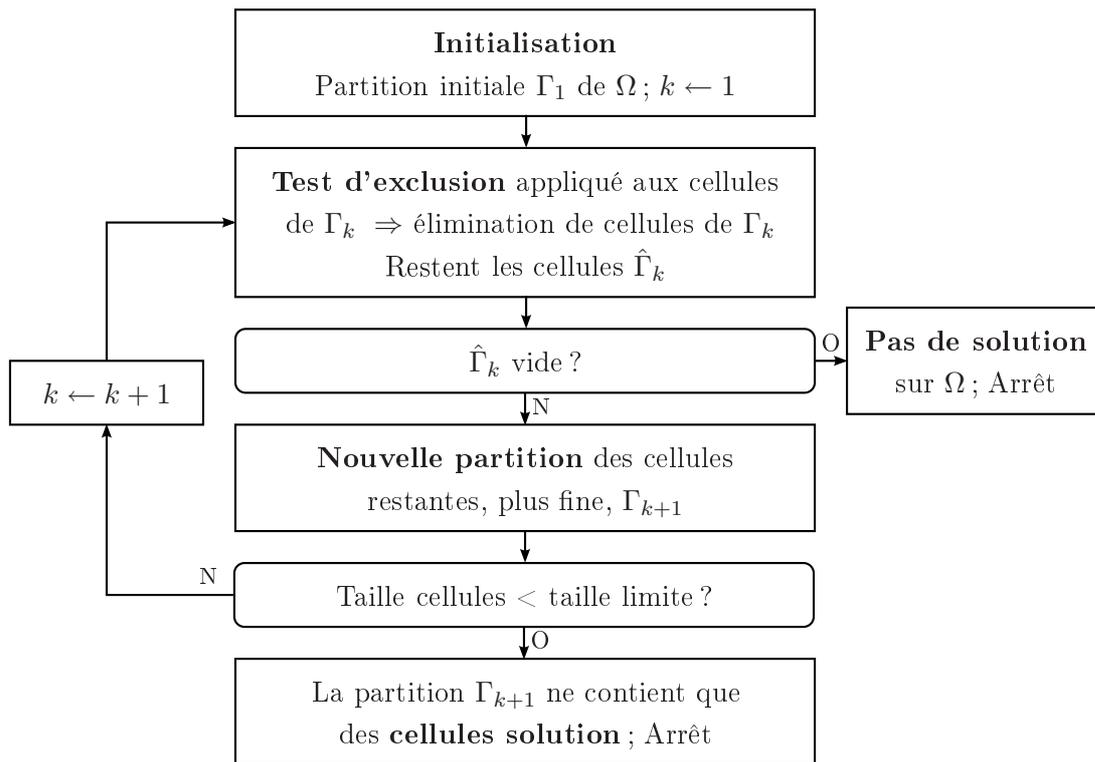


FIG. 4.1 – Méthodes d'exclusion - Algorithme général

types de tests sont proposés dans la littérature, adaptés à différentes situations. Nous donnons dans les sections suivantes les principaux tests d'exclusion et la littérature associée et nous conseillons dès à présent l'article [XZW96] qui constitue une bonne introduction à ces méthodes. Le second paramètre jouant sur l'efficacité de l'algorithme est la finesse de partition de la zone étudiée : si lorsque l'on passe de la partition  $\Gamma_k$  à une partition plus fine  $\Gamma_{k+1}$ , on subdivise chaque cellule en un trop grand nombre de sous-cellules, il faudra exécuter un très grand nombre de fois le test d'exclusion, ce qui sera coûteux ; si, à l'inverse, on subdivise trop peu  $\Gamma_k$ , le test d'exclusion est effectué un faible nombre de fois mais le domaine après exclusion n'est pas très différent du domaine précédent dans la mesure où les cellules testées sont peu différentes de celles du rang précédent.

Ce chapitre s'organise comme suit : nous détaillons tout d'abord les étapes du schéma 4.1, puis nous consacrons les sections 4.2 et 4.3 à des tests d'exclusion exploitant les éventuelles spécificités de la fonction  $H(\tilde{x})$  que l'on essaye d'annuler. Nous présentons ensuite en détail un test d'exclusion plus général basé sur la méthode dite « du simplexe ». Enfin, nous appliquons cette dernière méthode au même oscillateur de Duffing que celui du chapitre précédent.

On rappelle que l'on note  $P$  la dimension de  $\tilde{x}$  et on suppose, sauf mention contraire, que le système est carré (i.e. on possède autant d'équations que d'inconnues).

## 4.1 Etapes de l'algorithme général

Lorsque l'on veut mettre en œuvre l'algorithme décrit brièvement en introduction et représenté en figure 4.1, il faut commencer par créer une première partition  $\Gamma_1$  de l'espace  $\Omega$

qui représente le domaine de  $\mathbb{R}^P$  sur lequel les variables  $\tilde{x}_i$ ,  $1 \leq i \leq P$  peuvent prendre leurs valeurs. Les cellules  $\tilde{X}$  constituant une telle partition sont des pavés de  $\mathbb{R}^P$ . On peut donc choisir  $\Gamma_1 = \Omega$  et ne considérer qu'une cellule pour la première partition si la forme de  $\Omega$  s'y prête ou au contraire considérer plusieurs cellules, par exemple d'une taille maximale imposée si l'on possède des informations sur l'efficacité du test d'exclusion utilisé ensuite. Pour ce qui nous concerne, nous avons choisi la première possibilité.

L'étape sur laquelle on est ensuite susceptible de s'interroger est celle consacrée au raffinement de  $\hat{\Gamma}_k$  en  $\Gamma_{k+1}$ .  $\hat{\Gamma}_k$  contient uniquement des cellules que le test d'exclusion ne peut éliminer ; elles sont donc trop vastes, il faut en réduire la taille. Il est rare de trouver dans la littérature une autre méthode que celle qui vient assez naturellement de former une partition de l'espace restant en divisant les cellules en deux selon chacune des  $P$  directions. Il semble même que cette méthode soit généralement optimale selon [XZW96]. Notons quand même la proposition faite dans [YT02] d'exploiter les calculs réalisés pendant l'application du test d'exclusion dans le cas de la méthode du simplexe.

Examinons maintenant le critère important de la taille limite de la cellule en-deçà de laquelle on estime que si le test ne l'a pas exclue, c'est qu'elle contient une solution. Idéalement, le critère serait que la cellule soit réduite à un point. Ceci étant impossible à obtenir par dichotomie, on choisit que la cellule soit suffisamment petite, c'est-à-dire que chacune des bornes inférieures et supérieures de ses composantes soit suffisamment proche. Ceci peut se quantifier de façon relative - suffisamment proche par rapport à la valeur centrale - ou absolue. La première façon est séduisante sauf dans le cas où la cellule possède une de ses composantes centrée sur une valeur proche de zéro (le critère relatif imposant alors de diviser par zéro). Nous choisissons donc le critère suivant :

$$\text{Cellule assez petite} \Leftrightarrow \forall i \in \llbracket 1, P \rrbracket, \begin{cases} \text{Si } 0 \in \tilde{X}_i, w(\tilde{X}_i) < \epsilon_{Xa} \\ \text{Si } 0 \notin \tilde{X}_i, w(\tilde{X}_i) < \epsilon_{Xr} \min(|\underline{\tilde{X}}_i|, |\overline{\tilde{X}}_i|) \end{cases} \quad (4.1)$$

en notant  $\overline{\tilde{X}}_i$  et  $\underline{\tilde{X}}_i$  respectivement les bornes inférieure et supérieure et  $w(\tilde{X}_i) = \overline{\tilde{X}}_i - \underline{\tilde{X}}_i$  la largeur de la cellule selon la composante  $i$ .

Un choix qui semble raisonnable des valeurs  $\epsilon_{Xa}$  et  $\epsilon_{Xr}$  est  $10^{-6}$  pour le premier et  $10^{-3}$  pour le second.

Enfin, les deux étapes permettant l'arrêt de l'algorithme correspondent au cas où toutes les cellules de la partition  $\Gamma_k$  ont été éliminées par le test : il n'existe donc aucune solution sur le domaine  $\Omega$  étudié. L'autre cas, plus attendu, correspond à la configuration où  $\tilde{\Gamma}_k$  ne contenait plus que quelques cellules qui après subdivision sont d'une taille jugée raisonnable pour pouvoir être considérées comme représentantes de solutions.

## 4.2 Tests d'exclusion simples

Nous reportons dans cette section deux tests d'exclusion simples dans leur formulation, proposés par [XZW96] et qui s'appuient sur des propriétés fortes de la fonction  $H$  étudiée. On rappelle que l'objectif d'un test d'exclusion appliqué à une cellule  $\tilde{X}$  est de garantir la non-existence d'une solution sur cet espace. On utilise donc généralement la contraposée d'une

proposition de la forme :

$$\text{Il existe une solution } \tilde{x}^* \text{ sur } \tilde{X} \Rightarrow \text{Conditions à vérifier} \quad (4.2)$$

En testant les conditions et en trouvant qu'elles ne sont pas vérifiées, on peut affirmer qu'il n'existe aucune solution sur  $\tilde{X}$ . Nous formaliserons un test d'exclusion conformément à (4.2). Enfin, nous définissons les notations suivantes :

- Centre d'une cellule  $\tilde{X}$  :  $\text{mid}(\tilde{X})$ ,  $\forall i \in \llbracket 1, P \rrbracket$ ,  $\text{mid}(\tilde{X}_i) = \frac{\tilde{X}_i + \overline{\tilde{X}_i}}{2}$
- Largeur d'une cellule  $\tilde{X}$  :  $w(\tilde{X}_i) = \overline{\tilde{X}_i} - \underline{\tilde{X}_i}$  et  $w(\tilde{X}) = \max_{1 \leq i \leq P} (w(\tilde{X}_i))$

#### 4.2.1 Fonction lipschitzienne

Si  $H$  est lipschitzienne sur l'ouvert  $\Omega \subset \mathbb{R}^P$ , on a par définition :

$$\forall i \in \llbracket 1, P \rrbracket, \exists \alpha_i \in \mathbb{R}^+ \text{ tel que } \forall (\tilde{x}, \tilde{y}) \in \Omega^2, \|H_i(\tilde{x}) - H_i(\tilde{y})\| \leq \alpha_i \|\tilde{x} - \tilde{y}\|$$

Ainsi, s'il existe un zéro  $\tilde{x}^*$  sur la cellule  $\tilde{X} \subset \Omega$ , on peut écrire que l'on a nécessairement :

$$\forall i \in \llbracket 1, P \rrbracket, \|H_i(\tilde{y})\| \leq \alpha_i \|\tilde{x}^* - \tilde{y}\|$$

En particulierisant  $\tilde{y}$  comme étant le centre  $\text{mid}(\tilde{X})$  de la cellule étudiée et en notant  $\tilde{r}^t = \{w(\tilde{X}_1), \dots, w(\tilde{X}_P)\}$  le vecteur désignant les dimensions de la cellule selon ses  $P$  coordonnées, on aboutit au CDC suivant :

$$\exists \tilde{x}^* \in \tilde{X} / H(\tilde{x}^*) = 0 \Rightarrow \forall i \in \llbracket 1, P \rrbracket, \|H_i(\text{mid}(\tilde{X}))\| \leq \alpha_i \tilde{r}_i / 2 \quad (4.3)$$

La difficulté pour l'utilisation de ce test d'exclusion est de déterminer les plus petits  $\alpha_i$ , c'est-à-dire les constantes de Lipschitz de chacune des composantes de  $H$  qui permettent de réaliser un test efficace.

#### 4.2.2 Fonction monotone décomposable

En définissant la relation d'ordre suivante sur  $\mathbb{R}^P$  :

$$\tilde{x} \leq \tilde{y} \Leftrightarrow \forall i \in \llbracket 1, P \rrbracket, \tilde{x}_i \leq \tilde{y}_i,$$

on dit qu'une fonction  $C$  de  $\mathbb{R}^P$  dans  $\mathbb{R}^P$  est *isotone* si elle conserve l'ordre :

$$C \text{ isotone} \Leftrightarrow \forall (\tilde{x}, \tilde{y}) \in (\mathbb{R}^P)^2, \tilde{x} \leq \tilde{y} \Rightarrow C(\tilde{x}) \leq C(\tilde{y})$$

On dit alors que  $H$  est *monotonement décomposable* si et seulement s'il existe deux fonctions  $C$  et  $D$  isotones telles que  $H = C - D$ .

Si  $H$  est monotone décomposable, et possède un zéro  $\tilde{x}^*$  sur  $\tilde{X}$ , on peut alors utiliser la relation  $\underline{\tilde{X}} \leq \tilde{x}^* \leq \overline{\tilde{X}}$  pour écrire :

$$C(\underline{\tilde{X}}) \leq C(\tilde{x}^*) \leq C(\overline{\tilde{X}}) \text{ et } D(\underline{\tilde{X}}) \leq D(\tilde{x}^*) \leq D(\overline{\tilde{X}})$$

et par suite :

$$C(\underline{\tilde{X}}) - D(\overline{\tilde{X}}) \leq H(\tilde{x}^*) = 0 \leq C(\overline{\tilde{X}}) - D(\underline{\tilde{X}})$$

On obtient donc le CDC :

$$\exists \tilde{x}^* \in \tilde{X} / H(\tilde{x}^*) = 0 \Rightarrow \begin{cases} C(\underline{\tilde{X}}) \leq D(\overline{\tilde{X}}) \\ D(\underline{\tilde{X}}) \leq C(\overline{\tilde{X}}) \end{cases} \quad (4.4)$$

Il faut donc pour pouvoir utiliser ce test d'exclusion avoir déterminé - si cela est possible - les deux fonctions  $C$  et  $D$  qui décomposent  $H$ . On a donc un travail préalable d'existence et de construction des fonctions à effectuer, ce qui ne correspond pas à notre objectif de pouvoir traiter automatiquement une grande variété de non-linéarités.

### 4.3 Tests basés sur la domination de fonction

Nous abordons dans cette section des tests reposant sur le principe de la domination de fonction. L'objectif sera donc d'une façon générale de trouver des fonctions  $g_i$  pertinentes qui majorent la valeur absolue des fonctions  $H_i$ . Le théorème 4.1 donné plus loin permet alors de construire différents tests.

#### 4.3.1 Fonctions développables en série entière

Nous reportons dans cette section le test d'exclusion construit et justifié par [XZL97]. Donnons d'abord quelques définitions et notations :

1. Conventions d'écriture :

- $\forall \alpha \in \mathbb{N}^P, |\alpha| = \sum_{k=1}^P \alpha_k$

- $\forall (\tilde{x}, \alpha) \in (\mathbb{R}^P \times \mathbb{N}^P), \tilde{x}^\alpha = \prod_{k=1}^P \tilde{x}_k^{\alpha_k}$

- $\forall \tilde{x} \in \mathbb{R}^P, |\tilde{x}| = \{|\tilde{x}_1|, \dots, |\tilde{x}_P|\}^t$

2. On dira qu'une fonction  $h : \mathbb{R}^P \rightarrow \mathbb{R}$  est *développable en série entière* si et seulement si elle peut s'écrire pour tout  $\tilde{x} \in \mathbb{R}^P$  :

$$h(\tilde{x}) = \sum_{|\alpha|=0}^{\infty} c_\alpha \tilde{x}^\alpha, \alpha \in \mathbb{N}^P, c_\alpha \in \mathbb{R}$$

3. On en définit alors la valeur absolue  $|h|$  :

$$\forall \tilde{x} \in \mathbb{R}^P, |h|(\tilde{x}) = \sum_{|\alpha|=0}^{\infty} |c_\alpha| \tilde{x}^\alpha$$

4. Relation d'ordre partielle pour  $h = \sum c_\alpha \tilde{x}^\alpha$  et  $g = \sum d_\alpha \tilde{x}^\alpha$  :

$$h < g \Leftrightarrow \forall \alpha \in \mathbb{N}^P, |c_\alpha| < |d_\alpha|$$

5. On note  $\mathcal{E}_h$  l'ensemble des fonctions de  $\mathbb{R}^P$  dans  $\mathbb{R}$  pour lesquelles il existe  $k \in \mathbb{N}$  fonctions décomposables en séries entières  $h^{(k)}$  telles que

$$\forall \tilde{x} \in \mathbb{R}^P, h(\tilde{x}) = h^{(1)}(\tilde{x}) + \sum_{j=2}^k |h^{(j)}(\tilde{x})|$$

6. On note enfin  $\mathcal{E}_H = \{H : \mathbb{R}^P \rightarrow \mathbb{R}^P / \forall i \in \llbracket 1, P \rrbracket, H_i \in \mathcal{E}_h\}$

Une fois ces définitions établies, on peut énoncer le théorème suivant (démontré dans [XZL97]) :

### **Théorème 4.1 : Domination par une fonction décomposable en série entière**

Soit  $h \in \mathcal{E}_h$  et  $g$  une fonction décomposable en série entière telle que  $|h| < g$ .

Alors  $\forall (\tilde{x}, \tilde{y}) \in (\mathbb{R}^P)^2, |h(\tilde{x}) - h(\tilde{y})| \leq |g|(|\tilde{y}| + |\tilde{x} - \tilde{y}|) - |g|(|\tilde{y}|)$  □

Ainsi, s'il existe un zéro  $\tilde{x}^*$  de  $h$  sur  $\tilde{X}$  et que l'on choisit  $\tilde{y} = \text{mid}(\tilde{X})$ , on peut écrire :

$$|h(\text{mid}(\tilde{X}))| \leq |g|(\text{mid}(\tilde{X}) + \tilde{r}) - |g|(\text{mid}(\tilde{X}))$$

On obtient alors aisément le test d'exclusion suivant sur  $H \in \mathcal{E}_H$  telle qu'il existe  $G : \mathbb{R}^P \rightarrow \mathbb{R}^P$  développable en série entière et dominant  $H$  :

$$\exists \tilde{x}^* \in \tilde{X} / H(\tilde{x}^*) = 0 \Rightarrow |H(\text{mid}(\tilde{X}))| \leq |G|(|\text{mid}(\tilde{X})| + \tilde{r}) - |G|(|\text{mid}(\tilde{X})|) \quad (4.5)$$

On voit que les conditions de mise en œuvre d'un tel test sont fortes; elles nécessitent une forme particulière pour  $H$  et la construction d'une fonction  $G$  qui domine  $H$  « au plus proche ». Il faut en effet tenter d'obtenir une majoration qui ne soit pas trop grossière de peur d'obtenir un test totalement inefficace.

### **4.3.2 Développement de Taylor**

Reprenant les travaux de [XZW96] et [XZL97], Kurt Georg propose dans [Geo01] et [Geo03] plusieurs tests d'exclusion basés sur le développement de Taylor de la fonction  $H$ . Comme précédemment, nous donnons quelques définitions et notations avant d'énoncer les résultats principaux.

1. On appelle  $h : \Omega \subset \mathbb{R}^P \rightarrow \mathbb{R}$  une fonction *absolument continue* si et seulement si  $\forall \epsilon > 0, \exists \eta > 0, / \forall (\tilde{x}, \tilde{y}) \in (\Omega)^2, \|\tilde{x} - \tilde{y}\| < \eta \Rightarrow |h(\tilde{x}) - h(\tilde{y})| < \epsilon$
2. On note :  $\forall \alpha \in \mathbb{N}^P, \partial^\alpha h = \partial^{|\alpha|} h / \partial x_1^{\alpha_1} \dots \partial x_P^{\alpha_P}$
3. On définit aussi les deux ensembles suivant :
  - $\mathcal{A}_k(\tilde{X}) = \left\{ h : \tilde{X} \rightarrow \mathbb{R} / \forall \alpha \in \mathbb{N}^P, |\alpha| < k, \partial^\alpha h \text{ est absolument continue} \right\}$
  - $\mathcal{K}_k(\tilde{X}) = \left\{ g \in \mathcal{A}_k(\tilde{X}) / \forall |\alpha| < k, \forall (\tilde{x}, \tilde{y}) \in (\tilde{X})^2, \right.$   
 $\left. 0 \leq \tilde{x} \leq \tilde{y} \Rightarrow 0 \leq \partial^\alpha g(\tilde{x}) \leq \partial^\alpha g(\tilde{y}) \right\}$
4. Pour  $h \in \mathcal{A}_k(\tilde{X})$  et  $g \in \mathcal{K}_k(\tilde{X})$ , on dira que  $g$  *domine*  $h$  à l'ordre  $k$  sur  $\tilde{X}$  si et seulement si

$$\forall |\alpha| < k, \forall \tilde{x} \in \tilde{X}, |\partial^\alpha h(\tilde{x})| \leq \partial^\alpha g(|\tilde{x}|)$$

On notera cette relation  $h(\tilde{x}) \prec_k g(\tilde{x}), \tilde{x} \in \tilde{X}$  ou plus simplement  $h \prec_k g$  s'il n'y a pas d'ambiguïté.

5. On généralise cette définition de dominance à  $H : \Omega \rightarrow \mathbb{R}^P$  en posant que si  $\forall i \in \llbracket 1, P \rrbracket$ ,  $H_i \in \mathcal{A}_k(\tilde{X})$ , il existe  $G_i \in \mathcal{K}_k(\tilde{X})$ , telle que  $H_i \prec_k G_i$ , alors  $G = \{G_1, \dots, G_P\}^t$  domine  $H$  à l'ordre  $k$  sur  $\tilde{X} : H \prec_k G$ .

Fort de ces définitions préalables, on peut alors énoncer les différents tests d'exclusion suivants :

- 1 - Soit  $\tilde{X}$  une cellule de  $\mathbb{R}^P$ ,  $k > 0$  un entier et  $G$  tels que  $|H(\tilde{x})| \prec_k G(\tilde{x} - \text{mid}(\tilde{X}))|$  pour  $\tilde{x} \in \tilde{X}$ . Alors (4.6) est un test d'exclusion pour  $H$  sur  $\tilde{X}$ .

$$\begin{aligned} \exists \tilde{x}^* \in \tilde{X} / H(\tilde{x}^*) = 0 \Rightarrow \\ |H(\text{mid}(\tilde{X}))| \leq G(\tilde{r}/2) - G(0) - \underbrace{\sum_{0 < |\alpha| < k} \left( \partial^\alpha G(0) - |\partial^\alpha H(\text{mid}(\tilde{X}))| \right)}_{\geq 0} (\tilde{r}/2)^\alpha \end{aligned} \quad (4.6)$$

- 2 - Soit  $\tilde{X}$  une cellule de  $\mathbb{R}^P$ ,  $k > 0$  un entier et  $G$  tels que  $|H(\tilde{x})| \prec_k G(\tilde{x})$  pour  $\tilde{x} \in \tilde{X}$ . Alors (4.7) est un test d'exclusion pour  $H$  sur  $\tilde{X}$ .

$$\begin{aligned} \exists \tilde{x}^* \in \tilde{X} / H(\tilde{x}^*) = 0 \Rightarrow |H(\text{mid}(\tilde{X}))| \leq G(|\text{mid}(\tilde{X})| + \tilde{r}/2) - G(|\text{mid}(\tilde{X})|) \\ - \underbrace{\sum_{0 < |\alpha| < k} \left( \partial^\alpha G(\text{mid}(\tilde{X})) - |\partial^\alpha H(\text{mid}(\tilde{X}))| \right)}_{\geq 0} (\tilde{r}/2)^\alpha \end{aligned} \quad (4.7)$$

- 3 - *Constantes de Lipschitz pour  $H$*  : si  $H \in \mathcal{A}_1(\tilde{X})$ , considérant les constantes de Lipschitz  $C_\alpha$  :

$$\forall |\alpha| = 1, C_\alpha \geq \sup_{\tilde{y} \in \tilde{X}} |\partial^\alpha H(\tilde{y})|$$

Alors (4.8) est un test d'exclusion pour  $H$  sur  $\tilde{X}$ .

$$\exists \tilde{x}^* \in \tilde{X} / H(\tilde{x}^*) = 0 \Rightarrow |H(\text{mid}(\tilde{X}))| \leq \sum_{|\alpha|=1} C_\alpha (\tilde{r}/2)^\alpha \quad (4.8)$$

- 4 - *Constantes de Lipschitz pour  $H'$*  : si  $H \in \mathcal{A}_2(\tilde{X})$ , considérant les constantes de Lipschitz  $C_\beta$  :

$$\forall |\beta| = 2, C_\beta \geq \sup_{\tilde{y} \in \tilde{X}} \left| \partial^\beta H(\tilde{y}) \right|$$

Alors (4.9) est un test d'exclusion pour  $H$  sur  $\tilde{X}$ .

$$\exists \tilde{x}^* \in \tilde{X} / H(\tilde{x}^*) = 0 \Rightarrow |H(\text{mid}(\tilde{X}))| \leq \sum_{|\alpha|=1} |\partial^\alpha H(\text{mid}(\tilde{X}))| (\tilde{r}/2)^\alpha + \sum_{|\beta|=2} C_\beta (\tilde{r}/2)^\beta \quad (4.9)$$

Ces différents développements impliquent une étude préalable de la fonction pour pouvoir construire une fonction  $G$  appropriée. Encore une fois, étant donnée notre volonté de traiter des non-linéarités variées dans leurs expressions, nous laissons de côté la mise en œuvre de ces méthodes pour nous tourner vers une méthode dont le champ d'application est plus vaste et que nous appellerons « méthode du simplexe ». Nous lui consacrons la section suivante.

## 4.4 Méthode du simplexe

La méthode du simplexe est une méthode d'optimisation sous contraintes linéaires. Elle est expliquée en détail dans [PTVF92, Chap. 10.8]. Partant d'un système non-linéaire de la forme (4.10), que l'on cherche à résoudre sur  $\tilde{X}$ , intervalle borné de  $\mathbb{R}^P$ , on traduit les non-linéarités du système en inégalités (4.11) :

$$\bar{H}.\tilde{x} + \hat{H}(\tilde{x}) - H_c = 0, \tilde{x} \in \tilde{X} \quad (4.10)$$

$$\bar{H}.\tilde{x} + \tilde{y} = 0, \tilde{x} \in \tilde{X}, \underline{\hat{H}(\tilde{X}) - H_c} \leq \tilde{y} \leq \overline{\hat{H}(\tilde{X}) - H_c} \quad (4.11)$$

On n'a donc pas deux systèmes équivalents mais l'implication (4.10)  $\Rightarrow$  (4.11) ; (4.11) correspond bien à un test d'exclusion pour (4.10), comme l'ont démontré et illustré [YKT98, YT02].

Dans une première section, nous exposons la méthode du simplexe, puis, nous la relierons à notre problème initial (4.10).

### 4.4.1 Problème traité par une méthode de simplexe : forme et résolution

#### Forme générale

Une méthode de simplexe est capable de traiter les problèmes d'optimisation linéaire à contraintes linéaires. Ils peuvent se mettre d'une façon générale sous la forme suivante :

$$\text{Maximiser : } z = f^t.\tilde{x}, f^t = \{f_1, \dots, f_P\} \in \mathbb{R}^P \quad (4.12a)$$

$$\text{Sous les contraintes primaires : } \tilde{x} \geq 0 \quad (4.12b)$$

$$\text{Sous les contraintes supplémentaires : } C_1.\tilde{x} \leq b_1, C_1 \in \mathcal{M}_{m_1, P}, b_1 \geq 0 \quad (4.12c)$$

$$C_2.\tilde{x} \geq b_2, C_2 \in \mathcal{M}_{m_2, P}, b_2 \geq 0 \quad (4.12d)$$

$$C_3.\tilde{x} = b_3, C_3 \in \mathcal{M}_{m_3, P}, b_3 \geq 0 \quad (4.12e)$$

On entend bien sûr par  $\tilde{x} \leq \tilde{y}$  la relation  $\forall i \in \llbracket 1, P \rrbracket, \tilde{x}_i \leq \tilde{y}_i$  et on note  $\mathcal{M}_{m,n}$  l'ensemble des matrices réelles à  $m$  lignes et  $n$  colonnes.

$z$  est appelée *fonction objectif* (objective function).

#### Forme normale

Le problème ne peut être traité directement sous la forme précédente : il faut ramener les inégalités (4.12c-4.12d) à des égalités. Pour ceci, on augmente le système d'inconnues en ajoutant  $m_1 + m_2$  variables d'écart (slack variables),  $\lambda^{(1)} \in \mathbb{R}^{m_1}$  et  $\lambda^{(2)} \in \mathbb{R}^{m_2}$  définies par :

$$\lambda^{(1)} = b_1 - C_1.\tilde{x} \quad \text{et} \quad \lambda^{(2)} = C_2.\tilde{x} - b_2$$

En adoptant les notations suivantes, on peut exprimer le problème sous une forme normale (4.13) équivalente à (4.12) :

$$\hat{x} = \begin{Bmatrix} \tilde{x} \\ \lambda^{(1)} \\ \lambda^{(2)} \end{Bmatrix}, \hat{f} = \begin{Bmatrix} f \\ 0_{m_1,1} \\ 0_{m_2,1} \end{Bmatrix}, \hat{b} = \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix}, \hat{C} = \begin{bmatrix} C_1 & I_{m_1} & \\ C_2 & & -I_{m_2} \\ C_3 & & \end{bmatrix}$$

$$\text{Maximiser : } \quad \hat{z} = \hat{f}^t \cdot \hat{x} \quad (4.13a)$$

$$\text{Sous les contraintes primaires : } \quad \hat{x} \geq 0 \quad (4.13b)$$

$$\text{Sous les contraintes supplémentaires : } \quad \hat{C} \cdot \hat{x} = \hat{b}, \hat{b} \geq 0 \quad (4.13c)$$

On note  $\hat{P} = P + m_1 + m_2$  la dimension de  $\hat{x}$ .

### Forme normale restreinte

La résolution du système sous forme normale nécessite une dernière manipulation : il faut que l'on ait identifié des variables de gauche et des variables de droite. Ceci se fait en retravaillant l'expression des contraintes additionnelles (4.13c). Il faut en effet obtenir une écriture de ces contraintes telle que chaque égalité possède dans son membre de gauche au moins une variable affectée d'un coefficient strictement positif et qui n'apparaisse que dans cette égalité. On appelle alors ces variables *variables de gauche* (left-hand variables) et les autres, *variables de droite* (right-hand variables). Le problème est alors dit sous sa *forme normale restreinte* (restricted normal form).

En admettant que l'on ait réussi à mettre les contraintes sous cette forme, adoptons les notations suivantes :

$$\hat{C}' \cdot \hat{x} = \hat{b}', \hat{b}' \geq 0$$

$N_l$  : ensemble des indices  $i$  tels que  $\hat{x}_i$  soit une variable de gauche

$N_r$  : ensemble des indices  $i$  tels que  $\hat{x}_i$  soit une variable de droite

$$N_l \cup N_r = \{1, \dots, P + m_1 + m_2 + m_3\}, N_l \cap N_r = \emptyset$$

La distinction entre variables de gauche et variables de droite peut alors s'écrire ainsi :

$$\begin{aligned} \forall i \in \llbracket 1, m_1 + m_2 + m_3 \rrbracket, \exists j \in \llbracket 1, \hat{P} \rrbracket \text{ tel que} \\ \hat{C}'_{ij} > 0 \text{ et } \forall k \in \llbracket 1, m_1 + m_2 + m_3 \rrbracket, k \neq i, \hat{C}'_{kj} = 0 \\ \hat{x}_i \text{ est appelée « variable de gauche » .} \end{aligned}$$

Ce passage de la forme normale à la forme normale restreinte est illustré sur un exemple en équations (4.15) et (4.16).

En remplaçant les variables de gauche par leurs expressions en fonction des  $\hat{b}'_i$  et des variables de droite, seules subsistent alors les variables de droite et les  $\hat{b}'_i$  dans l'expression de  $\hat{z}$  (équation (4.17) de l'exemple).

S'il est possible de mettre le problème (4.13) sous forme restreinte, ceci veut dire qu'il existe au moins un vecteur respectant l'ensemble des contraintes. En effet, mettant à zéro les variables de droite, ce qui respecte les contraintes primaires associées, on obtient que les variables de gauche  $\hat{x}_i$ ,  $i \in N_l$  valent  $\hat{b}'_k / \hat{C}'_{ki} \geq 0$  par respect des contraintes additionnelles. Elles respectent ainsi elles aussi les contraintes primaires de non négativité. On possède donc un vecteur  $\hat{x}$  qui, sans forcément maximiser  $\hat{z}$ , respecte toutes les contraintes. Un tel vecteur est appelé *vecteur admissible* (feasible vector). La recherche systématique d'un vecteur admissible pour un problème sous forme normale donné constitue la *phase I de la méthode du simplexe*. Si aucun vecteur admissible ne peut être exhibé, c'est qu'il n'existe aucun vecteur satisfaisant à toutes les contraintes simultanément, et donc, que le problème n'admet pas de solution. La résolution du problème à partir d'une forme normale restreinte constitue la *phase*

*II de la méthode du simplexe.* Elle s'effectue à l'aide d'un tableau et de pivots successifs entre variables de gauche et variables de droite.

Nous donnons ci-après la méthode utilisée pour réaliser la phase II de la méthode du simplexe, puis, la méthode permettant la transformation systématique d'une formulation normale en une formulation normale restreinte, dite phase I.

### Phase II de la méthode du simplexe

Nous ne donnons pas ici d'explication théorique et préférons l'explication par l'exemple. Considérons le problème général :

$$\begin{aligned}
 &\text{Maximiser } z = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 \\
 &\text{Sous les contraintes :} \\
 &\quad \tilde{x}_1 \geq 0, \tilde{x}_2 \geq 0, \tilde{x}_3 \geq 0 \\
 &\quad \tilde{x}_1 + \tilde{x}_2 \leq 6 \\
 &\quad \tilde{x}_2 + 2\tilde{x}_3 \leq 7 \\
 &\quad \tilde{x}_1 - 2\tilde{x}_2 = 0
 \end{aligned} \tag{4.14}$$

On introduit  $\lambda_1 = 6 - (\tilde{x}_1 + \tilde{x}_2)$  et  $\lambda_2 = 7 - (\tilde{x}_2 + 2\tilde{x}_3)$  pour mettre le problème sous forme normale :

$$\begin{aligned}
 &\text{Maximiser } \hat{z} = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 \\
 &\text{Sous les contraintes :} \\
 &\quad \tilde{x}_1 \geq 0, \tilde{x}_2 \geq 0, \tilde{x}_3 \geq 0, \lambda_1 \geq 0, \lambda_2 \geq 0 \\
 &\quad \lambda_1 + \tilde{x}_1 + \tilde{x}_2 = 6 \\
 &\quad \lambda_2 + \tilde{x}_2 + 2\tilde{x}_3 = 7 \\
 &\quad \tilde{x}_1 - 2\tilde{x}_2 = 0
 \end{aligned} \tag{4.15}$$

Le problème étant très simple ici, il est aisé de le mettre sous forme normale restreinte en constatant que  $\tilde{x}_1 = 2\tilde{x}_2$  :

$$\begin{aligned}
 &\text{Maximiser } \hat{z} = 3\tilde{x}_2 + \tilde{x}_3 \\
 &\text{Sous les contraintes :} \\
 &\quad \tilde{x}_1 \geq 0, \tilde{x}_2 \geq 0, \tilde{x}_3 \geq 0, \lambda_1 \geq 0, \lambda_2 \geq 0 \\
 &\quad \lambda_1 + 3\tilde{x}_2 = 6 \\
 &\quad \lambda_2 + 2\tilde{x}_2 + 2\tilde{x}_3 = 7 \\
 &\quad \tilde{x}_1 - 2\tilde{x}_2 = 0
 \end{aligned} \tag{4.16}$$

Pour résoudre ce système, on construit le tableau suivant :

	$b$	$\tilde{x}_2$	$\tilde{x}_3$
$\hat{z}$	0	3	1
$\tilde{x}_1$	0	2	0
$\lambda_1$	6	-3	0
$\lambda_2$	7	-2	-2

$\leftarrow$  Variables de droite ( $\tilde{x}_2, \tilde{x}_3$ )

$L_0 \quad \hat{z} = 0 + 3.\tilde{x}_2 + 1.\tilde{x}_3$

$L_1 \quad \tilde{x}_1 = 0 + 2.\tilde{x}_2 + 0.\tilde{x}_3$

$L_2 \quad \lambda_1 = 6 - 3.\tilde{x}_2 + 0.\tilde{x}_3$

$L_3 \quad \lambda_2 = 7 - 2.\tilde{x}_2 - 2.\tilde{x}_3$

$$\tag{4.17}$$

$\uparrow$   
 Variables de gauche ( $\tilde{x}_1, \lambda_1, \lambda_2$ )

Ce tableau traduit la fonction à maximiser et les contraintes additionnelles en exprimant les variables de gauche en fonction des variables de droite. La colonne  $b$  donne la valeur de  $\hat{z}$  et des variables de gauche lorsque toutes les variables de droite sont mises à zéro.

L'objectif étant de maximiser  $\hat{z}$  en augmentant les variables de droite (étant à 0, elles ne peuvent qu'augmenter), on repère les variables pour lesquelles les coefficients de  $L_0$  sont positifs : si la variable augmente,  $\hat{z}$  fera de même.

Sur notre exemple, les deux variables de droite sont susceptibles d'augmenter  $\hat{z}$ .

Il faut, après sélection des variables, déterminer de combien elles peuvent augmenter au maximum. Le facteur limitant est cette fois déterminé par les lignes  $L_1$  à  $L_{m_1+m_2+m_3}$  qui traduisent la dépendance des variables de gauche avec celles de droite et qui doivent respecter les contraintes primaires de non-négativité. Ainsi, une variable de droite peut augmenter jusqu'à annuler une variable de gauche. Seules les variables de gauche diminuant lorsque les variables de droite augmentent sont limitantes : il faut examiner les lignes comportant des coefficients négatifs. Après avoir trouvé l'augmentation maximale pour chaque variable de droite sélectionnée, on regarde son impact sur la fonction objectif  $\hat{z}$  et on choisit d'utiliser celle provoquant le gain le plus important.

Examinons notre exemple :

- variable  $\tilde{x}_2$  :
  - limitée par  $\lambda_1$  à une variation de  $-(6/(-3)) = 2$
  - limitée par  $\lambda_2$  à une variation de  $-(7/(-2)) = 7/2$
  - $\Rightarrow \tilde{x}_2$  peut varier au plus de 2, ce qui augmenterait  $\hat{z}$  de  $2 \times 3 = 6$
- variable  $\tilde{x}_3$  :
  - limitée par  $\lambda_2$  à une variation de  $-(7/(-2)) = 7/2$
  - $\Rightarrow \tilde{x}_3$  peut varier au plus de  $7/2$ , ce qui augmenterait  $\hat{z}$  de  $7/2 \times 1 = 7/2$

C'est donc  $\tilde{x}_2$ , limité par  $\lambda_1$  que l'on va faire varier de +2.

Après variation, la variable de droite retenue est à sa valeur maximale pour la configuration actuelle et la variable de gauche qui a servi à sa limitation est annulée, donc est à sa valeur minimale ; elle ne peut qu'augmenter. On effectue donc un « échange » entre les deux variables : la première devient une variable de gauche et vice-versa. On répercute bien évidemment cet échange sur le reste du tableau en réexprimant la dépendance de  $\hat{z}$  et des autres variables de gauche avec la nouvelle variable de droite.

L'échange  $\tilde{x}_2 \leftrightarrow \lambda_1$  impacte le tableau de façon suivante :

	$b$	$\lambda_1$	$\tilde{x}_3$	
$\hat{z}$	6	-1	1	$\hat{z} = 0 + 3.(2 - 1/3.\lambda_1) + 1.\tilde{x}_3 = 6 - 1.\lambda_1 + 1.\tilde{x}_3$
$\tilde{x}_1$	4	-2/3	0	$\tilde{x}_1 = 0 + 2.(2 - 1/3.\lambda_1) + 0.\tilde{x}_3 = 4 - (2/3).\lambda_1 + 0.\tilde{x}_3$
$\tilde{x}_2$	2	-1/3	0	$\tilde{x}_2 = 2 - 1/3.\lambda_1 + 0.\tilde{x}_3$
$\lambda_2$	3	2/3	-2	$\lambda_2 = 7 - 2.(2 - 1/3.\lambda_1) - 2.\tilde{x}_3 = 3 + (2/3).\lambda_1 - 2.\tilde{x}_3$

En répétant cette opération jusqu'à ce que la ligne  $L_0$  relative à  $\hat{z}$  ne contiennent que des coefficients négatifs - ce qui signifie que  $\hat{z}$  ne peut plus augmenter, donc est maximisée - on obtient le tableau :

	$b$	$\lambda_1$	$\lambda_2$
$\hat{z}$	15/2	-1	1
$\tilde{x}_1$	4	-2/3	0
$\tilde{x}_2$	2	-1/3	0
$\tilde{x}_3$	3/2	1/3	-1/2

On lit alors que la valeur maximale de  $\hat{z}$  est 15/2 et est atteinte pour  $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (4, 2, 3/2)$ .

**Remarque :** Il peut arriver que le problème ne soit pas borné, et donc qu'une solution particulière ne puisse être obtenue. Ceci se traduit dans le tableau par l'existence d'une variable de droite susceptible de faire augmenter  $\hat{z}$  (son coefficient en  $L_0$  est strictement positif) qui n'est limitée par aucune variable de gauche (les coefficients de sa colonne sont tous positifs ou nuls). Si ce cas est détecté, la procédure est arrêtée et signale que le problème est non borné.

### Phase I de la méthode du simplexe

Comme la phase II de la méthode du simplexe ne sait résoudre que des problèmes sous forme normale restreinte, on s'assure d'avoir un tel système en l'augmentant encore à l'aide de *variables artificielles* (fake variables) :  $\mu = \{\mu_1, \dots, \mu_{m_1+m_2}\}^t$  définies par :

$$\mu = \hat{b} - \hat{C}.\hat{x}$$

et soumises aux contraintes primaires et additionnelles :

$$\mu \geq 0 \quad \text{et} \quad \mu + \hat{C}.\hat{x} = \hat{b}$$

Le nouveau problème s'écrit donc :

$$\text{Maximiser :} \quad \hat{z} = \hat{f}^t.\hat{x} \quad (4.18a)$$

$$\text{Sous les contraintes primaires :} \quad \hat{x} \geq 0, \mu \geq 0 \quad (4.18b)$$

$$\text{Sous les contraintes supplémentaires :} \quad \mu + \hat{C}.\hat{x} = \hat{b}, \hat{b} \geq 0 \quad (4.18c)$$

Ainsi, on voit se dessiner les variables de droite,  $\hat{x}$  servant à exprimer  $\hat{z}$  et les variables de gauche  $\mu$  (chaque  $\mu_i$  n'apparaît bien que dans une équation de contrainte supplémentaire, affecté du coefficient positif 1). En revanche, le nouveau problème (4.18) n'est pas équivalent à celui que l'on souhaite traiter (4.13), à moins que  $\mu = 0$ . On impose donc cette condition en introduisant une *fonction objectif auxiliaire* (auxiliary objective function)  $z' = -\mu_1 - \dots - \mu_{m_1+m_2}$ , que l'on réexprime en fonction des variables de droite  $\tilde{x}_i$  et que l'on va chercher à maximiser. Les  $\mu_i$  étant tenus d'être tous positifs ou nuls,  $z'$  est maximale lorsqu'elle est nulle, ainsi que tous les  $\mu_i$ .

La maximisation de  $z'$  s'effectue grâce à la phase II du simplexe. Si cette étape aboutit à  $z' = 0$  alors le tableau résultant, auquel on enlève les colonnes relatives aux  $\mu_i$  (qui sont passés de variables de gauche à variables de droite puisque nuls), fournit le tableau initial

pour la résolution du problème consistant à maximiser  $\hat{z}$ . Si cette étape aboutit à  $z' \neq 0$ , c'est qu'il n'existe pas de vecteur admissible pour notre problème initial : le problème n'admet pas de solution.

En termes de tableaux, on observe donc les étapes décrites en figure 4.2.

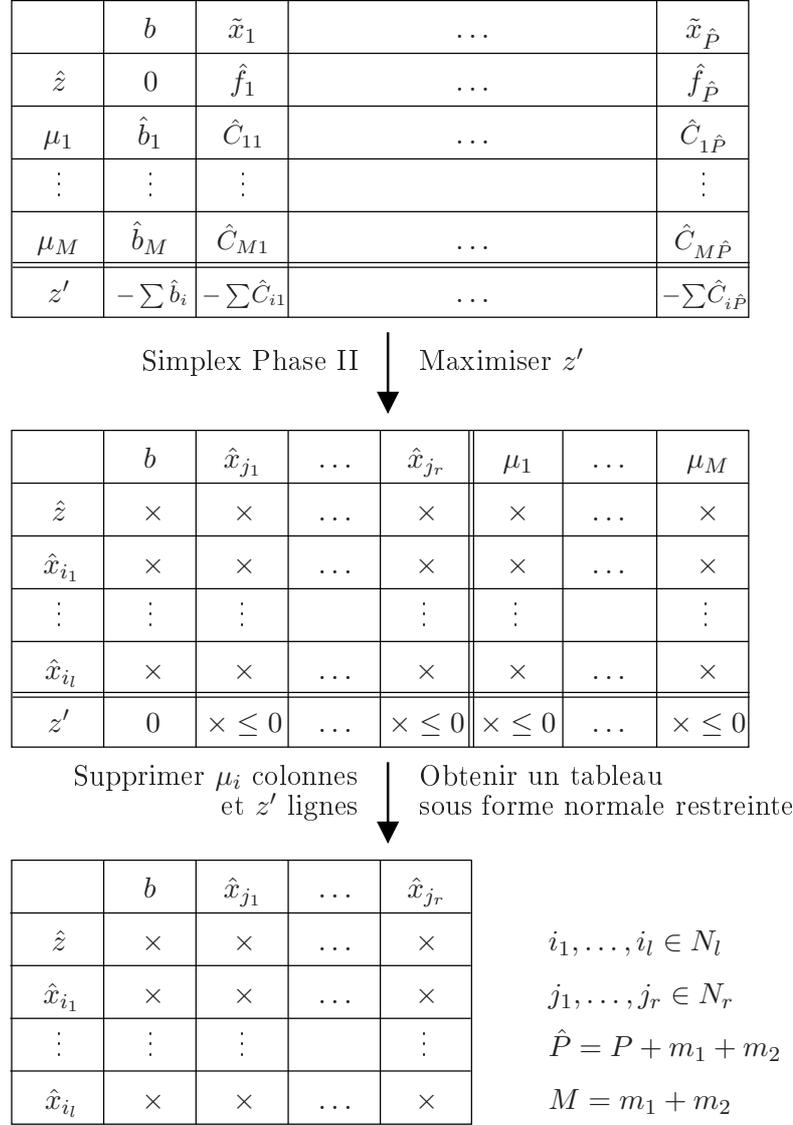


FIG. 4.2 – Méthodes d'exclusion - Méthode du simplexe - Tableaux de la Phase II

#### 4.4.2 Application de la méthode à un système non-linéaire

Comme évoqué en introduction, l'idée essentielle de cette méthode est de passer les non-linéarités en contraintes additionnelles de type inégalitaires. Les étapes de résolution du problème initial (4.10) sont décrites ci-dessous :

1 - On introduit  $\tilde{y} \in \mathbb{R}^P$  variables supplémentaires contraintes « remplaçant » les non-linéarités :

$$\begin{aligned}
 \text{Existent-ils } \tilde{x}, \tilde{y} \text{ tels que : } & \quad \bar{H} \cdot \tilde{x} + \tilde{y} = 0 \\
 & \quad \tilde{x} \leq \bar{X}, \tilde{y} \leq \hat{H}(\tilde{X}) - H_c \\
 & \quad \tilde{x} \geq \underline{X}, \tilde{y} \geq \underline{\hat{H}}(\tilde{X}) - H_c
 \end{aligned} \tag{4.19}$$

2 - On réalise ensuite le changement de variables (4.20) pour faire apparaître les contraintes primaires et additionnelles correctement formulées (4.21) :

$$\chi = \tilde{x} - \underline{\tilde{X}}, \quad \kappa = \tilde{y} - \underline{\hat{H}(\tilde{X}) - H_c} \quad (4.20)$$

$$\begin{aligned} \text{Existent-ils } \chi, \kappa \text{ tels que : } & \quad \bar{H}.\chi + \kappa = - \left( \bar{H}.\underline{\tilde{X}} + \underline{\hat{H}(\tilde{X}) - H_c} \right) \\ & \quad \chi \geq 0, \quad \kappa \geq 0 \\ & \quad \chi \leq b_\chi, \quad b_\chi = \overline{\tilde{X}} - \underline{\tilde{X}} \geq 0 \\ & \quad \kappa \leq b_\kappa, \quad b_\kappa = \hat{H}(\tilde{X}) - H_c - \underline{\hat{H}(\tilde{X}) - H_c} \geq 0 \end{aligned} \quad (4.21)$$

3 - On résout le problème de signe dans l'égalité de façon à avoir un terme constant à droite positif, conformément à la formulation des contraintes en égalité de la forme (4.12e). On pose donc :

$$s = -\text{sign} \left( \bar{H}.\underline{\tilde{X}} + \underline{\hat{H}(\tilde{X}) - H_c} \right) \text{ et } b_3 = -s \left( \bar{H}.\underline{\tilde{X}} + \underline{\hat{H}(\tilde{X}) - H_c} \right) \quad (4.22)$$

Ceci permet de réécrire (4.21) sous la forme :

$$\begin{aligned} \text{Existent-ils } \chi, \kappa \text{ tels que : } & \quad \chi \geq 0, \quad \kappa \geq 0 \\ & \quad \chi \leq b_\chi, \quad b_\chi \geq 0 \\ & \quad \kappa \leq b_\kappa, \quad b_\kappa \geq 0 \\ & \quad s(\bar{H}.\chi + \kappa) = b_3, \quad b_3 \geq 0 \end{aligned} \quad (4.23)$$

4 - On retrouve ainsi exactement la formulation des contraintes primaires et additionnelles d'un problème général résolu par la méthode du simplexe (voir système (4.12)). Il ne manque qu'une fonction à maximiser... dont nous n'avons a priori que faire étant donné que seule nous intéresse l'existence ou non d'un vecteur admissible qui satisfait à toutes les contraintes. Or, cette (non-)existence est donnée par la phase I de la méthode qui maximise une fonction objectif auxiliaire déterminée automatiquement.

Le test d'exclusion consiste donc en l'application de la phase I de la méthode du simplexe sur le système (4.23). Si la phase I se termine en retournant un vecteur admissible, la cellule est conservée puis divisée. En revanche, si la phase I renvoie l'information qu'il n'existe aucun vecteur admissible, c'est que le problème n'admet pas de solution sur la cellule  $\tilde{X}$  étudiée : elle est donc exclue.

## 4.5 Application à un oscillateur de Duffing

Afin de tester l'efficacité de la méthode du simplexe, nous l'appliquons à l'oscillateur de Duffing dont l'étude est détaillée en annexe A pour une pulsation d'excitation  $\omega = 1.2 \text{ rad.s}^{-1}$ . Cette méthode d'analyse globale permettant la résolution de systèmes algébriques, nous choisissons de transformer l'équation dynamique en équation algébrique en appliquant une méthode de balance harmonique. Nous utiliserons plusieurs décompositions fréquentielles du signal afin d'engendrer un nombre  $P$  d'inconnues algébriques variable. Les différentes décompositions et leur dénominations courtes sont résumées dans le tableau 4.1. Les équations en résultant sont aisées à retrouver en utilisant les expressions (A.24), (A.25) et (A.26) de l'annexe.

Dénomination	$P$	Expression
Cas 1H	2	$q(t) = a_1 \cos(\omega t) + b_1 \sin(\omega t)$
Cas C1H	3	$q(t) = a_0 + a_1 \cos(\omega t) + b_1 \sin(\omega t)$
Cas 1H3H	4	$q(t) = a_1 \cos(\omega t) + b_1 \sin(\omega t) + a_3 \cos(3\omega t) + b_3 \sin(3\omega t)$
Cas C1H3H	5	$q(t) = a_0 + a_1 \cos(\omega t) + b_1 \sin(\omega t) + a_3 \cos(3\omega t) + b_3 \sin(3\omega t)$

TAB. 4.1 – Méthode du simplexe - Cas d'application à un oscillateur de Duffing

Dans un premier temps, nous présentons les résultats obtenus pour les différents cas ; dans un second, nous nous appuyons sur ces résultats pour effectuer des extrapolations sur les coûts numériques.

#### 4.5.1 Résultats

Nous avons appliqué l'algorithme du simplexe aux trois cas du tableau 4.1. Nous avons choisi les valeurs suivantes pour les paramètres relatifs au critère d'arrêt :

$$\epsilon_{Xr} = 10^{-2}, \quad \epsilon_{Xa} = 10^{-3}, \quad \epsilon_H = 10^{-3}$$

L'espace initial  $\Omega$  est pris égal à :

$$\Omega = [-2, 2]^P$$

Chaque variable  $a_k$  ou  $b_k$  peut donc varier sur un intervalle  $[-2, 2]$ .

Les résultats chiffrés sont présentés dans le tableau 4.2 ; ils sont obtenus après calcul sur un serveur Intel Core 2 Duo E6600 (4Go de Ram).

Cas	$P$	Temps cpu (s)	Nombre de cellules traitées	Nombre de cellules retenues
1H	2	14	2073	42
C1H	3	143	9161	106
1H3H	4	1836	56481	153
C1H3H	5	9864	235425	306

TAB. 4.2 – Méthode du simplexe - Résultats

Les figures 4.3 à 4.5 donnent une représentation graphique des premiers résultats sous forme de pavés. Sur ces figures, les cellules ont été regroupées automatiquement sur le critère que si deux cellules possèdent au moins un coin en commun, elles font partie d'un même groupe. Nous avons ensuite représenté le barycentre des centres de chaque cellule d'un même groupe par un point vert. Puis, partant de ce point nous avons résolu de façon « exacte » l'équation en initialisant un solveur de type Newton avec ce barycentre. Le point obtenu est alors représenté en rouge.

En figure 4.4, la composante  $a_0$  n'a pas été représentée. Pour chaque rectangle en  $a_1, b_1$  représenté, deux cellules sont formées en ajoutant les intervalles  $[-4.883 \cdot 10^{-4}, 0]$  ou  $[0, 4.883 \cdot 10^{-4}]$  selon  $a_0$  pour le premier groupe de cellules et les intervalles  $[-2.441 \cdot 10^{-4}, 0]$  ou  $[0, 2.441 \cdot 10^{-4}]$

pour les trois autres. Par ailleurs, les groupes 3 et 4 sur les figures 4.3 et 4.4 font vraisemblablement référence à la même solution. Cependant, ne possédant aucun coin en commun, ils apparaissent comme distincts. De même pour les groupes 3, 4 et 5 sur la figure 4.5.

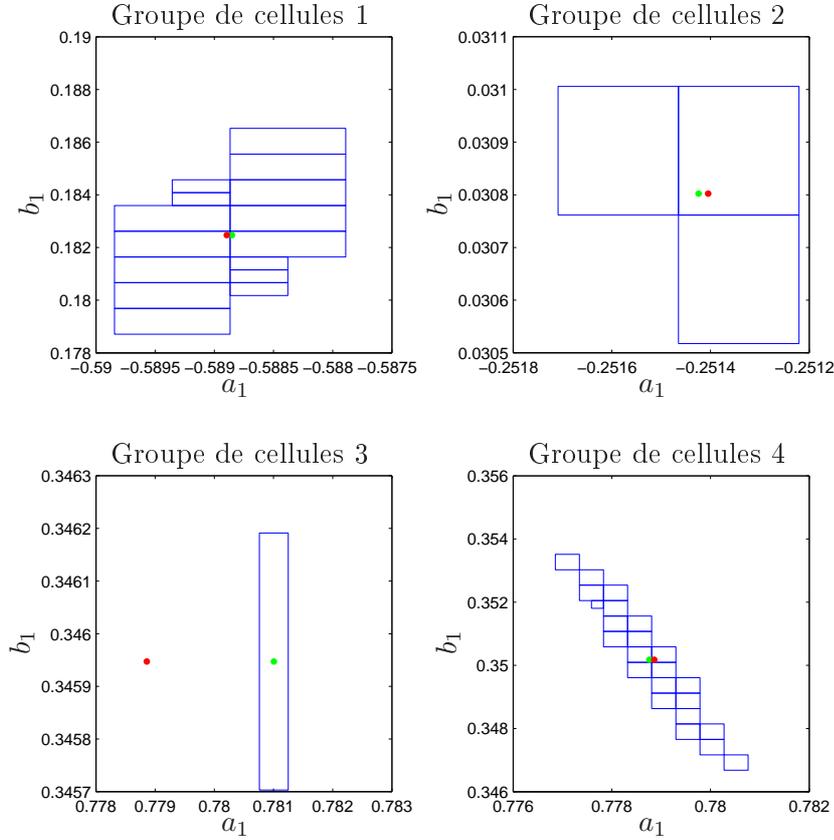


FIG. 4.3 – Méthode du simplexe - Cas 1H

Ces figures montrent que cette méthode permet de bien retrouver les trois solutions périodiques du système à peu de degré de liberté. Nous allons dans le paragraphe qui vient extrapoler les performances de cette méthode en termes de coût numérique.

#### 4.5.2 Extrapolations

Étant donné que la méthode procède par une division des cellules, il nous semble naturel pour établir une loi d'évolution du nombre de cellules de la chercher sous la forme  $N_c = a.b^P$  où  $N_c$  désigne le nombre de cellules traitées et  $P$  le nombre d'inconnues algébriques. Le graphique 4.6 valide cette hypothèse avec  $a = 82.4$  et  $b = 4.96$ .

Nous tentons maintenant une extrapolation du temps de calcul. Nous partons du principe que le temps de calcul  $T$  est de la forme  $T = N_c.t_u$ ,  $t_u$  représentant le temps de calcul par cellule. Nous cherchons ensuite la dépendance de  $t_u$  avec  $P$ . La courbe 4.7 semble indiquer qu'elle est linéaire :  $t_u = c.P + d$ ,  $c = 0.0123$  et  $d = -0.0189$ . Ceci est cohérent avec la méthode qui effectue une évaluation de la fonction  $H$  par cellule (ce qui implique le calcul de  $P$  termes). On obtient ainsi une loi reliant  $T$ , le temps de calcul et  $P$  le nombre d'inconnues algébriques :

$$T = (a.b^P).(c.P + d)$$

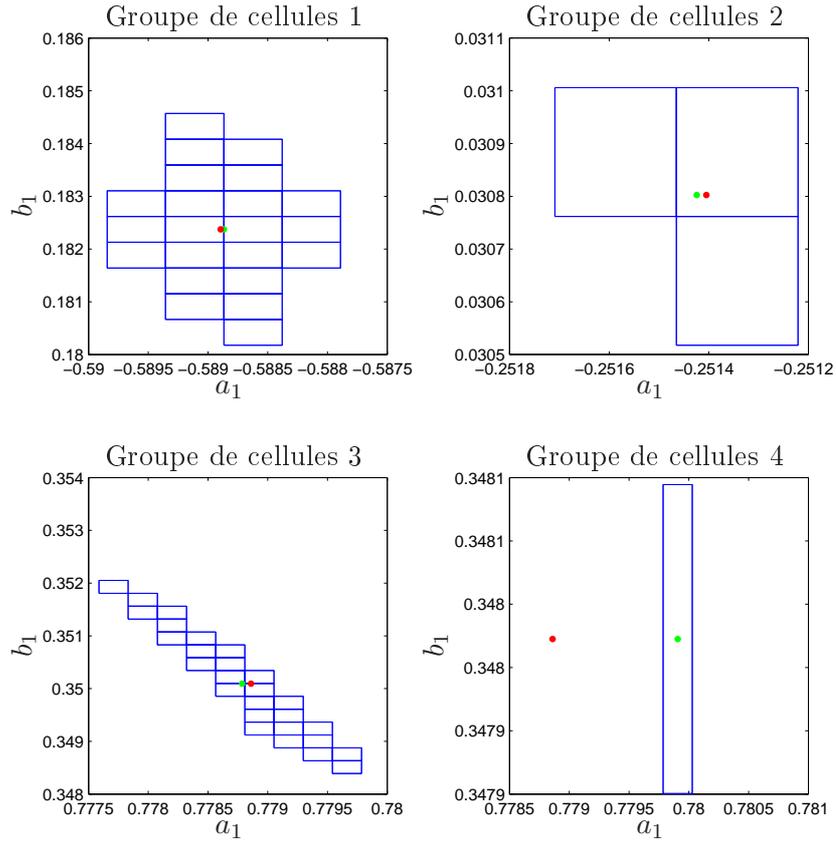


FIG. 4.4 – Méthode du simplexe - Cas C1H

Le graphique 4.8 présente une extrapolation du temps de calcul pour  $P$  variant de 1 à 10. Les étoiles rouges reportent les résultats du tableau 4.1. On lit alors que dès que  $P$  est supérieur ou égal à 7, les temps de calcul dépassent la semaine. Il est bon de noter que le temps de calcul par cellule  $t_u$  est « minimal » sur cet exemple dans le sens où, d'une part, le système est simple dans l'expression de sa non-linéarité qui est cubique et d'autre part, l'étape AFT (Alternate Frequency Time) qui oblige à reconstituer le signal temporel  $q(t)$  à partir des coefficients fréquentiels  $a_k, b_k$  pour pouvoir calculer  $\hat{H}$  conformément à l'équation (2.20) est évitée car nous avons mené analytiquement cette étape pour obtenir directement l'expression de  $\hat{H}$  en fonction des variables algébriques. Les temps donnés ici sont donc en quelque sorte minimaux et cette méthode ne semble pas prometteuse.

## 4.6 Conclusion

Cette méthode du simplexe qui s'inscrit dans la classe des algorithmes procédant par exclusion de cellules - c'est-à-dire de portions de l'espace  $\Omega$  sondé - présente l'avantage d'être très aisée à la mise en œuvre et de ne pas nécessiter une étude importante du système dynamique traité si ce n'est une transformation en un système d'équations algébriques. En revanche, bien que le coût numérique de l'application du test à une cellule soit peu élevé, sa faiblesse repose sur l'immense quantité de cellules générées lors de la subdivision d'une cellule non-éliminée par le test. Le coût global est ainsi trop important pour envisager l'utilisation de cette méthode sur des systèmes non-académiques à plus de trois degrés de liberté : dans le cas

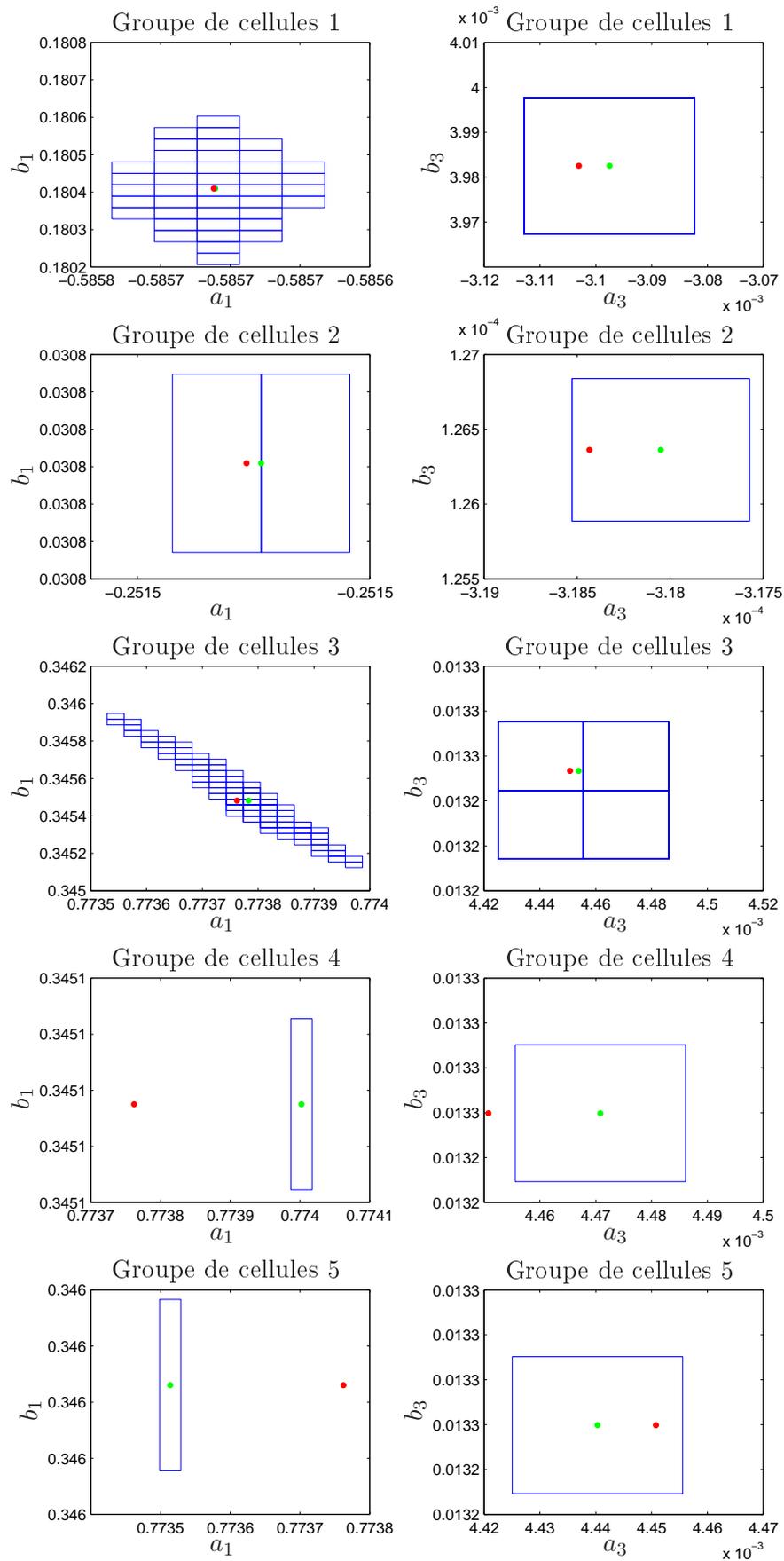


FIG. 4.5 – Méthode du simplexe - Cas 1H3H

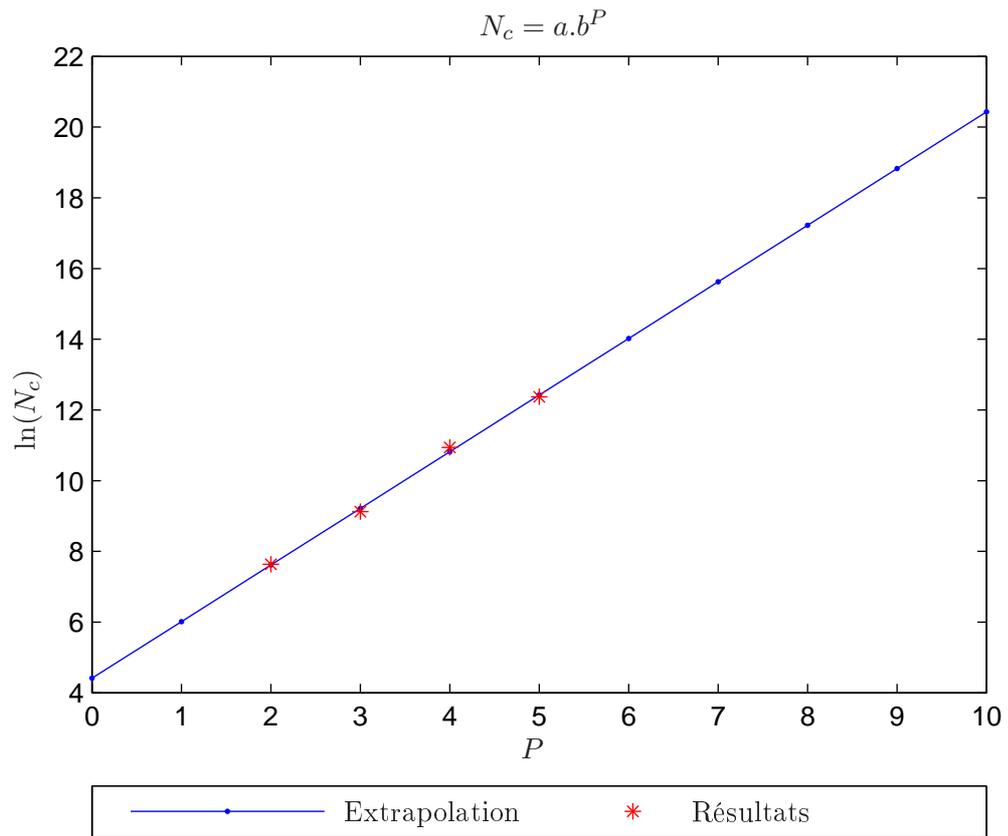


FIG. 4.6 – Méthode du simplexe - Extrapolation du nombre de cellules

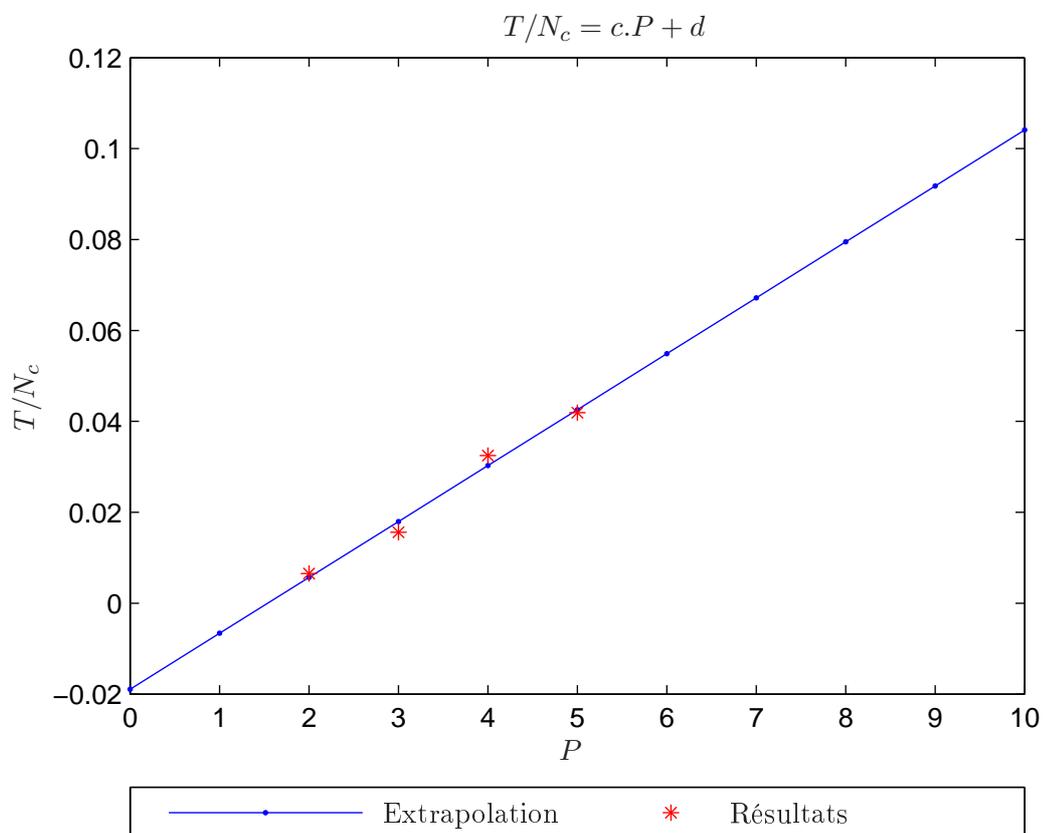


FIG. 4.7 – Méthode du simplexe - Extrapolation du temps par cellule

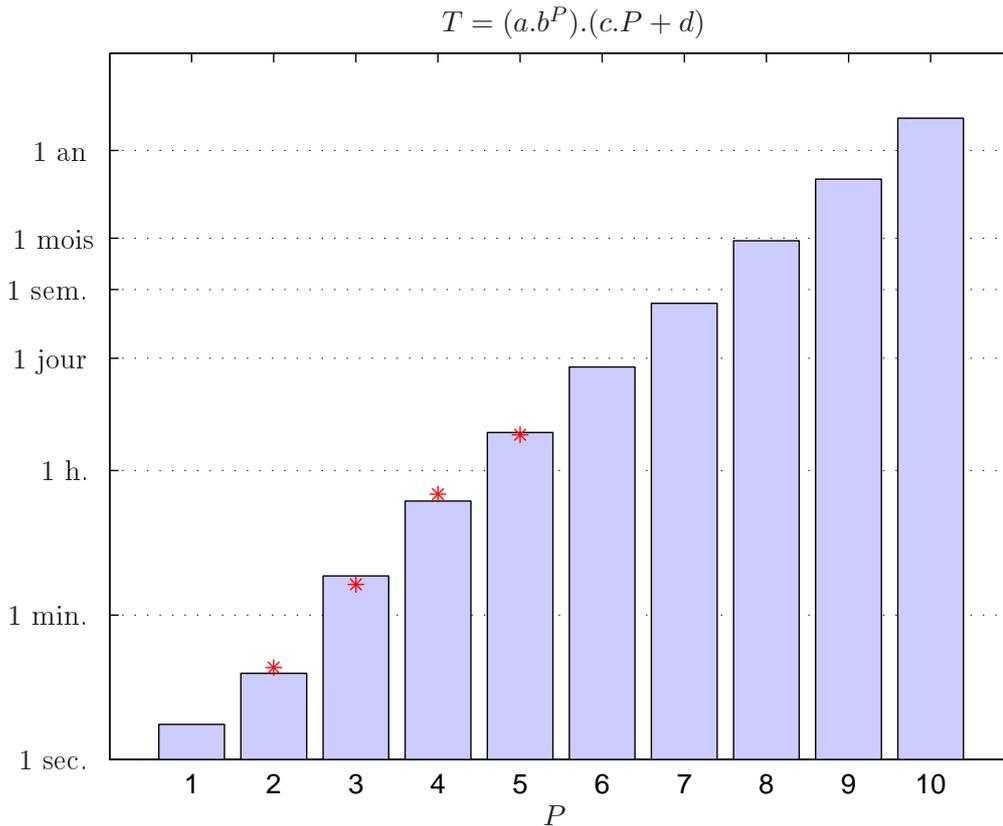


FIG. 4.8 – Méthode du simplexe - Extrapolation du temps de calcul

fréquent de la recherche de solution périodique, la décomposition du signal temporel requiert au moins l'utilisation d'un harmonique. En postulant même le terme constant comme nul pour des mouvements centrés, seuls des systèmes à moins de trois degrés de liberté peuvent être traités dans des délais de moins d'une journée selon notre extrapolation qui repose pourtant sur un exemple pour lequel les calculs nécessaires à une application du test sont réduits au minimum dans leur complexité (système initialement simple et expression analytique de la non-linéarité algébrique). Nous ne retiendrons donc pas cette méthode, ni en tant que solution viable à l'issue de ces travaux, ni en tant que méthode sur laquelle il semble intéressant de mener de nouvelles recherches étant donné que la construction somme toute rudimentaire de l'algorithme permet facilement d'affirmer qu'aucun gain majeur en temps ne peut être effectué.

Dans le chapitre suivant, nous étudierons une méthode reposant sur l'arithmétique des intervalles qui reprend le principe de découpage de l'espace en cellules mais qui ne procède pas uniquement par subdivision pour en réduire la taille.

## Chapitre 5

# Analyse globale utilisant l'arithmétique des intervalles

---

*L'algorithme décrit dans ce chapitre est basé sur l'utilisation de l'arithmétique des intervalles. Il est tiré de l'ouvrage [HW04] et a l'avantage de combiner deux approches : globale et locale. Il s'emploie comme les méthodes du chapitre précédent à la résolution d'un système algébrique de la forme (2.10). La tentative d'exclusion d'une cellule est ici intimement liée à sa réduction en cas d'échec. Il a été construit dans un souci d'efficacité, considérant toujours le gain d'espace éliminé après application d'une méthode et changeant d'approche si celui-ci est trop faible. Il existe une version très simple adaptée aux équations scalaires. Nous présentons ici, après une courte introduction au calcul sur les intervalles, la version appropriée aux problèmes à plusieurs variables.*

### 5.1 Théorie des intervalles

Les principaux points de la théorie des intervalles sont résumés dans l'article [AM00]. En quelques mots, elle vise à transposer les opérations courantes dans  $\mathbb{R}$  aux intervalles. Ainsi, si l'on connaît le domaine  $\tilde{X}_i$  sur lequel la variable  $\tilde{x}_i$  est susceptible de varier, on peut encadrer rigoureusement  $H_j(\tilde{x}_i)$ . Cette théorie est relativement triviale pour les opérations élémentaires. Elle se complique légèrement lorsqu'il s'agit de fonctions particulières, telles les fonctions trigonométriques. Nous donnons néanmoins les définitions des opérations élémentaires en annexe B.1. Une mise en œuvre informatique de l'arithmétique des intervalles est

mise à disposition sous forme d'une toolbox Matlab : INTLAB<sup>1</sup>.

Il est important de noter que les encadrements obtenus sont pessimistes : ils sont, au bout de quelques opérations, beaucoup plus larges que nécessaires. Ainsi, si cette approche est un bon point de départ, elle ne peut pas, à elle seule donner naissance à un outil de résolution efficace. De plus, il est important de tenir compte, lors de l'écriture des fonctions, de la sous-distributivité : en considérant trois intervalles  $A$ ,  $B$  et  $C$ , on a :

$$A \times (B + C) \subseteq A \times B + A \times C$$

Il est donc intéressant de factoriser un maximum les expressions des parties non-linéaires afin d'avoir l'encadrement le moins pessimiste possible.

Enfin, nous définissons ici trois notations fréquemment utilisées par la suite pour un pavé de  $\mathbb{R}^P$  noté indifféremment  $\tilde{X} = \{\tilde{X}_1, \dots, \tilde{X}_P\}$  ou  $\tilde{X} = \tilde{X}_1 \times \dots \times \tilde{X}_P$  :

- Borne inférieure et borne supérieure :  $\underline{\tilde{X}}_i$  et  $\overline{\tilde{X}}_i$ ,  $\tilde{X}_i = [\underline{\tilde{X}}_i, \overline{\tilde{X}}_i]$
- Largeur (width) :  $w(\tilde{X}_i) = \overline{\tilde{X}}_i - \underline{\tilde{X}}_i$  et  $w(\tilde{X}) = \max_{1 \leq i \leq P} (w(\tilde{X}_i))$
- Magnitude :  $\text{mag}(\tilde{X}_i) = \max(|\underline{\tilde{X}}_i|, |\overline{\tilde{X}}_i|)$

## 5.2 L'algorithme général d'Hansen et Walster

Cet algorithme présenté de façon très générale en figure 5.2 combine à la fois tests d'exclusion et méthodes de réduction de cellule adaptées à des grosses cellules dans un premier temps (Hull et Box Consistency) et à des cellules plus fines dans un second (Newton Généralisé). Nous présentons ici la démarche générale ; le détail des étapes importantes est donné dans les paragraphes suivants et on trouvera un schéma complet et détaillé de l'algorithme en annexe B.3.

En premier lieu, on se donne un espace de variation des variables  $\tilde{x}_i$  que l'on traduit sous forme d'un ensemble de pavés de  $\mathbb{R}^P$ . Ainsi, si l'on possède deux variables  $\tilde{x}_1$  et  $\tilde{x}_2$  et qu'on leur laisse comme espace de variation les aires bleutées en figure 5.1.(a) ou 5.1.(b), on obtiendra les ensembles suivants :

- Cas 1 :  $(\tilde{x}_1, \tilde{x}_2) \in \{[-1, 3], [-0.5, 2]\}$
- Cas 2 :  $(\tilde{x}_1, \tilde{x}_2) \in [-\frac{\sqrt{2}}{2}r, \frac{\sqrt{2}}{2}r] \times [-\frac{\sqrt{2}}{2}r, \frac{\sqrt{2}}{2}r] \cup [\frac{\sqrt{2}}{2}r, \frac{\sqrt{3}}{2}r] \times [-\frac{r}{2}, \frac{r}{2}] \cup [-\frac{\sqrt{3}}{2}r, \frac{\sqrt{2}}{2}r] \times [-\frac{r}{2}, \frac{r}{2}] \cup [-\frac{r}{2}, \frac{r}{2}] \times [\frac{\sqrt{3}}{2}r, \frac{\sqrt{2}}{2}r] \cup [-\frac{r}{2}, \frac{r}{2}] \times [-\frac{\sqrt{3}}{2}r, -\frac{\sqrt{2}}{2}r]$

On stocke chacun de ces pavés initiaux ou « cellules » dans une liste que nous noterons *intList*. L'algorithme extrait alors une première cellule  $\tilde{X}$  de cette liste pour la traiter. Il commence par appliquer les méthodes de Hull Consistency (cf. § 5.2.2) et Box Consistency (cf. § 5.2.3) qui tentent de réduire voire d'éliminer la cellule ; ces méthodes sont réitérées si le progrès est considéré comme important : bloc « méthode efficace » en figure 5.2. Le critère adopté ici pour juger d'un progrès satisfaisant est :

$$\exists i \in \llbracket 1, P \rrbracket / w(\tilde{X}_i^{(j)}) - w(\tilde{X}_i) > \frac{w(\tilde{X}^{(j)})}{4} \quad (5.1)$$

<sup>1</sup>Cette toolbox est accessible et gratuite si son utilisation est faite à des fins non commerciales, sur la page de l'Institute for Reliable Computing, <http://www.ti3.tu-harburg.de/>

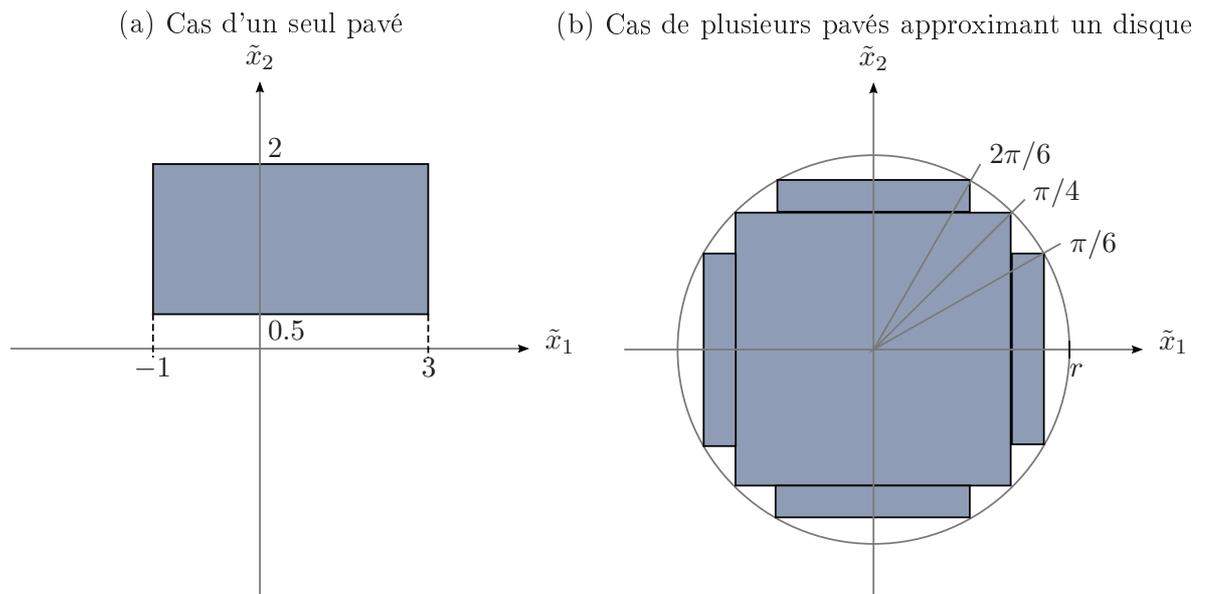


FIG. 5.1 – Arithmétique des intervalles - Exemple d'ensembles de départ

où  $\tilde{X}^{(j)}$  désigne la cellule avant application de la méthode.

On requiert donc qu'une composante au moins soit réduite d'un quart par rapport à la composante la plus large de la cellule initiale. Ces méthodes de Hull et Box Consistency présentent l'avantage d'être peu coûteuses numériquement; elles sont donc appliquées en premier lieu.

Lorsque ces méthodes n'ont plus une action efficace de réduction ou d'élimination de la cellule, et si celle-ci semble pouvoir s'y prêter (bloc « Cellule suffisamment petite pour une approche locale ? »), une approche plus locale est adoptée en utilisant une méthode que nous nommerons « Newton Généralisée » (cf. § 5.2.4) suivie éventuellement par une application partielle d'une méthode de Box Consistency.

Si l'ensemble des étapes Hull et Box Consistency, Newton Généralisée et Box Consistency Partiel a permis un net progrès de réduction de la cellule, elles sont appliquées une nouvelle fois; sinon, la cellule est subdivisée en huit sous-cellules au plus qui sont ajoutées à la liste *intList* puis traitées les unes après les autres.

Si en cours de route, une cellule s'avère satisfaire aux critères qui en font une cellule solution (globalement, une cellule suffisamment petite et dont l'image est de faible magnitude, voir § 5.2.8 pour le détail), elle est stockée dans une liste notée *resultList*.

On trouvera ci-dessous (sections 5.2.1 à 5.2.8) le détail des outils et étapes majeures de l'algorithme.

### 5.2.1 Étape de Newton

Nous exposons ici un des outils clé de l'algorithme; il est utilisé lors de la méthode de Box Consistency et le même principe apparaît dans la méthode dite « Newton Généralisée ». Il sert à la fois de test d'exclusion et de réduction d'un intervalle. On cherche à savoir s'il existe des solutions de :

$$H(\tilde{x}) = 0, \quad \tilde{x} \in \tilde{X} = \{\tilde{X}_1, \dots, \tilde{X}_P\}$$

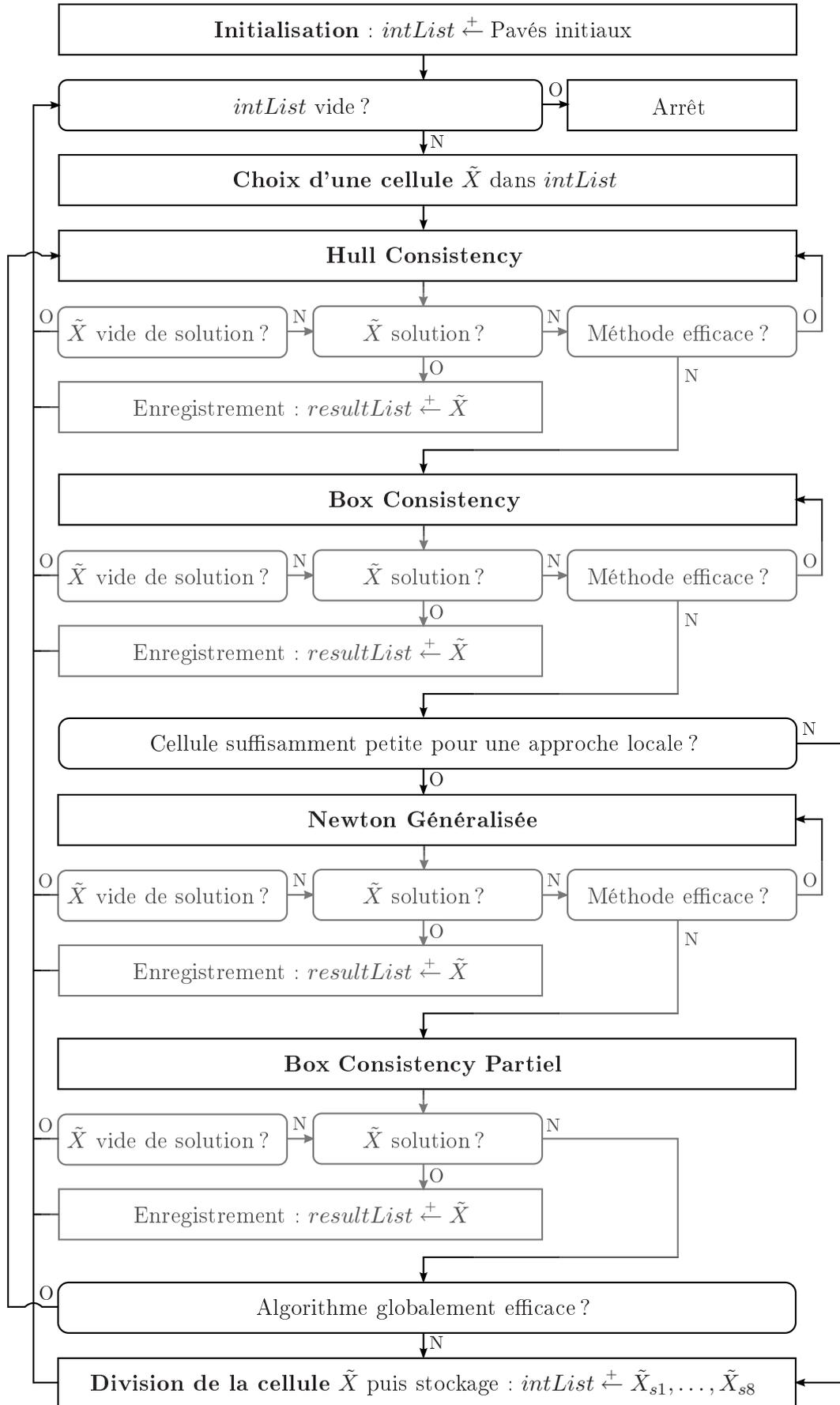


FIG. 5.2 – Arithmétique des intervalles - Algorithme Général

S'il en existe, il existe a fortiori des solutions de :

$$H_j(\tilde{x}_i) \ni 0, \begin{cases} \tilde{x}_i \in \tilde{X}_i \\ H_j(\tilde{x}_i) = H_j(\tilde{X}_1, \dots, \tilde{X}_{i-1}, \tilde{x}_i, \tilde{X}_{i+1}, \dots, \tilde{X}_P) \end{cases}$$

La fonction considérée va de  $\tilde{X}_i$  dans l'ensemble des intervalles de  $\mathbb{R}$  et correspond à la  $j$ -ième composante de la fonction  $H$  dont seule l'inconnue  $\tilde{x}_i$  est gardée comme variable réelle, les autres variables étant remplacées par leurs intervalles de valeurs. Utilisant alors la formule de Taylor sur un intervalle  $\tilde{Y}_i \subset \tilde{X}_i$ , on peut écrire que :

$$\forall (\tilde{x}_i^o, \tilde{x}_i^*) \in (\tilde{Y}_i)^2, \exists \xi \in (\tilde{x}_i^o, \tilde{x}_i^*) \text{ tel que } H_j(\tilde{x}_i^*) = H_j(\tilde{x}_i^o) + H_j'(\xi) \cdot (\tilde{x}_i^* - \tilde{x}_i^o)$$

La notation  $(\tilde{x}_i^o, \tilde{x}_i^*)$  sert à désigner indifféremment  $[\tilde{x}_i^o, \tilde{x}_i^*]$  ou  $[\tilde{x}_i^*, \tilde{x}_i^o]$  et  $H_j' = \partial H_j / \partial \tilde{x}_i$ .

En remplaçant  $\xi$  par  $\tilde{Y}_i$  qui contient toutes les valeurs que peut prendre  $\xi$ , cette formule devient :

$$\forall (\tilde{x}_i^o, \tilde{x}_i^*) \in (\tilde{Y}_i)^2, H_j(\tilde{x}_i^*) \subset H_j(\tilde{x}_i^o) + H_j'(\tilde{Y}_i) \cdot (\tilde{x}_i^* - \tilde{x}_i^o) \quad (5.2)$$

On s'affranchit ainsi de la valeur exacte de  $\xi$  qui est inconnue.

Ainsi, si l'on cherche les valeurs de  $\tilde{x}_i^*$  sur  $\tilde{Y}_i$  telles que  $0 \in H_j(\tilde{x}_i^*)$ , on peut écrire que pour ces valeurs, on a nécessairement :

$$0 \in H_j(\tilde{x}_i^o) + H_j'(\tilde{Y}_i) \cdot (\tilde{x}_i^* - \tilde{x}_i^o) \quad (5.3)$$

Ce qui implique :

$$\tilde{x}_i^* \in N_j(\tilde{x}_i^o, \tilde{Y}_i) = \tilde{x}_i^o - \frac{H_j(\tilde{x}_i^o)}{H_j'(\tilde{Y}_i)} \quad (5.4)$$

Or  $\tilde{x}_i^* \in \tilde{Y}_i$ . On en déduit donc que :

1. Si  $\tilde{Y}_i \cap N_j(\tilde{x}_i^o, \tilde{Y}_i) = \emptyset$ , alors  $0 \notin H_j(\tilde{Y}_i)$  et par conséquent,  $0 \notin H(\tilde{X}_1, \dots, \tilde{X}_{i-1}, \tilde{Y}_i, \tilde{X}_{i+1}, \dots, \tilde{X}_P)$ . On peut donc éliminer cette partie de l'espace de recherche.
2. Si  $\tilde{Y}_i \cap N_j(\tilde{x}_i^o, \tilde{Y}_i) \neq \emptyset$ , à défaut d'éliminer tout l'espace correspondant à  $\tilde{x}_i \in \tilde{Y}_i$ , on peut au moins restreindre nos recherches au sous-espace  $\tilde{Y}_i' = \tilde{Y}_i \cap N(\tilde{x}_i^o, \tilde{Y}_i)$  - en espérant qu'il soit effectivement plus petit que  $\tilde{Y}_i$ .

Notons que dans ce cas, si  $0 \in H_j'(\tilde{Y}_i)$ ,

- soit  $0 \in H_j(\tilde{x}_i^o)$ , ce qui implique que  $N_j(\tilde{x}_i^o, \tilde{Y}_i) = \mathbb{R} \cup \{-\infty, +\infty\}$ . Alors,  $\tilde{Y}_i' = \tilde{Y}$  et aucun progrès n'est réalisé.
- soit  $0 \notin H_j(\tilde{x}_i^o)$ , ce qui implique que  $N_j(\tilde{x}_i^o, \tilde{Y}_i) = [-\infty, n] \cup [p, +\infty]$ ,  $n < 0 < p$  : il existe un gap qui se répercutera peut-être sur  $\tilde{Y}_i'$ , le séparant en deux intervalles distincts.

### 5.2.2 Méthode de Hull Consistency

La méthode du Hull Consistency essaie d'exploiter au maximum la partie linéaire de la fonction  $H$ , mieux appréhendée mathématiquement, afin de réduire ou d'éliminer la cellule  $\tilde{X}$  traitée. En effet, si l'on cherche à résoudre  $H(\tilde{x}) = \bar{H} \cdot \tilde{x} + \hat{H}(\tilde{x}) - H_c = 0$  et que l'on cherche un tel zéro sur la cellule  $\tilde{X} = \{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_P\}$ , il est intéressant de remarquer que l'on peut

écrire :  $\forall i \in \llbracket 1, P \rrbracket$ ,  $\bar{H}_{ij} \cdot \tilde{X}_j + \hat{H}_i(\tilde{X}) - H_{ci} = 0$  et que si l'on cherche à réduire  $\tilde{X}_k$ , on peut affirmer que l'on peut le restreindre à son intersection avec  $\tilde{X}'_k$  défini par :

$$\bar{H}_{ik} \cdot \tilde{X}'_k = -\hat{H}_i(\tilde{X}) + H_{ci} \ominus \sum_{j \neq k} \bar{H}_{ij} \cdot \tilde{X}_j, \quad 1 \leq i \leq P \quad (5.5)$$

où  $\ominus$  désigne la soustraction étendue :  $X \ominus Y = [\underline{X} - \underline{Y}, \bar{X} - \bar{Y}]$

En appliquant ceci pour tout  $i$ , pour chacun des  $\tilde{X}_k$ , et en réinjectant les résultats au fur et à mesure de la réduction des intervalles  $\tilde{X}_k$ , on peut espérer diminuer de façon conséquente la taille de la cellule initiale.

Au final, on aura amené la cellule  $\tilde{X}$  à être cohérente par rapport à la formulation  $\bar{H} \cdot \tilde{x} = -\hat{H}(\tilde{x}) + H_c$  du problème ; c'est à dire à faire en sorte d'avoir le plus petit pavé  $\tilde{X}$  contenant toutes les solutions de  $-\hat{H}(\tilde{X}) + H_c$ . Ce plus petit pavé de  $\mathbb{R}^P$  contenant toutes les solutions d'une équation de la forme  $A \cdot \tilde{x} = b$ , avec  $A$  et  $b$  respectivement une matrice et un vecteur dont les éléments sont des intervalles, s'appelle en anglais, le *hull* - d'où le nom de la méthode. Il ne correspond pas forcément exactement à l'ensemble des solutions mais au plus petit pavé qui contiendra cet ensemble comme l'illustre la figure 5.3. Sur cette illustration, la partie bleutée correspond à l'ensemble solution de l'équation suivante :

$$\begin{bmatrix} [2, 3] & [0, 1] \\ [1, 2] & [2, 3] \end{bmatrix} \cdot \tilde{x} = \begin{Bmatrix} [0, 90] \\ [60, 200] \end{Bmatrix}$$

Le rectangle bleu correspond au hull de l'équation.

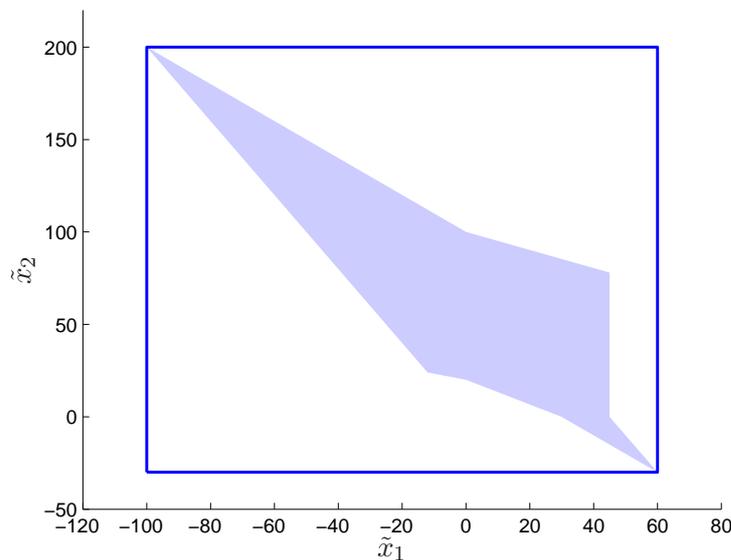


FIG. 5.3 – Arithmétique des intervalles - Illustration du hull

### 5.2.3 Méthode de Box Consistency

Tout comme la méthode de Hull Consistency, cette méthode a pour but principal de diminuer la taille de la cellule considérée ; il arrive cependant qu'elle permette de l'éliminer. Elle se base principalement sur l'utilisation d'étapes de Newton (cf § 5.2.1).

Considérons que nous cherchons à réduire  $\tilde{X}_i$ . En choisissant une des composantes de  $H$ , par exemple  $H_j$ , on sait que le calcul de  $N_j(\tilde{x}_i^o, \tilde{Y}_i), \tilde{Y}_i \subset \tilde{X}_i, \tilde{x}_i^o \in \tilde{Y}_i$  permet d'éliminer ou de réduire  $\tilde{Y}_i$ . Si l'on note  $\tilde{X}_i = [\underline{\tilde{X}}_i, \overline{\tilde{X}}_i]$  et que le but est de remonter  $\underline{\tilde{X}}_i$  et de descendre  $\overline{\tilde{X}}_i$ , il est alors assez naturel d'effectuer une étape de Newton pour  $\tilde{Y}_i = [\underline{\tilde{X}}_i, \tilde{c}_i], \tilde{c}_i \in \tilde{X}_i$  puis une autre pour  $\tilde{Y}_i = [\tilde{d}_i, \overline{\tilde{X}}_i], \tilde{d}_i \in \tilde{X}_i$ . Le choix de  $\tilde{c}_i$  et  $\tilde{d}_i$  est un compromis entre rendre les intervalles  $\tilde{Y}_i$  larges, donc éliminer éventuellement de grandes portions de  $\tilde{X}_i$ , et considérer des intervalles suffisamment petits pour que l'intervalle  $H'_j(\tilde{Y}_i)$  soit suffisamment petit lui aussi et ne rende pas  $N_j(\tilde{x}_i^o, \tilde{Y}_i)$  trop vaste. Les auteurs proposent de commencer par  $\tilde{c}_i = \underline{\tilde{X}}_i + \frac{1}{4}(\overline{\tilde{X}}_i - \underline{\tilde{X}}_i)$  puis d'augmenter la proportion de l'intervalle considérée si la méthode s'avère efficace. Enfin, le point d'expansion  $\tilde{x}_i^o$  est pris égal à la borne inférieure  $\underline{\tilde{X}}_i$  ou supérieure  $\overline{\tilde{X}}_i$  selon que l'on cherche à faire remonter la première ou redescendre l'autre.

L'algorithme utilisé pour remonter  $\underline{\tilde{X}}_i$  relativement à l'équation  $H_j$  est donné en figure 5.4. La quantité  $\epsilon_{Xa}$  est la même que celle du critère d'arrêt A détaillé en § 5.2.8 et jugeant de la suffisante « petitesse » d'une cellule. Une fois exécutée cette boucle pour remonter la borne inférieure, une boucle totalement similaire est utilisée pour descendre la borne supérieure, à moins qu'il ait été déterminé qu'il n'existe aucune solution sur l'intervalle considéré. Ceci est à appliquer pour chaque variable  $\tilde{x}_i, 1 \leq i \leq P$  au sein de chaque équation  $H_j, 1 \leq j \leq P$ .

#### 5.2.4 Méthode de Newton Généralisée

Cette méthode n'est appliquée que si « la cellule est assez petite pour une approche locale » selon l'algorithme général donné en figure 5.2. On comprend aisément qu'il n'est pas évident de répondre *a priori* à cette question lorsque le peu d'informations que l'on possède sur une cellule tient en ses bornes et éventuellement son image; c'est pourtant ce que l'on doit essayer de faire afin de ne pas consacrer de la ressource numérique à une méthode coûteuse et totalement inappropriée. Les auteurs proposent un critère basé sur la mémorisation de la réussite des tentatives précédentes : tout au long de l'exécution de l'algorithme, deux variables  $w_R$  et  $w_I$  sont réactualisées et désignent respectivement la largeur de la dernière cellule pour laquelle l'application d'une méthode de Newton Généralisée s'est « bien passée » et la largeur de la dernière cellule pour laquelle l'application d'une telle méthode s'est « mal passée ». La définition de ces qualifications « bien » ou « mal passée » repose globalement sur la régularité de la jacobienne (composée d'intervalles) calculée pour les cellules en question; le schéma détaillé de l'algorithme donné en annexe B.3 permet de mieux comprendre ces définitions. Possédant ces deux nombres réels  $w_R$  et  $w_I$  et devant décider si oui ou non une méthode de Newton Généralisée sera appliquée, on compare la largeur de la cellule à la moyenne des deux valeurs; si sa largeur est supérieure, elle est jugée mauvaise candidate à une telle méthode, sinon, l'approche par méthode de Newton Généralisée s'enclenche :

$$\begin{cases} \text{Si } w(\tilde{X}) > \frac{w_R + w_I}{2}, \text{ division de la cellule} \\ \text{Si } w(\tilde{X}) \leq \frac{w_R + w_I}{2}, \text{ application de la méthode de Newton Généralisée} \end{cases} \quad (5.6)$$

Nous entamons maintenant la description de cette méthode qui calque en quelque sorte une méthode de Newton classique à ceci près qu'elle s'effectue sur des matrices et vecteurs constitués d'intervalles ce qui rend les calculs beaucoup plus coûteux.

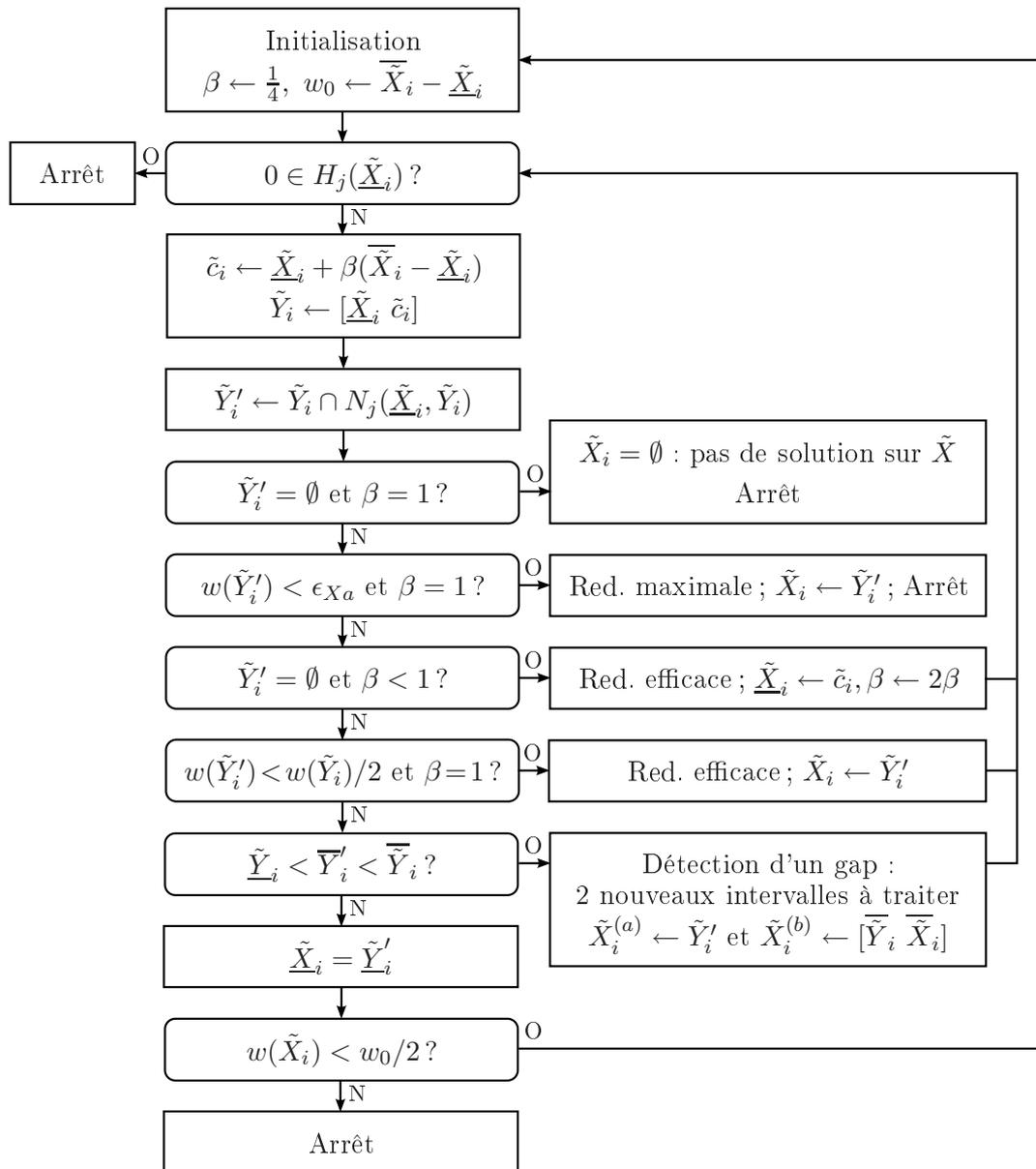


FIG. 5.4 – Arithmétique des intervalles - Algorithme de Box Consistency

Le point initial dans cette méthode est, comme dans le cas d'une étape de Newton, un développement limité autour d'un point d'expansion  $\tilde{x}_{exp}$  appartenant à  $\tilde{X}$ . Le choix de ce point d'expansion est discuté en section 5.2.5. S'il existe un zéro  $\tilde{x}^*$  de  $H$  sur  $\tilde{X}$ , on peut écrire :

$$0 \in H(\tilde{x}_{exp}) + J.(\tilde{x}^* - \tilde{x}_{exp}), \quad J = \left[ \frac{\partial H_i}{\partial \tilde{x}_j}(\tilde{X}) \right]_{ij} \quad (5.7)$$

Par suite, on peut affirmer que l'ensemble des  $\tilde{x}^*$  solution de  $H(\tilde{x}) = 0$  sur  $\tilde{X}$  est tel que :

$$J.(\tilde{x}^* - \tilde{x}_{exp}) = -H(\tilde{x}_{exp}) \quad (5.8)$$

On retrouve donc une écriture classique du problème, identique à celle présente dans les algorithmes dits de Newton.

Afin de minimiser les problèmes numériques, on préconditionne le problème en utilisant  $P$ , matrice inverse de la matrice « moyenne »  $J_c$  de  $J$ . Cette matrice moyenne est obtenue en prenant le centre de chaque élément de  $J$ . Ainsi,  $\tilde{J} = P.J$  est rapprochée de l'identité. En notant  $r = P.(-H(\tilde{x}_{exp}))$ , on obtient la formulation suivante du problème : trouver les  $\tilde{x}^*$  tels que

$$\tilde{J}.(\tilde{x}^* - \tilde{x}_{exp}) = r \quad (5.9)$$

Il faut alors distinguer les cas où  $\tilde{J}$  est régulière (i.e. ne contient aucune matrice singulière) et où elle ne l'est pas. Si elle l'est, une méthode de recherche du hull est utilisée pour résoudre exactement le problème. Si elle ne l'est pas, on s'en remet à une méthode de type Gauss-Seidel pour tenter de diminuer  $\tilde{X}$  malgré tout.

### Méthode de recherche du hull

La démonstration de cette méthode permettant de trouver le hull de l'équation linéaire  $\tilde{J}.\tilde{Y} = r$  est relativement longue. Nous ne donnons donc que les étapes principales. On rappelle que hull est le plus petit pavé contenant l'ensemble des solutions d'une équation de la forme  $A.\tilde{x} = b$ ,  $A$  et  $b$  étant composés d'intervalles.

1. Il faut faire en sorte que  $\forall k \in \llbracket 1, P \rrbracket$ ,  $|\underline{r}_k| \leq \bar{r}_k$ . On utilise pour cela une matrice  $S$  diagonale, comportant  $-1$  sur sa diagonale lorsque l'inégalité précédente n'est pas vérifiée par  $r_k$ , et  $1$  sinon. On considère donc  $r' = S.r$ .

2. Puis, on définit :

- $\tilde{P} = (\tilde{J})^{-1}$
- $\forall k \in \llbracket 1, P \rrbracket$ ,  $C_k = 1/(2\tilde{P}_{kk} - 1)$
- $\forall k \in \llbracket 1, P \rrbracket$ ,  $Z_k = (\underline{r}'_k + \bar{r}'_k)\tilde{P}_{kk} - (\tilde{P}.\bar{r}')_k$

Les formules suivantes permettent alors de calculer  $\tilde{Y}' = S.\tilde{Y}$  :

- $\forall k \in \llbracket 1, P \rrbracket$ ,  $\tilde{Y}'_k = \begin{cases} C_k Z_k & \text{si } Z_k > 0 \\ Z_k & \text{si } Z_k \leq 0 \end{cases}$
- $\bar{\tilde{Y}}' = \tilde{P}.\bar{r}'$

3. Il ne reste plus qu'à récupérer  $\tilde{Y}$  en réappliquant  $S$  à  $\tilde{Y}'$  puis à utiliser  $\tilde{Y} = (N - \tilde{x}_{exp})$  :  $N = \tilde{x}_{exp} + S.\tilde{Y}'$ . On espère alors réduire ou éliminer  $\tilde{X}$  en appliquant l'intersection  $\tilde{X}' = N \cap \tilde{X}$ .

### Gauss-Seidel Method

On cherche ici à résoudre approximativement  $\tilde{J}.\tilde{Y} = r$ . Tout comme dans la méthode de Hull Consistency, il est possible d'écrire :

$$\forall k \in \llbracket 1, P \rrbracket, \tilde{J}_{kk}.\tilde{Y}'_k = r_k \ominus \sum_{j \neq k} \tilde{J}_{kj}.\tilde{Y}'_j$$

Puis  $\tilde{X}'_k = \tilde{X}_k \cap N_k$ ,  $N_k = \tilde{Y}'_k + \tilde{x}_{exp_k}$ . En réinjectant dans la somme de droite les  $\tilde{Y}'_k$  déjà calculés, on peut rendre cette procédure plus efficace.

Une autre remarque permettant de rendre la procédure plus efficace est de l'appliquer d'abord aux  $k$  tels que  $0 \notin \tilde{J}_{kk}$ . Cette quantité venant diviser le membre de droite, elle peut, si elle contient 0, aboutir à un gap dans le meilleur des cas, ou rendre l'étape sans intérêt si le numérateur contient 0 lui aussi (c.f. la section 5.2.1 dévolue à l'étape de Newton). En traitant d'abord les lignes  $k$  ne rencontrant pas ce problème, on diminue les chances d'obtenir 0 aussi au numérateur dans lequel certains  $\tilde{X}_k$  auront été remplacés par les  $\tilde{X}'_k$ , moins larges.

Enfin, l'apparition de gap ayant été mentionnée, le choix des auteurs est de les ignorer au cours de la procédure, mais de stocker l'information pour une utilisation ultérieure lors de l'étape de division de cellule. Ainsi, s'il apparaît que  $\tilde{Y}'_k = [-\infty, n_k] \cup [p_k, +\infty]$ ,  $n_k < 0 < p_k$ , ce qui se répercute sur une décomposition de  $\tilde{X}'_k$  en deux intervalles au plus :  $\tilde{X}'_k = [a_k, b_k] \cup [c_k, d_k]$ , on utilise dans la suite de la procédure  $\tilde{X}'_k = [a_k, d_k]$  et l'on stocke l'existence d'un gap  $[b_k, c_k]$  sur la  $k$ -ième composante de  $\tilde{X}$ .

#### 5.2.5 Point d'expansion

Le choix du point d'expansion durant la méthode de Newton Généralisée est important dans la mesure où plus la linéarisation du système se fait proche du zéro recherché, meilleure sera la correction. Il est donc intéressant de « perdre » un peu de temps de calcul à essayer de l'optimiser.

Si l'on se trouve dans le cas d'une première application de la méthode de Newton Généralisée, le choix arbitraire du centre de la cellule est retenu :

$$x_{exp} = mid(\tilde{X}) \quad (5.10)$$

En revanche, lorsqu'une première passe de type méthode de Newton Généralisée a déjà été appliquée à la cellule, on possède une matrice réelle  $P$ , correspondant à l'inverse de la matrice jacobienne centrée du problème linéarisé autour d'un point d'expansion  $\tilde{x}_{exp}$ . Si ce point est toujours dans la cellule  $\tilde{X}$  à l'étude, on peut le réutiliser directement ou tenter de l'optimiser en le corrigeant via l'équation linéarisée dans  $\mathbb{R}$  :

$$H(\tilde{x}'_{exp}) \approx H(\tilde{x}_{exp}) + (P)^{-1} . (\tilde{x}'_{exp} - \tilde{x}_{exp})$$

Cherchant  $\tilde{x}'_{exp}$  tel que  $H(\tilde{x}'_{exp}) = 0$ , on obtient :

$$\tilde{x}'_{exp} = \tilde{x}_{exp} - P.H(\tilde{x}_{exp}) \quad (5.11)$$

Si  $\|H(\tilde{x}'_{exp})\| < \|H(\tilde{x}_{exp})\|$ , ce qui est le but recherché, on réitère la procédure en remplaçant  $\tilde{x}_{exp}$  par  $\tilde{x}'_{exp}$ . Le choix est fait par les auteurs de ne réitérer cette procédure que 3 fois au plus

afin de ne pas non plus effectuer un trop grand nombre d'itérations - même peu coûteuses - sur un système qui n'est qu'une approximation grossière du problème véritable. Si, en revanche, le nouveau point  $\tilde{x}'_{exp}$  n'est pas « meilleur » que le précédent, la correction est arrêtée et c'est  $\tilde{x}_{exp}$  qui est gardé. Enfin, il faut s'assurer que le point retenu est dans  $\tilde{X}$ . En effet, s'il est issu de corrections, il peut être sorti de la cellule considérée. Le choix des auteurs est alors de remplacer toute composante  $\tilde{x}_{exp_i}$  se trouvant à l'extérieur de  $\tilde{X}_i$  par la borne de  $\tilde{X}_i$  la plus proche.

### 5.2.6 Méthode de Box Consistency Partiel

A la suite de la méthode de Newton Généralisée, on profite d'avoir le système préconditionné, dont normalement la diagonale de la jacobienne est dominante, pour appliquer une méthode de Box Consistency. Supposant cette dominance de la diagonale, seules les équations la faisant intervenir sont considérées, c'est-à-dire les équations considérant la variable  $\tilde{x}_i$  pour la composante  $H_i$ ,  $1 \leq i \leq P$ . On nomme donc cette méthode « Box Consistency Partiel ».

### 5.2.7 Division des cellules (splitting)

Lorsque les différentes méthodes n'ont plus une efficacité suffisante sur la réduction de la cellule, il faut se résoudre à subdiviser celle-ci en cellules plus petites. Une idée simple est de couper la cellule en deux dans chacune des directions. On obtient alors  $2^P$  cellules... ce qui rend vite cette pratique impossible, dû au trop grand nombre de sous-cellules générées. Les auteurs nous proposent de générer au plus 8 sous-cellules, obtenues en découpant selon les composantes les plus larges ou les plus influentes et en utilisant le fait que des gaps ont pu être repérés mais encore non exploités. Dissocions les démarches selon qu'une passe de la méthode de Newton Généralisée ait été effectuée ou non avant cette étape de division.

Supposons qu'aucune passe de cette méthode n'ait été effectuée au préalable. Seules les méthodes de Hull et de Box Consistency ont été appliquées à la cellule. Cela nous donne peu d'informations quant à l'importance de telle ou telle composante de  $\tilde{X}$  par rapport aux autres. On affecte donc un poids très simple à chaque composante qui est sa largeur,  $w(\tilde{X}_i)$ . On choisit ensuite de couper la cellule en deux le long de chacune des trois composantes les plus larges. On génère ainsi au plus 8 nouvelles cellules.

Examinons maintenant le cas où la méthode de Newton Généralisée a été appliquée à la cellule. On possède la matrice jacobienne  $J$  évaluée sur une cellule  $\tilde{X}^o$  contenant la cellule  $\tilde{X}$ . Il est intéressant d'utiliser cette matrice qui reflète l'influence locale de chaque composante sur l'évolution locale de  $H$ . Sur chaque composante  $H_i(\tilde{X})$ ,  $\tilde{X}_j$  a l'influence  $J_{ij}$ . La variation absolue engendrée sur  $H_i(\tilde{X})$  par  $\tilde{X}_j$  est donc  $\text{mag}(J_{ij}) \cdot w(\tilde{X}_j)$ , où  $\text{mag}(X) = \max(|\underline{X}|, |\overline{X}|)$  pour  $X$  intervalle de  $\mathbb{R}$ . Il semble donc intéressant d'affecter comme poids à chaque composante  $\tilde{X}_i$  la somme :

$$T_i = w(\tilde{X}_i) \sum_{j=1}^N \text{mag}(J_{ji}) \quad (5.12)$$

Il se peut aussi dans ce cas qu'il existe des gaps qui n'ont pas été encore exploités. Cependant, couper selon un gap n'est pas forcément plus intéressant que couper en deux selon une composante très large et influente. Les auteurs définissent le critère suivant pour décider de

l'intérêt d'un gap : supposons qu'un gap  $[n_i, p_i]$  ait été détecté sur la composante  $\tilde{X}_i$ . On ne considérera son exploitation pertinente que s'il satisfait :

$$\min(p_i - \underline{\tilde{X}}_i, \overline{\tilde{X}}_i - n_i) \geq 0.2w(\tilde{X}) \quad (5.13)$$

Les gaps satisfaisant à cette condition sont ordonnés selon les poids  $T_i$  décroissants précédemment calculés.

Ainsi, on procède dans l'ordre suivant pour les trois coupes (au plus) que l'on souhaite réaliser : d'abord selon les gaps « intéressants » puis, s'il existe moins de trois gaps, en coupant en deux l'intervalle  $\tilde{X}_i$ , par ordre de poids  $T_i$  décroissant.

### 5.2.8 Critères d'arrêt

Nous décrivons dans ce dernier paragraphe trois critères d'arrêt permettant de répondre à la question «  $\tilde{X}$  est-elle une cellule solution ? ». Ils sont dénommés Critère A, B ou C dans la description détaillée de l'annexe B.3 et correspondent pour les deux premiers à la question «  $\tilde{X}$  solution ? » de l'algorithme général présenté en figure 5.2. Le critère C vient ajouter une possibilité de réponse positive à cette question lorsqu'elle est posée à la suite de l'application d'une méthode de Newton Généralisée. Notons que lorsque la réponse donnée à cette question est positive, cela n'affirme pas qu'une solution existe au sein de la cellule mais que la cellule a de fortes chances (au vu de sa taille, de son image, etc) d'en contenir une.

#### Critère d'arrêt A : taille de la cellule

Le critère A est lié à la taille de la cellule  $\tilde{X}$  considérée. Il n'est pas facile de construire un critère universel adapté à tous les cas de figure et tous les ordres de grandeurs. Un critère qui s'adapte tout seul à l'ordre de grandeur de la quantité recherchée fait tendre vers un critère type erreur relative ; malheureusement, ce type de critère devient invérifiable si la solution recherchée est nulle. Nous avons donc fait le choix d'un critère « double » pour chaque composante de la cellule considérée, selon qu'elle contient ou non zéro :

$$\text{Critère A vérifié} \Leftrightarrow \forall i \in \llbracket 1, P \rrbracket, \begin{cases} \text{Si } 0 \in \tilde{X}_i, w(\tilde{X}_i) < \epsilon_{Xa} \\ \text{Si } 0 \notin \tilde{X}_i, w(\tilde{X}_i) < \epsilon_{Xr} \min(|\underline{\tilde{X}}_i|, |\overline{\tilde{X}}_i|) \end{cases} \quad (5.14)$$

Le premier cas impose que si 0 est contenu dans la composante  $i$ , alors, tous les éléments de l'intervalle  $\tilde{X}_i$  sont « quasiment nuls » : si ce n'est pas le cas, c'est que soit 0 n'est pas la solution et sera exclu de l'intervalle en raffinant la composante, soit 0 est bien la solution cherchée et les bornes de cet intervalle devront en être rapprochées.

Le deuxième cas impose une largeur relative de la cellule et peut être transformé de façon à montrer que pour  $\tilde{x}^*$  solution dans  $\tilde{X}$  :

$$\forall \tilde{x}_i \in \tilde{X}_i, \frac{|\tilde{x}_i - \tilde{x}_i^*|}{|\tilde{x}_i^*|} < \epsilon_{Xr} \quad (5.15)$$

On a donc une majoration de l'erreur relative commise en prenant n'importe quel point  $\tilde{x}_i$  de l'intervalle pour la solution exacte  $\tilde{x}_i^*$ .

### Critère d'arrêt B : magnitude de l'image

Le critère B vérifie quant à lui que l'image de la cellule, censée représenter un zéro donc d'image nulle, est suffisamment petite :

$$\text{Critère B vérifié} \Leftrightarrow \|H(\tilde{X})\| < \epsilon_H, \quad \text{avec } \|\tilde{Y}\| = \max_{1 \leq i \leq P} (\text{mag}(\tilde{Y}_i)) \quad (5.16)$$

Tous les éléments de l'intervalle image de la cellule  $\tilde{X}$  doivent donc être suffisamment proches de zéro pour que  $\tilde{X}$  soit considérée comme un bon encadrement du zéro recherché.

### Critère d'arrêt C : cellule irréductible

Ce dernier critère est lié à l'application d'une passe de la méthode de recherche du hull dans le cadre de la méthode de Newton Généralisée. Il se formule ainsi :

$$\text{Critère C vérifié} \Leftrightarrow \begin{cases} 0 \in H(\tilde{X}) \\ \tilde{J} \text{ régulière} \\ \tilde{X} \subset N \end{cases} \quad (5.17)$$

Les deux premières conditions sont déjà vérifiées lorsque l'on en arrive à ce critère. La troisième condition est à interpréter selon les notations utilisées en 5.2.4, au paragraphe décrivant la méthode de recherche du hull :  $\tilde{X}$  désigne la cellule avant application de la méthode de recherche du hull et  $N$  l'intervalle solution de  $\tilde{J} \cdot (N - \tilde{x}_{exp}) = r$ . Ainsi, si la troisième condition est vérifiée, c'est que l'on ne peut plus réduire la cellule. Elle constitue donc une bonne approximation de la solution cherchée.

## 5.3 Application à un oscillateur de Duffing

Étant donné que cette méthode traite l'espace à sonder sous forme de cellules comme la précédente, nous proposons la même démarche qu'au chapitre 4. Nous appliquons donc l'algorithme aux cas décrits par le tableau 4.1 avec les mêmes paramètres pour le critère d'arrêt :

$$\epsilon_{Xr} = 10^{-2}, \quad \epsilon_{Xa} = 10^{-3}, \quad \epsilon_H = 10^{-3}$$

et le même espace initial :

$$\Omega = [-2, 2]^P$$

### 5.3.1 Résultats

Les résultats de l'application de l'algorithme pour chaque cas sont donnés dans le tableau 5.1 (calculs réalisés sur un serveur Intel Core 2 Duo E6600 (4Go de Ram)). Les figures 5.5, 5.6 et 5.7 donnent une représentation graphique des trois premiers cas (1H, C1H et 1H3H). Nous avons adopté les mêmes symboles qu'en section 4.5.1 : un point vert pour le barycentre des centres des cellules d'un même groupe et un point rouge pour la valeur « exacte » obtenue après application d'un algorithme de résolution locale à l'équation  $H(\tilde{x}) = 0$  initialisé avec le barycentre.

Cas	P	Temps cpu (s)	Nombre de cellules traitées	Nombre de cellules retenues
1H	2	9	60	3
C1H	3	52	142	6
1H3H	4	1004	1362	3
C1H3H	5	2910	2267	6

TAB. 5.1 – Arithmétique des intervalles - Résultats

Seules trois cellules sont représentées en figure 5.6 alors que le tableau 5.1 en donne six. L'algorithme a en effet retenu deux fois chacune des cellules avec  $a_0 = 0$  à chaque fois. Ceci s'explique de la façon suivante : l'intervalle initial sur  $a_0$  est symétrique. Une première division selon cette direction donne donc des cellules dont la composante  $a_0$  varie dans les intervalles  $[-2, 0]$  et  $[0, 2]$ . Deux de ces cellules, notamment, peuvent s'écrire ainsi  $[-2, 0] \times \tilde{X}_1 \times \tilde{X}_2$  et  $[0, -2] \times \tilde{X}_1 \times \tilde{X}_2$ ; les intervalles pour  $a_1$  et  $b_1$  sont donc identiques. Une réduction sur ces deux cellules amène à trouver que la variation de  $a_0$  se réduit au point 0 pour chacune de ces deux cellules et aux mêmes intervalles finaux  $\tilde{X}'_1$  et  $\tilde{X}'_2$  pour les variables  $a_1$  et  $b_1$ . Le dédoublement observé provient donc du fait que la variation composante  $a_0$  soit réduite à un point et qui plus est un point constituant une borne inférieure pour une cellule et supérieure pour l'autre. Nous en profitons pour préciser qu'une fois cette composante réduite à un point pour une cellule, plus aucune division selon cette direction n'a été effectuée pour la cellule en question afin de ne pas générer des cellules totalement identiques mais traitées successivement et donc de consommer inutilement de la ressource numérique. C'est le même phénomène qui génère de nouveau six cellules au lieu de trois dans le cas C1H3H.

Ces figures montrent que la méthode est efficace étant donné le peu de différence entre le barycentre des cellules retenues et le point « vrai ».

### 5.3.2 Extrapolations

Nous allons maintenant réaliser le même type d'extrapolations que pour la méthode du simplexe. Comme on peut s'y attendre, l'évolution du nombre de cellules avec la dimension  $P$  évolue selon la loi  $N_c = a.b^P$ ,  $a = 4.03$  et  $b = 3.73$  (figure 5.8). On constate que le nombre de cellules traitées est plus faible pour  $P = 3$  et  $P = 5$ . Ceci est dû au fait que ces deux cas correspondent à l'adjonction de la composante  $a_0$ . L'algorithme va en effet réduire très vite les intervalles de cette composante à 0 ce qui permet de se ramener rapidement à des systèmes non pas à 3 et 5 inconnues mais à 2 et 4 inconnues seulement.

Nous tentons alors d'établir une loi sur le temps de calcul par cellule  $t_u = T/N_c$ . La méthode est basée soit sur l'évaluation de la jacobienne (méthode de Newton Généralisée) soit sur l'application  $P^2$  fois d'une méthode (telles les méthodes de Hull Consistency ou Box Consistency réitérant la même opération pour chaque équation  $H_i$ ,  $1 \leq i \leq P$  et chaque inconnue  $\tilde{x}_j$ ,  $1 \leq j \leq P$ ). Nous cherchons donc une loi quadratique  $t_u = c.P^2 + d.P + e$  et nous trouvons  $c = 0.0814$ ,  $d = -0.1896$  et  $e = 0.1963$  qui permettent un bon lissage des résultats (figure 5.9).

Ces deux lois nous permettent d'effectuer une extrapolation sur le temps de calcul  $T$  en

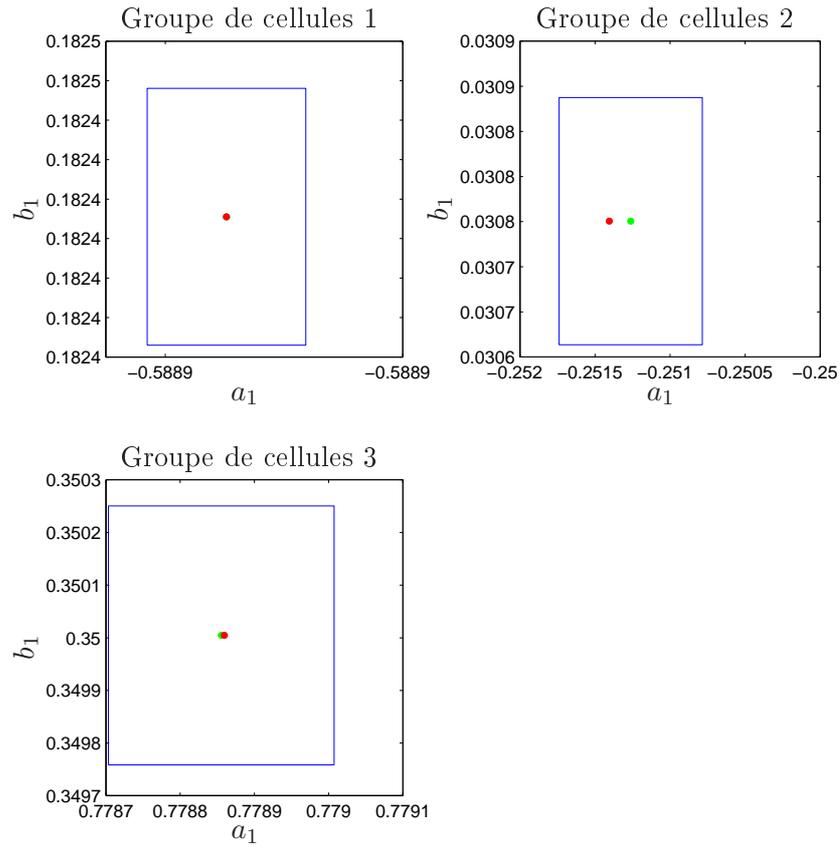


FIG. 5.5 – Arithmétique des intervalles - Cas 1H

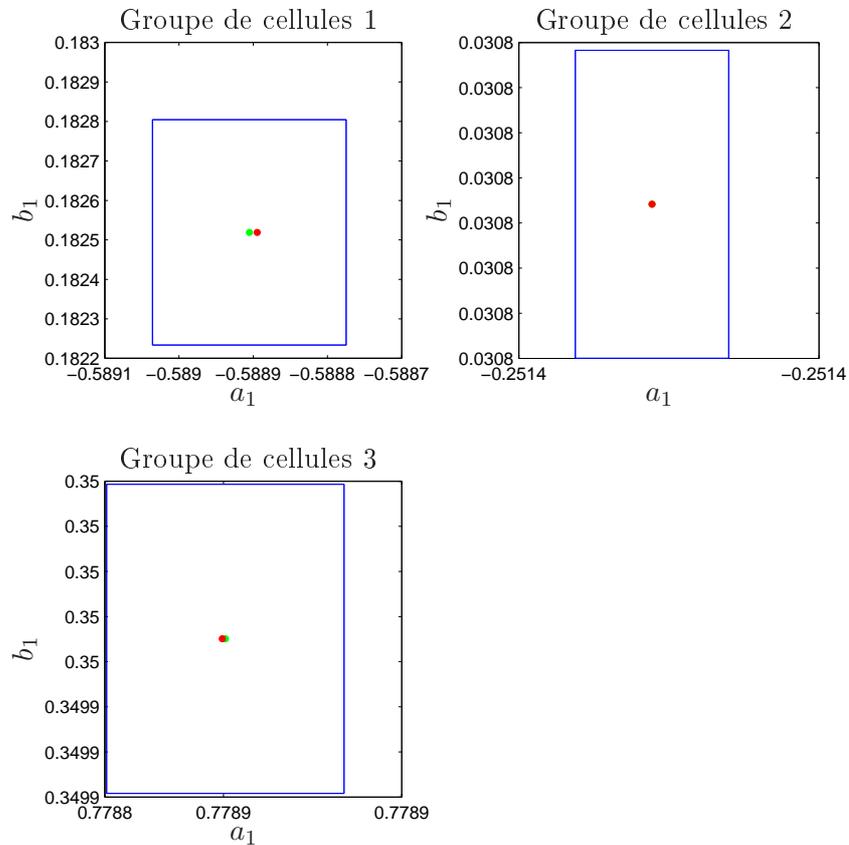


FIG. 5.6 – Arithmétique des intervalles - Cas C1H

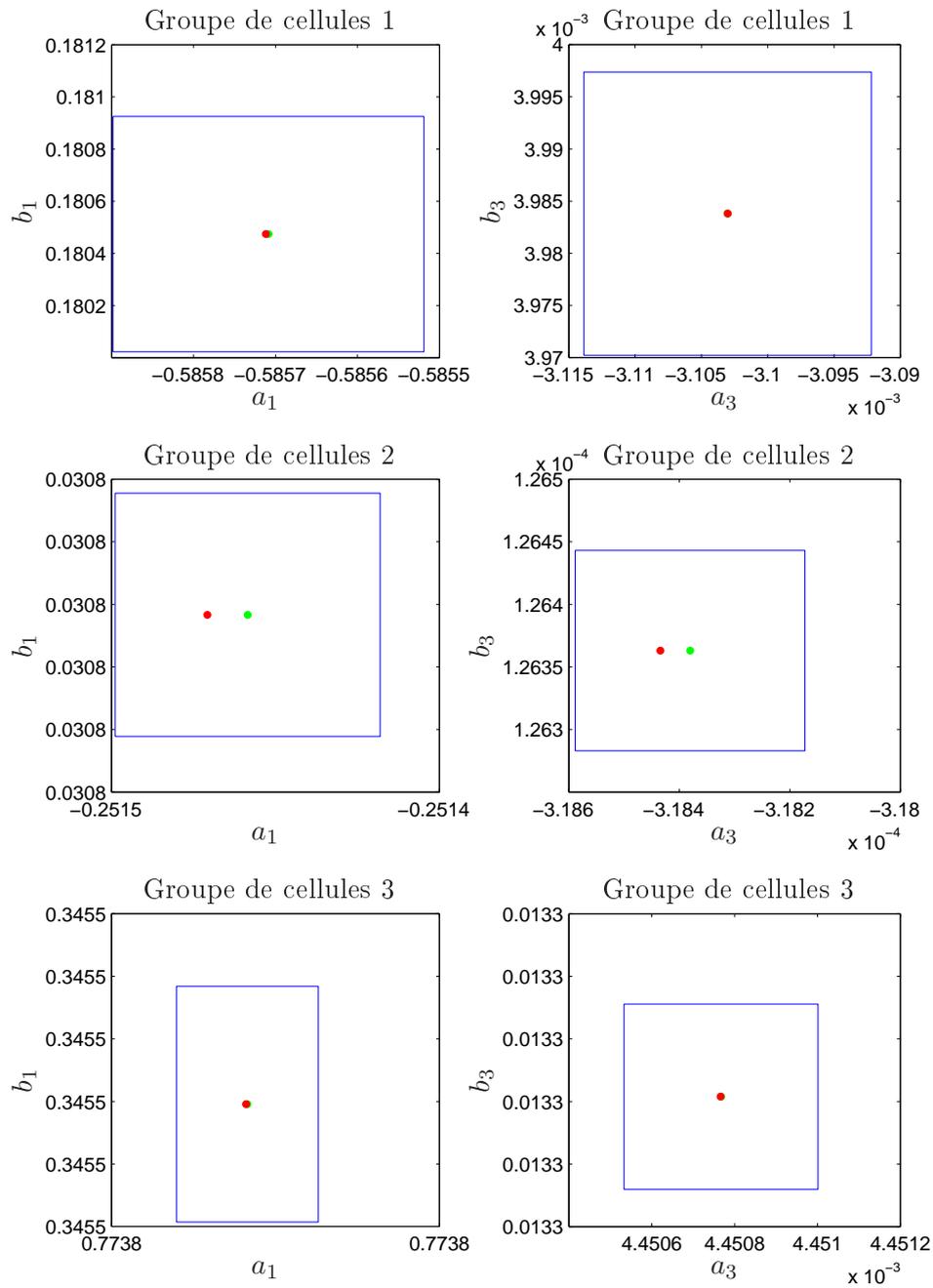


FIG. 5.7 – Arithmétique des intervalles - Cas 1H3H

fonction du nombre d'inconnues  $P$  :

$$T = (a.b^P).(a.P^2 + b.P + c)$$

représentée en figure 5.10. On voit donc qu'un temps de calcul d'une semaine environ est obtenu dès  $P = 8$ . Si cette méthode donne de meilleurs résultats que la précédente, il semble donc qu'elle soit elle aussi trop coûteuse pour être envisagée sur de gros systèmes.

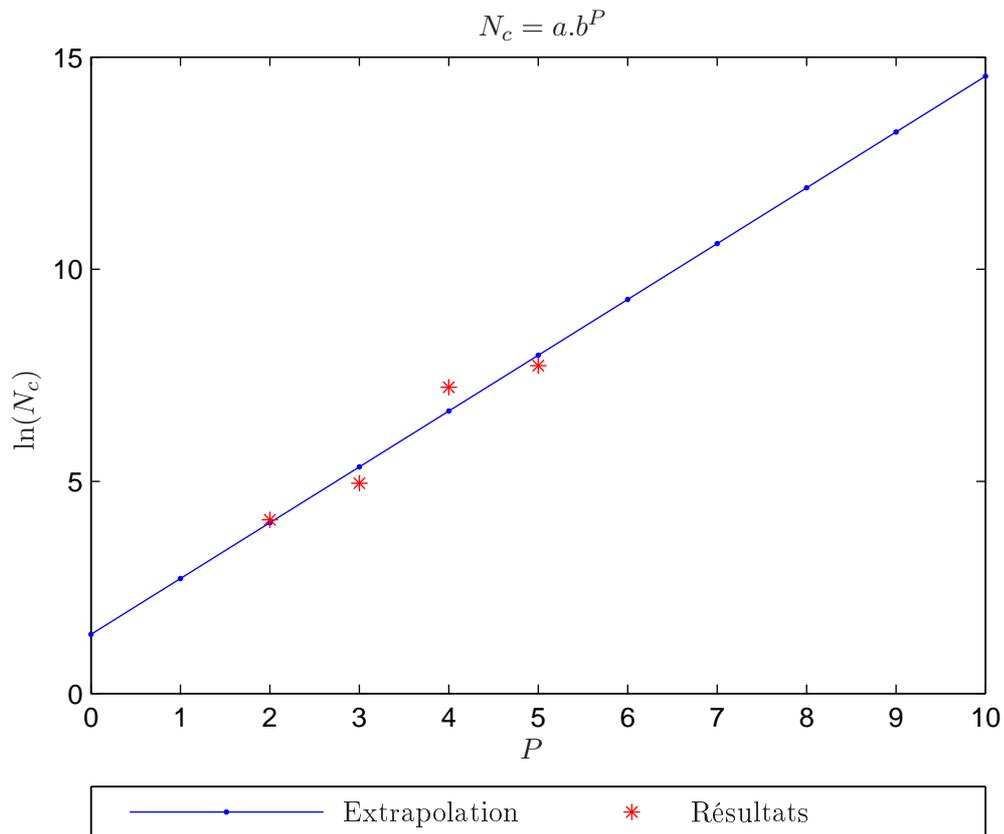


FIG. 5.8 – Arithmétique des intervalles - Extrapolation du nombre de cellules

## 5.4 Conclusion

L'utilisation d'une méthode élaborée pour traiter le domaine initial  $\Omega$  de variation des inconnues algébriques permet donc de gagner du temps par rapport aux méthodes d'exclusion mais n'offre toujours pas la possibilité de traiter des systèmes à plus de trois degrés de liberté.

D'une façon générale, le principe consistant à définir un domaine que l'on scrute à la recherche de solutions semble aboutir à des démarches trop coûteuses. En effet, les trois premières méthodes abordées, le cell-mapping qui cherche les solutions sur une partie bornée de l'espace d'états  $\mathcal{E}_{IC}$  ainsi que les méthodes reposant sur les tests d'exclusion ou l'analyse par intervalles qui cherchent les inconnues algébriques traduisant des solutions sur un domaine  $\Omega$ , ont présenté des résultats satisfaisants du point de vue de leur objectif d'analyse globale mais obtenus dans des délais trop importants dans le cas de systèmes à plus de quelques degrés de liberté. Nous proposons donc, dans un quatrième volet, une méthode qui ne procède

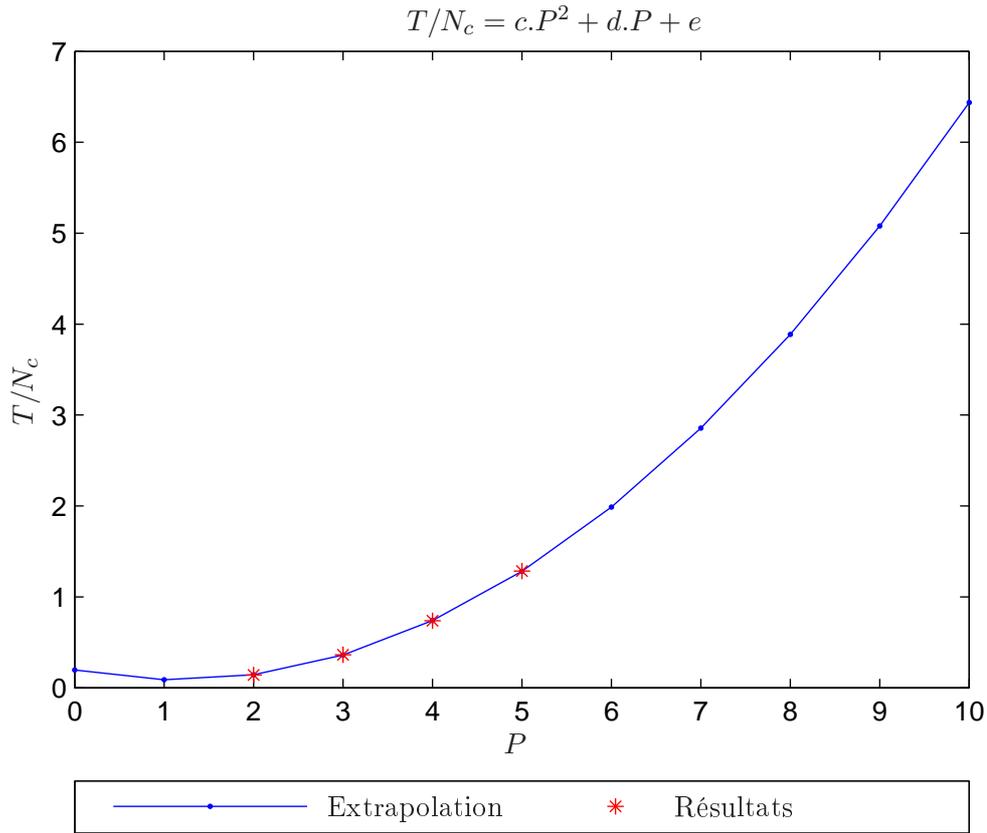


FIG. 5.9 – Arithmétique des intervalles - Extrapolation du temps par cellule

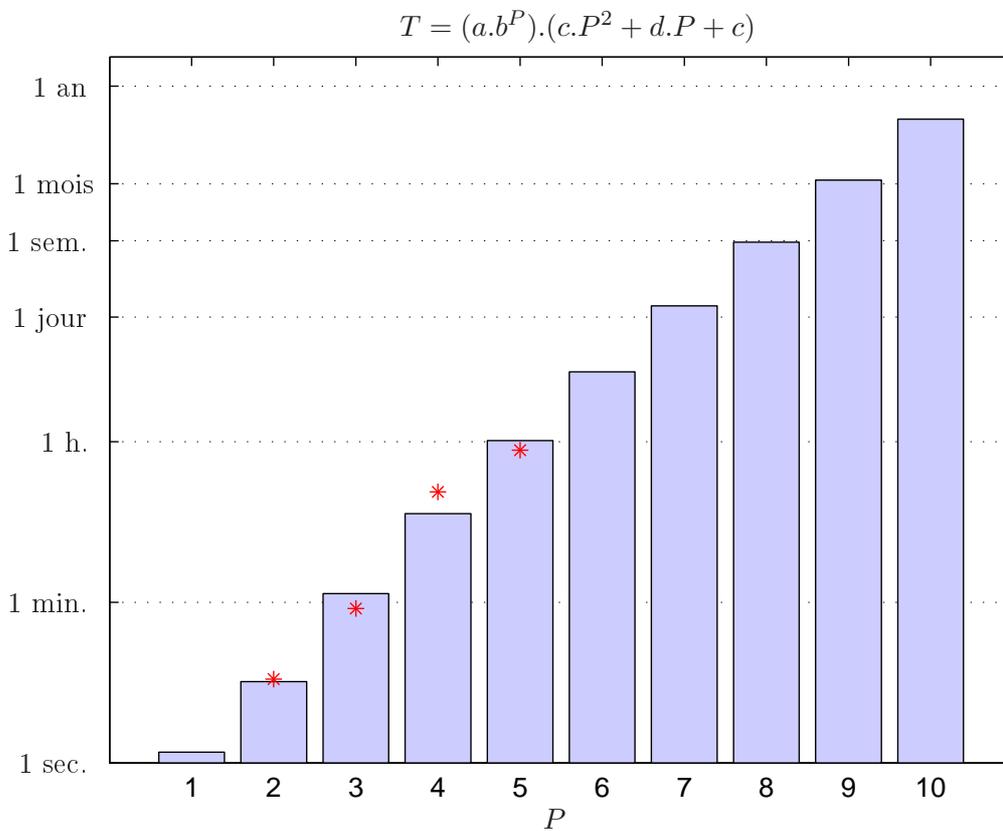


FIG. 5.10 – Arithmétique des intervalles - Extrapolation du temps de calcul

---

pas par analyse d'un espace continu préalablement borné. La contrepartie sera de faire une nouvelle hypothèse ou approximation sur le système traité, en l'occurrence, celle d'un système polynomial.



## Chapitre 6

# Analyse globale par homotopies polynomiales

---

*Les trois méthodes précédemment traitées ne font pas d'hypothèse sur la ou les non-linéarité(s) mise(s) en jeu dans le problème. Elles en font éventuellement sur la forme de la solution recherchée pour passer des équations dynamiques aux équations algébriques, mais non sur la forme mathématique de la non-linéarité. Cette absence de cadre de travail mathématiquement particularisé empêche de faire un travail analytique préliminaire qui permettrait de gagner du temps à la résolution. Nous avons donc eu l'idée d'approximer la partie non-linéaire des équations algébriques par des formes polynomiales afin de se donner une classe de problèmes aux propriétés mathématiques fortes sur laquelle travailler. Nous avons ensuite cherché dans la littérature les méthodes permettant de résoudre les systèmes ainsi obtenus, à savoir des systèmes multi-polynomiaux à plusieurs variables. Nous nous sommes alors aperçus que cette question était traitée assez largement depuis la fin des années 70 (les premiers travaux ont été publiés par Drexler en 1977 et Garcia et Zangwill en 1979). Une série d'articles, parus entre 1987 et 2002 ([Wri85, MS87b, MS87a, LSY87, Zul88, Wam92, Li97, Wat02]) permet notamment de suivre l'évolution des principales techniques mises en place pour résoudre ces problèmes. Enfin, un ouvrage de référence, rédigé par A.J. Sommese et C.W. Wampler II, [SW05], synthétise les résultats de ces publications et les intègre au contexte numérique.*

*Le principe général de la méthode imaginée est en premier lieu, d'approximer la partie non-linéaire des équations algébriques par une famille de polynômes. On obtient ainsi un système d'équations multi-polynomiales à plusieurs variables,  $\mathbf{P}(\tilde{x}) = 0$ . Pour résoudre ce système, on utilise ensuite une approche par homotopie qui consiste à partir d'un système*

multi-polynomial à plusieurs variables plus simple,  $\mathbf{Q}(\tilde{x}) = 0$ , que l'on sait résoudre, et à l'amener progressivement vers  $\mathbf{P}(\tilde{x}) = 0$  par une homotopie  $\mathbf{H}(\tilde{x}, \lambda) = \lambda\mathbf{Q}(\tilde{x}) + (1 - \lambda)\mathbf{P}(\tilde{x}) = 0$ , où l'on fait varier progressivement  $\lambda$  de 1 à 0. La difficulté majeure réside dans le choix du polynôme  $\mathbf{Q}(\tilde{x})$  : il faut le choisir suffisamment proche de  $\mathbf{P}(\tilde{x})$  structurellement de façon à pouvoir passer continûment et sans bifurcation des solutions de  $\mathbf{Q}(\tilde{x}) = 0$  à celles de  $\mathbf{P}(\tilde{x}) = 0$ , mais suffisamment simple pour pouvoir en évaluer les solutions.

Nous exposerons tout d'abord les points théoriques essentiels qui sous-tendent ces méthodes puis nous présenterons les différents choix possibles de polynômes  $\mathbf{Q}(\tilde{x})$  et ceux qui nous semblent les plus appropriés aux types de systèmes que nous traitons.

Afin de lever toute ambiguïté ultérieure, nous donnons dès à présent la définition d'un polynôme à plusieurs variables dans  $\mathcal{C}$  et les notations associées :

### Définition 6.1 : Polynôme à plusieurs variables

Une fonction  $\mathbf{P} : \mathcal{C}^P \rightarrow \mathcal{C}$  de  $P$  variables  $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_P)$  est un polynôme s'il peut s'exprimer comme une somme de termes, chaque terme étant le produit d'un coefficient et d'un monôme, chaque coefficient étant un nombre complexe et chaque monôme étant un produit des variables élevées à une puissance entière positive ou nulle. En formulant ceci avec une notation multi-degré et en notant  $\alpha = (\alpha_1, \dots, \alpha_P) \in \mathbb{N}^P$ , on écrit les monômes de la façon suivante :

$$\tilde{x}^\alpha = \prod_{i=1}^P \tilde{x}_i^{\alpha_i}$$

Alors, un polynôme  $\mathbf{P}$  est une fonction qui peut s'écrire :

$$\begin{aligned} \mathbf{P} : \quad \mathcal{C}^P &\rightarrow \mathcal{C} \\ \tilde{x} &\mapsto \mathbf{P}(\tilde{x}) = \sum_{\alpha \in \mathcal{I}} a_\alpha \tilde{x}^\alpha \end{aligned} \tag{6.1}$$

où  $\mathcal{I}$  est un ensemble d'index fini et  $a_\alpha \in \mathcal{C}$ . La notation  $\mathbf{P} \in \mathcal{C}[\tilde{x}_1, \dots, \tilde{x}_P] = \mathcal{C}[\tilde{x}]$ , signifie que  $\mathbf{P}$  est un polynôme des variables  $\tilde{x}_i$  avec ses coefficients dans  $\mathcal{C}$ .

Le degré total d'un monôme  $\tilde{x}^\alpha$  est  $|\alpha| = \alpha_1 + \dots + \alpha_P$  et le degré total du polynôme  $\mathbf{P}$  vaut :

$$\deg(\mathbf{P}) = \max_{\alpha \in \mathcal{I}/a_\alpha \neq 0} |\alpha|$$

Quand aucune confusion n'est possible, on parle simplement de degré du polynôme et non de son degré total.  $\square$

## 6.1 Passage d'un système algébrique non-linéaire à un système multi-polynomial

Pour passer d'une équation algébrique de la forme (2.10) à une équation multi-polynômiale de la forme (6.2),

$$\mathbf{P}(\tilde{x}) = 0, \quad \mathbf{P}(\tilde{x}) = \left\{ \begin{array}{l} \mathbf{P}_1(\tilde{x}_1, \dots, \tilde{x}_P) = \sum_{\alpha^1 \in \mathcal{I}_1} a_{1\alpha^1} \tilde{x}^{\alpha^1} \\ \vdots \\ \mathbf{P}_P(\tilde{x}_1, \dots, \tilde{x}_P) = \sum_{\alpha^P \in \mathcal{I}_P} a_{P\alpha^P} \tilde{x}^{\alpha^P} \end{array} \right\} \quad (6.2)$$

deux possibilités s'offrent à nous : soit la non-linéarité présente dans l'équation algébrique se développe naturellement bien polynomialement et il est intéressant de faire ce développement analytiquement (comme ce sera le cas du Duffing que nous traiterons en section 6.6), soit la non-linéarité n'a pas véritablement une forme polynomiale qui la sous-tend et l'on réalise alors un lissage de la non-linéarité sur une famille choisie de polynômes. Le premier cas est transparent et ne nécessite pas plus d'explication ; nous donnons pour le second les rudiments permettant d'effectuer l'approximation polynomiale de la partie non linéaire de l'équation.

Dans la cas où une écriture « à la main » de la non-linéarité sous forme polynomiale n'est pas aisée, l'objectif est d'approximer « au mieux » la partie non-linéaire  $\hat{H}(\tilde{x})$  par une forme polynomiale  $\mathbf{P}_{\mathbf{nl}}(\tilde{x})$  pour un espace de variation de  $\tilde{x}$ , noté  $\mathcal{D}$ , généralement borné dans les problèmes d'ingénierie qui nous concernent. Un moyen simple de déterminer  $\mathbf{P}_{\mathbf{nl}}$  est d'utiliser une approximation aux moindres carrés qui s'écrit comme suit :

- on se donne  $n$  points  $\tilde{x}^{(j)} \in \mathcal{D}$  pour lesquels on calcule  $b^{(j)} = \hat{H}(\tilde{x}^{(j)}) \in \mathbb{R}^P$ ,
- on se donne pour chaque polynôme  $\mathbf{P}_{\mathbf{nl}i}$  les monômes qui le composeront (choix des  $\alpha^i$ ),
- puis, pour chaque composante  $\mathbf{P}_{\mathbf{nl}i}$ , on écrit que l'on cherche à minimiser la quantité dépendant des  $a_{i\alpha^i}$  :

$$r^{(i)}(a_{i\alpha^i}) = \sum_{j=1}^n (\mathbf{P}_{\mathbf{nl}i}(\tilde{x}^{(j)}) - b_i^{(j)})^2 = \sum_{j=1}^n \left[ \left( \sum_{k=1}^{|\mathcal{I}_i|} a_{i\alpha^i, k} (\tilde{x}^{(j)})^{\alpha^i, k} \right) - b_i^{(j)} \right]^2 \geq 0$$

On écrit donc que la dérivée par rapport à chacun des coefficients  $a_{i\alpha^i, k}$  (au nombre de  $|\mathcal{I}_i|$ ) est nulle, ce qui nous donne le système carré suivant pour chacun des  $\mathbf{P}_{\mathbf{nl}i}$  :

$$R^{(i)} \begin{Bmatrix} a_{i\alpha^i, 1} \\ \vdots \\ a_{i\alpha^i, |\mathcal{I}_i|} \end{Bmatrix} = \begin{Bmatrix} \sum_{j=1}^n b_i^{(j)} (\tilde{x}^{(j)})^{\alpha^i, 1} \\ \vdots \\ \sum_{j=1}^n b_i^{(j)} (\tilde{x}^{(j)})^{\alpha^i, |\mathcal{I}_i|} \end{Bmatrix},$$

$$R^{(i)} \in \mathcal{M}_{|\mathcal{I}_i|}(\mathbb{R}), \quad R_{pq}^{(i)} = \sum_{j=1}^n (\tilde{x}^{(j)})^{\alpha^i, p} (\tilde{x}^{(j)})^{\alpha^i, q}$$

Une fois que l'on possède l'approximation polynomiale de la partie non-linéaire, il ne reste plus qu'à ajouter la partie linéaire  $\bar{H}$  et le terme constant  $H_c$  provenant généralement de

l'excitation pour obtenir le polynôme final :

$$\mathbf{P}(\tilde{x}) = \begin{cases} -H_{c1} + \sum_{k=1}^P \bar{H}_{1k} \tilde{x}_k + \mathbf{P}_{nl1}(\tilde{x}) \\ \vdots \\ -H_{cP} + \sum_{k=1}^P \bar{H}_{Pk} \tilde{x}_k + \mathbf{P}_{nlP}(\tilde{x}) \end{cases}$$

## 6.2 Nature des solutions d'un système multi-polynomial

Lorsque l'on résout un système de la forme (6.2), on le fait généralement dans  $\mathcal{C}$ . En effet, il est plus simple de s'appuyer sur le corps des complexes plutôt que sur celui des réels (tout polynôme est scindé dans  $\mathcal{C}$ , non dans  $\mathbb{R}$ ). Il nous faudra donc extraire dans un second temps les valeurs réelles qui seules, généralement, nous intéresseront.

Par ailleurs, un système multi-polynomial peut avoir différents types de solutions : des points (isolés), mais aussi des courbes, des surfaces, etc. On imagine aisément qu'il est difficile numériquement de décrire des espaces de dimension non-nulle (non réduits à des points) et qu'il faut faire appel à des techniques relativement sophistiquées pour y parvenir (pour plus de détails, on peut se reporter au dernier chapitre de [SW05] et consulter les articles [SV00] et [SVW03]). Les solutions de dimension non-nulle n'ayant généralement pas de pertinence physique, nous ne nous y intéresserons pas dans la suite de notre étude.

Nous chercherons donc uniquement les solutions de type points isolés dont nous extrairons les éléments réels.

## 6.3 Principe et difficultés d'une résolution par méthode homotopique

Le principe général de cette méthode, comme énoncé dans l'introduction de ce chapitre, est de partir d'un polynôme  $\mathbf{Q}(\tilde{x})$  que l'on sait résoudre et de le faire évoluer continûment vers le polynôme dont on cherche les zéros en résolvant un système de la forme (6.3) pour des valeurs de  $\lambda$  allant de 1 vers 0. On réalise donc une continuation des solutions du système (6.3).

$$\mathbf{H}(\tilde{x}, \lambda) = 0, \quad \mathbf{H}(\tilde{x}, \lambda) = \lambda \mathbf{Q}(\tilde{x}) + (1 - \lambda) \mathbf{P}(\tilde{x}) \quad (6.3)$$

Pour que la méthode soit efficace, chaque zéro  $\tilde{x}^{*(j)}$  de  $\mathbf{P}(\tilde{x})$  doit être relié par un chemin continu à un zéro de  $\mathbf{Q}(\tilde{x})$ . Il faut donc choisir  $\mathbf{Q}(\tilde{x})$  avec au moins autant de zéros que  $\mathbf{P}(\tilde{x})$ . La continuité requise du chemin, impose quant à elle qu'il soit structurellement proche de  $\mathbf{P}(\tilde{x})$  afin d'éviter toute bifurcation le long du chemin suivi. La question de l'élaboration de  $\mathbf{Q}(\tilde{x})$  est traitée en section 6.5.

Par ailleurs, il est fréquent que l'on réalise une surestimation du nombre de zéros de  $\mathbf{P}(\tilde{x})$ , donc que  $\mathbf{Q}(\tilde{x})$  possède des zéros « en trop ». Les chemins partant de ces zéros divergent alors en norme vers  $+\infty$  lorsque  $\lambda$  tend vers 0. Cette divergence étant difficile à traiter numériquement, on a recours à un artefact : on homogénéise le polynôme  $\mathbf{P}(\tilde{x})$  à l'aide d'une variable supplémentaire  $\tilde{x}_0$  pour se placer dans l'espace projectif. Nous donnons le détail de la procédure dans la section 6.4.

La figure 6.1 illustre le principe de cette méthode homotopique pour un cas trivial où le système non-linéaire serait à une seule inconnue ( $P = 1$ ). On peut lire sur cette figure que le choix du polynôme  $\mathbf{Q}(\tilde{x})$  est tel qu'il possède deux zéros  $\tilde{x}^{(1)}$  et  $\tilde{x}^{(2)}$ . Le chemin partant du premier conduit à l'unique zéro de  $\mathbf{P}(\tilde{x})$ ,  $\tilde{x}^*$  tandis que l'autre, que l'on peut qualifier a posteriori d'« inutile », diverge vers  $+\infty$  lorsque  $\lambda$  tend vers 0.

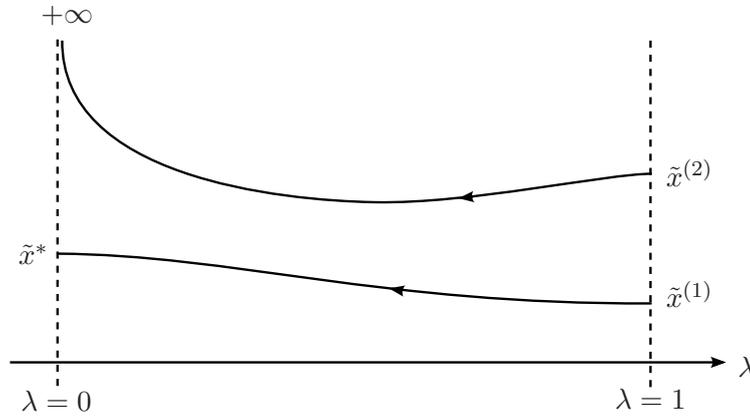


FIG. 6.1 – Illustration du principe de la méthode homotopique

## 6.4 Espace projectif et homogénéisation

Afin de parer au problème des chemins divergeant lorsque  $\lambda$  tend vers 0, on peut remarquer que si un polynôme  $\mathbf{Q}(\tilde{x})$  a tous ses monômes d'un même degré  $d$ , alors pour tout  $\beta \in \mathcal{C}$ ,  $\mathbf{Q}(\beta\tilde{x}) = \beta^d \mathbf{Q}(\tilde{x})$ . On dit d'un tel polynôme qu'il est *homogène*. Pour une telle catégorie de polynômes, on voit que si  $\tilde{x}^*$  est un zéro de  $\mathbf{Q}$ , alors  $\forall \beta \in \mathcal{C}$ ,  $\beta\tilde{x}^*$  en est un aussi. Ainsi, si la norme de  $\tilde{x}^*$  est trop grande pour permettre un calcul numérique assez précis, on peut considérer à la place  $\beta\tilde{x}^*$  en prenant par exemple  $\beta = 1/\|\tilde{x}^*\|$  pour ramener le zéro considéré à une norme de 1. On sent qu'il y a là une notion de classe d'équivalence et que l'on choisit finalement le représentant de cette classe qui nous convient le mieux numériquement.

Le fait est que pour la plupart des polynômes, tous les monômes n'ont pas même degré. Pour pouvoir utiliser malgré cela la remarque précédente on réalise alors l'*homogénéisation* du polynôme  $\mathbf{P}(\tilde{x})$  considéré en ajoutant une  $P + 1^{\text{ème}}$  variable  $\tilde{x}_0$  par laquelle on vient multiplier avec un degré adéquat chaque monôme constituant  $\mathbf{P}(\tilde{x})$  de façon à obtenir un degré identique de tous les monômes (constitués alors des variables  $\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_P$  regroupées dans le vecteur  $\tilde{x}_h$  et non plus seulement  $\tilde{x}_1, \dots, \tilde{x}_P$ ). Donnons deux exemples pour bien illustrer cette technique :

– Exemple 1 :  $\mathbf{P}(\tilde{x}_1, \tilde{x}_2) = 2\tilde{x}_1^2 + \tilde{x}_2$

Le premier monôme est de degré 2 quand le second est de degré 1. Le polynôme  $\mathbf{P}$  homogénéisé correspondant est  $\mathbf{P}_h(\tilde{x}_0, \tilde{x}_1, \tilde{x}_2) = 2\tilde{x}_1^2 + \tilde{x}_0\tilde{x}_2$ . Ainsi les deux monômes sont de degré 2.

– Exemple 2 :  $\mathbf{P}(\tilde{x}_1, \tilde{x}_2) = 2\tilde{x}_1^2 + 3\tilde{x}_1\tilde{x}_2 + 4$

Les deux premiers monômes sont de degré 2 et le dernier est de degré 0. Le polynôme  $\mathbf{P}$  homogénéisé correspondant est  $\mathbf{P}_h(\tilde{x}_0, \tilde{x}_1, \tilde{x}_2) = 2\tilde{x}_1^2 + 3\tilde{x}_1\tilde{x}_2 + 4\tilde{x}_0^2$ .

En formalisant la remarque d'introduction, on aboutit à la définition de l'espace projectif  $\mathbb{P}^P$  déduit de  $\mathcal{C}^{P+1}$  et la relation d'équivalence « pressentie » permettant de choisir le meilleur représentant numérique :

**Définition 6.2 : Espace projectif  $\mathbb{P}^P$**

L'espace projectif complexe de dimension  $P$  noté  $\mathbb{P}^P$  est l'espace des lignes complexes passant par l'origine dans  $\mathcal{C}^{P+1}$ . Les points de  $\mathbb{P}^P$  sont des  $P + 1$ -uplets  $[\tilde{x}] = (\tilde{x}_0, \dots, \tilde{x}_P)$  tels que  $\forall i \in \llbracket 0, P \rrbracket$ ,  $\tilde{x}_i \in \mathcal{C}$  et non tous nuls, avec la relation d'équivalence suivante :

$$[\tilde{x}] = (\tilde{x}_0, \dots, \tilde{x}_P) \sim [\tilde{x}'] = (\tilde{x}'_0, \dots, \tilde{x}'_P) \text{ si et seulement si } \exists \beta \in \mathcal{C}^* / (\tilde{x}'_0, \dots, \tilde{x}'_P) = (\beta\tilde{x}_0, \dots, \beta\tilde{x}_P)$$

D'une façon plus générale, pour  $\mathbb{E}$  un espace vectoriel sur un corps commutatif  $\mathbb{K}$  et en notant  $G$  la relation d'équivalence précédemment définie, l'ensemble quotient  $\mathbb{E}^*/G$  est appelé espace projectif déduit de  $\mathbb{E}$ .  $\square$

Ainsi, si l'on cherche numériquement les zéros d'un polynôme  $\mathbf{P}(\tilde{x}_1, \dots, \tilde{x}_P)$ , et que l'on prend la peine de le rendre homogène de degré  $d$  (notons le polynôme obtenu  $\mathbf{P}_h(\tilde{x}_0, \dots, \tilde{x}_P)$ ), il est plus avantageux de mener la recherche sur  $\mathbb{P}^P$  en faisant les deux remarques suivantes :

$$\forall \tilde{x} \in \mathcal{C}^P, \mathbf{P}(\tilde{x}_1, \dots, \tilde{x}_P) = \mathbf{P}_h(1, \tilde{x}_1, \dots, \tilde{x}_P) \quad (6.4)$$

$$\forall \tilde{x}^\circ \in \mathcal{C}^{P+1} / \mathbf{P}_h(\tilde{x}^\circ) = 0, \forall \beta \in \mathcal{C}, \mathbf{P}_h(\beta\tilde{x}^\circ) = \beta^d \mathbf{P}_h(\tilde{x}^\circ) = 0 \quad (6.5)$$

La seconde remarque se traduit de la façon suivante : si  $\tilde{x}^\circ$  est solution de  $\mathbf{P}_h(\tilde{x}_h) = 0$ , alors tout élément équivalent - au sens de la relation d'équivalence donnée dans la définition d'un espace projectif - l'est aussi. On peut donc chercher, en ce sens, les zéros non égaux à l'élément nul de  $\mathbf{P}_h$  dans  $\mathbb{P}^P$  en choisissant alors le représentant de la classe d'équivalence qui convient le mieux numériquement. La première remarque, triviale, permet quant à elle de relier les zéros  $\tilde{x}^*$  de  $\mathbf{P}$  et ceux  $\tilde{x}^\circ$  du polynôme homogène associé  $\mathbf{P}_h$  pourvu que leur première coordonnée soit non nulle :

$$\forall \tilde{x}^\circ = (\tilde{x}_0^\circ, \dots, \tilde{x}_P^\circ) \in \mathcal{C}^{P+1} / \mathbf{P}_h(\tilde{x}^\circ) = 0 \text{ et } \tilde{x}_0^\circ \neq 0, \tilde{x}^* = \frac{1}{\tilde{x}_0^\circ}(\tilde{x}_1^\circ, \dots, \tilde{x}_P^\circ) \text{ est tel que } \mathbf{P}(\tilde{x}^*) = 0$$

et réciproquement

$$\forall \tilde{x}^* = (\tilde{x}_1^*, \dots, \tilde{x}_P^*) \in \mathcal{C}^P / \mathbf{P}(\tilde{x}^*) = 0, \tilde{x}^\circ = (1, \tilde{x}_1^*, \dots, \tilde{x}_P^*) \text{ est tel que } \mathbf{P}_h(\tilde{x}^\circ) = 0 \quad (6.6)$$

Nous avons considéré jusqu'ici des polynômes  $\mathbf{P}$  à valeur dans  $\mathcal{C}$ ; lorsque l'on considère - comme c'est le cas dans les applications qui nous intéressent - un système multi-polynomial  $\mathbf{P}$  à valeurs dans  $\mathcal{C}^P$ , on homogénéise séparément chaque composante  $\mathbf{P}_i$  dont on amène tous les monômes à un degré  $d_i$ . Si l'on considère par exemple le système à deux variables suivant :

$$\mathbf{P}(\tilde{x}) = \left\{ \begin{array}{l} \mathbf{P}_1(\tilde{x}_1, \tilde{x}_2) = \tilde{x}_1^3 + \tilde{x}_1\tilde{x}_2 + \tilde{x}_2^2 \\ \mathbf{P}_2(\tilde{x}) = \tilde{x}_1 + \tilde{x}_2 - 2 \end{array} \right\}$$

on obtient après homogénéisation :

$$\mathbf{P}_h(\tilde{x}_h) = \left\{ \begin{array}{l} \mathbf{P}_{1h}(\tilde{x}_0, \tilde{x}_1, \tilde{x}_2) = \tilde{x}_1^3 + \tilde{x}_0\tilde{x}_1\tilde{x}_2 + \tilde{x}_0\tilde{x}_2^2 \\ \mathbf{P}_{2h}(\tilde{x}_0, \tilde{x}_1, \tilde{x}_2) = \tilde{x}_1 + \tilde{x}_2 - 2\tilde{x}_0 \end{array} \right\}$$

Le degré de la première composante est  $d_1 = 3$  tandis que celui de la deuxième est  $d_2 = 1$ . Ainsi on peut écrire que :

$$\forall \beta \in \mathcal{C}, \mathbf{P}_h(\beta\tilde{x}_h) = \left\{ \begin{array}{l} \beta^{d_1}\mathbf{P}_{1h}(\tilde{x}_h) \\ \beta^{d_2}\mathbf{P}_{2h}(\tilde{x}_h) \end{array} \right\}$$

De la même façon, les équations et relations (6.4), (6.5) et (6.6) se transposent aisément lorsque l'on passe d'un simple polynôme à un système multi-polynomial.

Remettons maintenant cette technique dans le contexte qui nous intéresse, à savoir celui de la résolution du système multi-polynomial à  $P$  variables  $\mathbf{P}(\tilde{x}_1, \dots, \tilde{x}_P) = 0$  par une homotopie  $\mathbf{H}(\tilde{x}, \lambda)$  partant du système  $\mathbf{Q}(\tilde{x})$  pour  $\lambda = 1$  et allant vers  $\mathbf{P}(\tilde{x})$  pour  $\lambda = 0$ . Conformément à ce que nous venons d'expliquer, on ne travaille pas directement avec  $\mathbf{P}(\tilde{x})$  mais avec son équivalent homogénéisé à  $P + 1$  variables. On modifiera donc  $\mathbf{Q}(\tilde{x})$  conformément à cette homogénéisation en lui adjoignant la variable supplémentaire  $\tilde{x}_0$ . Ainsi, on effectue la continuation des solutions du système suivant :

$$\begin{aligned} \mathbf{H}_h(\tilde{x}_h, \lambda) &= 0, \\ \mathbf{H}_h(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_P, \lambda) &= \lambda\mathbf{Q}_h(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_P) + (1 - \lambda)\mathbf{P}_h(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_P) \end{aligned} \quad (6.7)$$

On est tenté d'objecter que si l'on a ajouté une inconnue, on n'a pas simultanément introduit une équation. Il faut alors se rappeler ce pour quoi nous avons introduit la variable  $\tilde{x}_0$  : dans le but de ramener la norme de l'inconnue  $\tilde{x}$  à une grandeur permettant des calculs numériques satisfaisants. Une équation supplémentaire est donc implicitement introduite et s'explique de différentes façons selon le choix que l'on préfère :

$$\max_{i=0, \dots, P} |\tilde{x}_i| = 1 \quad (6.8a)$$

$$\langle \tilde{x}_h, z \rangle = 1, \text{ où } z \in \mathcal{C} \text{ est un vecteur choisi de façon adéquate} \quad (6.8b)$$

Ainsi, lors du processus de continuation, on procède comme indiqué dans le tableau suivant selon la contrainte choisie :

Contrainte (6.8a)	Contrainte (6.8b)
Point connu au rang $n$ : $(\tilde{x}_h^{(n)}, \lambda^{(n)})$ tel que	
$\ \mathbf{H}_h(\tilde{x}_h^{(n)}, \lambda^{(n)})\  < \epsilon$ $\max_{i=0, \dots, P}  \tilde{x}_i^{(n)}  =  \tilde{x}_j^{(n)}  = 1$	$\ \mathbf{H}_h(\tilde{x}_h^{(n)}, \lambda^{(n)})\  < \epsilon$ $\langle \tilde{x}_h^{(n)}, z \rangle = 1$
Prédiction/Correction pour obtenir un point $(\tilde{x}_h^{(n+1)}, \lambda^{(n+1)})$ au rang $n + 1$	
en imposant $\tilde{x}_j^{(n+1)} = \tilde{x}_j^{(n)}$ donc en bloquant une variable	sur le système augmenté : $\begin{cases} \mathbf{H}_h(\tilde{x}_h, \lambda) = 0 \\ \langle \tilde{x}_h, z \rangle = 1 \end{cases}$
Choix du meilleur représentant numérique pour $\tilde{x}_h^{(n+1)}$ dans $\mathbb{P}^P$	
en remplaçant $\tilde{x}_h^{(n+1)}$ par l'élément équivalent $(\max_{i=0, \dots, P}  \tilde{x}_i^{(n+1)} )^{-1} \tilde{x}_h^{(n+1)}$	en modifiant éventuellement $z$ dans le cas où $\tilde{x}_h^{(n+1)}$ et $z$ seraient proches de l'orthogonalité : ceci entraîne alors des coordonnées $\tilde{x}_i^{(n+1)}$ très grandes en norme et n'est pas souhaitable ; on génère donc par exemple aléatoirement un nouveau vecteur $z$ de $\mathcal{C}^{P+1}$ puis on remplace $\tilde{x}_h^{(n+1)}$ par l'élément équivalent $(\langle \tilde{x}_h^{(n+1)}, z \rangle)^{-1} \tilde{x}_h^{(n+1)}$

On peut donc contenir la norme des vecteurs considérés au long de la continuation et l'on détecte aisément les points partant à l'infini lorsque  $\lambda$  tend vers 0 en faisant la remarque suivante : si dans  $\mathcal{C}^P$   $\|(\tilde{x}_1, \dots, \tilde{x}_P)\| \rightarrow +\infty$  lorsque  $\lambda \rightarrow 0$ , le procédé d'homogénéisation et l'utilisation qui en est préconisée vont nous amener à contenir cette norme notamment en faisant tendre  $\tilde{x}_0$  vers 0 : schématiquement, si l'on choisit la contrainte (6.8a) et que  $|\tilde{x}_j| \rightarrow 0$  lorsque  $\lambda \rightarrow 0$  ( $j \in \llbracket 1, P \rrbracket$ ), le procédé proposé imposant une norme égale à 1 pour  $\tilde{x}_j$ , on divisera  $\tilde{x}_0$  par  $\tilde{x}_j$ , donc on aura  $\tilde{x}_0 \rightarrow 0$ . Le passage de  $\mathcal{C}^P$  à  $\mathbf{P}^P$  via le processus d'homogénéisation permet donc de faciliter le suivi numérique des valeurs et de détecter en fin de continuation les points allant à l'infini qui sont alors caractérisés par :

$$\frac{|\tilde{x}_0|}{\max_{i=1, \dots, P} |\tilde{x}_i|} < \epsilon_d \quad (6.9)$$

où  $\epsilon_d$  est choisi « petit » (typiquement, entre  $10^{-3}$  et  $10^{-6}$ ). Par souci d'alléger les notations, on notera identiquement par la suite le système polynomial initial  $\mathbf{P}$  et le système homogénéisé  $\mathbf{P}_h$  lorsque la distinction n'est pas essentielle, réservant toutefois l'indice 0 à la variable d'homogénéisation.

## 6.5 Choix d'un polynôme initial $\mathbf{Q}(\tilde{x})$

Comme évoqué en § 6.3, la principale difficulté dans la mise en œuvre efficace d'une méthode de résolution par homotopie est le choix du polynôme  $\mathbf{Q}(\tilde{x})$  de départ. Il faut en effet être capable de le résoudre pour pouvoir initialiser la continuation, donc qu'il soit d'une structure relativement simple, tout en étant structurellement proche du polynôme visé  $\mathbf{P}(\tilde{x})$  de

façon à ce que la continuation ne comporte pas de bifurcation. Cette « proximité » structurelle consiste en premier lieu en le fait de posséder au moins autant de zéros que  $\mathbf{P}(\tilde{x})$  de façon à ce que chaque zéro de  $\mathbf{P}(\tilde{x})$  (comptés distinctement s'ils sont d'un ordre de multiplicité supérieur à 1) soit relié à un unique zéro de  $\mathbf{Q}(\tilde{x})$ . Il faut pour cela pouvoir évaluer le nombre de zéros de  $\mathbf{P}(\tilde{x})$  avant même de les connaître et ceci se fait en examinant plus ou moins finement la structure de  $\mathbf{P}(\tilde{x})$ . Selon la finesse de l'analyse, on obtient un polynôme  $\mathbf{Q}(\tilde{x})$  qui possède un nombre de zéros de plus en plus proche de celui de  $\mathbf{P}(\tilde{x})$  (et toujours supérieur). Si chacun des zéros surnuméraires crée un chemin qui diverge en l'infini, et donc un investissement en calcul inutile, l'analyse préliminaire peut devenir elle-même-coûteuse et difficilement automatisable. Le graphique 6.5 (issu de [SW05, Sec 8.1]) présente de façon hiérarchisée, les noms des différentes méthodes actuellement dans la littérature. Étant donnée la diversité des systèmes que nous souhaitons pouvoir traiter, nous n'expliquons dans les paragraphes suivants que les trois premières dont la mise en œuvre peut s'automatiser aisément. Dans un dernier paragraphe, nous donnerons quelques préconisations et éléments chiffrés pour aider au choix de la méthode.

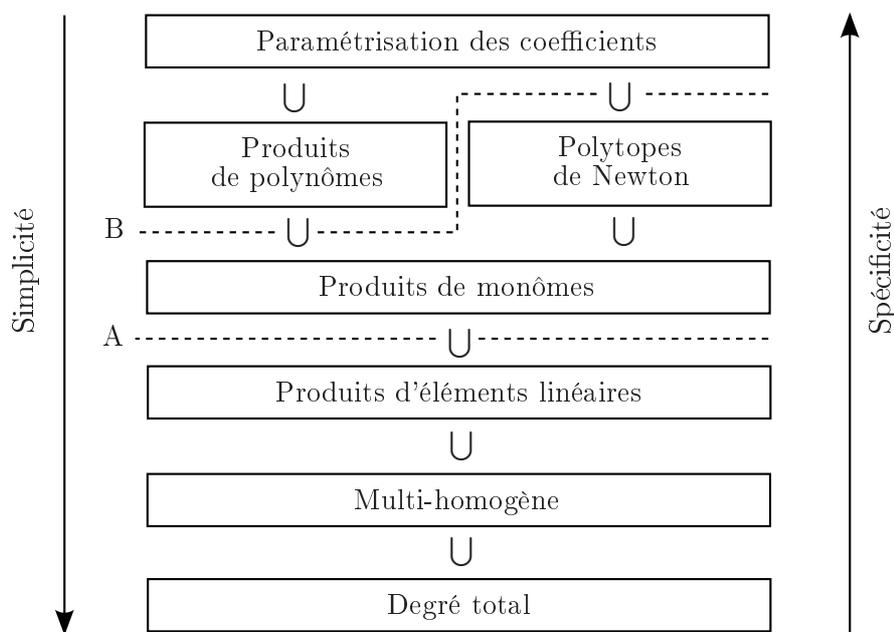


FIG. 6.2 – Graphique des différentes méthodes permettant d'élaborer  $\mathbf{Q}(\tilde{x})$

Sous la ligne A, seules des techniques de résolution de systèmes linéaires sont nécessaires.

Au dessus de la ligne B, des méthodes spéciales doivent être conçues au cas par cas.

### 6.5.1 Méthode du degré total

Une façon certaine et simple de tenir compte de tous les zéros de  $\mathbf{P}(\tilde{x})$  est de s'appuyer sur le théorème de Bezout qui nous dit que pour un système  $\mathbf{P}(\tilde{x})$  dont chaque composante  $\mathbf{P}_i(\tilde{x})$  est un polynôme de degré total  $d_i$ , le nombre maximal de solutions est  $\mathcal{N} = d_1 \times \dots \times d_P$ . Il

est alors facile de concevoir l'utilisation du polynôme initial  $\mathbf{Q}(\tilde{x})$  suivant :

$$\mathbf{Q}(\tilde{x}) = \begin{pmatrix} \tilde{x}_1^{d_1} - 1 \\ \vdots \\ \tilde{x}_P^{d_P} - 1 \end{pmatrix}$$

dont les racines sont bien connues puisque sont celles de l'unité. Le théorème suivant permet alors de garantir que l'on trouvera toutes les solutions de  $\mathbf{P}(\tilde{x})$  :

### **Théorème 6.1 : Homotopie de degré total**

Soit un système multi-polynomial  $\mathbf{P}(\tilde{x})$  de  $\mathcal{C}^P$  dans  $\mathcal{C}^P$  et  $d_i$  le degré de chacune de ses composantes  $\mathbf{P}_i(\tilde{x})$ ,  $i = 1, \dots, P$ . Soit  $\mathbf{Q}(\tilde{x})$  un système multi-polynomial dont chacune des composantes  $\mathbf{Q}_i(\tilde{x})$  est de degré  $d_i$  et possédant  $\mathcal{N} = \prod_{i=1}^P d_i$  solutions non-singulières. Alors les  $\mathcal{N}$  chemins solutions de l'homotopie

$$\mathbf{H}(\tilde{x}, \lambda) = \lambda \gamma \mathbf{Q}(\tilde{x}) + (1 - \lambda) \mathbf{P}(\tilde{x}) = 0$$

prenant source en les solutions de  $\mathbf{Q}(\tilde{x}) = 0$  sont non-singuliers pour  $\lambda \in ]0, 1]$  et l'ensemble de leurs points finaux lorsque  $\lambda \rightarrow 0$  inclut toutes les solutions non-singulières de  $\mathbf{P}(\tilde{x}) = 0$  pour presque tout  $\gamma \in \mathcal{C}$ , excepté en un nombre fini de droites de dimension réelle égale à 1 passant par l'origine. En particulier, si l'on choisit  $\gamma = e^{i\theta}$  sur le cercle unité, ces exceptions se limitent à un nombre fini de points  $\theta \in [0, 2\pi]$ .  $\square$

Si cette méthode est très facile à mettre en œuvre, on sent bien que le calcul de  $\mathcal{N} = \prod_{i=1}^P d_i$  chemins sera souvent inutilement coûteux dans la mesure où peu des systèmes considérés posséderont en réalité autant de solutions. En particulier pour ce qui nous concerne, les variables  $\tilde{x}_i$  se séparent en deux groupes de tailles inégales : celles relatives aux ddl linéaires, donc participant uniquement à des monômes de degré 1 et celles relatives à des ddl non-linéaires, en faible nombre par rapport aux précédentes, mais pouvant participer à des monômes de degré élevé. Cela semblerait donc dommageable de ne pas pouvoir tenir compte de cette distinction évidente et c'est pourquoi nous énonçons les deux méthodes suivantes.

### **6.5.2 Méthode par multi-homogénéisation**

Pour tenir compte de groupes de variables distincts - disons  $m$  groupes - une possibilité est d'utiliser un procédé de multi-homogénéisation : de la même façon qu'expliqué en § 6.4, on introduit des variables visant à amener tous les monômes d'une composante  $\mathbf{P}_i(\tilde{x})$  à un même degré  $d_i$ , mais au lieu d'introduire une seule variable supplémentaire  $\tilde{x}_0$ , on en introduit autant que de groupes constitués, c'est-à-dire  $m$ , notées  $\tilde{z}_1, \dots, \tilde{z}_m$ . On partitionne donc les  $P$  variables initiales  $\tilde{x}_1, \dots, \tilde{x}_P$  en  $m$  groupes  $\tilde{X}_j$  possédant  $k_j$  éléments ( $j = 1, \dots, m$ ,  $k_1 + \dots + k_m = P$ ), puis on associe à chaque groupe une variable d'homogénéisation  $\tilde{z}_j$ . On homogénéise ensuite chacun des polynômes  $\mathbf{P}_i$  conformément à cette multi-homogénéisation et on forme une matrice de degrés  $D = (d_{ij})_{1 \leq i \leq P, 1 \leq j \leq m}$  dont le terme  $d_{ij}$  correspond au degré maximal observé dans le polynôme  $\mathbf{P}_i$  pour les monômes formés par les variables du groupe  $\tilde{X}_j$ .

Considérons le système suivant pour illustrer cette méthode de multi-homogénéisation :

$$\mathbf{P}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = \left\{ \begin{array}{l} \tilde{x}_1^2 + 2 \\ \tilde{x}_2\tilde{x}_3 - 4 \\ \tilde{x}_1\tilde{x}_2 + \tilde{x}_3 - 6 \end{array} \right\}$$

On voit que seule la variable  $\tilde{x}_1$  intervient à un degré 2. On choisit donc de faire les deux groupes suivants auxquels on associe les variables d'homogénéisation  $\tilde{z}_1$  et  $\tilde{z}_2$  :

$$\begin{aligned} \tilde{X}_1 &= \{\tilde{x}_1\} \quad (k_1 = 1) \leftarrow \tilde{z}_1 \\ \tilde{X}_2 &= \{\tilde{x}_2, \tilde{x}_3\} \quad (k_2 = 2) \leftarrow \tilde{z}_2 \end{aligned}$$

On effectue alors la multi-homogénéisation suivante en remplaçant chaque variable  $\tilde{x}_i$  du groupe  $\tilde{X}_j$ ,  $j = 1, \dots, m$  par  $\tilde{x}_i/\tilde{z}_j$ ,  $i = 1, \dots, k_j$  puis en ramenant les dénominateurs de chaque  $\mathbf{P}_i$  à 1 (donc en ramenant aux numérateurs des monômes de la forme  $\prod_j \tilde{z}_j^{\alpha_{ij}}$  ; on obtient ainsi :

$$\mathbf{P}_h(\tilde{z}_1, \tilde{x}_1, \tilde{z}_2, \tilde{x}_2, \tilde{x}_3) = \left\{ \begin{array}{l} \tilde{x}_1^2 + 2\tilde{z}_1^2 \\ \tilde{x}_2\tilde{x}_3 - 4\tilde{z}_2^2 \\ \tilde{x}_1\tilde{x}_2 + \tilde{x}_3\tilde{z}_1 - 6\tilde{z}_1\tilde{z}_2 \end{array} \right\} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$$

Pour une homogénéisation classique introduisant uniquement la variable  $\tilde{x}_0$ , on aurait eu :

$$\mathbf{P}_h(\tilde{x}_0, \tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = \left\{ \begin{array}{l} \tilde{x}_1^2 + 2\tilde{x}_0^2 \\ \tilde{x}_2\tilde{x}_3 - 4\tilde{x}_0^2 \\ \tilde{x}_1\tilde{x}_2 + \tilde{x}_3\tilde{x}_0 - 6\tilde{x}_0^2 \end{array} \right\} \quad D = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

Une fois le système multi-homogénéisé, on peut utiliser le théorème de Bezout pour structures multi-homogènes qui donne que le nombre de zéros d'un tel système est au plus  $\text{Bez}(D, K)$  avec la formule suivante permettant de l'évaluer (une autre formule basée sur une approche combinatoire est possible ; nous lui préférons celle-ci, plus simple à énoncer) :

$$\text{Bez}(D, K) = \text{coeff} \left( \tilde{z}_1^{k_1} \dots \tilde{z}_m^{k_m}, \prod_{i=1}^P \sum_{j=1}^m d_{ij} \tilde{z}_j \right)$$

où l'expression  $\text{coeff}(\tilde{x}^\alpha, \mathbf{P}(\tilde{x}))$  désigne le coefficient du monôme  $\tilde{x}^\alpha$  dans le polynôme  $\mathbf{P}(\tilde{x})$ . Pour notre exemple, on voit donc que le nombre maximal de zéros à rechercher est :

$$\text{Bez}(D, K) = \text{coeff}(\tilde{z}_1^1 \tilde{z}_2^2, (2\tilde{z}_1 + 0\tilde{z}_2)(0\tilde{z}_1 + 2\tilde{z}_2)(\tilde{z}_1 + \tilde{z}_2)) = 4$$

alors qu'utilisant une homogénéisation classique, le majorant du nombre de zéros serait :

$$\mathcal{N} = 2 \times 2 \times 2 = 8$$

En réalisant une multi-homogénéisation, on divise donc par deux, sur cet exemple, le nombre de chemins à suivre.

Enfin, pour respecter la structure du polynôme initial  $\mathbf{P}(\tilde{x})$  conformément à cette multi-homogénéisation, il suffit de construire un polynôme  $\mathbf{Q}(\tilde{x})$  ayant même matrice de degré  $D$  et que l'on sait résoudre. Pour notre exemple, nous pouvons donc proposer le polynôme suivant :

$$\mathbf{Q}(\tilde{x}) = \left\{ \begin{array}{l} \tilde{x}_1^2 - 1 \\ (\tilde{x}_2 - 2)(\tilde{x}_3 + 3) \\ (\tilde{x}_1 + 4)(\tilde{x}_2 + \tilde{x}_3 - 1) \end{array} \right\}, \text{ de solutions } \begin{array}{l} \tilde{x}^{*(1)} = (1, 2, -1), \quad \tilde{x}^{*(2)} = (1, 4, -3), \\ \tilde{x}^{*(3)} = (-1, 2, -1), \quad \tilde{x}^{*(4)} = (-1, 4, -3) \end{array}$$

Le théorème énoncé ci-dessous nous permet de conclure en garantissant l'obtention de toutes les solutions isolées et non-sigulières de  $\mathbf{P}(\tilde{x})$ .

**Théorème 6.2 : Homotopie de multi-homogénéisation**

Soit un système multi-polynomial  $\mathbf{P}(\tilde{x})$  de  $\mathcal{C}^P$  dans  $\mathcal{C}^P$  possédant une matrice de degré  $D$  pour ses variables partitionnées en sous-ensembles de tailles  $K$  comme présenté précédemment. Soit  $\mathbf{Q}(\tilde{x})$  un système multi-polynomial ayant même matrice de degré  $D$  et possédant  $\text{Bez}(D, K)$  solutions non-singulières. Alors les  $\text{Bez}(D, K)$  chemins solutions de l'homotopie

$$\mathbf{H}(\tilde{x}, \lambda) = \lambda \gamma \mathbf{Q}(\tilde{x}) + (1 - \lambda) \mathbf{P}(\tilde{x}) = 0$$

prenant source en les solutions de  $\mathbf{Q}(\tilde{x}) = 0$  sont non-singuliers pour  $\lambda \in ]0, 1]$  et l'ensemble de leurs points finaux lorsque  $\lambda \rightarrow 0$  inclut toutes les solutions non-singulières de  $\mathbf{P}(\tilde{x}) = 0$  pour presque tout  $\gamma \in \mathcal{C}$ , excepté en un nombre fini de droites de dimension réelle égale à 1 passant par l'origine. En particulier, si l'on choisit  $\gamma = e^{i\theta}$  sur le cercle unité, ces exceptions se limitent à un nombre fini de points  $\theta \in [0, 2\pi]$ .  $\square$

En dernier lieu, nous proposons une formule « simple » permettant d'évaluer  $\text{Bez}(D, K)$  pour les systèmes qui nous intéressent : des structures majoritairement linéaires auxquelles on ajoute quelques ddl non-linéaires représentant un organe, un contact, etc. Nous proposons les notations suivantes :

- $P$  le nombre total d'inconnues  $\tilde{x}_i$ , égal au nombre d'équations
- $P_l$  le nombre d'équations linéaires (supposées ne dépendre que des  $V_l$  variables linéaires)
- $P_{nl}$  le nombre d'équations non-linéaires (pouvant dépendre de toutes les variables) :  
 $P_l + P_{nl} = P$
- $V_l$  le nombre de variables  $\tilde{x}_i$  n'intervenant que linéairement dans les équations (ne participant qu'à des monômes de degré 1)
- $V_{nl}$  le nombre de variables  $\tilde{x}_i$  intervenant non-linéairement dans les équations (partici-  
pant à au moins un monôme de degré supérieur ou égal à 2) :  $V_l + V_{nl} = P$
- $d$  le degré maximal des monômes contenant des variables non-linéaires

Si l'on ordonne par suite les  $\tilde{x}_i$  en faisant apparaître d'abord les  $V_l$  variables purement linéaires puis les  $V_{nl}$  variables non-linéaires et de même en écrivant d'abord les  $P_l$  équations purement linéaires et les  $V_{nl}$  autres ensuite, on peut former les deux groupes :

$$\tilde{X}_1 = \{\tilde{x}_1, \dots, \tilde{x}_{V_l}\}, \quad k_1 = V_l, \quad \text{homogénéisé par } \tilde{z}_1$$

$$\tilde{X}_2 = \{\tilde{x}_{V_l+1}, \dots, \tilde{x}_P\}, \quad k_2 = V_{nl}, \quad \text{homogénéisé par } \tilde{z}_2$$

La réorganisation des équations permet ensuite d'écrire une forme générale du polynôme homogénéisé :

$$\mathbf{P}_h(\tilde{z}_1, \tilde{x}_1, \dots, \tilde{x}_{V_l}, \tilde{z}_2, \tilde{x}_{V_l+1}, \dots, \tilde{x}_P) = \left\{ \begin{array}{l} \text{Eq linéaire 1} \\ \vdots \\ \text{Eq linéaire } P_l \\ \text{Eq non-linéaire 1} \\ \vdots \\ \text{Eq non-linéaire } P_{nl} \end{array} \right\}, \quad D = \begin{bmatrix} 1 & 0 \\ \vdots & \\ 1 & 0 \\ 1 & d \\ \vdots & \\ 1 & d \end{bmatrix}$$

On écrit donc facilement par suite que :

$$\begin{aligned}
\text{Bez}(D, K) &= \text{coeff} \left( \tilde{z}_1^{V_l} \tilde{z}_2^{V_{nl}}, \prod_{i=1}^{P_l} \tilde{z}_1 \times \prod_{j=1}^{P_{nl}} (\tilde{z}_1 + d\tilde{z}_2) \right) \\
&= \text{coeff} \left( \tilde{z}_1^{V_l} \tilde{z}_2^{V_{nl}}, \tilde{z}_1^{P_l} \sum_{j=0}^{P_{nl}} C_{P_{nl}}^j \tilde{z}_1^j (d\tilde{z}_2)^{P_{nl}-j} \right) \\
&= C_{P_{nl}}^{P_{nl}-V_{nl}} d^{V_{nl}} \\
&= \frac{P_{nl}!}{(P_{nl} - V_{nl})! V_{nl}!} d^{V_{nl}}
\end{aligned}$$

Cette formule n'est pas universelle dans la mesure où tous les systèmes non-linéaires que nous pouvons rencontrer ne sont pas exactement de cette forme : certaines équations linéaires peuvent faire intervenir des variables non-linéaires et certaines équations non-linéaires ne font intervenir aucune variable linéaire. Néanmoins, cette formule permet de donner rapidement une idée du nombre de chemins sans même avoir posé précisément les équations.

### 6.5.3 Méthode par produits d'éléments linéaires

Nous venons d'étudier un moyen d'approcher mieux la structure du système à résoudre  $\mathbf{P}(\tilde{x})$  en regroupant des variables. Le regroupement compte alors pour toutes les équations ; on peut donc envisager comme étape suivante des groupements différents pour chaque équation en gardant à l'esprit la contrainte de savoir résoudre le système multi-polynomial  $\mathbf{Q}(\tilde{x})$  qui résulte de l'étude. La méthode proposée consiste en la décomposition de chaque équation en un produit d'éléments linéaires. Il faut donc pour chaque équation reconstituer les espaces vectoriels de monômes de degré 1 dont le produit aboutit à l'espace vectoriel auquel appartient la solution. Afin de rendre ceci plus clair, nous posons quelques notations :

- $\langle e_1, \dots, e_k \rangle$  dénote le  $\mathcal{C}$ -espace vectoriel engendré par les éléments  $e_1, \dots, e_k$ .
- $\{e_1, \dots, e_k\} \otimes \{f_1, \dots, f_l\} = \{e_i \otimes f_j, 1 \leq i \leq k, 1 \leq j \leq l\}$  désigne l'ensemble produit des deux autres ensembles
- $\langle E \rangle \times \langle F \rangle = \{ef | e \in \langle E \rangle, f \in \langle F \rangle\}$  ensemble produit des espaces vectoriels  $\langle E \rangle$  et  $\langle F \rangle$  est inclus dans l'espace vectoriel engendré par le produit des ensembles  $E$  et  $F$ ,  $\langle E \otimes F \rangle$ .
- $E^{(n)}$  et  $\langle E \rangle^{(n)}$  désignent naturellement les ensembles produits respectifs  $E \otimes \dots \otimes E$  et  $\langle E \rangle \times \dots \times \langle E \rangle$  chacun comportant  $n$  facteurs identiques.

Pour un système multi-polynomial  $\mathbf{P}(\tilde{x}) = 0$  à  $P$  inconnues et  $P$  équations  $\mathbf{P}_i(\tilde{x}) = 0$ ,  $1 \leq i \leq P$ , on cherche à décomposer le  $\mathcal{C}$ -espace vectoriel  $\mathbb{E}_i$  auquel appartient  $\mathbf{P}_i(\tilde{x})$  de degré  $d_i$  en un produit de  $d_i$  ensembles  $E_{ij}$  ne contenant que des polynômes de degré 1 :

$$\begin{aligned}
\mathbf{P}_i(\tilde{x}) \in \mathbb{E}_i, \mathbb{E}_i &= \langle E_{i1} \otimes \dots \otimes E_{id_i} \rangle, 1 \leq i \leq P, \\
\forall i \in [1, P], E_{ij} &\subset \{\tilde{x}_1, \dots, \tilde{x}_P, 1\}, 1 \leq j \leq d_i
\end{aligned} \tag{6.10}$$

On peut alors montrer ([SW05, Sec. 8.3.4]) qu'il suffit de choisir le polynôme initial  $\mathbf{Q}(\tilde{x})$  dans l'ensemble  $E_i$  produit des espaces vectoriels  $\mathbb{E}_{ij}$  engendrés par les ensembles  $E_{ij}$  :

$$\forall i \in [1, P], \mathbf{Q}_i(\tilde{x}) \in E_i = \langle E_{i1} \rangle \times \dots \times \langle E_{id_i} \rangle = \mathbb{E}_{i1} \times \dots \times \mathbb{E}_{id_i} \tag{6.11}$$

Chaque composante  $\mathbf{Q}_i(\tilde{x})$  du système initial s'écrit donc comme un produit de polynômes  $p_{ij}(\tilde{x}) \in \mathbb{E}_{ij}$  de degré 1 au plus, d'où le nom de la méthode « produits d'éléments linéaires » :

$$\mathbf{Q}(\tilde{x}) = \left\{ \begin{array}{c} p_{11}(\tilde{x}) \dots p_{1d_1}(\tilde{x}) \\ \vdots \\ p_{P1}(\tilde{x}) \dots p_{Pd_P}(\tilde{x}) \end{array} \right\}, \forall i \in \llbracket 1, P \rrbracket, \forall j \in \llbracket 1, d_i \rrbracket, \deg(p_{ij}(\tilde{x})) \leq 1$$

Il est alors facile de résoudre le système initial : chaque zéro doit annuler au moins l'un des facteurs linéaires de chaque composante. Il suffit donc de tester tous les systèmes linéaires obtenus en choisissant un facteur linéaire par équation. Comme on cherche uniquement les solutions non-singulières, on peut avant de résoudre chaque système linéaire en calculer le rang et éliminer d'office les systèmes ayant un rang non égal à  $P$ .

Examinons maintenant les exemples suivants pour illustrer cette décomposition en produits d'éléments linéaires :

- Exemple 1 : soit  $\mathbf{P}(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = \tilde{x}_1^2 + 4\tilde{x}_1\tilde{x}_2 - \tilde{x}_3$ . Le degré total de ce polynôme est 2 et tous ses monômes appartiennent à l'ensemble produit  $\{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3\} \otimes \{\tilde{x}_1, 1\} = \{\tilde{x}_1^2, \tilde{x}_1\tilde{x}_2, \tilde{x}_1\tilde{x}_3, \tilde{x}_1, \tilde{x}_2, \tilde{x}_3\}$ . En termes de structure, ce polynôme appartient donc à l'espace vectoriel  $\langle \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3\} \otimes \{\tilde{x}_1, 1\} \rangle$ .
- Exemple 2 : soit le système multi-polynomial et les ensembles  $E_i$  correspondants

$$\mathbf{P}(\tilde{x}_1, \dots, \tilde{x}_6) = \left\{ \begin{array}{c} \tilde{x}_1 + \tilde{x}_2 - 3 \\ \tilde{x}_3 + \tilde{x}_4 - 5 \\ \tilde{x}_2 + \tilde{x}_5 + \tilde{x}_5^3 - 2 \\ \tilde{x}_3 + \tilde{x}_6 + \tilde{x}_6^3 - 1 \\ \tilde{x}_3 + \tilde{x}_5 \\ \tilde{x}_4 + \tilde{x}_6 \end{array} \right\}, \left\{ \begin{array}{l} E_1 = \langle \tilde{x}_1, \tilde{x}_2, 1 \rangle \\ E_2 = \langle \tilde{x}_3, \tilde{x}_4, 1 \rangle \\ E_3 = \langle \tilde{x}_2, \tilde{x}_5, 1 \rangle \times \langle \tilde{x}_5, 1 \rangle^{(2)} \\ E_4 = \langle \tilde{x}_3, \tilde{x}_6, 1 \rangle \times \langle \tilde{x}_6, 1 \rangle^{(2)} \\ E_5 = \langle \tilde{x}_3, \tilde{x}_5 \rangle \\ E_6 = \langle \tilde{x}_4, \tilde{x}_6 \rangle \end{array} \right.$$

On peut proposer d'après les  $E_i$  le polynôme  $Q(\tilde{x})$  suivant :

$$\mathbf{Q}(\tilde{x}) = \left\{ \begin{array}{c} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_2 + \tilde{x}_5 + 1)(\tilde{x}_5 - 1)(\tilde{x}_5 + 2) \\ (\tilde{x}_3 + \tilde{x}_6 + 1)(\tilde{x}_6 - 3)(\tilde{x}_6 + 5) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\}$$

Les différents systèmes linéaires à résoudre pour trouver les zéros du système initial sont

au nombre de 9 dont trois seulement ont un rang de 6 :

$$\left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_2 + \tilde{x}_5 + 1) \\ (\tilde{x}_3 + \tilde{x}_6 + 1) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1} \quad \left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_5 - 1) \\ (\tilde{x}_3 + \tilde{x}_6 + 1) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1} \quad \left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_5 + 2) \\ (\tilde{x}_3 + \tilde{x}_6 + 1) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1}$$

rang = 6                      rang = 5                      rang = 5

$$\left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_2 + \tilde{x}_5 + 1) \\ (\tilde{x}_6 - 3) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1} \quad \left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_5 - 1) \\ (\tilde{x}_6 - 3) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1} \quad \left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_5 + 2) \\ (\tilde{x}_6 - 3) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1}$$

rang = 6                      rang = 5                      rang = 5

$$\left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_2 + \tilde{x}_5 + 1) \\ (\tilde{x}_6 + 5) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1} \quad \left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_5 - 1) \\ (\tilde{x}_6 + 5) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1} \quad \left. \begin{array}{l} (\tilde{x}_1 + \tilde{x}_2 + 1) \\ (\tilde{x}_3 + \tilde{x}_4 + 1) \\ (\tilde{x}_5 + 2) \\ (\tilde{x}_6 + 5) \\ (\tilde{x}_3 + \tilde{x}_5) \\ (\tilde{x}_4 + \tilde{x}_6) \end{array} \right\} = 0_{6,1}$$

rang = 6                      rang = 5                      rang = 5

Ainsi, on n'aura que 3 chemins à suivre partant des solutions des systèmes linéaires écrits en colonne de gauche. Si l'on avait utilisé une méthode du degré total, on aurait dû étudier  $1 \times 1 \times 3 \times 3 \times 1 \times 1 = 9$  chemins et par une méthode de multi-homogénéisation utilisant les deux groupes « intuitifs »  $\tilde{X}_1 = \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4\}$  et  $\tilde{X}_2 = \{\tilde{x}_5, \tilde{x}_6\}$ , on aurait généré 22 chemins, c'est-à-dire plus que dans le cas du degré total.

Comme on vient de le voir, l'évaluation du nombre de zéros du système multi-polynomial initial passe par l'évaluation du rang de chacun des systèmes linéaires que l'on peut former, au nombre de  $\prod_{i=1}^P d_i$ . On ne peut donc pas donner de formule générale mais affirmer qu'il est nécessairement inférieur ou égal à celui résultant de la méthode du degré total. Le théorème suivant permet de garantir l'efficacité de la méthode.

### **Théorème 6.3 : Homotopie par produit d'éléments linéaires**

Soit un système multi-polynomial  $\mathbf{P}(\tilde{x})$  de  $\mathcal{C}^P$  dans  $\mathcal{C}^P$  appartenant à la famille de systèmes définie par l'équation (6.10). Soit  $\mathbf{Q}(\tilde{x})$  un système multi-polynomial appartenant à la famille de système définie par l'équation (6.11). Alors les chemins solutions de l'homotopie

$$\mathbf{H}(\tilde{x}, \lambda) = \lambda \mathbf{Q}(\tilde{x}) + (1 - \lambda) \mathbf{P}(\tilde{x}) = 0$$

prenant source en les solutions non-singulières de  $\mathbf{Q}(\tilde{x}) = 0$  sont non-singuliers pour  $\lambda \in ]0, 1]$  et l'ensemble de leurs points finaux lorsque  $\lambda \rightarrow 0$  inclut toutes les solutions non-singulières de  $\mathbf{P}(\tilde{x}) = 0$  pour presque tout  $\gamma \in \mathcal{C}$ , excepté en un nombre fini de droites de dimension réelle égale à 1 passant par l'origine. En particulier, si l'on choisit  $\gamma = e^{i\theta}$  sur le cercle unité, ces exceptions se limitent à un nombre fini de points  $\theta \in [0, 2\pi]$ .  $\square$

### 6.5.4 Comparaison des trois méthodes

Nous venons de présenter différentes méthodes permettant de construire un système initial  $\mathbf{Q}(\tilde{x})$  d'après des analyses de plus en plus fines de la structure du système initial  $\mathbf{P}(\tilde{x})$ . Nous proposons de les comparer sur un exemple possédant une non-linéarité cubique; le développement polynomial sera donc ici exact.

#### Description du système et écriture des équations dynamiques

Le système se compose de trois masses et ressorts linéaires amortis dont le dernier maillon comporte en plus une raideur non-linéaire cubique (figure 6.3).

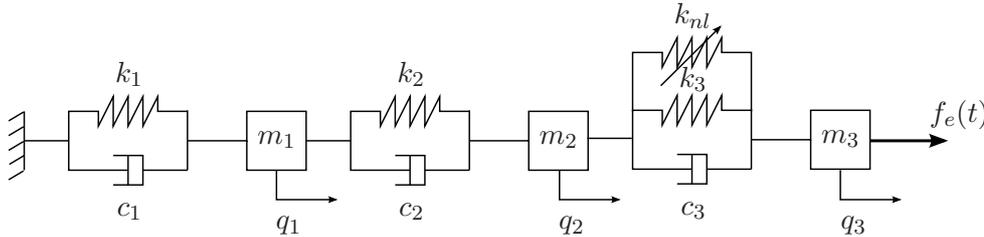


FIG. 6.3 – Système étudié

Les équations s'écrivent ainsi simplement :

$$\begin{cases} m_1 \ddot{q}_1 + c_1 \dot{q}_1 - c_2(\dot{q}_2 - \dot{q}_1) + k_1 q_1 - k_2(q_2 - q_1) = 0 \\ m_2 \ddot{q}_2 + c_2(\dot{q}_2 - \dot{q}_1) - c_3(\dot{q}_3 - \dot{q}_2) + k_2(q_2 - q_1) - k_3(q_3 - q_2) - k_{nl}(q_3 - q_2)^3 = 0 \\ m_3 \ddot{q}_3 + c_3(\dot{q}_3 - \dot{q}_2) + k_3(q_3 - q_2) + k_{nl}(q_3 - q_2)^3 = f_e \cos(\omega t) \end{cases} \quad (6.12)$$

Pour alléger les écritures, on passe ensuite en déplacements relatifs en posant :

$$\begin{cases} r_1 = q_1 \\ r_2 = q_2 - q_1 \\ r_3 = q_3 - q_2 \end{cases}, \quad U = \begin{Bmatrix} r_1 \\ r_2 \\ r_3 \end{Bmatrix}$$

puis on utilise l'écriture matricielle suivante :

$$\underbrace{\begin{bmatrix} m_1 & & \\ m_2 & m_2 & \\ m_3 & m_3 & m_3 \end{bmatrix}}_M \ddot{U} + \underbrace{\begin{bmatrix} c_1 & -c_2 & \\ & c_2 & -c_3 \\ & & c_3 \end{bmatrix}}_C \dot{U} + \underbrace{\begin{bmatrix} k_1 & -k_2 & \\ & k_2 & -k_3 \\ & & k_3 \end{bmatrix}}_K U + \begin{Bmatrix} 0 \\ -k_{nl}r_3^3 \\ k_{nl}r_3^3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ f_e \cos(\omega t) \end{Bmatrix} \quad (6.13)$$

Nous choisissons enfin les valeurs numériques suivantes :

$$\begin{aligned} m_1 &= 1 \text{ (kg)}, k_1 = 1 \text{ (N.m}^{-1}\text{)}, c_1 = \eta k_1 \text{ (kg.s}^{-1}\text{)}, \eta = 0.04 \text{ (s)} \\ m_2 &= 2 \text{ (kg)}, k_2 = 1 \text{ (N.m}^{-1}\text{)}, c_2 = \eta k_2 \text{ (kg.s}^{-1}\text{)}, f_e = 0.1 \text{ (N)} \\ m_3 &= \frac{3}{2} \text{ (kg)}, k_3 = 1 \text{ (N.m}^{-1}\text{)}, c_3 = \eta k_3 \text{ (kg.s}^{-1}\text{)}, k_{nl} = 100 \text{ (N.m}^{-1}\text{)} \end{aligned}$$

#### Écriture des équations algébriques

Nous allons maintenant rechercher les solutions périodiques par une méthode de balance harmonique avec une troncature du développement à 1 harmonique. On passe donc de 3

inconnues dynamiques à  $3 \times (2 \times 1 + 1) = 9$  inconnues algébriques :

$$U = a_0 + a_1 \cos(\omega t) + b_1 \sin(\omega t) \quad (6.14)$$

$$a_0 = \begin{Bmatrix} a_{01} \\ a_{02} \\ a_{03} \end{Bmatrix}, \quad a_1 = \begin{Bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{Bmatrix}, \quad b_1 = \begin{Bmatrix} b_{11} \\ b_{12} \\ b_{13} \end{Bmatrix}, \quad \tilde{x} = \begin{Bmatrix} a_0 \\ a_1 \\ b_1 \end{Bmatrix} \in \mathbb{R}^9$$

En substituant la forme (6.14) de  $U$  dans les équations dynamiques (6.13), on obtient les équations algébriques suivantes :

$$\underbrace{\begin{bmatrix} K & & \\ & K - \omega^2 M & \omega C \\ & -\omega C & K - \omega^2 M \end{bmatrix}}_{\bar{H}} \tilde{x} + \underbrace{\hat{H}(\tilde{x})}_{H_c} = \begin{Bmatrix} 0_{3,1} \\ \left( \begin{matrix} 0_{2,1} \\ f_e \end{matrix} \right) \\ 0_{3,1} \end{Bmatrix} \quad (6.15)$$

$$\hat{H}(\tilde{x}) = k_{nl} \begin{Bmatrix} 0 \\ -a_{03}(a_{03}^2 + \frac{3}{2}a_{13}^2 + \frac{3}{2}b_{13}^2) \\ +a_{03}(a_{03}^2 + \frac{3}{2}a_{13}^2 + \frac{3}{2}b_{13}^2) \\ 0 \\ -3a_{13}(a_{03}^2 + \frac{1}{4}a_{13}^2 + \frac{1}{4}b_{13}^2) \\ +3a_{13}(a_{03}^2 + \frac{1}{4}a_{13}^2 + \frac{1}{4}b_{13}^2) \\ 0 \\ -3b_{13}(a_{03}^2 + \frac{1}{4}a_{13}^2 + \frac{1}{4}b_{13}^2) \\ +3b_{13}(a_{03}^2 + \frac{1}{4}a_{13}^2 + \frac{1}{4}b_{13}^2) \end{Bmatrix}$$

### Écriture du système multi-polynomial

Étant donnée l'utilisation d'une non-linéarité cubique, les équations algébriques sont déjà sous forme polynomiale ; il n'est donc pas nécessaire de réaliser de lissage pour obtenir  $\mathbf{P}(\tilde{x})$  et l'on obtient directement la forme exacte (6.16).

$$\mathbf{P}(\tilde{x}) = \left\{ \begin{array}{l} k_1 a_{01} - k_2 a_{02} \\ k_2 a_{02} - k_3 a_{03} - k_{nl} a_{03} (a_{03}^2 + \frac{3}{2} a_{13}^2 + \frac{3}{2} b_{13}^2) \\ k_3 a_{03} + k_{nl} a_{03} (a_{03}^2 + \frac{3}{2} a_{13}^2 + \frac{3}{2} b_{13}^2) \\ (k_1 - \omega^2 m_1) a_{11} - k_2 a_{12} + \omega c_1 b_{11} - \omega c_2 b_{12} \\ -\omega^2 m_2 a_{11} + (k_2 - \omega^2 m_2) a_{12} - k_3 a_{13} + \omega c_2 b_{12} - \omega c_3 b_{13} - 3k_{nl} a_{13} (a_{03}^2 + \frac{1}{4} a_{13}^2 + \frac{1}{4} b_{13}^2) \\ -\omega^2 m_3 a_{11} - \omega^2 m_3 a_{12} + (k_3 - \omega^2 m_3) a_{13} + \omega c_3 b_{13} + 3k_{nl} a_{13} (a_{03}^2 + \frac{1}{4} a_{13}^2 + \frac{1}{4} b_{13}^2) - f_e \\ -\omega c_1 a_{11} + \omega c_2 a_{12} + (k_1 - \omega^2 m_1) b_{11} - k_2 b_{12} \\ -\omega c_2 a_{12} + \omega c_3 a_{13} - \omega^2 m_2 b_{11} + (k_2 - \omega^2 m_2) b_{12} - k_3 b_{13} - 3k_{nl} b_{13} (a_{03}^2 + \frac{1}{4} a_{13}^2 + \frac{1}{4} b_{13}^2) \\ -\omega c_3 a_{13} - \omega^2 m_3 b_{11} - \omega^2 m_3 b_{12} + (k_3 - \omega^2 m_3) b_{13} + 3k_{nl} b_{13} (a_{03}^2 + \frac{1}{4} a_{13}^2 + \frac{1}{4} b_{13}^2) \end{array} \right\} \quad (6.16)$$

### Étude préalable par continuation

Avant de procéder à l'analyse globale de ce système, on procède à une étude par continuation en faisant varier la pulsation d'excitation entre 0 et  $5 \text{ rad.s}^{-1}$  afin de déterminer une pulsation pour laquelle on possède plusieurs solutions. Nous obtenons la figure 6.4 représentant l'amplitude du déplacement relatif  $r_3$  en fonction de la pulsation  $\omega$  et choisissons de travailler pour  $\omega = 2 \text{ rad.s}^{-1}$ .

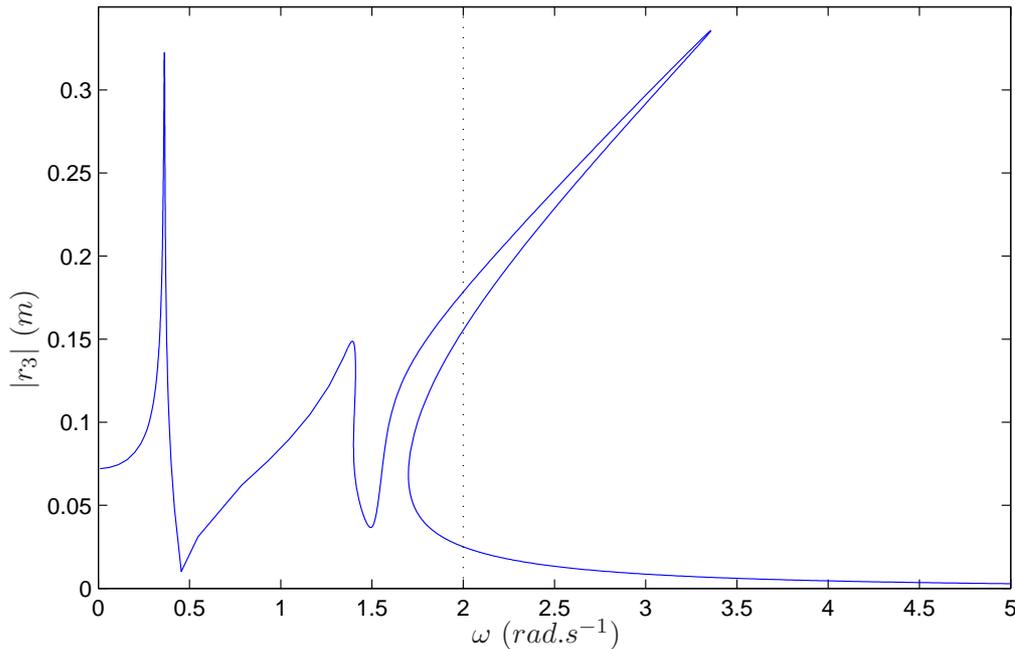


FIG. 6.4 – Continuation du système non-linéaire cubique à 3 ddl : amplitude de  $r_3$

### Analyse globale

Afin d'illustrer et de chiffrer les performances des trois méthodes, nous donnons les éléments caractéristiques de chacune d'elle permettant la construction du polynôme initial  $\mathbf{Q}(\tilde{x})$  :

- **Méthode du degré total** : il suffit de relever le degré total de chaque composante  $\mathbf{P}_i(\tilde{x})$  :

$$D = \left[ 1 \ 3 \ 3 \ 1 \ 3 \ 3 \ 1 \ 3 \ 3 \right]^t \Rightarrow \mathcal{N} = 3^6 = 729$$

- **Méthode par multi-homogénéisation** : on sépare les variables en deux groupes,  $\tilde{X}_1$  contenant les variables « linéaires » qui seront homogénéisées par  $\tilde{z}_1$  et  $\tilde{X}_2$  qui contient les variable « non-linéaires » qui seront homogénéisées par  $\tilde{z}_2$  :

$$\tilde{X}_1 = \{a_{01}, a_{02}, a_{11}, a_{12}, b_{11}, b_{12}\}, \quad k_1 = 6 \leftarrow \tilde{z}_1$$

$$\tilde{X}_2 = \{a_{03}, a_{13}, b_{13}\}, \quad k_2 = 3 \leftarrow \tilde{z}_2$$

On construit ensuite la matrice de degré  $D$  pour calculer le nombre de Bezout  $\text{Bez}(D, K)$  :

$$D = \left[ \begin{array}{cccccccc} 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 3 & 3 & 0 & 3 & 3 & 0 & 3 & 3 \end{array} \right]^t$$

$$\Rightarrow \text{Bez}(D, K) = \text{coeff}(\tilde{z}_1^6 \tilde{z}_2^3, \tilde{z}_1^3 (\tilde{z}_1 + 3\tilde{z}_2)^5 (3\tilde{z}_2)) = 270$$

- **Méthode par produits d'éléments linéaires** : on écrit une décomposition des ensembles  $E_i$  en produit d'espaces vectoriels ne contenant que des monômes de degré inférieur ou égal à 1 :

$$\begin{aligned}
E_1 &= \langle a_{01}, a_{02} \rangle \\
E_2 &= \langle a_{02}, a_{03}, a_{13}, b_{13}, 1 \rangle \otimes \langle a_{03}, a_{13}, b_{13}, 1 \rangle \otimes \langle a_{03}, 1 \rangle \\
E_3 &= \langle a_{03}, a_{13}, b_{13}, 1 \rangle^{(2)} \otimes \langle a_{03} \rangle \\
E_4 &= \langle a_{11}, a_{12}, b_{11}, b_{12} \rangle \\
E_5 &= \langle a_{11}, a_{12}, a_{13}, b_{12}, 1 \rangle \otimes \langle a_{03}, a_{13}, b_{13}, 1 \rangle^{(2)} \\
E_6 &= \langle a_{11}, a_{12}, a_{13}, 1 \rangle \otimes \langle a_{03}, a_{13}, b_{13}, 1 \rangle^{(2)} \\
E_7 &= \langle a_{11}, a_{12}, b_{11}, b_{12} \rangle \\
E_8 &= \langle a_{12}, b_{11}, b_{12}, b_{13}, 1 \rangle \otimes \langle a_{03}, a_{13}, b_{13}, 1 \rangle^{(2)} \\
E_9 &= \langle b_{11}, b_{12}, b_{13}, 1 \rangle \otimes \langle a_{03}, a_{13}, b_{13}, 1 \rangle^{(2)}
\end{aligned}$$

⇒ 99 systèmes linéaires de rang 9

Les trois méthodes permettent d'aboutir à neuf solutions complexes dont trois seulement sont réelles. Nous en donnons une représentation composante par composante en superposition des courbes obtenues par continuation en figure 6.5. Les temps de calcul et le nombre de chemins sont récapitulés dans le tableau 6.1. Les calculs ont été réalisés sur un serveur Intel Core 2 Duo E6600 (4Go de Ram).

	Degré Total	Multi-homogénéisation	Produit d'éléments linéaires
Nombre de chemins	729	270	99
Temps cpu (s)	467	200	74
Temps cpu/chemin (s)	0.641	0.741	0.748

TAB. 6.1 – Comparaison des temps de calcul sur les différentes méthodes par homotopie

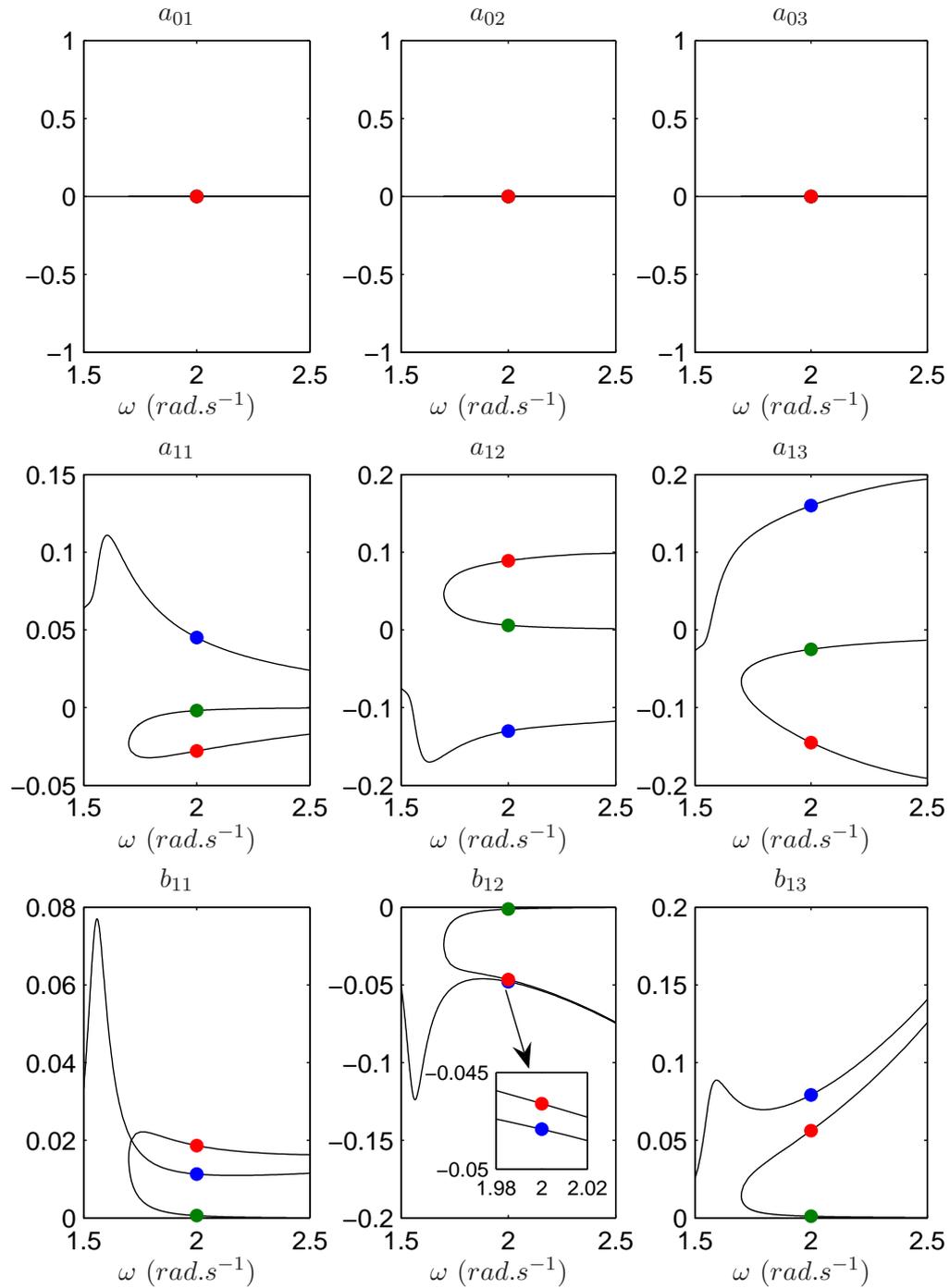


FIG. 6.5 – Solutions réelles multiples obtenues par homotopie

## 6.6 Application à un oscillateur de Duffing

Afin de pouvoir comparer les méthodes entre elles, nous appliquons cette dernière approche aux quatre cas présentés dans le tableau 4.1 ainsi qu'aux deux cas supplémentaires du tableau 6.2 dont le temps de calcul est rendu raisonnable par l'utilisation de cette méthode. Comme pour les méthodes précédentes, nous présentons d'abord les résultats puis nous proposons une extrapolation du temps de calcul.

Dénomination	$P$	Expression
Cas 1H2H3H	6	$q(t) = a_1 \cos(\omega t) + b_1 \sin(\omega t) + a_2 \cos(2\omega t) + b_2 \sin(2\omega t) \dots$ $\dots + a_3 \cos(3\omega t) + b_3 \sin(3\omega t)$
Cas C1H2H3H	6	$q(t) = a_0 + a_1 \cos(\omega t) + b_1 \sin(\omega t) + a_2 \cos(2\omega t) + b_2 \sin(2\omega t) \dots$ $\dots + a_3 \cos(3\omega t) + b_3 \sin(3\omega t)$

TAB. 6.2 – Homotopies polynomiales - Cas supplémentaires d'application à un oscillateur de Duffing

### 6.6.1 Résultats

Étant donné que le système dynamique ne comporte qu'un seul degré de liberté et que toutes les inconnues algébriques résultant de la balance harmonique sont touchées par la non-linéarité, seule une méthode de degré total est appliquée. Le tableau 6.3 résume les caractéristiques principales des six calculs menés. Les calculs sont encore une fois menés sur un serveur Intel Core 2 Duo E6600 (4Go de Ram). On voit dès à présent l'énorme supériorité

Cas	P	Temps cpu (s)	Nombre de chemins
1H	2	1.09	9
C1H	3	4.51	27
1H3H	4	29.4	81
C1H3H	5	121.45	243
1H2H3H	6	521.86	729
C1H2H3H	7	2140.42	2187

TAB. 6.3 – Homotopies polynomiales - Résultats

de cette méthode sur les précédentes.

A titre d'illustration, nous proposons les figures 6.6 à 6.9. Elles représentent les chemins suivis pour les trois solutions réelles et un des chemins partant d'une solution surnuméraire qui diverge lorsque  $\lambda$  tend vers 0 lors de la résolution du cas 1H. Le chemin bleu représente  $a_1$  et le chemin vert représente  $b_1$ .

### 6.6.2 Extrapolations

Nous possédons pour cette catégorie de méthode une évaluation a priori du nombre de chemin  $N_c$  utilisés par l'algorithme. Utilisant une méthode du degré total, on sait que l'on a  $N_c = 3^P$ , ce que l'on vérifie sur la figure 6.10 pour laquelle  $a = 1$  et  $b = 3$ . Le temps de

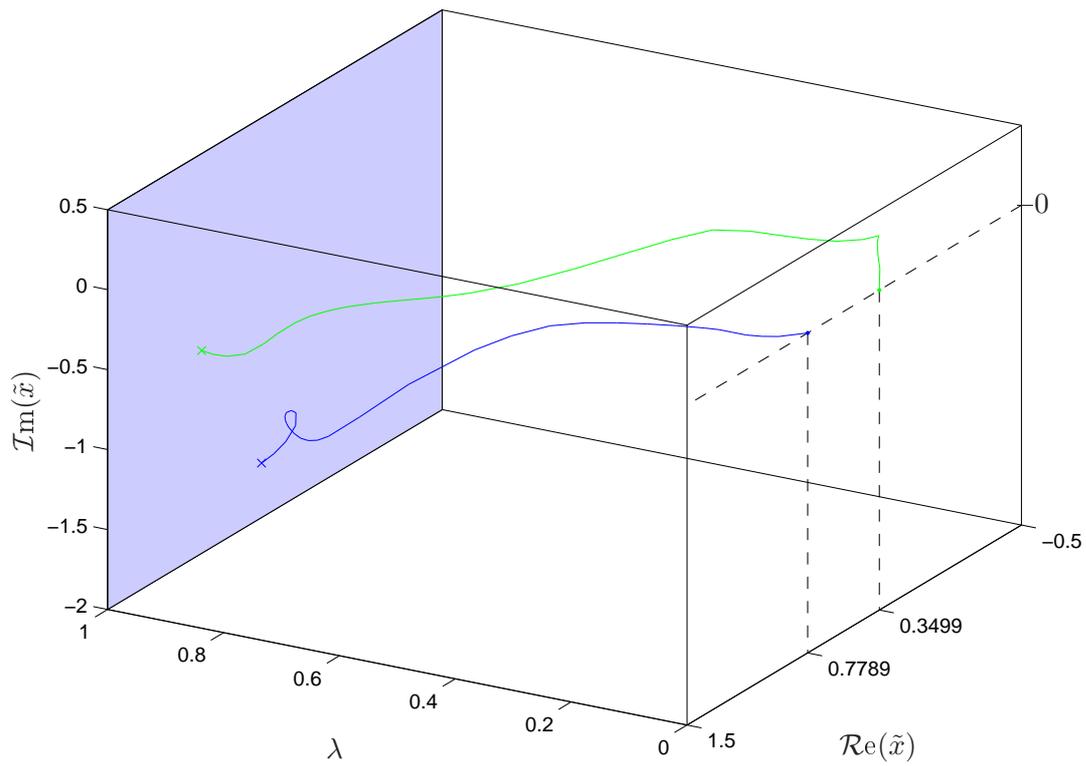


FIG. 6.6 – Homotopies polynomiales - Cas 1H - Solution 1

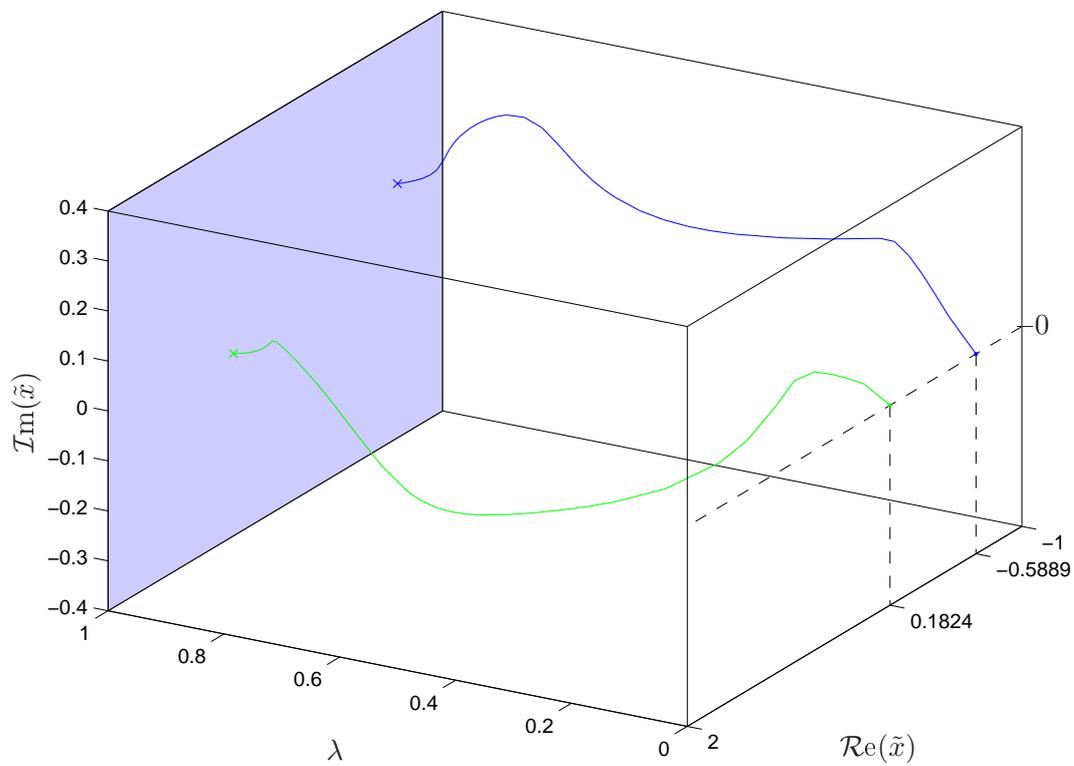


FIG. 6.7 – Homotopies polynomiales - Cas 1H - Solution 2

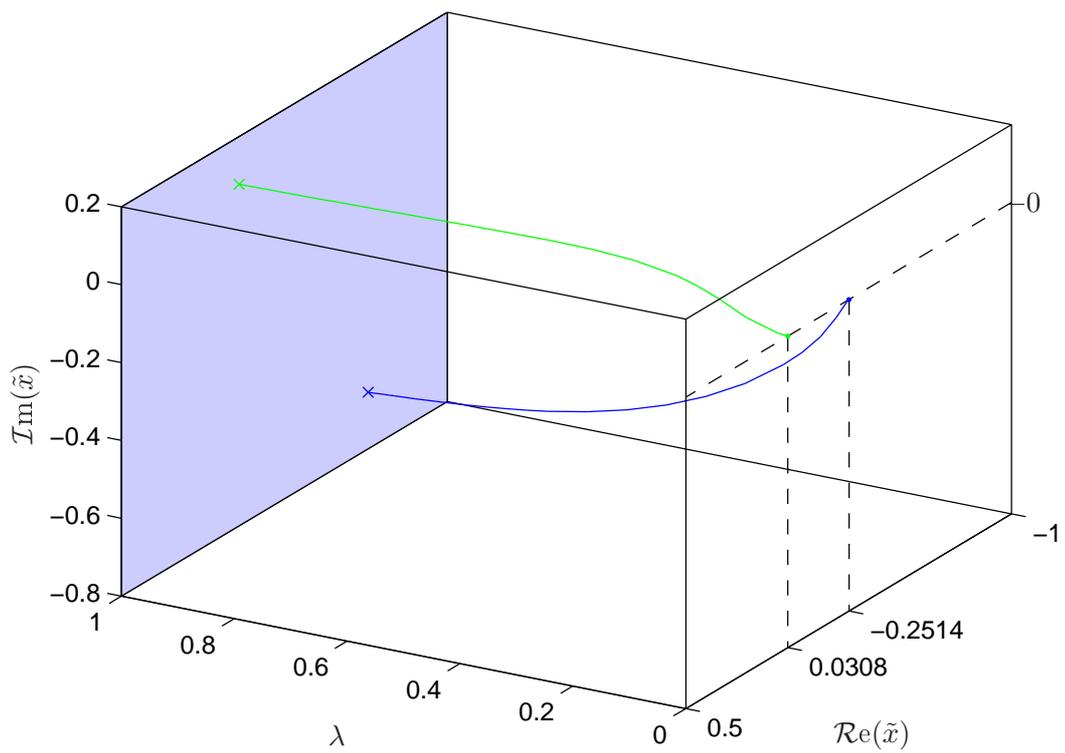


FIG. 6.8 – Homotopies polynomiales - Cas 1H - Solution 3

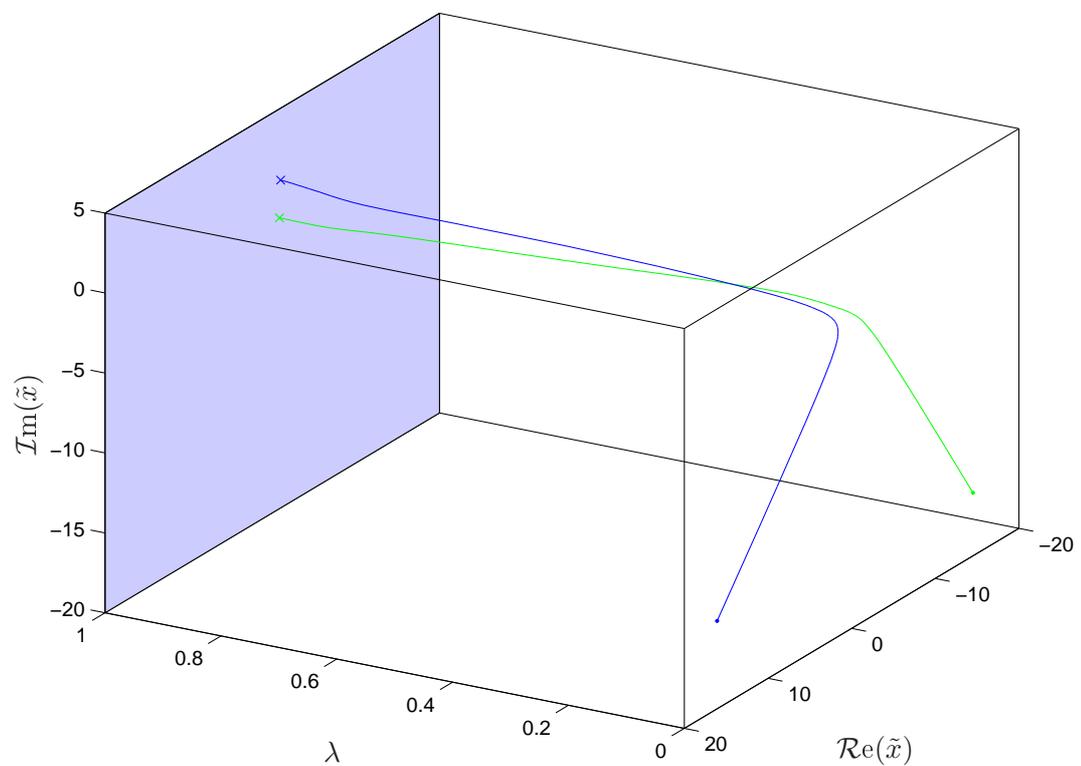


FIG. 6.9 – Homotopies polynomiales - Cas 1H - Chemin divergent

calcul unitaire, qui correspond au temps nécessaire pour passer d'un zéro de  $Q(\tilde{x})$  à un zéro de  $P(\tilde{x})$  est lissé en utilisant un polynôme d'ordre 2 en  $P$  dans la mesure où les méthodes de continuation utilisent l'évaluation d'une jacobienne. On trouve  $t_u = cP^2 + dP + e$  avec  $c = 0.0208$ ,  $d = -0.0138$  et  $e = 0.0543$  en figure 6.11. Ces deux évaluations nous permettent d'établir des prédictions pour le temps de calcul total en figure 6.12. On constate alors que c'est la seule des quatre méthodes testées qui permette de traiter des systèmes à non-linéarité cubique à plus de trois ddl dans la journée.

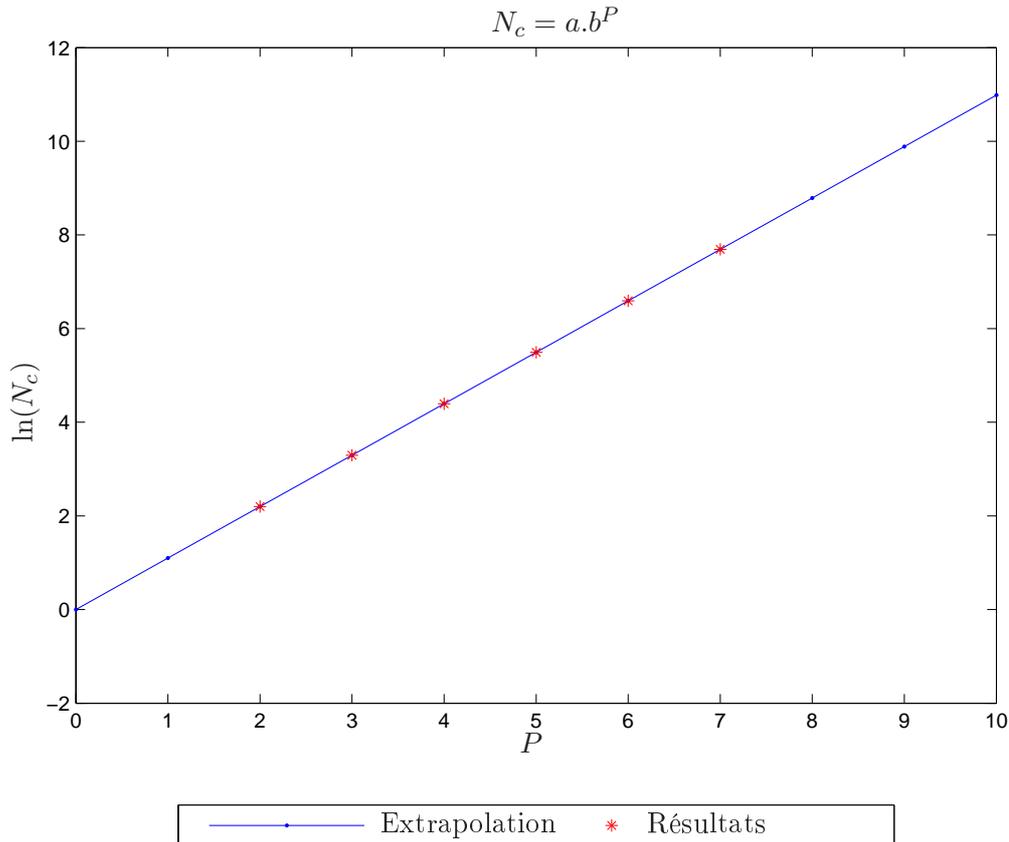


FIG. 6.10 – Homotopies polynomiales - Extrapolation du nombre de cellules

## 6.7 Conclusion

Cette dernière catégorie de méthodes, basée sur l'approximation polynomiale des équations algébriques non-linéaires résultant de la transformation du système dynamique, présente des temps de calcul intéressants. Le fait de ne pas partir d'un espace continu que l'on essaie de réduire à un ensemble de points mais de partir directement de points discrets que l'on fait évoluer lors d'une continuation est à notre sens à l'origine de ce gain de temps. Dans le dernier chapitre, nous proposons une application de cette méthode à différents systèmes, moins triviaux et plus proches dans leur formulation de ce que l'on peut rencontrer en dynamique des machines tournantes.

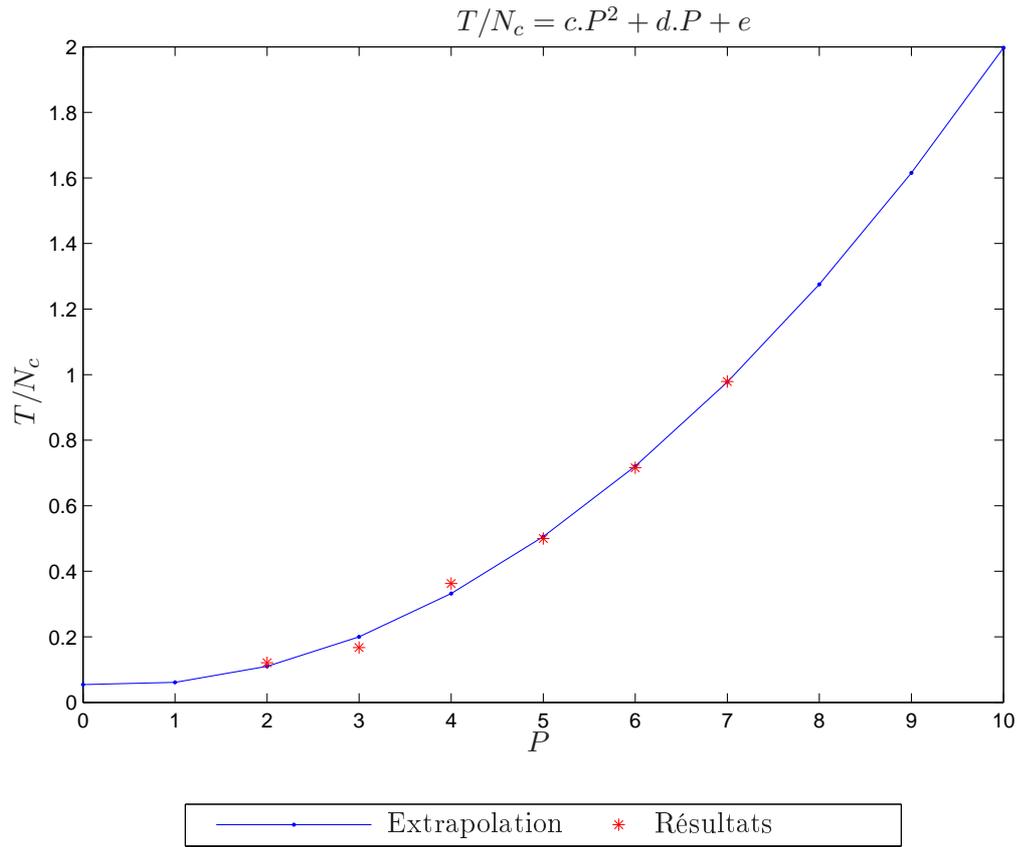


FIG. 6.11 – Homotopies polynomiales - Extrapolation du temps par chemin

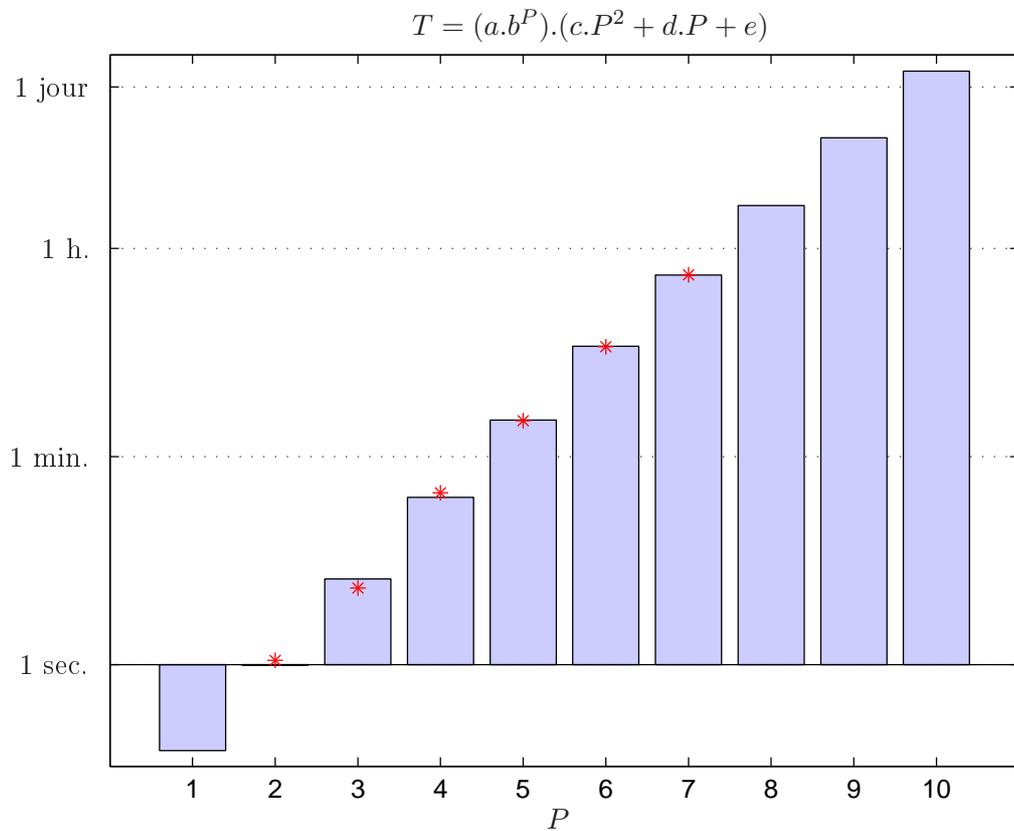


FIG. 6.12 – Homotopies polynomiales - Extrapolation du temps de calcul



## Chapitre 7

# Application à la dynamique des rotors

---

*Nous avons exposé différentes façons de réaliser l'analyse globale d'un système dynamique. Les méthodes présentées ont de façon évidente des avantages et des inconvénients intrinsèques. Cependant, nous n'avons pas trouvé dans la littérature de réelle évaluation de leurs performances lorsqu'elles sont appliquées à des systèmes dynamiques issus de la formulation de problèmes mécaniques. Nous proposons donc d'abord une comparaison rapide des résultats obtenus par chacune des méthodes sur l'oscillateur de Duffing puis une application de certaines des méthodes à des cas moins académiques, pouvant s'approcher du besoin industriel. Nous avons choisi ces exemples dans le domaine de la dynamique des rotors dont on trouvera des études abondantes dans les ouvrages [Ehr99], [Chi93], [LF98] ou encore [Van88].*

### 7.1 Comparaisons des méthodes sur un cas simple

Nous proposons ici de dresser un bilan de l'application des quatre méthodes d'analyse globale étudiée dans les chapitre 3 à 6 au même exemple simple que constitue l'oscillateur de Duffing.

La première constatation, évidente, est que toutes s'avèrent coûteuses en termes de ressources numériques, la moins coûteuse étant de loin la famille de méthodes basée sur l'homotopie d'un système polynomial. Le cell-mapping, tentant pour sa richesse d'informations rendues et le peu d'hypothèses qu'il nécessite pour être mis en œuvre serait la solution idéale si l'on ne considérait que des systèmes à un ou deux degrés de liberté. Au delà, sa consommation en temps de calcul et en espace de stockage rend cette méthode impraticable. Les deux méthodes reposant sur les tests d'exclusion et sur l'arithmétique des intervalles se sont

avérées peu performantes : elles ne cherchent qu'un type de solution et sont consommatrices en temps, avec de plus, une difficulté à estimer a priori le temps nécessaire. En revanche, les méthodes fondées sur la résolution de systèmes polynomiaux par homotopie, si elles obligent à une approximation de la non-linéarité, permettent d'évaluer rapidement le nombre de chemins et le temps nécessaire à la continuation d'un chemin (il suffit d'en suivre quelques-uns, ce qui prend quelques secondes au plus). On peut donc, dès le début de sa mise en œuvre, donner une estimation correcte du temps requis. De plus, la recherche de solutions périodiques se faisant fréquemment par une méthode de balance harmonique pour obtenir des équations algébriques, elles permettent, via l'approximation polynomiale, de s'affranchir de l'aspect AFT de cette méthode en utilisant directement une évaluation polynomiale de  $\hat{H}(\tilde{x})$ ; les allers et retours entre l'espace fréquentiel et l'espace temporel étant généralement coûteux, on réalise là un gain de temps important. Enfin, contrairement aux trois autres méthodes, elle ne nécessite pas de déterminer a priori l'espace de recherche initial. On est ainsi certain de ne pas manquer une solution pour avoir trop voulu restreindre cet espace ( $\mathcal{E}_{IC}$  ou  $\Omega$ ) afin d'économiser du temps de calcul.

Dans la suite de ce chapitre, nous appliquerons donc préférentiellement cette méthode pour l'étude des exemples possédant plus d'un degré de liberté.

## 7.2 Contact permanent rotor-stator avec stator non-linéaire

Nous proposons dans cette section d'étudier un modèle phénoménologique de contact permanent et sans frottement issu de [Les07, Chap 5.1] (dont on trouvera aussi une étude dans [LST07]). Le stator, déformable, est modélisé par un anneau élastique. Il est excité par une aube dont seule la traction-compression est prise en compte étant donné qu'en l'absence de frottement, la flexion ne joue pas ; elle est donc modélisée par un simple système masse-ressort formant un angle fixe  $\phi_b$  avec le repère tournant. Le contact entre l'aube et le stator est supposé permanent. Le schéma de ce système est donné en figure 7.1. Les équations sont écrites dans le repère tournant  $(O, \vec{X}, \vec{Y})$  lié au rotor et animé d'une vitesse de rotation par rapport au référentiel fixe  $(O, \vec{x}, \vec{y})$  valant  $\Omega$  ; on repère donc  $w(\phi, t)$  le déplacement tangentiel de l'anneau en un point de sa circonférence formant un angle  $\phi$  avec le repère tournant et  $u(\phi, t)$  son déplacement radial.

Un tel modèle permet de décomposer le déplacement le long de l'anneau de la façon suivante [LST07] :

$$w(\phi, t) = \sum_{p=2}^{p_{tot}} \alpha_p(t) \cos(p\phi) + \beta_p(t) \sin(p\phi), \quad u(\phi, t) = \frac{\partial w}{\partial \phi} \quad (7.1)$$

Cette décomposition du mouvement est appelée décomposition selon les  $p$ -èmes modes d'anneaux (dont le corps rigide,  $p = 1$ , a été éliminé).

L'étude du modèle linéaire fait alors apparaître des plages d'instabilité comme le montre la figure 7.2 qui rend compte de l'étude aux valeurs propres du système linéaire pour le troisième mode d'anneau ( $p = 3$ ) : une première entre 304.7 et 322.5 tr/min et une seconde à partir de 3840 tr/min. La première plage correspond à une divergence du mode direct de l'anneau

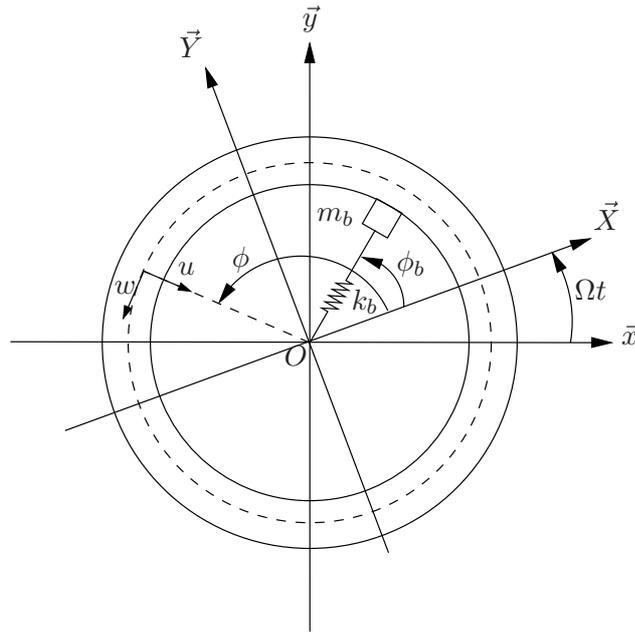


FIG. 7.1 – Contact permanent non-frottant aube-carter - Modèle

(fréquence nulle dans le repère tournant) tandis que la seconde correspond à une coalescence entre le mode direct et le mode rétrograde de l'anneau ; c'est une instabilité de flottement.

Il est donc proposé de rendre compte des non-linéarités qui limitent les déplacements en couplant la traction-compression et la flexion du stator pour des déplacements relativement importants. On introduit ainsi un effort non-linéaire sur son pourtour de la forme :

$$F_{nl}(\phi, t) = \tilde{k}_{nl}u^3(\phi, t)$$

Nous nous consacrons donc par la suite à l'étude des cycles limites apparaissant dans les zones d'instabilité. Nous donnons dans un premier temps les équations complètes du modèle ainsi que les valeurs numériques utilisées puis les résultats de leur résolution par différentes méthodes d'analyse globale.

### 7.2.1 Équations

Si l'on néglige les couplages entre les différents modes d'anneau, l'écriture des énergies en utilisant la forme (7.1) des déplacements, puis leurs dérivations conformément aux équations de Lagrange permet d'écrire le système différentiel gouvernant le  $p$ -ème mode suivant :

$$M\ddot{q} + (G + D)\dot{q} + Kq + \hat{f}(q) = f_e(t), \quad q = \begin{Bmatrix} \alpha_p \\ \beta_p \end{Bmatrix} \quad (7.2)$$

avec respectivement les matrices de masse, gyroscopique, d'amortissement visqueux et de raideur suivantes :

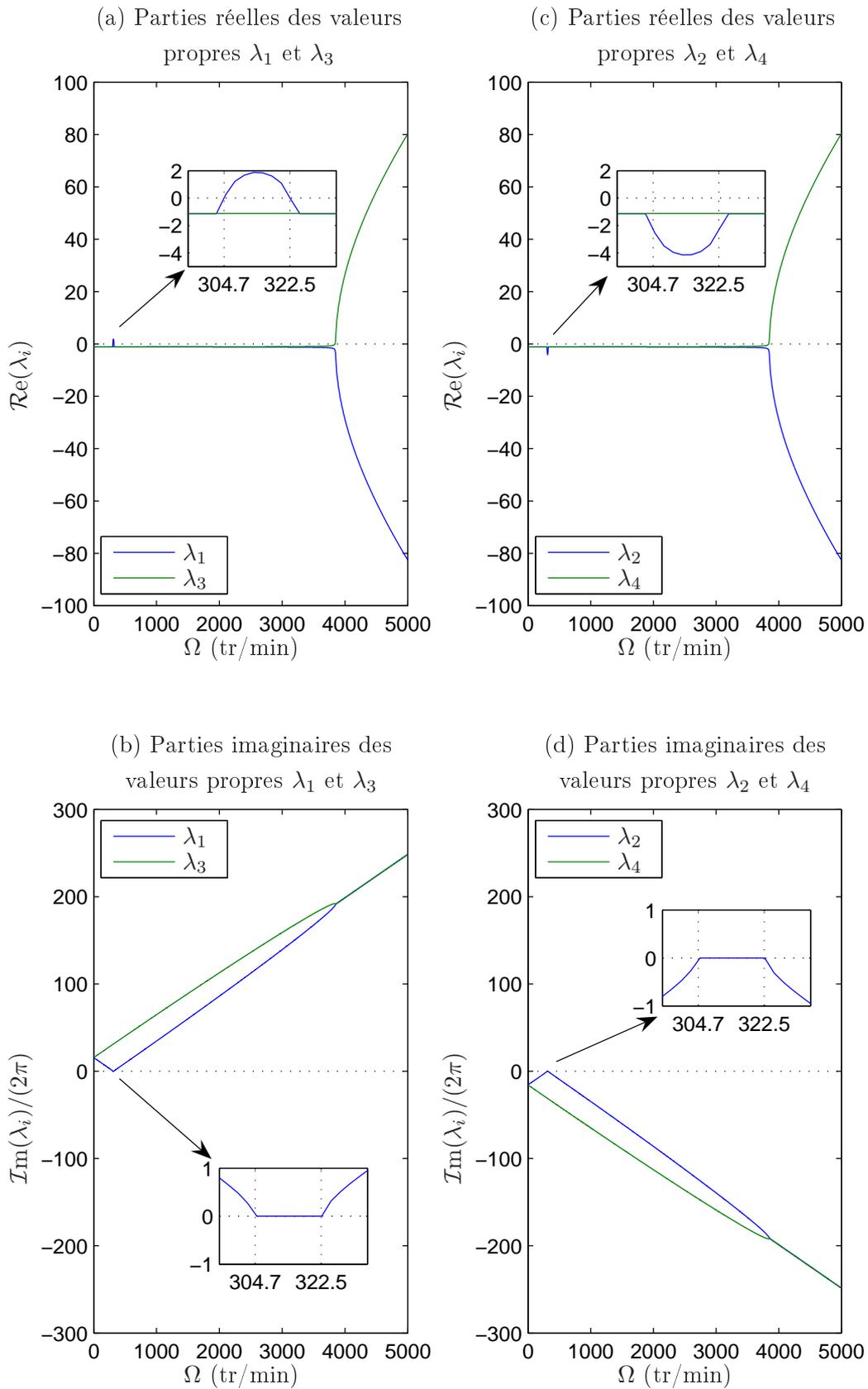


FIG. 7.2 – Contact rotor-stator non frottant - Analyse linéaire

$$\begin{aligned}
M_s &= (\rho_s S_s R_s \pi)(n^2 + 1) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & K_e &= (E_s I_s \pi / R_s^3)(n^3 - n)^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
G &= (2\rho_s S_s R_s \pi \Omega)(n(n^2 + 1)) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} & K_g &= (-\rho_s S_s R_s \pi \Omega^2)(n^2(n^2 + 1)) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
D &= (\eta_s E_s I_s \pi / R_s^3)(n^3 - n)^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & K_d &= (\eta_s E_s I_s \pi / R_s^3 \Omega)(n(n^3 - n)^2) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\
M_b &= (m_b p^2) \begin{bmatrix} \sin(p\phi_b)^2 & 0 \\ 0 & \cos(p\phi_b)^2 \end{bmatrix} & K_b &= ((k_b - m_b \Omega^2)p^2) \begin{bmatrix} \sin(p\phi_b)^2 & 0 \\ 0 & \cos(p\phi_b)^2 \end{bmatrix}
\end{aligned}$$

$$M = M_s + M_b$$

$$K = K_e + K_g + K_d + K_b$$

Force non-linéaire :

$$\hat{f}(\alpha_p, \beta_p) = \int_0^{2\pi} F_{nl}(\phi, t) \begin{Bmatrix} -p \sin(p\phi) \\ p \cos(p\phi) \end{Bmatrix} d\phi = p^4 k_{nl} \begin{Bmatrix} \alpha_p^3 + \alpha_p \beta_p^2 \\ \beta_p^3 + \alpha_p^2 \beta_p \end{Bmatrix}, \quad k_{nl} = \tilde{k}_{nl} \frac{3}{4}$$

Force due à l'aube :

$$f_e(t) = -pm_b R_s \Omega^2 \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$$

**Forme augmentée :**

$$\underbrace{\begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix}}_A \dot{x} + \underbrace{\begin{bmatrix} 0 & -K \\ K & G + D \end{bmatrix}}_B x + \underbrace{\begin{Bmatrix} 0 \\ \hat{f} \end{Bmatrix}}_{\hat{F}(x)} = \underbrace{\begin{Bmatrix} 0 \\ f_e(t) \end{Bmatrix}}_{F_e(t)}, \quad x = \begin{Bmatrix} q \\ \dot{q} \end{Bmatrix}$$

**Forme algébrique obtenue par balance harmonique :** Nous effectuerons pour les coefficients  $\alpha_p$  et  $\beta_p$  une recherche de solution périodique à un harmonique pour une pulsation égale à  $3\Omega$ . Afin d'obtenir le système d'équations algébriques traitable par les algorithmes de recherche globale, on applique une méthode de balance harmonique à l'équation (7.2). Nous utiliserons les notations suivantes en travaillant à  $p$  fixé :

$$\begin{Bmatrix} \alpha_p \\ \beta_p \end{Bmatrix} = \begin{Bmatrix} a_{0\alpha} \\ a_{0\beta} \end{Bmatrix} + \begin{Bmatrix} a_{1\alpha} \\ a_{1\beta} \end{Bmatrix} \cos(\omega t) + \begin{Bmatrix} b_{1\alpha} \\ b_{1\beta} \end{Bmatrix} \sin(\omega t), \quad \omega = 3\Omega \quad (7.3)$$

**Valeurs numériques**

**Stator**

$\rho_s$	7800 kg.m <sup>-3</sup>	masse volumique
$b$	1 m	longueur
$h$	2 10 <sup>-2</sup> m	épaisseur
$S_s$	$b.h$ m <sup>2</sup>	section longitudinale
$g$	1.4 10 <sup>-2</sup> m	gap
$R_s$	$R_r + h/2 + g$ m	rayon du stator
$E_s$	2 10 <sup>11</sup> Pa	module d'Young
$I_s$	$bh^3/12$ m <sup>4</sup>	inertie longitudinale
$\eta_s$	2.75 10 <sup>-4</sup> s	coefficient d'amortissement visqueux

**Autre**

$k_{nl}$	$5.10^8 \text{ N.m}^{-3}$	raideur non-linéaire
$p$	3	numéro du mode observé
$\Omega$	310 ou 4474 tr/min	vitesse de rotation

### 7.2.2 Application des méthodes d'analyse par intervalles et d'homotopie polynomiale

On applique la méthode exposée au chapitre 5 reposant sur l'analyse des intervalles ainsi que la méthode d'homotopie polynomiale du chapitre 6 pour la résolution du système algébrique à 6 inconnues.

Pour la première, nous choisissons le domaine de recherche suivant :

$$[-5.10^{-2}, +5.10^{-2}] \text{ sur chacune des 6 composantes harmoniques}$$

et les valeurs suivantes pour les critères d'arrêt :

$$\epsilon_{Xr} = 10^{-2}, \epsilon_{Xa} = 10^{-3} \text{ et } \epsilon_H = 10^{-3}$$

Pour la méthode par homotopie polynomiale, nous utilisons une méthode de degré total qui, étant donnée la forme cubique de la non-linéarité, conduit à la continuation de  $3^6 = 729$  chemins. Les résultats sont donnés dans le tableau 7.1.

	Homotopie polynomiale		Analyse intervalles	
	Temps cpu (s)	Nombre de solutions	Temps cpu (s)	Nombre de solutions
$\Omega = 310 \text{ tr/min}$	249	1	$\infty$	x
$\Omega = 4471 \text{ tr/min}$	269	5	$\infty$	x

TAB. 7.1 – Contact rotor-stator non frottant - Résultats

Le temps de calcul pour la méthode intervalles est donné comme étant infini car le calcul n'avait pas abouti au bout de 24 heures, temps maximal que nous nous étions fixé. Les solutions trouvées dans le cas de la méthode par homotopie polynomiale sont données dans le tableau 7.2. Nous avons pour chacune d'elles réalisé une intégration temporelle représentée en figure 7.3. Les solutions 1 et 2 d'une part et 3 et 5 d'autre part donnent des résultats très similaires ; en effet, les coefficients du premier harmonique de ces deux couples sont de signes opposés. Ceci s'explique par la forme de la fonction non-linéaire  $\hat{H}(\tilde{x})$  et l'absence de terme constant sur les équations algébriques gouvernant ces variables : les deux premières composantes de cette fonction,  $\hat{H}_1(\tilde{x})$  et  $\hat{H}_2(\tilde{x})$  qui représentent la projection des équations dynamiques selon le terme constant sont paires par rapport au premier harmonique tandis que les suivantes  $\hat{H}_i(\tilde{x})$ ,  $3 \leq i \leq 6$  sont impaires. On peut donc écrire :

$$\begin{pmatrix} \hat{H}_1(\tilde{x}_1, \tilde{x}_2, -\tilde{x}_3, -\tilde{x}_4, -\tilde{x}_5, -\tilde{x}_6) \\ \hat{H}_2(\tilde{x}_1, \tilde{x}_2, -\tilde{x}_3, -\tilde{x}_4, -\tilde{x}_5, -\tilde{x}_6) \\ \hat{H}_3(\tilde{x}_1, \tilde{x}_2, -\tilde{x}_3, -\tilde{x}_4, -\tilde{x}_5, -\tilde{x}_6) \\ \hat{H}_4(\tilde{x}_1, \tilde{x}_2, -\tilde{x}_3, -\tilde{x}_4, -\tilde{x}_5, -\tilde{x}_6) \\ \hat{H}_5(\tilde{x}_1, \tilde{x}_2, -\tilde{x}_3, -\tilde{x}_4, -\tilde{x}_5, -\tilde{x}_6) \\ \hat{H}_6(\tilde{x}_1, \tilde{x}_2, -\tilde{x}_3, -\tilde{x}_4, -\tilde{x}_5, -\tilde{x}_6) \end{pmatrix} = \begin{pmatrix} \hat{H}_1(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5, \tilde{x}_6) \\ \hat{H}_2(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5, \tilde{x}_6) \\ -\hat{H}_3(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5, \tilde{x}_6) \\ -\hat{H}_4(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5, \tilde{x}_6) \\ -\hat{H}_5(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5, \tilde{x}_6) \\ -\hat{H}_6(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5, \tilde{x}_6) \end{pmatrix}$$

Ainsi, s'il existe une solution dont la composante harmonique  $(\tilde{x}_3, \dots, \tilde{x}_6)$  est non nulle, il en existe nécessairement une deuxième comportant le même terme constant  $(\tilde{x}_1, \tilde{x}_2)$  et une composante harmonique opposée. Les deux solutions présentent alors les mêmes cycles limites à l'opposition de phase près. Les solutions 1,2 et 3 correspondent vraisemblablement à des solutions instables.

$\Omega$ (tr/min)	310	4474 Sol. 1	4474 Sol. 2	4474 Sol. 3	4474 Sol. 4	4474 Sol. 5
$a_{0\alpha}$	$-8.52.10^{-3}$	$-1.24.10^{-6}$	$-1.24.10^{-6}$	$-1.23.10^{-6}$	$-1.23.10^{-6}$	$-1.23.10^{-6}$
$a_{0\beta}$	$-3.39.10^{-3}$	$6.86.10^{-4}$	$6.86.10^{-4}$	$6.85.10^{-4}$	$6.89.10^{-4}$	$6.85.10^{-4}$
$a_{1\alpha}$	0	$-2.81.10^{-2}$	$2.81.10^{-2}$	$7.76.10^{-7}$	0	$-7.76.10^{-7}$
$a_{1\beta}$	0	$-7.16.10^{-7}$	$7.16.10^{-7}$	$2.53.10^{-2}$	0	$-2.53.10^{-2}$
$b_{1\alpha}$	0	$-6.33.10^{-7}$	$6.33.10^{-7}$	$2.53.10^{-2}$	0	$-2.53.10^{-2}$
$b_{1\beta}$	0	$2.80.10^{-2}$	$-2.80.10^{-2}$	$-7.01.10^{-7}$	0	$7.01.10^{-7}$

TAB. 7.2 – Contact rotor-stator non frottant - Solutions

### 7.2.3 Conclusion

Cet exemple réaffirme la supériorité de la méthode par homotopie polynomiale sur la méthode par analyse intervalles. Elle permet de trouver toutes les solutions d'un problème dans des temps de l'ordre de la minute quand les autres méthodes requièrent plus d'une journée - les méthodes de cell-mapping et du simplexe étant plus longues encore, nous pouvons affirmer qu'elles n'auraient pas donné de meilleurs résultats que celle basée sur l'analyse intervalles.

Nous utiliserons donc uniquement cette méthode pour étudier les exemples suivants.

## 7.3 Rotor sur palier squeeze-film

Nous souhaitons appliquer la méthode retenue basée sur la résolution de systèmes polynomiaux par homotopie à une non-linéarité très actuelle dans la conception des rotors, à savoir un amortisseur de type squeeze-film. Après une courte présentation de cet organe, nous présentons le modèle utilisé, les résultats obtenus par une analyse classique et enfin ceux obtenus par analyse globale qui permettent d'en souligner l'intérêt.

### 7.3.1 Généralités sur l'organe squeeze-film

Un squeeze-film est un organe amortissant constitué d'un film fluide écrasé (généralement une huile) s'insérant entre le palier et le bâti (fig. 7.4). Sa seule fonction est la dissipation d'énergie de façon à réduire les amplitudes vibratoires. Le palier souple réalise toujours la fonction de portance de l'arbre auquel il est lié par l'intermédiaire de roulements. Il est à noter que la bague extérieure du roulement ne doit pas comporter de mouvement de rotation. Le squeeze-film ne réagit donc qu'aux déplacements transverses de l'arbre et non à sa rotation propre.

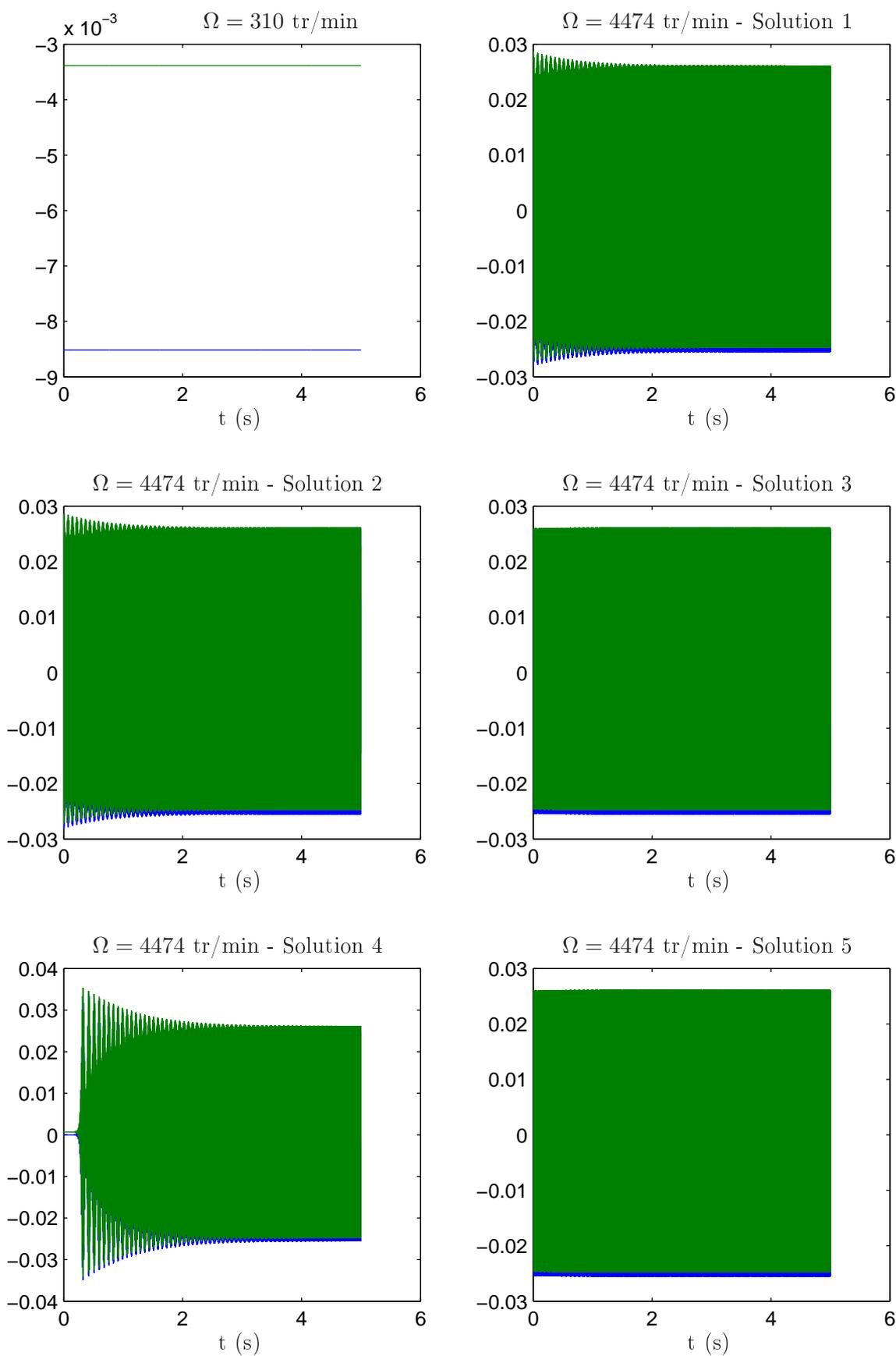


FIG. 7.3 – Contact rotor-stator non frottant - Intégration temporelle des solutions

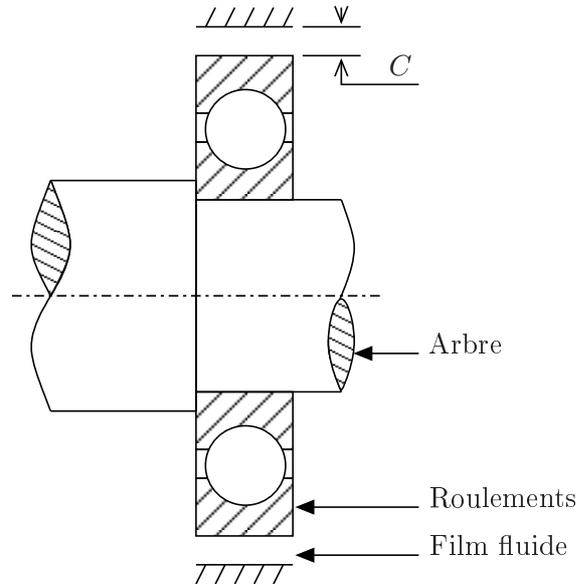


FIG. 7.4 – Schéma simplifié d'un palier comportant un amortisseur squeeze-film

Un grand nombre de modélisations différentes existent pour rendre compte de la physique d'un tel amortisseur. En effet, les forces hydrodynamiques mises en œuvre nécessitent de faire certaines hypothèses quant à la géométrie et à la plage de fonctionnement de l'organe pour pouvoir être simplifiées et ne pas rendre les temps de calcul prohibitifs. Pour une étude approfondie de cet organe, on pourra se reporter à la thèse réalisée par F. Assis Robriguez [AR03].

Nous avons pour notre part choisi d'utiliser une modélisation  $\pi$ -film comme suggéré dans l'article [ZH95]. Cette modélisation repose sur une hypothèse de paliers courts et de mouvements quasi circulaires et centrés de l'arbre dans la cage du squeeze-film.

### 7.3.2 Système dynamique étudié

On considère un modèle à palier souple ( $k_1$ ) et éventuellement dissipatif ( $c_1$ ) comportant un squeeze-film monté lui aussi sur palier souple ( $k_2$ ) et dissipatif ( $c_2$ ) (voir figure 7.5). Le rotor est modélisé par une masse  $m_1$  de coordonnées  $(x_1, y_1)$  et la bague extérieure du squeeze-film est modélisée par une masse  $m_2$  de coordonnées  $(x_2, y_2)$ .

On obtient aisément le système d'équations suivant :

$$\begin{cases} m_1 \ddot{x}_1 + c_1 \dot{x}_1 + k_1 x_1 = f_x + m_1 u \omega^2 \cos(\omega t) \\ m_1 \ddot{y}_1 + c_1 \dot{y}_1 + k_1 y_1 = f_y + m_1 u \omega^2 \sin(\omega t) \\ m_2 \ddot{x}_2 + c_2 \dot{x}_2 + k_2 x_2 = -f_x \\ m_2 \ddot{y}_2 + c_2 \dot{y}_2 + k_2 y_2 = -f_y \end{cases} \quad (7.4)$$

$f_x$  et  $f_y$  représentent l'action de l'huile sur l'arbre et dépendent des 8 variables  $x_1, x_2, y_1, y_2, \dot{x}_1, \dot{x}_2, \dot{y}_1, \dot{y}_2$ . Le terme en  $m_1 u \omega^2$  correspond quant à lui à l'excitation par balourd.

Puis, en divisant le système (7.4) par  $/(m_1 C \omega^2)$  et avec les notations précisées ensuite, on

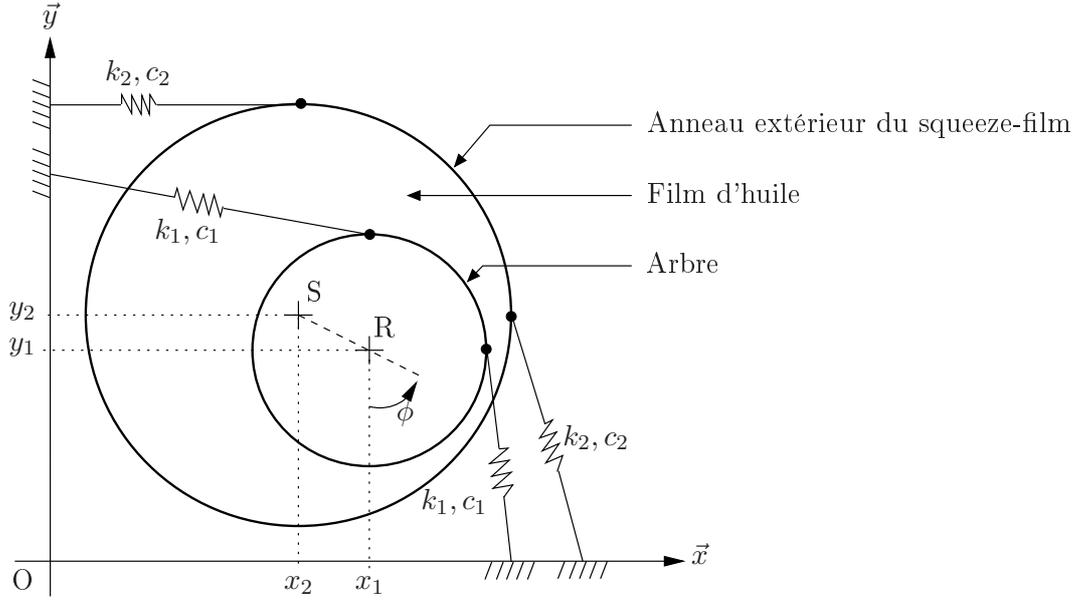


FIG. 7.5 – Squeeze-film - Modèle dynamique étudié

obtient le système adimensionné ci-dessous :

$$M\ddot{q} + D\dot{q} + Kq = \begin{Bmatrix} F_x \\ F_y \\ -F_x \\ -F_y \end{Bmatrix} + \begin{Bmatrix} U \cos \tau \\ U \sin \tau \\ 0 \\ 0 \end{Bmatrix} \quad (7.5)$$

### Notations

$$\tau = \omega t, \quad \dot{z} = \partial z / \partial \tau, \quad s = \omega / \omega_1$$

$$\alpha = m_2 / m_1, \quad \beta = k_2 / k_1$$

$$\xi_1 = c_1 / (m_1 \omega_1), \quad \xi_2 = c_2 / (m_2 \omega_2)$$

$$X = x / C : X_1 / C, \quad U = u / C, \quad F_x = f_x / (m_1 C \omega^2)$$

$$q = \{X_1, Y_1, X_2, Y_2\}^t$$

$$\epsilon = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

Matrices :

$$M = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \alpha & \\ & & & \alpha \end{bmatrix}, \quad D = \frac{1}{s} \begin{bmatrix} \xi_1 & & & \\ & \xi_1 & & \\ & & \xi_2 \sqrt{\alpha \beta} & \\ & & & \xi_2 \sqrt{\alpha \beta} \end{bmatrix}, \quad K = \frac{1}{s^2} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \beta & \\ & & & \beta \end{bmatrix}$$

Hypothèse palier court en  $\pi$ -film :

$$F_r = \left( -\frac{B}{s} \right) \left[ \frac{\pi}{2} \frac{1 + 2\epsilon^2}{(1 - \epsilon^2)^{5/2}} \dot{\epsilon} + 2 \frac{\epsilon^2}{(1 - \epsilon^2)^2} \dot{\phi} \right]$$

$$F_t = \left( -\frac{B}{s} \right) \left[ 2 \frac{\epsilon}{(1 - \epsilon^2)^2} \dot{\epsilon} + \frac{\pi}{2} \frac{\epsilon}{(1 - \epsilon^2)^{3/2}} \dot{\phi} \right]$$

$$F_x = F_t \cos \phi + F_r \sin \phi$$

$$F_y = -F_r \cos \phi + F_t \sin \phi$$

### 7.3.3 Quelques résultats sous hypothèses de mouvement circulaire synchrone centré

Nous avons souhaité dans un premier temps retrouver les résultats présentés dans [ZH95] dans le cas particulier des mouvements circulaires synchrones centrés. Sous ce lot d'hypothèses, les expressions du mouvement et de l'effort squeeze-film se simplifient comme suit :

$$\begin{cases} q(t) = \begin{cases} a_1 \\ b_1 \\ a_2 \\ b_2 \end{cases} \cos \tau + \begin{cases} b_1 \\ -a_1 \\ b_2 \\ -a_2 \end{cases} \sin \tau \\ F_r = \left(-\frac{B}{s}\right) \left(2 \frac{\epsilon^2}{(1-\epsilon^2)^2} \omega\right), F_t = \left(-\frac{B}{s}\right) \left(\frac{\pi}{2} \frac{\epsilon}{(1-\epsilon^2)^{3/2}} \omega\right) \end{cases} \quad (7.6)$$

Ainsi, bien qu'utilisant une décomposition harmonique des déplacements, on ne récupère que quatre inconnues algébriques. Seule une projection de l'équation dynamique sur  $\cos \tau$  sera donc utilisée, une projection sur  $\sin \tau$  étant dans ce cas redondante. On obtient donc le système d'équations algébriques à résoudre suivant :

$$\left( K - M + D \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 0 & -1 \\ 1 & 0 \end{bmatrix} \right) \tilde{x} + \hat{H}(\tilde{x}) = H_c \quad (7.7)$$

$$\text{avec } \hat{H}(\tilde{x}) = - \begin{cases} H_x \\ H_y \\ -H_x \\ -H_y \end{cases}, \begin{cases} H_x(\tilde{x}) = \frac{1}{\pi} \int_0^{2\pi} F_x(\tilde{x}) \cos(\tau) d\tau \\ H_y(\tilde{x}) = \frac{1}{\pi} \int_0^{2\pi} F_y(\tilde{x}) \cos(\tau) d\tau \end{cases} \quad \text{et } H_c = \begin{cases} U \\ 0 \\ 0 \\ 0 \end{cases}.$$

En utilisant l'ensemble de valeurs numériques suivant, on obtient alors les courbes 7.6 et 7.7 pour lesquelles on a fait varier la fréquence d'excitation (à travers le paramètre  $s$ ) et le rapport des raideurs  $\beta$  de 0.01 jusqu'à  $10^6$  qui représente le cas d'un support squeeze-film raide (configuration classique) :  $B = 0.1$ ,  $U = 0.4$ ,  $\xi_1 = 0$ ,  $\xi_2 = 0.05$ ,  $\alpha = 0.05$

### 7.3.4 Analyse globale par homotopies polynomiales

Nous retranscrivons maintenant l'étude globale réalisée par la méthode d'homotopie polynomiale. Nous commençons par expliquer le lissage, puis nous comparons les temps de calcul nécessaires pour l'application de différentes versions de la méthode homotopique.

#### Lissage polynomial de l'effort non-linéaire

Afin d'appliquer une méthode par homotopie polynomiale, il faut commencer par approximer les efforts du squeeze-film. Nous réalisons donc un lissage. Étant donné que l'effort squeeze-film ne fait intervenir les quatre variables  $X_1$ ,  $X_2$ ,  $Y_1$  et  $Y_2$  que sous la forme de différence  $X_1 - X_2$  et  $Y_1 - Y_2$ , on effectue le changement de variables 7.8 pour que l'effort ne

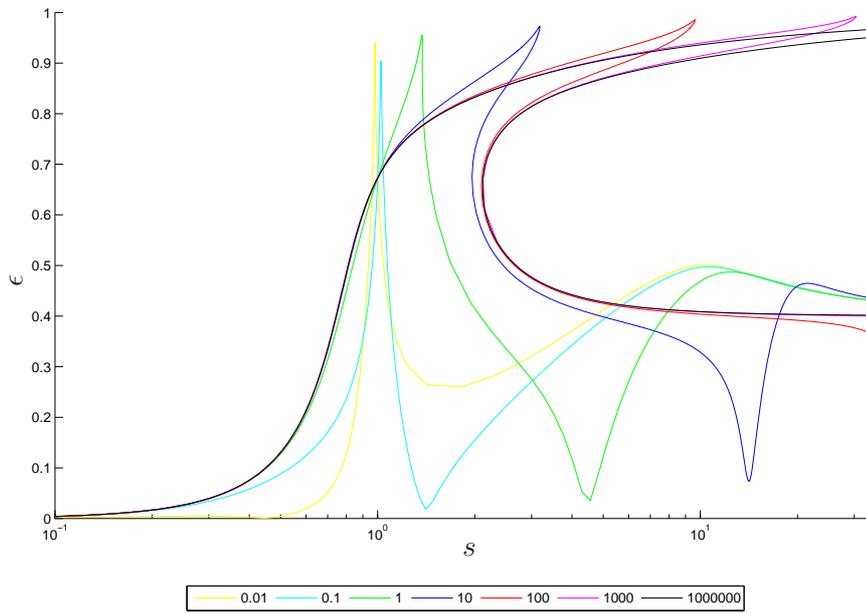


FIG. 7.6 – Squeeze-film - Mouvement circulaire synchrone centré  
Influence de  $\beta$  sur  $\epsilon = |RS|/C$

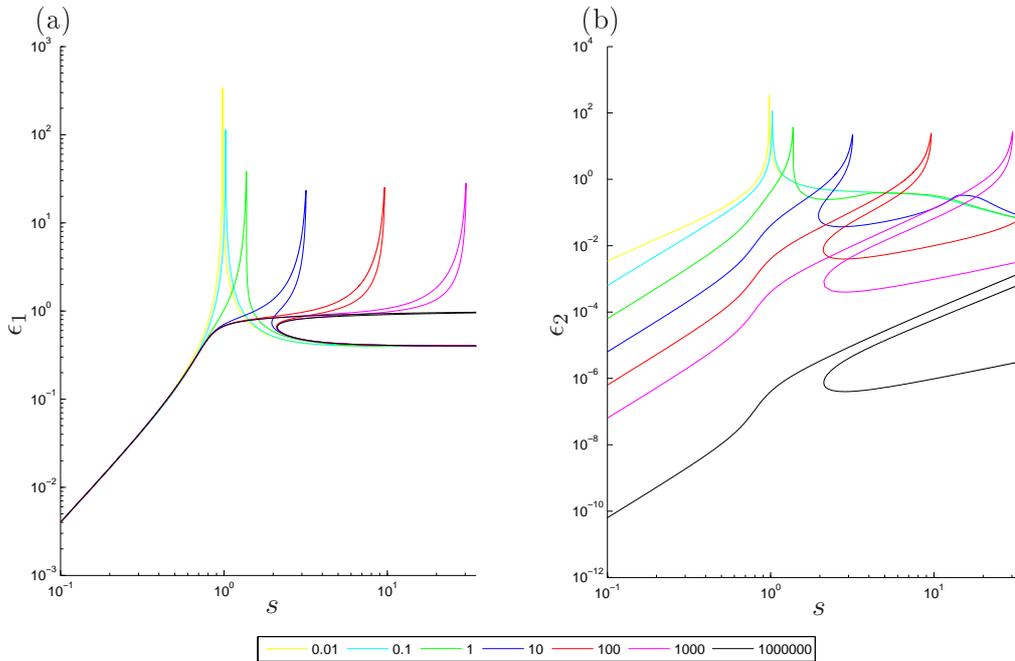


FIG. 7.7 – Squeeze-film - Mouvement circulaire synchrone centré - Influence de  $\beta$  sur :  
(a)  $\epsilon_1 = |OR|/C$  et (b)  $\epsilon_2 = |OS|/C$

dépende que des deux variables de déplacement relatif  $X_r$  et  $Y_r$  et non des quatre initiales.

$$q_r = \begin{Bmatrix} X_1 \\ Y_1 \\ X_2 - X_1 \\ Y_2 - Y_1 \end{Bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ -1 & & 1 & \\ & -1 & & 1 \end{bmatrix} q \quad (7.8)$$

La décomposition harmonique selon l'hypothèse de mouvement circulaire synchrone centré permet ensuite d'écrire le système avec les quatre inconnues algébriques  $a_1$ ,  $b_1$ ,  $a_r$  et  $b_r$  :

$$q_r = \begin{Bmatrix} a_1 \\ b_1 \\ a_r \\ b_r \end{Bmatrix} \cos(\tau) + \begin{Bmatrix} b_1 \\ -a_1 \\ b_r \\ -a_r \end{Bmatrix} \sin(\tau) \quad (7.9)$$

Nous effectuons le lissage en utilisant 3000 points  $P_n = (a_r, b_r)_n$  générés aléatoirement sous la contrainte que  $0 \leq \epsilon = \sqrt{a_r^2 + b_r^2} \leq 0.6$ . En effet, les efforts du squeeze-film tendent vers l'infini lorsque  $\epsilon$  tend vers 1, ce qui se représente mal polynomialement et ce type de modèle n'est pas représentatif de la physique au-delà de  $\epsilon = 0.6$ . Il ne sert donc à rien d'effectuer un lissage sur des points reflétant une excentricité supérieure à cette valeur. Nous calculons pour ces 3000 points les décompositions harmoniques  $H_x$  et  $H_y$  de  $F_x$  et  $F_y$  sur  $\cos \tau$ . Nous réalisons donc enfin le lissage de ces efforts par deux polynômes  $P_x(a_r, b_r)$  et  $P_y(a_r, b_r)$ .

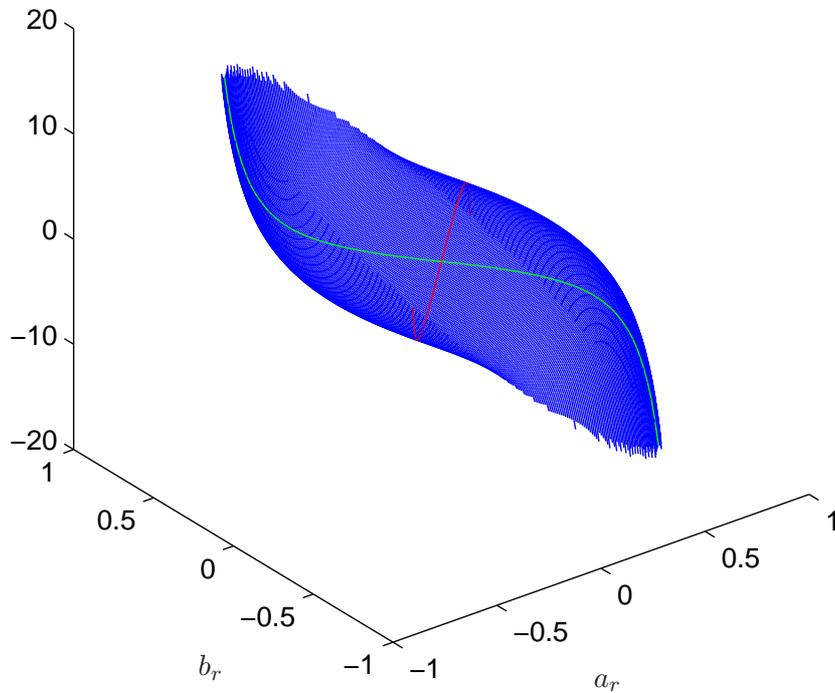
Nous avons testé deux approximations ; une utilisant des polynômes de degré 6 et une seconde utilisant des polynômes de degré 12. Nous estimons alors l'erreur commise par l'approximation polynomiale pour chacune de ces deux configurations : la figure 7.8.(a) représente en 3 dimensions la surface de réponse  $H_x$  en fonction de  $a_r$  et  $b_r$ . Les figures 7.8.(b) et 7.8.(c) représentent respectivement les valeurs obtenues par les deux approximations degré 6 et degré 12 selon les lignes de coupe rouge et verte représentées en figure 7.8.(a). Nous affinons cette estimation de l'erreur avec les figures 7.9 et 7.10 qui représentent l'erreur relative moyenne commise par l'approximation selon des couronnes de largeur 0.05 en excentricité. Sur la plage d'excentricités pour laquelle le lissage a été effectué, l'erreur est en moyenne de 12% pour le degré 6 et de 4% pour le degré 12. Ces moyennes sont dégradées par l'erreur sur les très faibles excentricités ( $\epsilon < 0.2$ ) ce qui veut dire que l'amplitude de la force non-linéaire est faible ; ces erreurs n'auront donc qu'un faible impact sur la solution trouvée.

### Application de la méthode homotopique

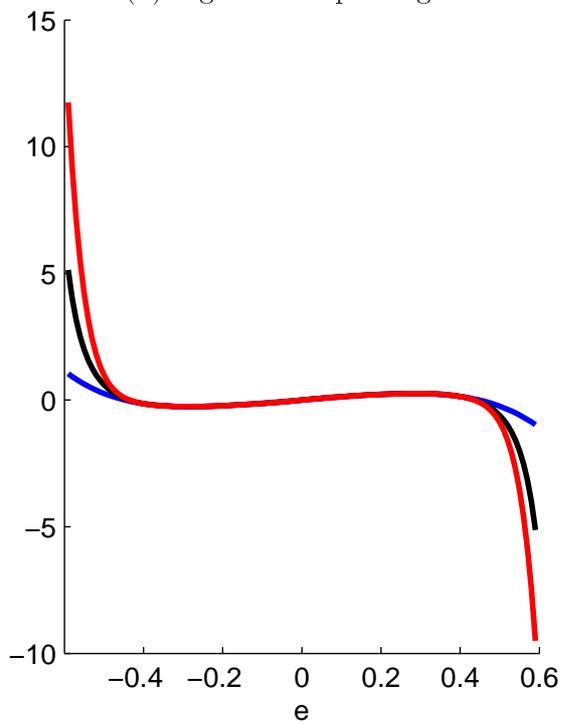
Nous pouvons appliquer la méthode homotopique de différentes façons :

- homotopie de degré total sur système complet
- homotopie de degré total sur système réduit
- homotopie par multi-homogénéisation sur système complet

La méthode homotopique de degré total considère les quatre inconnues  $a_1$ ,  $b_1$ ,  $a_r$  et  $b_r$  comme « identiques ». On obtient alors selon le degré  $d$  utilisé pour le lissage  $d^4$  chemins. La méthode homotopique par multi-homogénéisation consiste à partitionner les variables en deux groupes :  $a_1$  et  $b_1$ , linéaires et  $a_r$  et  $b_r$  non-linéaires ; on obtient alors  $6d^2$  chemins à suivre. Enfin, on

(a) Fonction vraie  $H_x$ 

(b) Ligne de coupe rouge



(c) Ligne de coupe verte

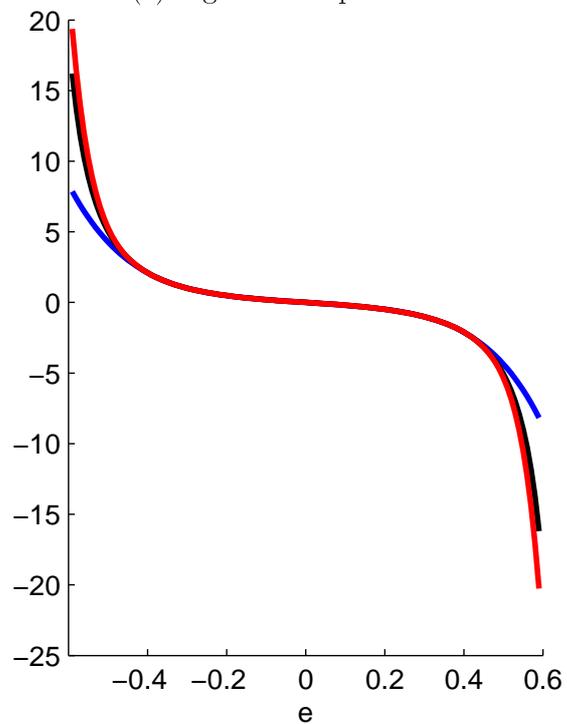


FIG. 7.8 – Squeeze-film - Erreur de l'approximation polynomiale

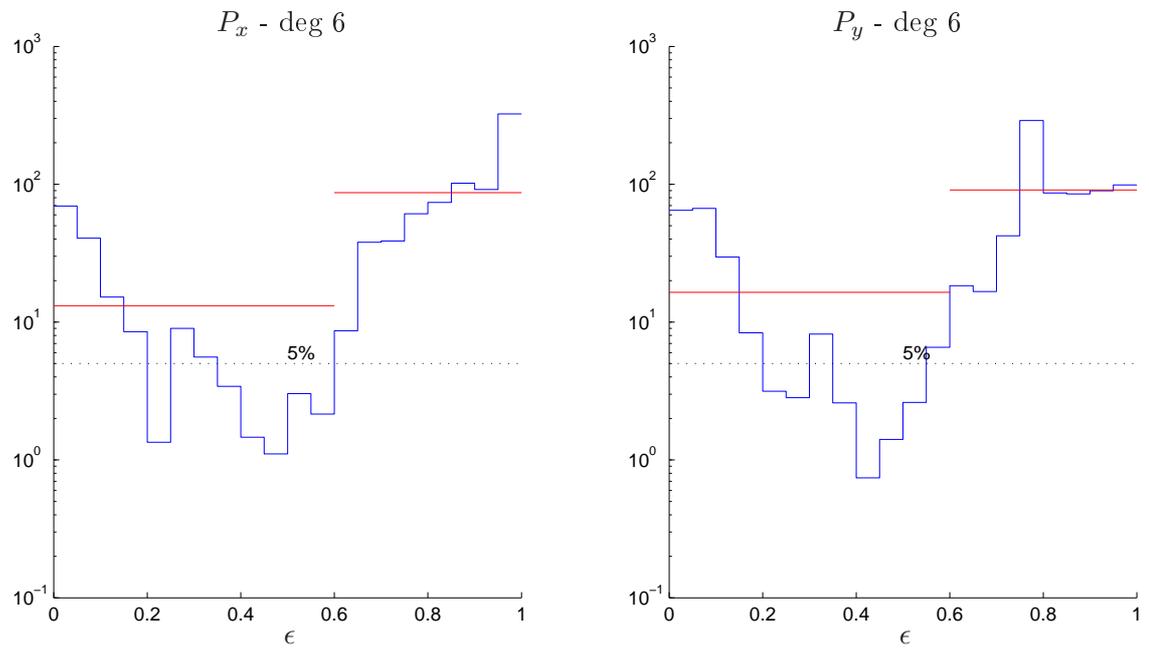


FIG. 7.9 – Squeeze-film - Erreur relative moyenne par couronne (%) - Degré 6

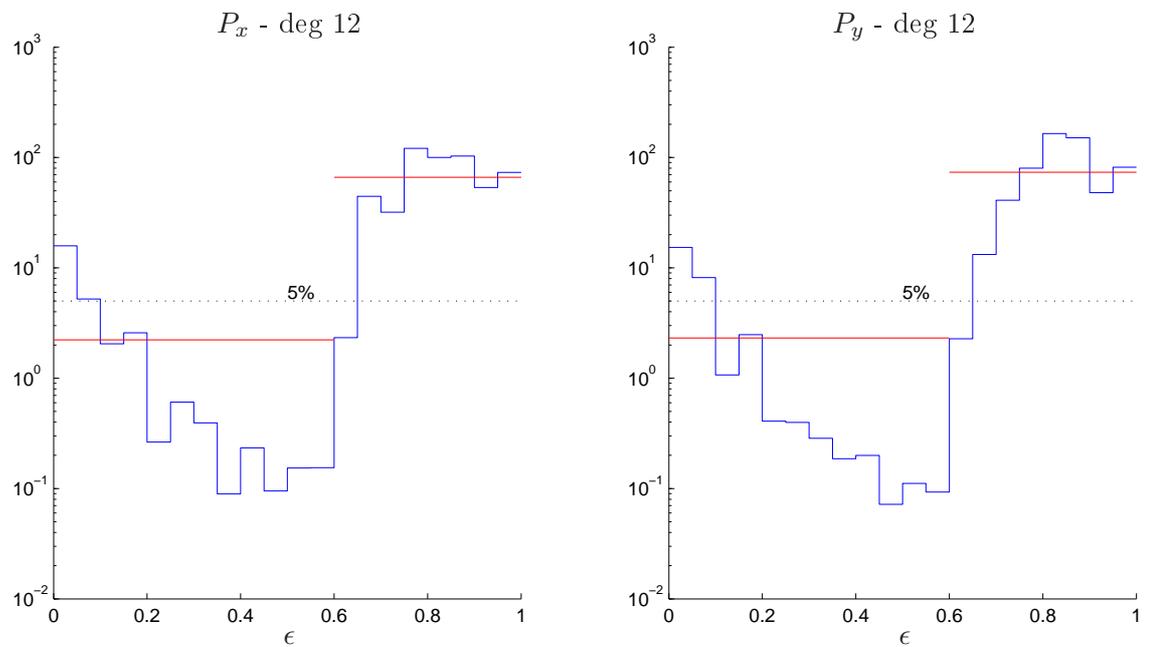


FIG. 7.10 – Squeeze-film - Erreur relative moyenne par couronne (%) - Degré 12

peut réaliser une réduction comme proposé en section 2.5. On ne traite alors le problème qu'à l'aide de deux équations à deux inconnues non-linéaires générant donc  $d^2$  chemins.

Le tableau 7.3 récapitule les différents résultats obtenus pour une même valeur de  $s$ .

	Deg 6		Deg 12	
	Chemins	Temps	Chemins	Temps
Degré total - Système complet	1296	7 min	20736	4 h
Degré total - Système réduit	36	4 s	144	36 s
2-homogénéisation - Système complet	216	69 s	864	10 min

TAB. 7.3 – Squeeze-film - Résultats des différentes méthodes homotopiques

La méthode reposant sur une réduction du système est donc sans conteste la plus intéressante ; cependant, la méthode par 2-homogénéisation affiche des temps tout à fait raisonnables. La méthode donne donc pleine satisfaction sur les temps de calcul. Elle nous a de surcroît permis de découvrir une seconde branche de solutions : en l'appliquant pour différentes fréquences d'excitation ( $s$ ) et pour  $\beta = 1$ , nous avons découvert une seconde branche de solutions qui n'apparaissait pas sur la figure 7.6. La figure 7.11 rend compte de ces résultats : les points verts représentent les solutions obtenues par analyse globale pour diverses valeurs de  $s$  et les courbes bleues représentent les solutions obtenues par continuation en résolvant le système sans approximation polynomiale (7.7). Il existe un décalage entre les résultats par homotopie polynomiale et la continuation pour la seconde branche de solutions (en haut à droite). Ceci est dû d'une part à l'erreur commise par l'approximation pour les valeurs d'excentricité élevées et d'autre part à la forme « plane » de la surface de réponse  $H(\tilde{x})$  dans ces zones, c'est-à-dire une faible sensibilité. Une erreur importante sur  $\tilde{x}$  n'entraîne donc qu'une variation faible de  $H(\tilde{x})$  et permet les écarts observés. Nous avons enfin réalisé une analyse de stabilité des deux branches dont nous rendons compte en figure 7.12. Les figures (a) et (b) représentent les coefficients de Floquet  $\rho$  calculés le long de chacune des courbes pour statuer sur la stabilité. La sortie du cercle unité de ces coefficients par 1 correspond bien aux bifurcations de type point de retournement observées en figure (c). Cette seconde branche était donc « indétectable » sans analyse globale dans la mesure où elle n'est pas issue d'une bifurcation. Rien ne nous aurait donc naturellement poussés à chercher d'autres solutions que celles obtenues par continuation. Enfin, nous avons réalisé des intégrations temporelles pour différentes valeurs de l'excitation ; les quatre graphiques de la figure 7.13 représentent l'évolution temporelle des centres rotor (en noir) et squeeze-film (en bleu), en utilisant des points de plus en plus gros au cours du temps. On voit ainsi que la seconde branche (figure 7.13.(d)) se distingue par une excentricité plus importante sur la bague intérieure du squeeze-film que l'excentricité du rotor alors que c'est l'inverse dans les autres cas.

### 7.3.5 Conclusion

Cet exemple de rotor sur palier à amortisseur squeeze-film montre tout l'intérêt de posséder un outil d'analyse globale. En appliquant les méthodes homotopiques, nous avons mis à jour une branche de solutions déconnectée de celle obtenue classiquement par continuation d'une solution trouvée par recherche locale. Sans cet outil, l'existence de cette branche de solutions

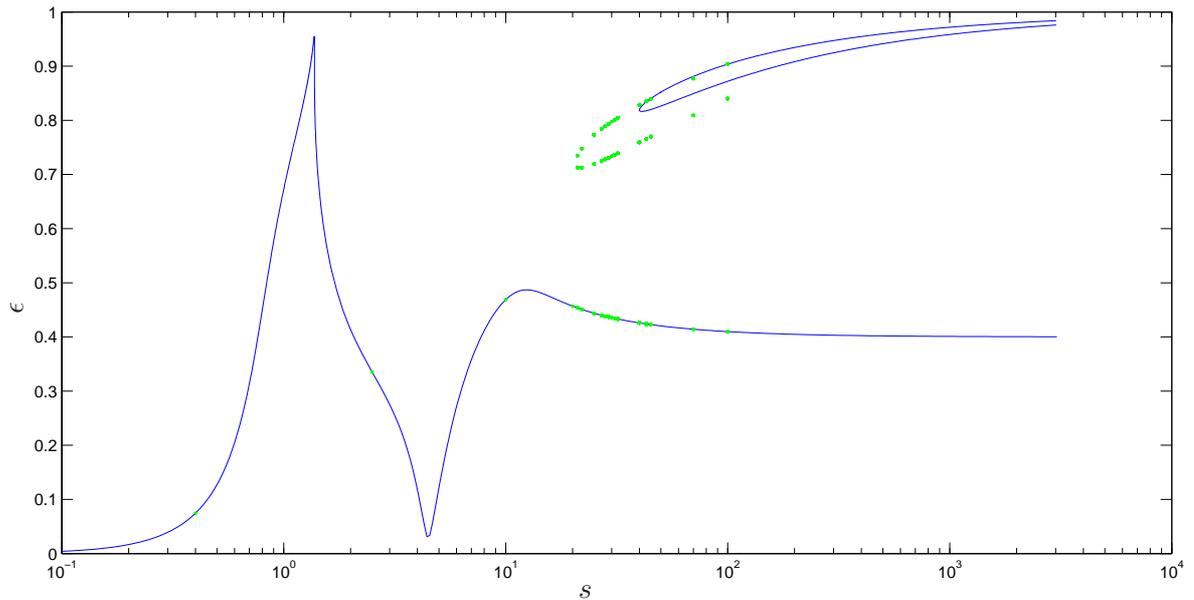


FIG. 7.11 – Squeeze-film - Découverte d'une nouvelle branche de solutions

n'aurait pu être décelée étant donné qu'elle ne prend pas naissance après une bifurcation de l'autre branche ; rien n'aurait donc pu laisser imaginer son existence.

Nous avons aussi, au travers de cette application, mis en œuvre une méthode de réduction du système qui permet d'abaisser encore les coûts de calcul, les faisant passer de 10 minutes dans le cas d'un système complet traité par 2-homogénéisation - ce qui est très raisonnable - à 36 secondes.

On entrevoit donc la possibilité de traiter des systèmes d'une taille adaptée à celle des systèmes non-linéaires rencontrés dans le milieu industriel.

## 7.4 Contact rotor-stator rigide

Cette dernière illustration consiste à étudier un système tournant à trois degrés de liberté. Outre l'intérêt que peuvent présenter les résultats, ce système qui, bien que petit, s'est avéré riche en comportements inattendus, nous a permis de mettre en application la plupart des techniques de recherche de solutions, continuation, analyse de stabilité et bifurcation exposées au chapitre 2.

Nous exposons dans un premier temps le système, ses caractéristiques et sa mise en équations. Puis, nous effectuons l'étude des points fixes, ainsi que celle de leur stabilité et de leurs bifurcations. Dans une dernière section, nous effectuons une ébauche d'analyse globale.

### 7.4.1 Système étudié

On cherche à étudier le contact entre un rotor et son stator. Ce contact arrive le plus souvent en présence d'un balourd provoquant une excentricité importante du rotor qui entre alors en contact avec le stator. Nous ne cherchons pas à étudier la non-linéarité forte engendrée par le passage d'un état sans contact à un état avec contact, mais uniquement le comportement

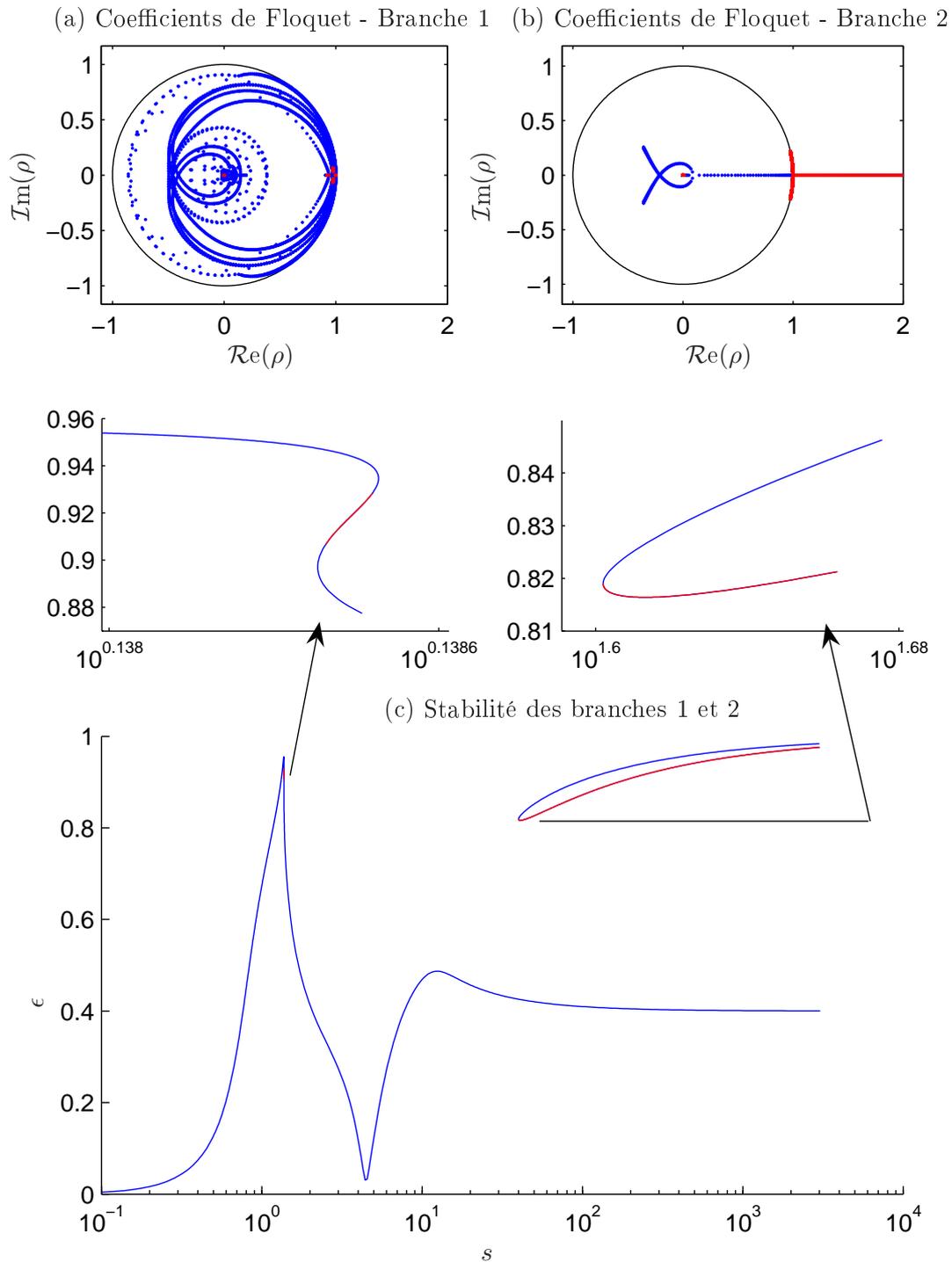
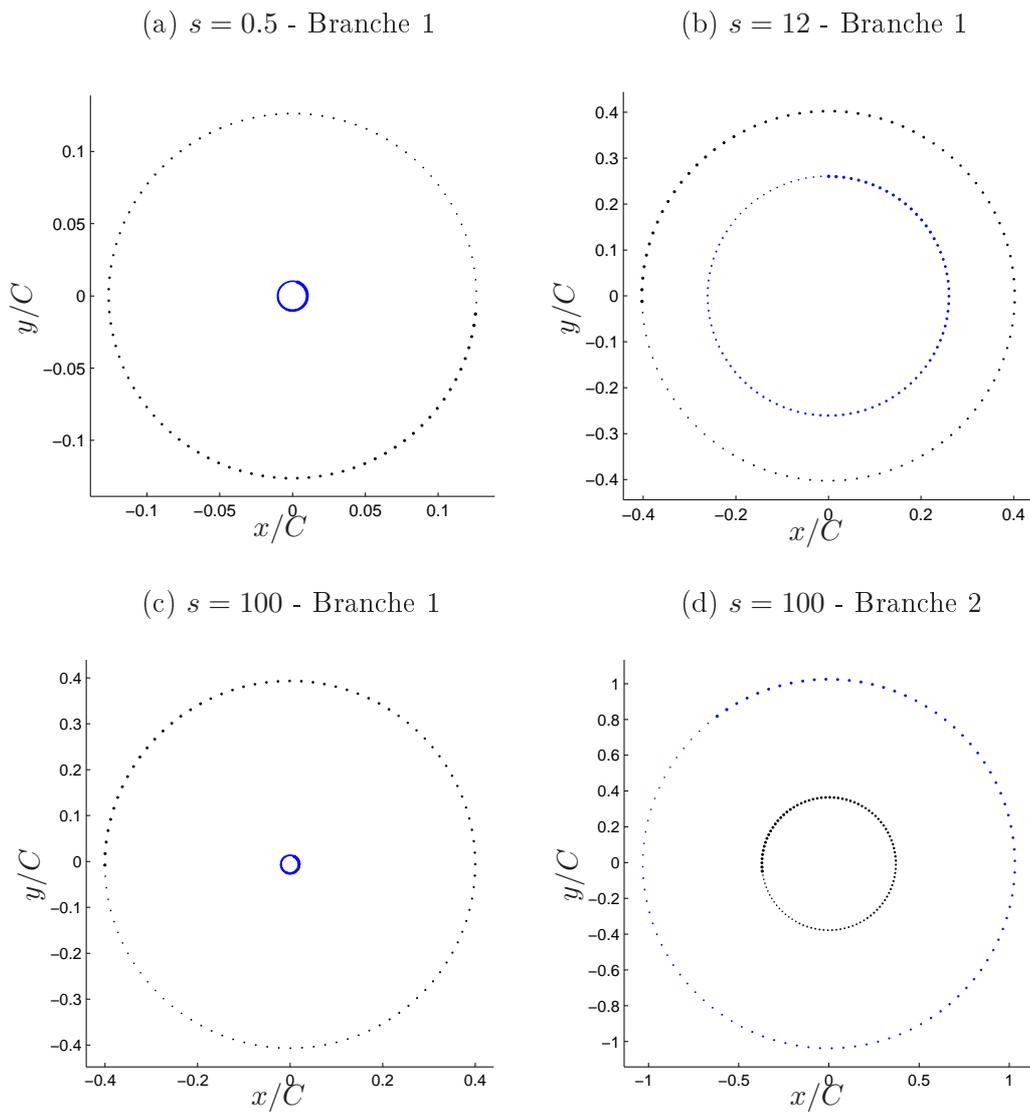


FIG. 7.12 – Squeeze-film - Stabilité des branches



dans le cas où le contact, établi, est gardé.

Nous faisons les hypothèses suivantes :

- contact rotor-stator permanent
- pas de frottement dans le contact
- effets gyroscopiques négligés

Comme on le voit sur les figures 7.14 et 7.15, le rotor et le stator sont soumis à une force de rappel élastique et à une dissipation que l'on modélisera sous sa forme hystérétique ou visqueuse selon les cas dans la suite.  $\vec{F}_c$  est la force de contact exercée par le rotor sur le stator. Le contact étant sans frottement, elle est alignée avec  $S\vec{R}$ .  $\vec{F}_e$  est la force d'excitation, de module  $F_e$  et de phase  $\theta_e$  par rapport au stator. Enfin, l'hypothèse de contact impose la relation géométrique suivante :  $\vec{r} = \vec{s} + g(\cos \theta_s \vec{X} + \sin \theta_s \vec{Y})$ . Les notations de la figure 7.15 sont définies ci-dessous :

$$\begin{aligned} s &= \overline{OS} & r &= \overline{OR} & g &= SR = cste \\ \alpha &= \angle(\vec{x}, \vec{X}) & \theta_s &= \angle(\vec{X}, S\vec{R}) = \angle(\vec{X}, \vec{F}_c) & \theta_e &= \angle(\vec{X}, \vec{F}_e) \end{aligned}$$

Deux cas sont étudiés :

- stator rigide : sa fréquence propre est supérieure à celle du rotor ( $\omega_S > \omega_R$ )
- stator souple : sa fréquence propre est inférieure à celle du rotor ( $\omega_S < \omega_R$ )

On a donc les deux jeux de valeurs suivants :

**Stator « rigide » :**

$$\begin{aligned} m_R &= 10 \text{ kg} \\ f_R &= \omega_R/2\pi = 10 \text{ Hz} \\ \eta_R &= 0.001 \text{ s (coefficient visqueux)} \\ \eta_R &= 0.02 \text{ (coefficient hystérétique)} \\ m_S &= 10 \text{ kg} \\ f_S &= \omega_S/2\pi = 15 \text{ Hz} \\ g &= 2 \text{ mm (jeu radial)} \end{aligned}$$

**Stator « souple » :**

$$\begin{aligned} m_R &= 10 \text{ kg} \\ f_R &= \omega_R/2\pi = 15 \text{ Hz} \\ \eta_R &= 0.001 \text{ s (coefficient visqueux)} \\ \eta_R &= 0.02 \text{ (coefficient hystérétique)} \\ m_S &= 10 \text{ kg} \\ f_S &= \omega_S/2\pi = 10 \text{ Hz} \\ g &= 2 \text{ mm (jeu radial)} \end{aligned}$$

## Équations générales

L'écriture des équations de ce système avec contact n'étant pas triviale, nous en donnons le détail. Nous choisissons une description du mouvement dans le référentiel tournant  $(O, \vec{X}, \vec{Y})$ .

**Quantités dynamiques :** Pour établir les équations, on a besoin des expressions de  $\vec{r}$  et  $\vec{s}$  et de leurs dérivées premières et secondes par rapport au temps. Nous écrivons ici ces quantités dans la base  $(\vec{X}, \vec{Y})$  en tenant compte de la relation géométrique entre  $\vec{r}$  et  $\vec{s}$ .

$$\vec{r} = \begin{Bmatrix} s + g \cos \theta_s \\ g \sin \theta_s \end{Bmatrix} \quad (7.10a)$$

$$\dot{\vec{r}} = \begin{Bmatrix} \dot{s} - g(\dot{\theta}_s + \dot{\alpha}) \sin \theta_s \\ s\dot{\alpha} + g(\dot{\theta}_s + \dot{\alpha}) \cos \theta_s \end{Bmatrix} \quad (7.10b)$$

$$\ddot{\vec{r}} = \begin{Bmatrix} \ddot{s} - g(\ddot{\theta}_s + \ddot{\alpha}) \sin \theta_s - s\dot{\alpha}^2 - g(\dot{\theta}_s + \dot{\alpha})^2 \cos \theta_s \\ s\ddot{\alpha} + g(\ddot{\theta}_s + \ddot{\alpha}) \cos \theta_s + 2\dot{s}\dot{\alpha} - g(\dot{\theta}_s + \dot{\alpha})^2 \sin \theta_s \end{Bmatrix} \quad (7.10c)$$

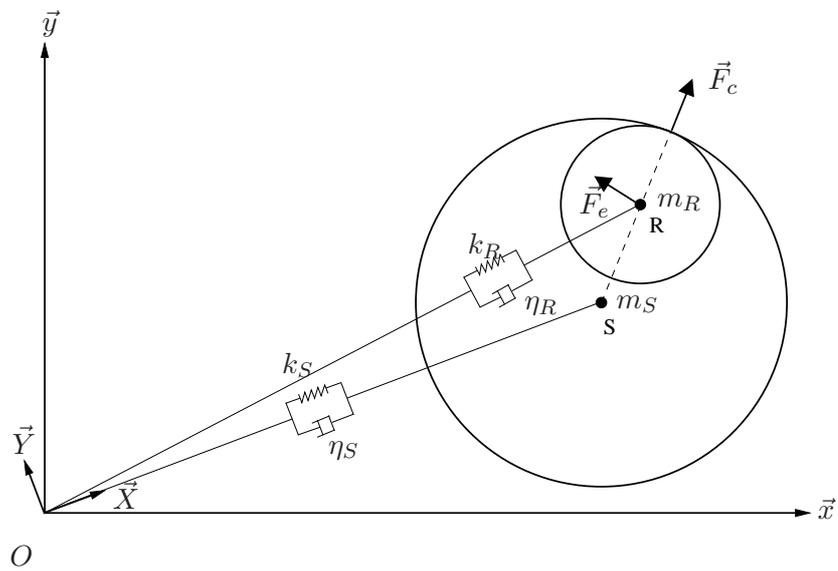


FIG. 7.14 – Contact rotor-stator rigide - Modèle

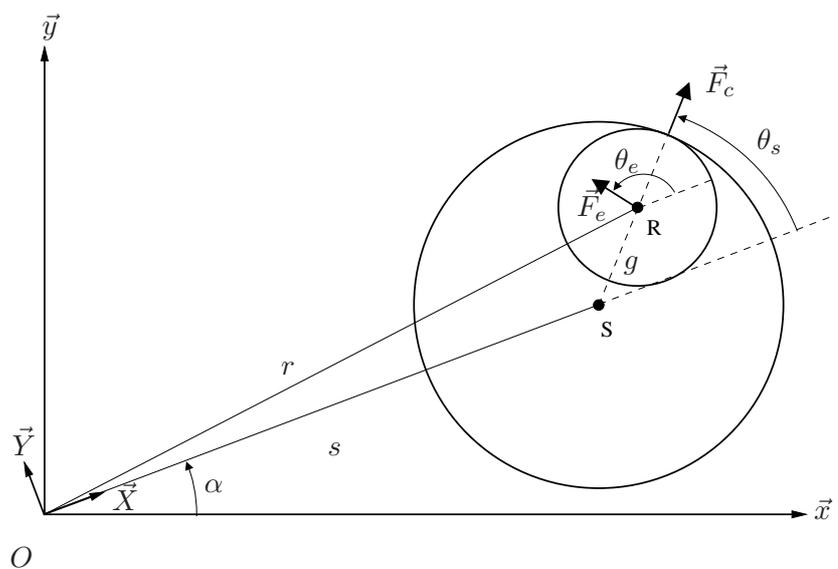


FIG. 7.15 – Contact rotor-stator rigide - Notations

$$\vec{s} = \begin{Bmatrix} s \\ 0 \end{Bmatrix} \quad (7.11a)$$

$$\dot{\vec{s}} = \begin{Bmatrix} \dot{s} \\ s\dot{\alpha} \end{Bmatrix} \quad (7.11b)$$

$$\ddot{\vec{s}} = \begin{Bmatrix} \ddot{s} - s\dot{\alpha}^2 \\ s\ddot{\alpha} + 2\dot{s}\dot{\alpha} \end{Bmatrix} \quad (7.11c)$$

**Excitation :** L'excitation, choisie de façon analogue à celle d'un balourd en ce sens qu'elle tourne avec le rotor, génère la force suivante :

$$\vec{F}_e(t) = Fe \begin{Bmatrix} \cos(\omega_e t) \\ \sin(\omega_e t) \end{Bmatrix}_{(\vec{x}, \vec{y})} = Fe \begin{Bmatrix} \cos(\omega_e t - \alpha) \\ \sin(\omega_e t - \alpha) \end{Bmatrix}_{(\vec{X}, \vec{Y})} \quad (7.12)$$

Ceci implique la relation :

$$\theta_e(t) = \omega_e t - \alpha(t) \quad (7.13)$$

**Équations :** L'écriture du principe fondamental de la dynamique appliqué aux deux sous-structures  $R$  et  $S$  donne pour un amortissement visqueux :

$$\begin{aligned} m_R \left( \ddot{s} - g(\ddot{\theta}_s + \ddot{\alpha}) \sin \theta_s - g(\dot{\theta}_s + \dot{\alpha})^2 \cos \theta_s - s\dot{\alpha}^2 \right) \\ = -(k_R \eta_R) \left( \dot{s} - g(\dot{\theta}_s + \dot{\alpha}) \sin \theta_s \right) - k_R (s + g \cos \theta_s) \\ + F_e \cos \theta_e - F_c \cos \theta_s \end{aligned} \quad (7.14a)$$

$$\begin{aligned} m_R \left( s\ddot{\alpha} + g(\ddot{\theta}_s + \ddot{\alpha}) \cos \theta_s + 2\dot{s}\dot{\alpha} - g(\dot{\theta}_s + \dot{\alpha})^2 \sin \theta_s \right) \\ = -(k_R \eta_R) \left( s\dot{\alpha} + g(\dot{\theta}_s + \dot{\alpha}) \cos \theta_s \right) - k_R (g \sin \theta_s) \\ + F_e \sin \theta_e - F_c \sin \theta_s \end{aligned} \quad (7.14b)$$

$$m_S (\ddot{s} - s\dot{\alpha}^2) = -(k_S \eta_S) \dot{s} - k_S s + F_c \cos \theta_s \quad (7.14c)$$

$$m_S (s\ddot{\alpha} + 2\dot{s}\dot{\alpha}) = -(k_S \eta_S) s\dot{\alpha} + F_c \sin \theta_s \quad (7.14d)$$

Ces écritures font apparaître  $F_c$  qui est une « fausse » inconnue dans la mesure où c'est une force de contact interne. On l'élimine en utilisant les combinaisons suivantes : (7.14a)+(7.14c), (7.14b)+(7.14d) et (7.14c)tan  $\theta_s$ -(7.14d).

On obtient alors le système d'équations suivant, dont les inconnues sont  $s$ ,  $\theta_s$  et  $\alpha$  :

$$\begin{aligned} (m_R + m_S)(\ddot{s} - s\dot{\alpha}^2) - m_R g \left( (\ddot{\theta}_s + \ddot{\alpha}) \sin \theta_s + (\dot{\theta}_s + \dot{\alpha})^2 \cos \theta_s \right) \\ = -(k_R \eta_R + k_S \eta_S) \dot{s} + (k_R \eta_R) g (\dot{\theta}_s + \dot{\alpha}) \sin \theta_s \\ - (k_R + k_S) s - k_R g \cos \theta_s + F_e \cos \theta_e \end{aligned} \quad (7.15a)$$

$$\begin{aligned} (m_R + m_S)(s\ddot{\alpha} + 2\dot{s}\dot{\alpha}) + m_R g \left( (\ddot{\theta}_s + \ddot{\alpha}) \cos \theta_s - (\dot{\theta}_s + \dot{\alpha})^2 \sin \theta_s \right) \\ = -(k_R \eta_R + k_S \eta_S)(s\dot{\alpha}) - (k_R \eta_R) g (\dot{\theta}_s + \dot{\alpha}) \cos \theta_s \\ - k_R g \sin \theta_s + F_e \sin \theta_e \end{aligned} \quad (7.15b)$$

$$m_S ((\ddot{s} - s\dot{\alpha}^2) \tan \theta_s - (s\ddot{\alpha} + 2\dot{s}\dot{\alpha})) = -(k_S \eta_S)(\dot{s} \tan \theta_s - s\dot{\alpha}) - k_S (s \tan \theta_s) \quad (7.15c)$$

$$\text{avec } \theta_e = \omega_e t - \alpha$$

La seule hypothèse ayant été faite ici est qu'il y a contact permanent entre le rotor et le stator, ce qui est équivalent à une force de contact  $F_c$  positive. Une quatrième équation peut être utilisée pour calculer  $F_c$  et vérifier a posteriori cette hypothèse, par exemple, l'équation (7.14d).

**Forme augmentée des équations :** Pour étudier complètement ce système nous aurons besoin de son expression sous forme augmentée de façon à le ramener à une formulation identique à (1.5).

Notre système est actuellement sous la forme :  $B(q)\ddot{q} = \tilde{F}(q, \dot{q}, t, \omega_e)$ , avec :

$$q = \begin{Bmatrix} s \\ \theta_s \\ \alpha \end{Bmatrix}, \quad B(q) = \begin{bmatrix} m_R + m_S & -m_{RG} \sin \theta_s & -m_{RG} \sin \theta_s \\ 0 & m_{RG} \cos \theta_s & m_{RG} \cos \theta_s + (m_R + m_S) s \\ m_S \tan \theta_s & 0 & -m_S s \end{bmatrix}$$

et

$$\tilde{F}(q, \dot{q}, \omega_e, t) = \left\{ \begin{array}{l} (m_R + m_S) s \dot{\alpha}^2 + g m_R (\dot{\theta}_s + \dot{\alpha})^2 \cos(\theta_s) - (k_R \eta_R + k_S \eta_S) \dot{s} \dots \\ \dots + k_R \eta_R g (\dot{\theta}_s + \dot{\alpha}) \sin(\theta_s) - (k_R + k_S) s - k_R g \cos(\theta_s) + F_e \cos(\theta_e) \\ (-2m_R - 2m_S) \dot{s} \dot{\alpha} + g m_R (\dot{\theta}_s + \dot{\alpha})^2 \sin(\theta_s) - (k_R \eta_R + k_S \eta_S) s \dot{\alpha} \dots \\ \dots - k_R \eta_R g (\dot{\theta}_s + \dot{\alpha}) \cos(\theta_s) - k_R g \sin(\theta_s) + F_e \sin(\theta_e) \\ m_S s \dot{\alpha}^2 \tan(\theta_s) + 2m_S \dot{s} \dot{\alpha} - k_S \eta_S (\dot{s} \tan(\theta_s) - s \dot{\alpha}) - k_S s \tan(\theta_s) \end{array} \right\}$$

Or on a besoin de le mettre sous la forme  $\dot{x} = F(x, t, \omega_e)$ . Il faut donc écrire  $G(x, t, \omega_e) = B(q)^{-1} \tilde{F}(q, \dot{q}, t, \omega_e)$  où  $x = \{q, \dot{q}\}^t$  est le vecteur d'inconnues augmenté.

On peut alors poser notre problème de deux façons :

a) Sous forme d'un système non-autonome :

$$x = \begin{Bmatrix} s \\ \theta_s \\ \alpha \\ \hat{s} \\ \hat{\theta}_s \\ \hat{\alpha} \end{Bmatrix}, \quad \dot{x} = F(x, t, \omega_e), \quad F(x, t, \omega_e) = \begin{Bmatrix} \hat{s} \\ \hat{\theta}_s \\ \hat{\alpha} \\ G(x, \omega_e, t) \end{Bmatrix} \quad (7.16)$$

où l'excitation s'écrit  $F_e \cos(\omega_e t - \alpha)$ ,  $F_e \sin(\omega_e t - \alpha)$

b) Sous forme d'un système autonome :

$$x = \begin{pmatrix} s \\ \theta_s \\ \theta_e \\ \hat{s} \\ \hat{\theta}_s \\ \hat{\alpha} \end{pmatrix}, \quad \dot{x} = F(x, \omega_e), \quad F(x, \omega_e) = \begin{pmatrix} \hat{s} \\ \hat{\theta}_s \\ \omega_e - \hat{\alpha} \\ G(x, \omega_e) \end{pmatrix} \quad (7.17)$$

où l'excitation s'écrit  $F_e \cos(\theta_e)$ ,  $F_e \sin(\theta_e)$

La seconde forme - autonome - nous a semblé plus judicieuse pour l'étude de nos solutions dans la mesure où, le problème ainsi posé, ces solutions correspondent à des points fixes, alors qu'en utilisant la formulation (7.16), les solutions sont périodiques moyennant la considération de  $\alpha[2\pi]$ , ce qui est numériquement plus long à traiter (étape d'intégration temporelle).

Le reste de l'étude se fera donc en utilisant la forme (7.17) des équations.

#### 7.4.2 Étude des points fixes

On cherche tout d'abord les points fixes du système (7.17) car ce sont les plus simples à trouver numériquement et qu'ils possèdent de surcroît une signification physique particulière. C'est en effet le cas le plus fréquemment rencontré en pratique : le mouvement de whirl est synchrone avec la vitesse de rotation utile  $\Omega$  du rotor. On a alors le rotor qui entraîne le stator dans son mouvement et ainsi  $\dot{\alpha} = \omega_e = \pm\Omega = cste$ . On a donc, une fois une configuration d'équilibre en  $F_e, \theta_e, s, \theta_s, F_c$  et  $\omega$  trouvée,  $\dot{s} = 0$ ,  $\dot{\theta}_s = 0$ ,  $\dot{\theta}_e = 0$  et  $\dot{\alpha} = \omega_e$  comme le montrent les équations (7.18) et (7.19).

$$\dot{x} = 0 \Leftrightarrow F(x, \omega_e) = 0 \Leftrightarrow \begin{pmatrix} \hat{s} \\ \hat{\theta}_s \\ \omega_e - \hat{\alpha} \\ G(x, \omega_e) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0_{3,1} \end{pmatrix} \quad (7.18)$$

dont on peut déduire depuis les trois premières lignes que :

$$\begin{cases} \dot{s} = \hat{s} = 0 & \Rightarrow \ddot{s} = 0 \\ \dot{\theta}_s = \hat{\theta}_s = 0 & \Rightarrow \ddot{\theta}_s = 0 \\ \dot{\theta}_e = \omega_e - \hat{\alpha} = 0 & \Leftrightarrow \dot{\alpha} = \omega_e \Rightarrow \ddot{\alpha} = 0 \end{cases} \quad (7.19)$$

Ces résultats ont pour conséquence de transformer les équations (7.14) de la façon suivante pour un amortissement de type visqueux :

$$(m_R \omega_e^2 - k_R)(s + g \cos \theta_s) + (k_R \eta_R g) \omega_e \sin \theta_s - F_c \cos \theta_s + F_e \cos \theta_e = 0 \quad (7.20a)$$

$$(m_R \omega_e^2 - k_R)g \sin \theta_s - (k_R \eta_R) \omega_e (s + g \cos \theta_s) - F_c \sin \theta_s + F_e \sin \theta_e = 0 \quad (7.20b)$$

$$(m_S \omega_e^2 - k_S)s + F_c \cos \theta_s = 0 \quad (7.20c)$$

$$-(k_S \eta_S) \omega_e s + F_c \sin \theta_s = 0 \quad (7.20d)$$

Dans le cas d'un amortissement de type hystérétique, on obtient :

$$(m_R \omega_e^2 - k_R)(s + g \cos \theta_s) + k_R \eta_R g \sin \theta_s - F_c \cos \theta_s + F_e \cos \theta_e = 0 \quad (7.21a)$$

$$(m_R \omega_e^2 - k_R)g \sin \theta_s - k_R \eta_R (s + g \cos \theta_s) - F_c \sin \theta_s + F_e \sin \theta_e = 0 \quad (7.21b)$$

$$(m_S \omega_e^2 - k_S)s + F_c \cos \theta_s = 0 \quad (7.21c)$$

$$-\eta_S k_S s + F_c \sin \theta_s = 0 \quad (7.21d)$$

### Fonctions de réponse en fréquence

On s'intéresse au tracé des solutions de ce système pour une amplitude d'excitation  $F_e$  donnée lorsque la pulsation d'excitation  $\omega_e$  varie. Nous utilisons pour cela une méthode de continuation comme expliquée en section 2.4 en utilisant une prédiction par tangente et un correcteur par pseudo-inverse de Moore-Penrose. Nous étudions d'abord le système à amortissement hystérétique pour les deux configurations stator souple et stator rigide. Nous nous intéressons ensuite au cas visqueux pour le cas du stator rigide.

**Résultats en amortissement hystérétique :** Nous considérons d'abord une modélisation de l'amortissement sous forme hystérétique ; ce sont donc les équations (7.21) qui sont utilisées. Nous avons tracé pour cette modélisation les fonctions de réponse en fréquence pour le cas où le stator est considéré comme rigide (plus rigide que le rotor) et celui où il est considéré comme souple. Dans chacun des cas, nous appliquons différentes amplitudes de la force d'excitation correspondant à différents niveaux de balourd. Les figures 7.16 et 7.17 rendent compte des résultats dans le cas du stator rigide tandis que les figures 7.18 et 7.19 rendent compte du jeu de paramètres où le stator est plus souple que le rotor ( $f_S = 10 \text{ Hz} < f_R = 15 \text{ Hz}$ ).

Dans les deux cas, mais de façon plus prononcée lorsque le stator est souple, on voit apparaître pour les faibles balourds, des îlots de solutions qui se replient voire se referment sur eux-mêmes. Ces îlots indiquent qu'il existe des solutions d'un autre type que celui étudié pour certaines valeurs de  $\omega_e$ .

Sans étude ultérieure de stabilité, nous ne pouvons pas développer plus l'analyse de ces résultats. Cependant, ceci nous a permis de valider nos codes dédiés à la continuation, réalisant avec succès la description de courbes présentant des retournements.

**Résultats en amortissement visqueux :** Pour la suite de l'étude, nous utiliserons un amortissement visqueux car nous aurons besoin de réaliser des intégrations temporelles et l'amortissement hystérétique s'y prête mal. Nous utiliserons donc dorénavant le système d'équations (7.20). Enfin, nous choisissons un niveau de balourd de 200  $N$ . L'étude ultérieure sera donc basée sur les points fixes représentés sur la figure 7.20 (en unités S.I.) correspondant à un rotor rigide.

### Étude de stabilité

Après avoir obtenu les courbes de solutions de type point fixe dans le cas d'un amortissement visqueux et d'un stator rigide, nous étudions leur stabilité. Cette étude ne peut pas se faire sur les équations simplifiées (7.20) : en effet, ces simplifications ne laissent pas de place

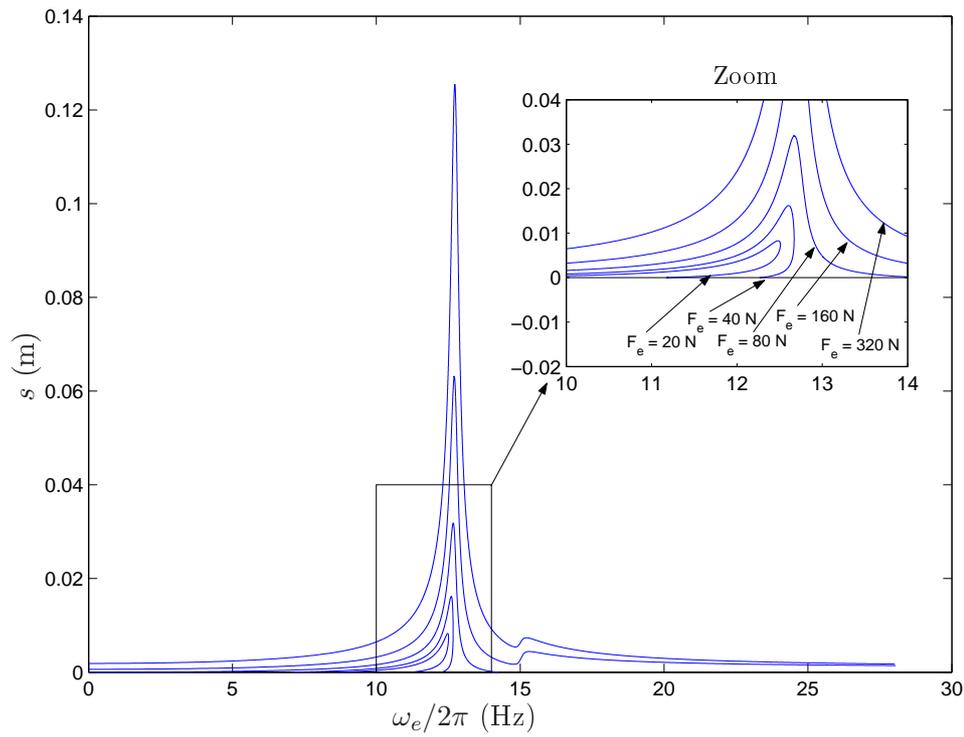


FIG. 7.16 – Contact rotor-stator rigide

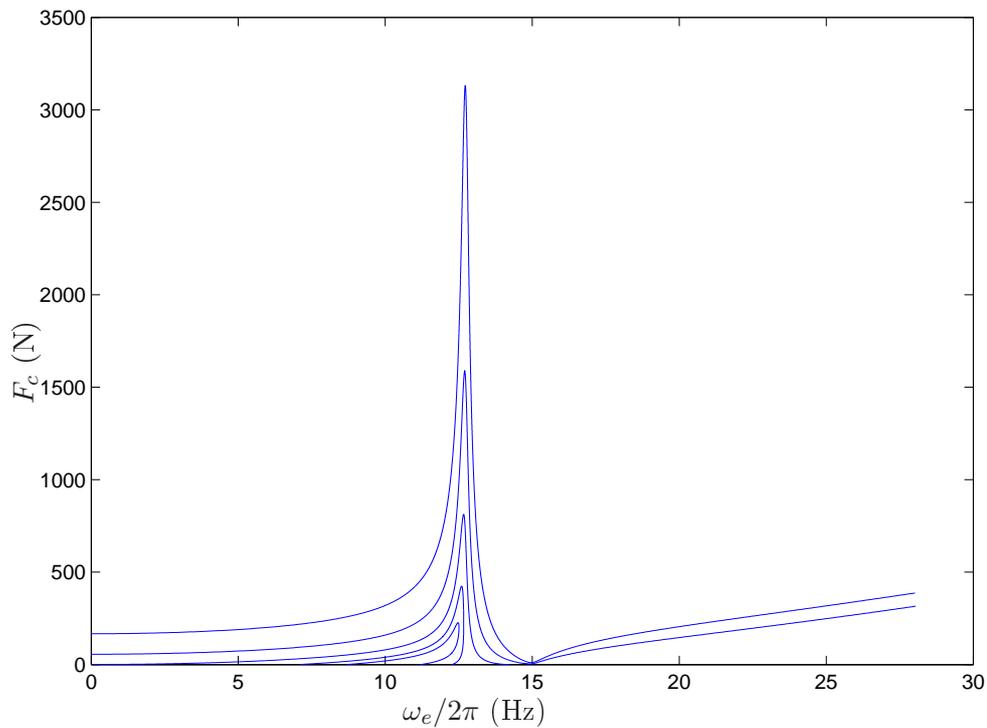
Étude fréquentielle sous amortissement hystérétique - Déplacement  $s$ 

FIG. 7.17 – Contact rotor-stator rigide

Étude fréquentielle sous amortissement hystérétique - Force de contact  $F_c$

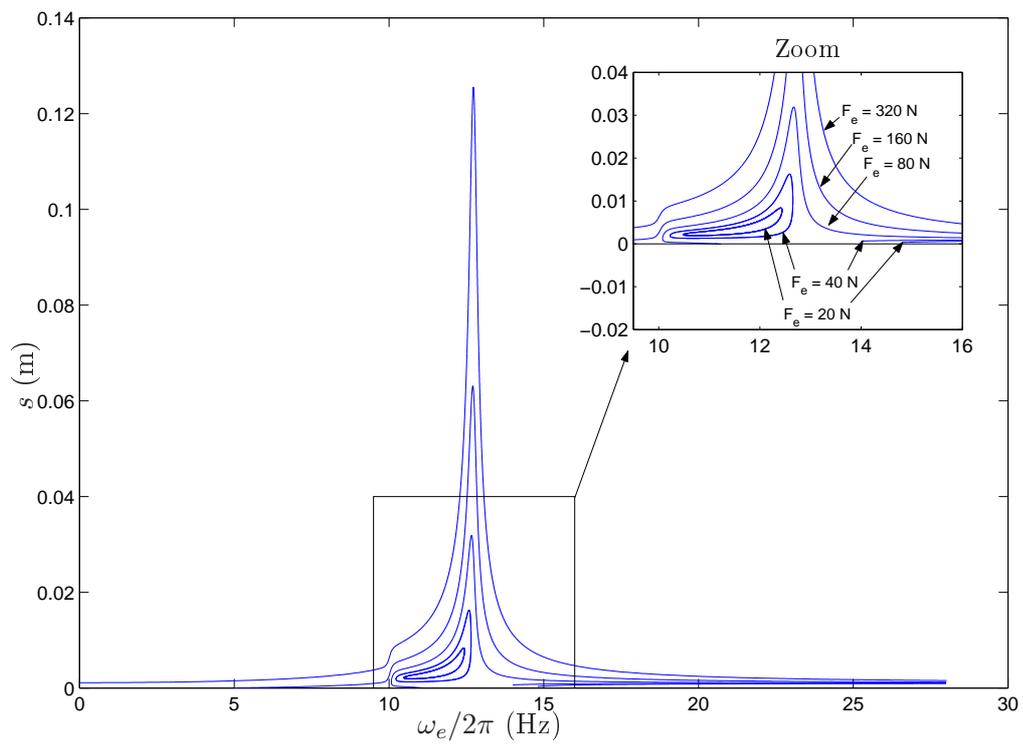


FIG. 7.18 – Contact rotor-stator souple  
Étude fréquentielle sous amortissement hystérétique - Déplacement  $s$

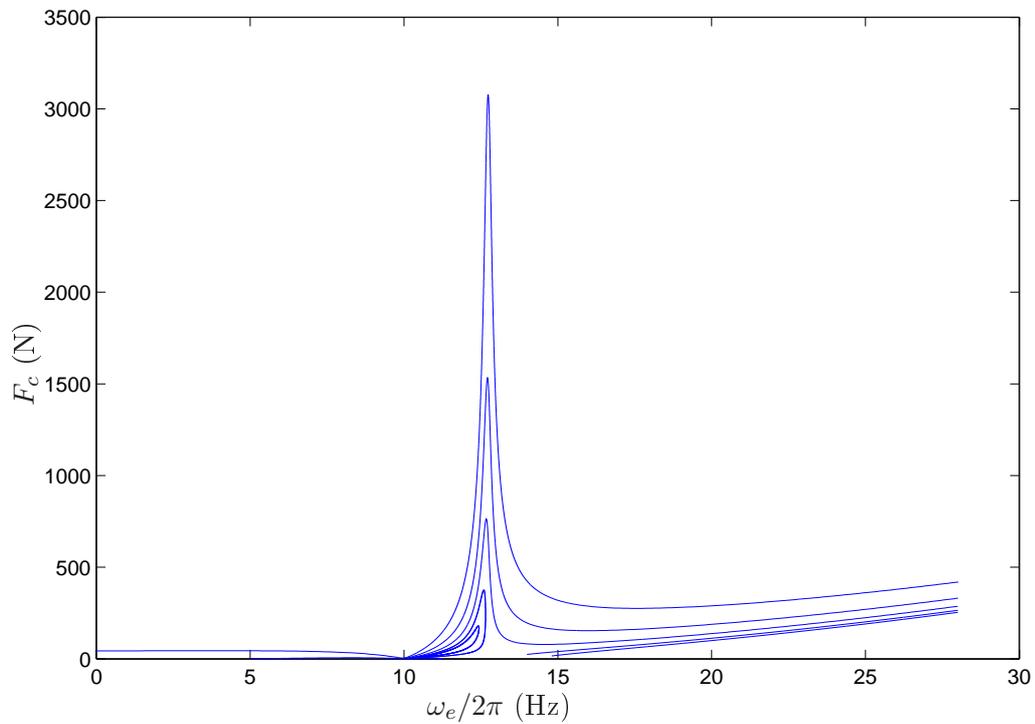


FIG. 7.19 – Contact rotor-stator souple  
Étude fréquentielle sous amortissement hystérétique - Force de contact  $F_c$

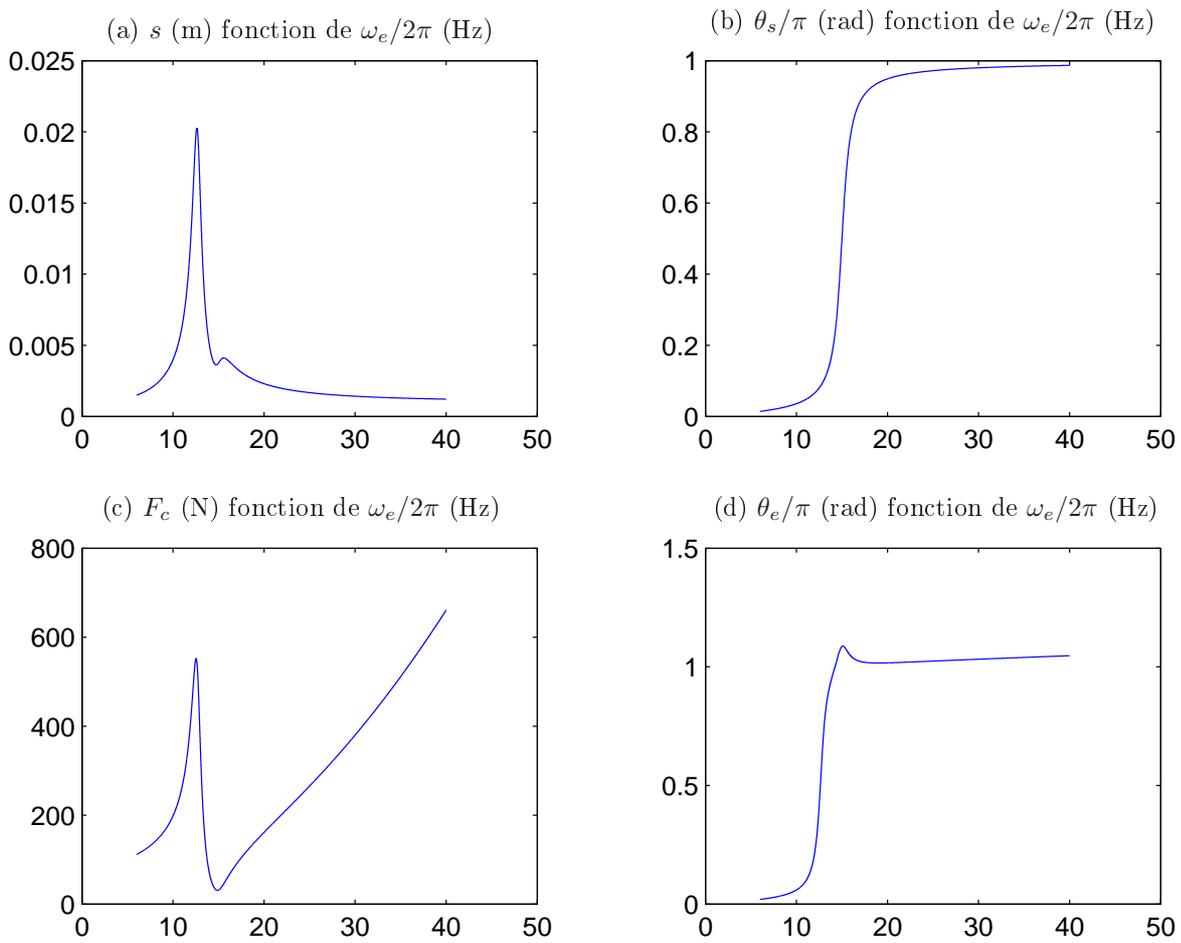


FIG. 7.20 – Contact rotor-stator rigide - Étude fréquentielle sous amortissement visqueux

à l'évolution d'une perturbation. Il faut donc reprendre la formulation complète (7.17) du problème et appliquer l'étude proposée au paragraphe 2.3.1 en réinjectant dans ces équations les solutions trouvées perturbées.

On obtient alors la courbe 7.21, qui donne deux zones d'instabilité par lecture du signe de la partie réelle des valeurs propres  $\lambda_i$  :  $\omega_e/2\pi \in [14.49 \ 16.16]$  Hz et  $\omega_e/2\pi \in [23.54 \ 30.72]$  Hz.

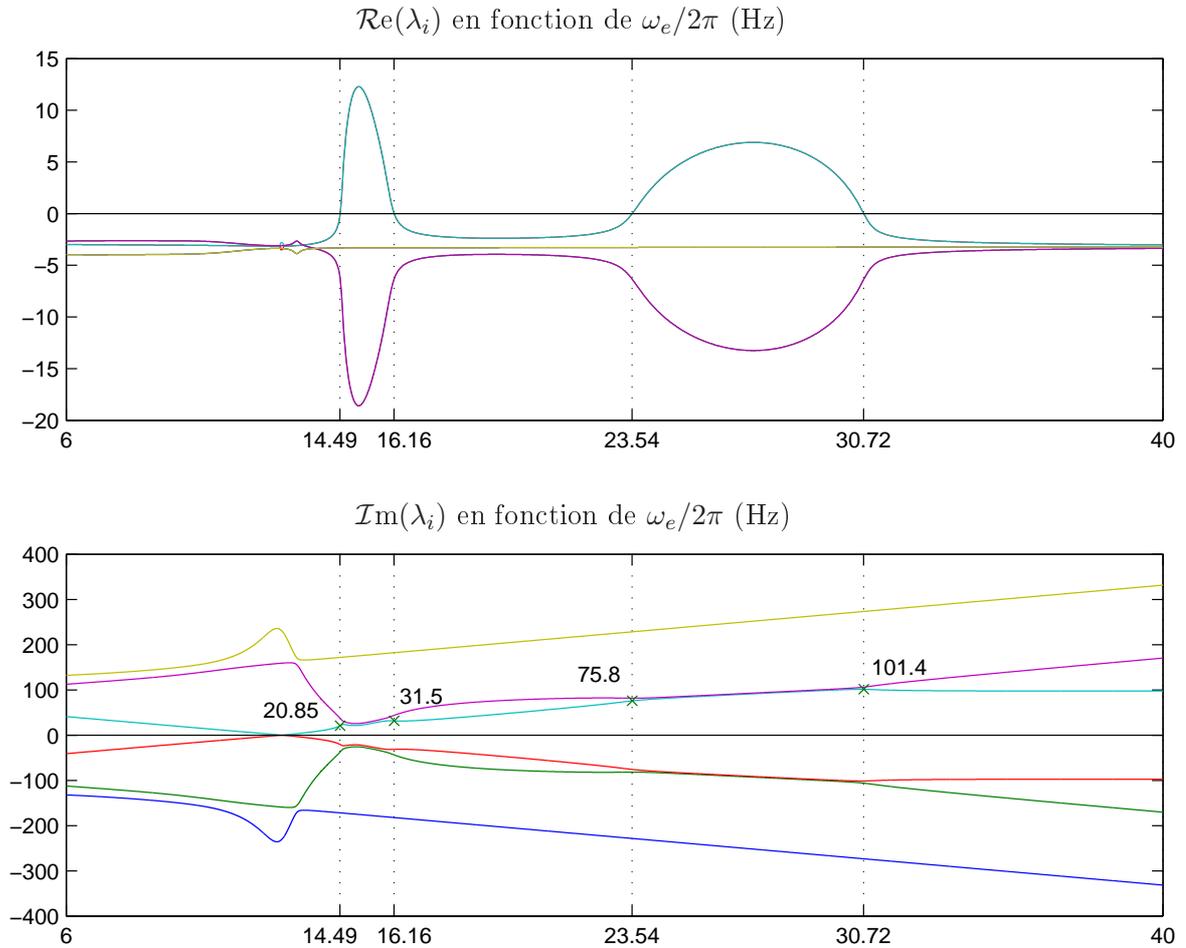


FIG. 7.21 – Contact rotor-stator rigide - Étude de la stabilité des solutions  
Valeurs propres de la jacobienne  $D_x F$  en fonction de la fréquence d'excitation

Chacune des transitions entre un état stable et instable se fait avec deux valeurs propres conjuguées de partie imaginaire non nulle lorsque la partie réelle s'annule : on est donc dans le cas de bifurcations de Hopf. On doit trouver aux voisinages de ces points de bifurcation des solutions périodiques de pulsation égale au module de la partie imaginaire des valeurs propres concernées, soit, par ordre d'apparition : 20.85, 31.50, 75.80 et 101.4 rad/s.

Nous allons dans la suite de l'étude nous concentrer sur la première zone d'instabilité située entre 14.49 et 16.16 Hz. Nous avons donc calculé les deux nombres de Lyapunov (voir paragraphe 2.4.3) relatifs à chacun des points de bifurcation. Comme précisé dans [Kuz04, Sec. 10.2.3], ce n'est pas la valeur de ce nombre mais uniquement son signe qui détermine la qualité sur ou sous-critique de la bifurcation. Dans le premier cas, le signe a été trouvé positif et dans le second, négatif. Ceci signifie que l'on a une bifurcation de Hopf sous-critique

en 14.49 Hz et une bifurcation de Hopf sur-critique en 16.16 Hz. Ainsi, il existe pour des fréquences légèrement inférieures à 14.49 Hz des solutions périodiques instables de pulsation voisine de 20.84 rad/s et pour des fréquences légèrement inférieures à 16.16 Hz, il existe des solutions périodiques stables de pulsation 31.50 rad/s qui disparaissent une fois passé le point de bifurcation. Nous détaillons l'étude de ces deux points de bifurcation dans le paragraphe 7.4.3.

Afin de valider cette étude, nous avons mis en place un algorithme d'intégration temporelle qui nous permet de suivre l'évolution d'une solution dans le temps pour un jeu de conditions initiales donné.

### Contrôle par intégration temporelle

L'objectif est ici de construire pas à pas l'évolution temporelle d'un jeu de conditions initiales. On utilise donc les équations (7.15), où l'amortissement est modélisé de façon visqueuse.

**Schéma d'intégration temporelle par différences finies :** On utilise un schéma aux différences finies, centré pour l'approximation des dérivées secondes et arrière pour l'approximation des dérivées premières de façon à ne pas avoir des équations non-linéaires à résoudre à chaque pas de temps.

En notant  $t_i = h(i - 1)$ ,  $i = 1, \dots, m$  où  $h$  est le pas de temps et  $q_i = q(t_i)$  la valeur de la grandeur  $q$  à l'instant  $t_i$ , on a donc :

$$q(t_i) = q_i, \quad \dot{q}(t_i) \approx \frac{q_i - q_{i-1}}{h} \quad \text{et} \quad \ddot{q}(t_i) \approx \frac{q_{i+1} - 2q_i + q_{i-1}}{h^2}$$

Ceci nous mène au système suivant :

$$\begin{bmatrix} m_R + m_S & -m_{RG} \sin \theta_{si} & -m_{RG} \sin \theta_{si} \\ 0 & m_{RG} \cos \theta_{si} & m_{RG} \cos \theta_{si} + (m_R + m_S)s_i \\ m_S \tan \theta_{si} & 0 & -m_{SSi} \end{bmatrix} \begin{Bmatrix} s_{i+1} \\ \theta_{si+1} \\ \alpha_{i+1} \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix} \quad (7.22)$$

avec :

$$\begin{aligned} b_1 &= -(m_R + m_S) \left( (-2s_i + s_{i-1}) - s_i(\alpha_i - \alpha_{i-1})^2 \right) \\ &\quad + m_{RG} \left( (\theta_{si} - \theta_{si-1} - 1 + \alpha_i - \alpha_{i-1})^2 \cos \theta_{si} - (-2\theta_{si} + \theta_{si-1} - 2\alpha_i + \alpha_{i-1}) \sin \theta_{si} \right) \\ &\quad - (k_R \eta_R + k_S \eta_S) h (s_i - s_{i-1}) + (k_R \eta_R) h g (\theta_{si} - \theta_{si-1} + \alpha_i - \alpha_{i-1}) \sin \theta_{si} \\ &\quad - (k_R + k_S) h^2 s_i - k_R h^2 g \cos \theta_{si} + h^2 F_e(t_i) \cos \theta_e(t_i) \\ b_2 &= -(m_R + m_S) (s_i(-2\alpha_i + \alpha_{i-1}) + 2(s_i - s_{i-1})(\alpha_i - \alpha_{i-1})) \\ &\quad + m_{RG} \left( (-2\theta_{si} + \theta_{si-1} - 2\alpha_i + \alpha_{i-1}) \cos \theta_{si} + (\theta_{si} - \theta_{si-1} + \alpha_i - \alpha_{i-1})^2 \sin \theta_{si} \right) \\ &\quad - (k_R \eta_R + k_S \eta_S) h s_i (\alpha_i - \alpha_{i-1}) - (k_R \eta_R) h g (\theta_{si} - \theta_{si-1} + \alpha_i - \alpha_{i-1}) \cos \theta_i \\ &\quad - k_R h^2 g \sin \theta_{si} + h^2 F_e(t_i) \sin \theta_e(t_i) \\ b_3 &= -m_s \left( (-2s_i + s_{i-1} - s_i(\alpha_i - \alpha_{i-1})^2) \tan \theta_{si} \right) \\ &\quad - s_i(-2\alpha_i + \alpha_{i-1}) - 2(s_i - s_{i-1})(\alpha_i - \alpha_{i-1}) \\ &\quad - (k_S \eta_S) h \left( (s_i - s_{i-1}) \tan \theta_{si} - s_i(\alpha_i - \alpha_{i-1}) \right) - k_S h^2 s_i \tan \theta_{si} \end{aligned}$$

avec  $\theta_e(t_i) = \omega_e \cdot t_i - \alpha_i$

Le calcul a posteriori de la force de contact se fait par :

$$F_{ci} \sin \theta_{si} = m_S \left( s_i \frac{\alpha_{i+1} - 2\alpha_i + \alpha_{i-1}}{h^2} + 2 \frac{s_i - s_{i-1}}{h} \frac{\alpha_i - \alpha_{i-1}}{h} \right) + (k_S \eta_S) s_i \frac{\alpha_i - \alpha_{i-1}}{h}$$

**Résultats - Comparaison avec l'étude fréquentielle :** On se place d'abord à une fréquence d'excitation pour laquelle le point fixe est stable. On choisit  $\omega_e/2\pi = 12.63 \text{ Hz}$ . La fenêtre temporelle étudiée est  $t \in [0 \ 12] \text{ s}$  et 100000 points de calcul sont effectués, ce qui donne un pas  $h = 12.10^{-6} \text{ s}$ .

On utilise deux jeux de conditions initiales. L'un, proche du point fixe et l'autre plus éloigné. Si le point fixe est stable et que son bassin d'attraction englobe ces deux points, on devrait pour un temps assez grand retrouver une réponse stationnaire constante en  $s$ ,  $\theta_s$ ,  $F_c$ ,  $\theta_e$  et  $\dot{\alpha}$  ( $= \omega_e$ ) avec les mêmes valeurs que pour l'étude fréquentielle. On lit sur la figure 7.20 que pour  $F_e = 200 \text{ N}$  et  $\omega_e/2\pi = 12.63 \text{ Hz}$ , on doit trouver une telle réponse avec  $s = 0.0203 \text{ m}$ ,  $\theta_s = 0.0847\pi \text{ rad}$ ,  $F_c = 544 \text{ N}$ ,  $\theta_e = \pi/2 \text{ rad}$  et  $\dot{\alpha} = \omega_e = 12.63 \times 2\pi \text{ rad.s}^{-1}$ . En donnant alors les conditions initiales suivantes qui permettent d'initialiser l'algorithme pour les rangs  $i=1$  ( $t=0$ ) et  $i=2$  ( $t=h$ ), on obtient les résultats présentés en figure 7.22 :

$$\begin{aligned} s_1 &= 0.005 \text{ m} & \theta_{s1} &= 0 \text{ rad} & F_{c1} &= 50 \text{ N} & \alpha_1 &= 0 \text{ rad} \\ s_2 &= 0.005 \text{ m} & \theta_{s2} &= 0 \text{ rad} & F_{c2} &= 50 \text{ N} & \alpha_2 &= 12.63/2 \times 2\pi \times h \text{ rad} \end{aligned}$$

Ces résultats coïncident avec ceux de l'étude fréquentielle.

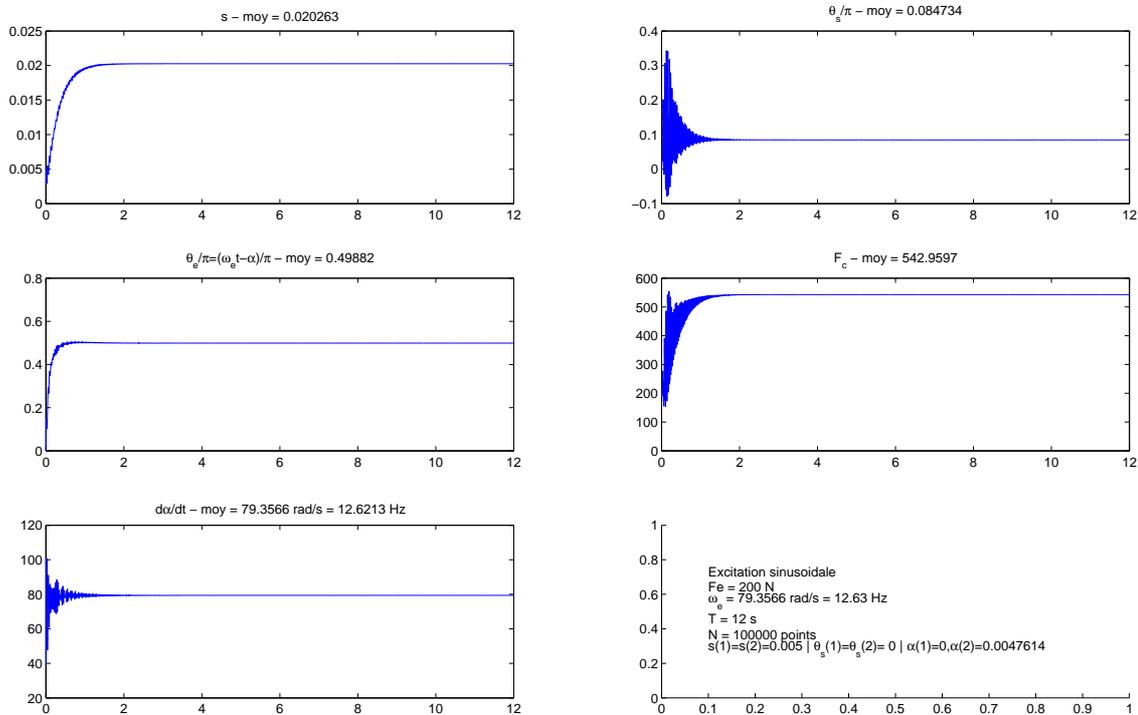


FIG. 7.22 – Contact rotor-stator rigide - Intégration temporelle pour  $\omega_e/2\pi = 12.63 \text{ Hz}$

De même, en choisissant des conditions éloignées de l'état stationnaire :

$$\begin{aligned} s_1 &= 0.01 \text{ m} & \theta_{s1} &= \pi \text{ rad} & F_{c1} &= 50 \text{ N} & \alpha_1 &= 0 \text{ rad} \\ s_2 &= 0.01 \text{ m} & \theta_{s2} &= 3\pi/2 \text{ rad} & F_{c2} &= 50 \text{ N} & \alpha_2 &= 5 \times 12.63 \times 2\pi \times h \text{ rad} \end{aligned}$$

on retrouve encore la solution prévue par l'analyse précédente comme le prouve la figure 7.23. Ce point fixe est donc bien stable.

En revanche, si l'on essaie d'accéder aux solutions prédites par le modèle sous hypothèses pour les fréquences d'excitation de la première zone d'instabilité, par exemple  $\omega_e/2\pi = 14.6031 \text{ rad/s}$ , il apparaît que les solutions ne sont pas stables, même en donnant pour conditions initiales les valeurs devant être obtenues pour un temps infini (figure 7.24).

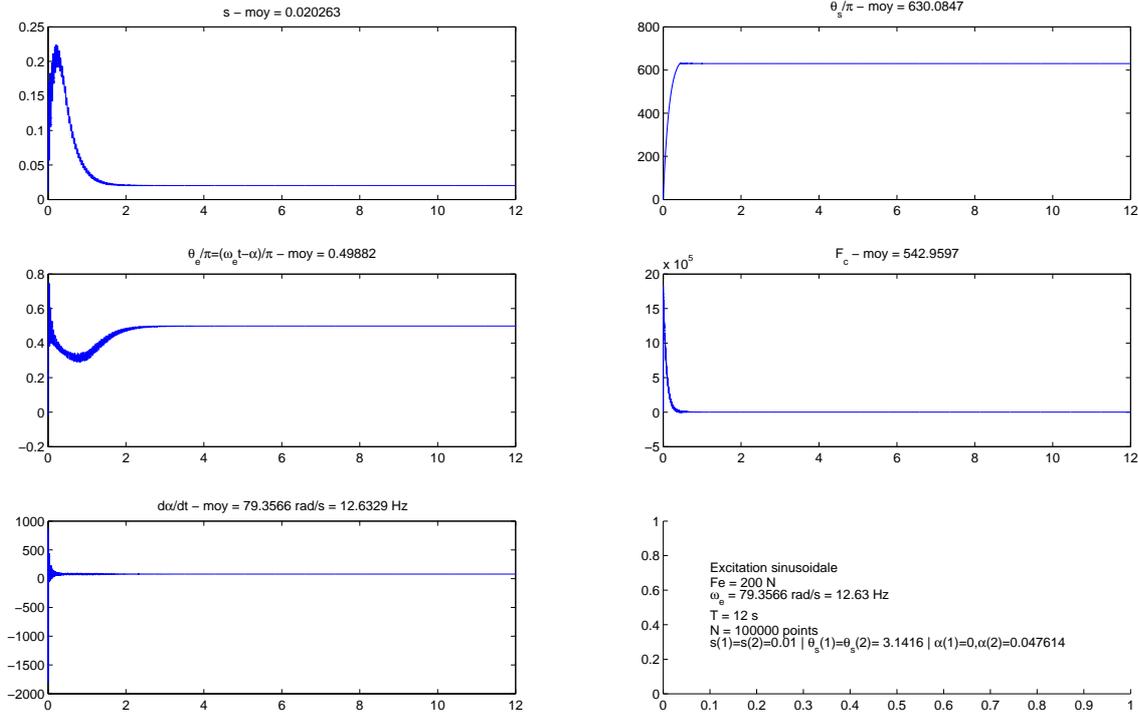


FIG. 7.23 – Contact rotor-stator rigide -  $\omega_e/2\pi = 12.63 \text{ Hz}$

Conditions initiales éloignées de la solution constante

L'analyse fréquentielle est donc corroborée par ces quelques résultats obtenus par intégration temporelle.

### 7.4.3 Étude des points de bifurcation de Hopf

Nous cherchons maintenant à exhiber les solutions périodiques naissant aux voisinages des deux points de bifurcation du point fixe en  $\omega_{e1}/2\pi = 14.49 \text{ Hz}$  et  $\omega_{e2}/2\pi = 16.16 \text{ Hz}$ . Nous utilisons pour cela la méthode donnée en 2.4.3 : pour chacun de ces points  $(x_{0i}, \omega_{ei})$ ,  $i = 1, 2$ , nous avons donc recherché les vecteurs propres  $v_i$  et  $\bar{v}_i$  associés aux valeurs propres  $\lambda_i = j\omega_{ei}$  et  $\bar{\lambda}_i$ , puis nous avons réalisé une intégration temporelle pour  $\tilde{\omega}_{ei} = \omega_{ei} - \Delta\omega$ ,  $0 < \Delta\omega \ll 1$  et  $x_i(0) = x_{0i} + \epsilon \mathcal{R}e(v_i)$ ,  $\epsilon \ll 1$  (nous rappelons que nous cherchons des solutions « à gauche » de chacune des pulsations, ce qui explique le signe moins devant  $\Delta\omega$ ). Les résultats de ces intégrations temporelles sont donnés en figures 7.25 et 7.26.

Une fois en possession d'un point du cycle limite, il est possible d'effectuer une continuation et une étude de stabilité des solutions périodiques. Étant donné que la nature des bifurcations n'est pas la même pour chacun des points, nous les traitons séparément.

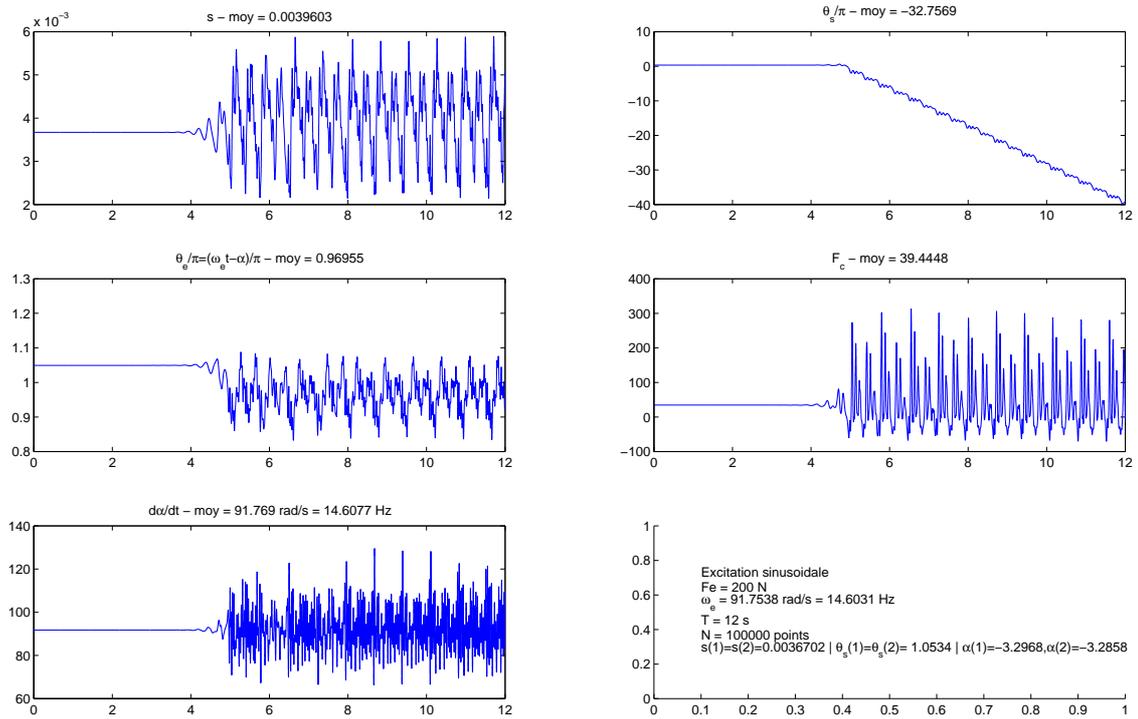


FIG. 7.24 – Contact rotor-stator rigide -  $\omega_e/2\pi = 14.6031$  Hz  
Instabilité de la solution constante

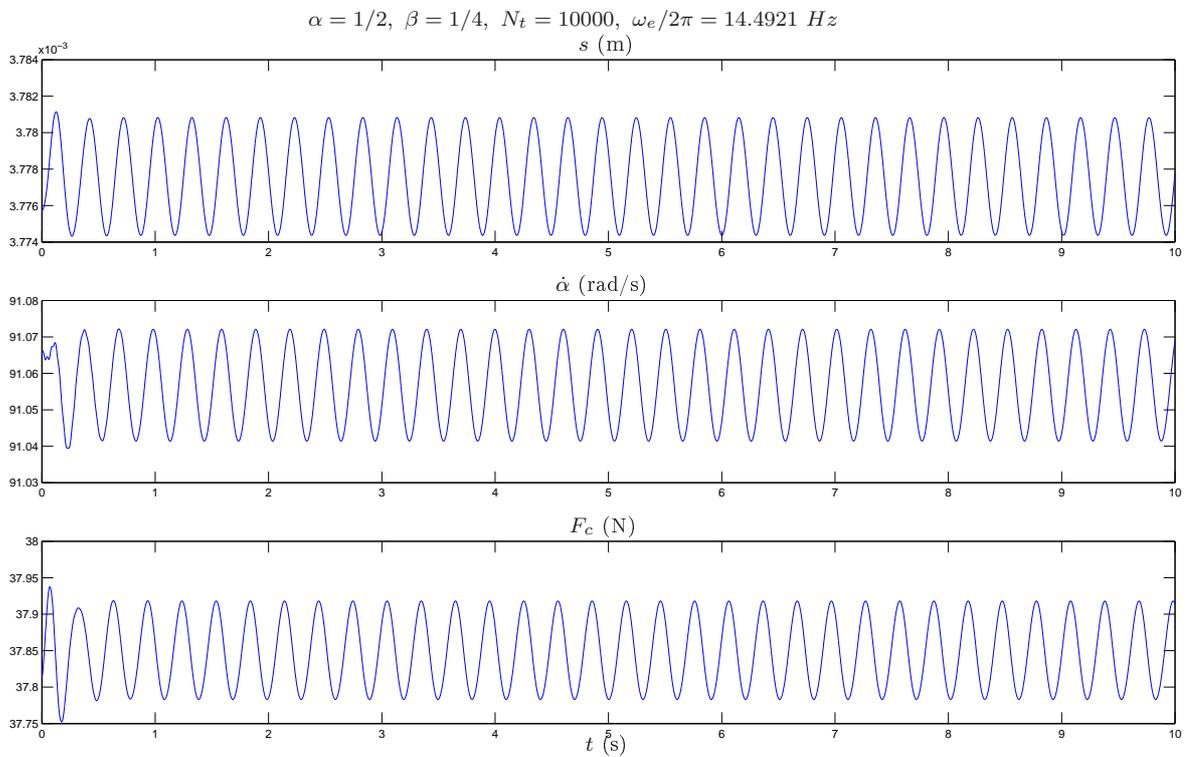


FIG. 7.25 – Contact rotor-stator rigide  
Intégration temporelle pour le point de bifurcation en  $\omega_e \approx \omega_{e1}$

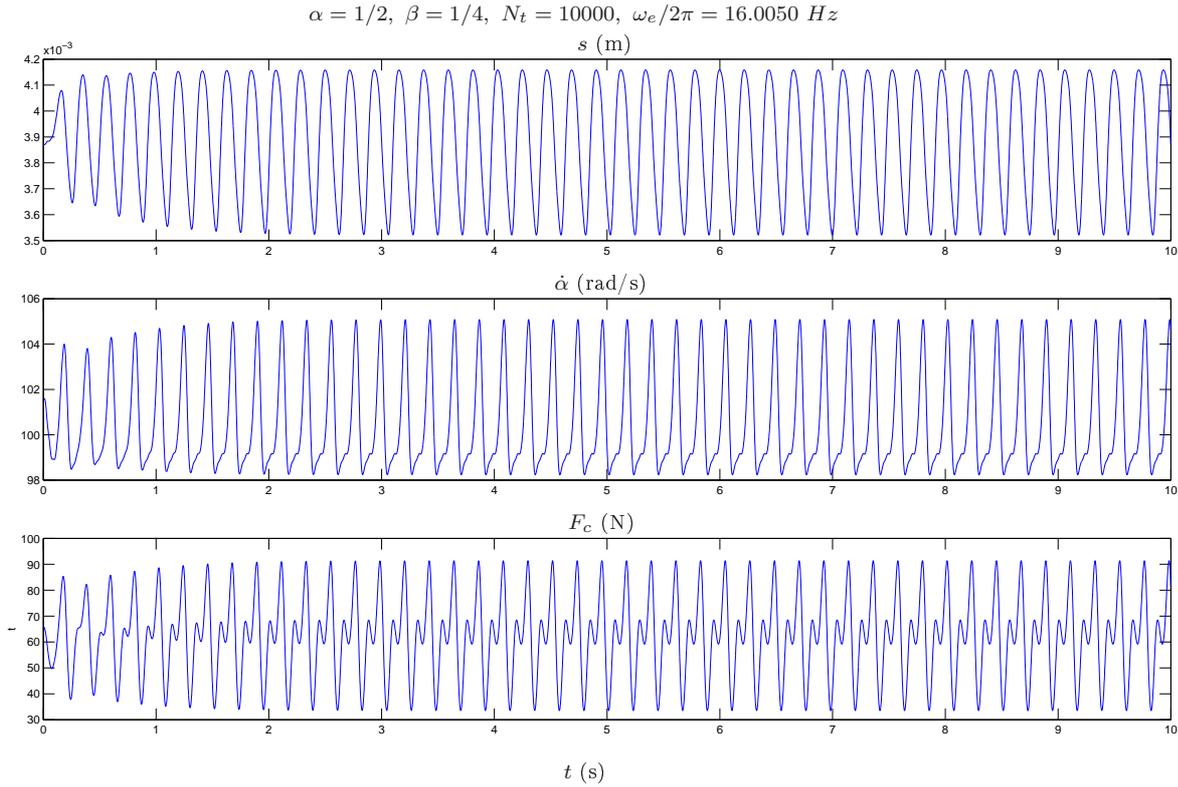


FIG. 7.26 – Intégration temporelle pour le point de bifurcation en  $\omega_e \approx \omega_{e2}$

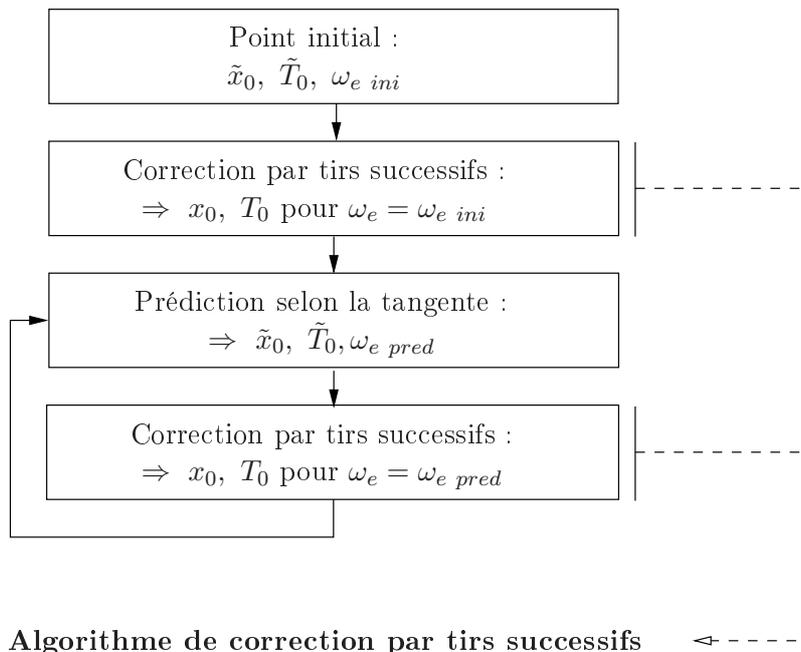
### Étude de la bifurcation de Hopf sous-critique en 14.49 Hz

Nous voulons suivre l'évolution de la solution périodique exhibée en figure 7.25 lorsque  $\omega_e$  varie. Nous avons pour cela choisi d'utiliser une méthode de continuation basée sur une méthode de tir pour la recherche de la solution périodique correspondant à chaque valeur du paramètre  $\omega_e$ , car cette méthode permet de rechercher une solution périodique de période inconnue. Il nous faut donc choisir une méthode d'intégration temporelle. Nous avons utilisé dans la section précédente une méthode basée sur les différences finies ; très facile à écrire, elle est cependant peu efficace, nécessitant un grand nombre de points de calcul. Nous préférons donc utiliser un schéma de type Newmark adapté aux problèmes non-linéaires (c.f. paragraphe 2.1.3). Nous choisissons une approximation de l'accélération par valeur moyenne :  $\gamma = 1/2$  et  $\beta = 1/4$ . Nous détaillons ci-dessous la méthode de continuation. Nous faisons ensuite une parenthèse sur le test de la périodicité qui n'est pas trivial pour ce problème. Enfin, nous donnons le résultat obtenu pour la solution qui nous intéresse. Les algorithmes utilisés sont résumés dans la figure 7.27.

### Continuation d'une solution périodique

On cherche à réaliser une continuation sur le problème  $\dot{x}(t, x_0, \omega_e) = F(x(t, x_0, \omega_e), \omega_e)$  en ayant pour objectif d'obtenir  $x(T_0, x_0, \omega_e) - x_0 = 0$ . Si l'on possède un jeu de paramètres  $x_0, T_0, \omega_{e0}$  satisfaisant cet objectif, on cherche ensuite les incréments  $\Delta x, \Delta T$  et  $\Delta \omega_e$  tels que  $x(T_0 + \Delta T, x_0 + \Delta x, \omega_{e0} + \Delta \omega_e) - (x_0 + \Delta x) = 0$  de façon à réaliser une prédiction. On écrit

### Algorithme général



### Algorithme de correction par tirs successifs

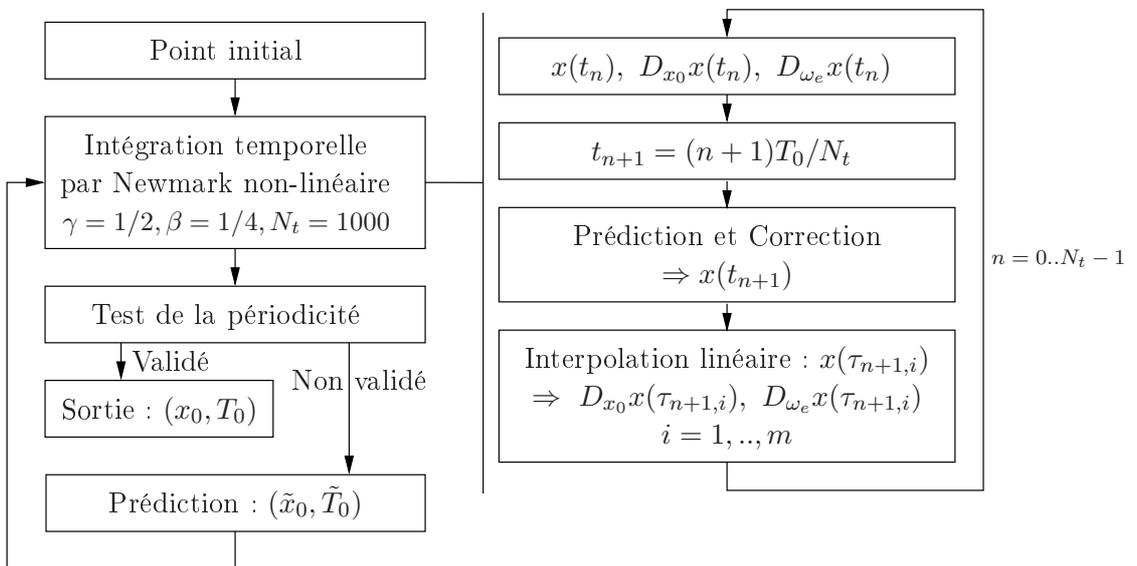


FIG. 7.27 – Algorithmes utilisés pour la continuation périodique

donc :

$$\begin{aligned} & x(T_0 + \Delta T, x_0 + \Delta x, \omega_{e0} + \Delta\omega_e) - (x_0 + \Delta x) \\ &= x(T_0, x_0, \omega_{e0}) + \frac{\partial x}{\partial T_0} \Delta T + \frac{\partial x}{\partial x_0} \Delta x + \frac{\partial x}{\partial \omega_e} \Delta\omega_e - (x_0 + \Delta x) \\ &= \frac{\partial x}{\partial T_0} \Delta T + \left( \frac{\partial x}{\partial x_0} - I \right) \Delta x + \frac{\partial x}{\partial \omega_e} \Delta\omega_e \end{aligned}$$

Comme vu au paragraphe 2.3.2,  $\frac{\partial x}{\partial T_0} = F(x(T_0, x_0, \omega_{e0}), \omega_{e0})$  et  $\frac{\partial x}{\partial x_0}$  est solution de l'équation différentielle :

$$\dot{Y} = D_x F(x(t, x_0, \omega_{e0}), \omega_{e0}) Y, \quad Y(0) = I$$

De la même façon, on montre que  $\frac{\partial x}{\partial \omega_e}$  est solution de l'équation différentielle :

$$\dot{Y} = D_x F(x(t, x_0, \omega_{e0}), \omega_{e0}) Y + D_{\omega_e} F(x(t, x_0, \omega_{e0}), \omega_{e0}), \quad Y(0) = 0 \text{ (vecteur nul)}$$

$$\text{On doit donc résoudre : } \frac{\partial x}{\partial T_0} \Delta T + \left( \frac{\partial x}{\partial x_0} - I \right) \Delta x + \frac{\partial x}{\partial \omega_e} \Delta\omega_e = 0.$$

On choisit enfin de forcer la correction  $\Delta x$  à être orthogonale à  $F(x(T_0, x_0, \omega_{e0}), \omega_{e0})$ . On peut donc mettre le système à résoudre sous la forme suivante :

$$\underline{A} \begin{Bmatrix} \Delta T \\ \Delta x \end{Bmatrix} + \begin{Bmatrix} \frac{\partial x}{\partial \omega_e} \\ 0 \end{Bmatrix} \Delta\omega_e = 0, \quad \underline{A} = \begin{bmatrix} \frac{\partial x}{\partial T_0} & \frac{\partial x}{\partial x_0} - I \\ 0 & F(x(T_0, x_0, \omega_{e0}), \omega_{e0}) \end{bmatrix} \quad (7.23)$$

En posant  $z = -\underline{A}^{-1} \begin{Bmatrix} \frac{\partial x}{\partial \omega_e} \\ 0 \end{Bmatrix}$ , on obtient par la condition de normalisation

$$\begin{Bmatrix} \Delta T \\ \Delta x \end{Bmatrix}^t \begin{Bmatrix} \Delta T \\ \Delta x \end{Bmatrix} + \Delta\omega_e^2 = 1 \text{ que } \Delta\omega_e = \pm \frac{1}{\sqrt{1+z^t z}} \text{ puis } \begin{Bmatrix} \Delta T \\ \Delta x \end{Bmatrix} = z \cdot \Delta\omega_e$$

Le signe est enfin déterminé par l'orientation de la tangente avec la condition :

$$\det \left( \begin{bmatrix} \underline{A} & \begin{Bmatrix} \frac{\partial x}{\partial \omega_e} \\ 0 \end{Bmatrix} \\ z & 1 \end{bmatrix} \right) \Delta\omega_e > 0.$$

Ainsi, toutes les prédictions sont basées sur les valeurs des dérivées partielles de  $x$  par rapport à  $x_0$ ,  $T_0$  et  $\omega_e$ , qui sont obtenues pour deux d'entre elles ( $D_{x_0}x$  et  $D_{\omega_e}x$ ), par intégration temporelle. Il est donc important que le calcul approché de ces grandeurs soit « bon ». S'il n'est pas nécessaire d'avoir un nombre de points très important pour bien décrire le signal  $x(t)$  par intégration temporelle, ces dérivées partielles, elles, nécessitent un nombre de points conséquent. Il faut donc augmenter le nombre de pas utilisés pour l'intégration de ces dérivées. Pour ne pas augmenter simultanément le nombre de pas de calcul de  $x(t)$ , ce qui rend les calculs coûteux, nous effectuons une interpolation linéaire entre deux points  $x(t_{n-1})$  et  $x(t_n)$  de façon à obtenir :

$$x(\tau_{n,i}) \approx x(t_{n-1}) + (x(t_n) - x(t_{n-1})) \cdot (i/m), \quad i = 0..m, \quad \tau_{n,i} = t_{n-1} + (t_n - t_{n-1}) \cdot (i/m)$$

Nous utilisons alors ces points intermédiaires pour l'intégration des dérivées partielles.

### Test de la périodicité

Ajoutons maintenant le moyen de nous affranchir de la « pseudo-périodicité » des angles. En effet, ce qui nous intéresse n'est pas de tester leur égalité mais leur égalité modulo  $2\pi$ , ce que l'on peut écrire  $\exists k \in \mathbb{Z}, \theta_2 = \theta_1 + k(2\pi)$  mais qui ne se dérive pas. On préfère donc écrire ceci de la façon suivante :

$$\begin{aligned} \theta_1 = \theta_2 [2\pi] &\Leftrightarrow \begin{cases} \cos(\theta_1) = \cos(\theta_2) \\ \sin(\theta_1) = \sin(\theta_2) \end{cases} \Leftrightarrow \begin{cases} \cos(\theta_1) - \cos(\theta_2) = 0 \\ \sin(\theta_1) - \sin(\theta_2) = 0 \end{cases} \\ &\Leftrightarrow (\cos(\theta_1) - \cos(\theta_2))^2 + (\sin(\theta_1) - \sin(\theta_2))^2 = 0 \Leftrightarrow \cos(\theta_1 - \theta_2) - 1 = 0 \end{aligned}$$

Ainsi, la condition d'égalité que nous vérifierons ne sera pas  $x(T_0, x_0) - x_0 = 0$ , mais

$$\varphi(x(T_0, x_0) - x_0) = 0$$

$$\text{avec } \varphi(x) = \varphi \left( \begin{pmatrix} s \\ \theta_s \\ \theta_e \\ \hat{s} \\ \hat{\theta}_s \\ \hat{\alpha} \end{pmatrix} \right) = \begin{pmatrix} s \\ \cos(\theta_s) - 1 \\ \sin(\theta_e) - 1 \\ \hat{s} \\ \hat{\theta}_s \\ \hat{\alpha} \end{pmatrix}$$

Le développement au premier ordre au voisinage de  $(T_0, x_0)$  devient donc :

$$\varphi(x(T_0, x_0) - x_0) + D_x \varphi(x(T_0, x_0) - x_0) \cdot \left( \frac{\partial x}{\partial T_0}(T_0, x_0) \Delta T_0 + \frac{\partial x}{\partial x_0}(T_0, x_0) \cdot \Delta x_0 \right) = 0 \quad (7.24)$$

$$\text{avec la jacobienne de } \varphi(x) \text{ qui s'écrit } D_x \varphi(x) = \begin{bmatrix} 1 & & & & & \\ & -\sin(\theta_s) & & & & \\ & & -\sin(\theta_e) & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

Elle est donc toujours inversible sauf si la différence entre les angles à  $t = T_0$  et  $t = 0$  vaut 0 ou  $\pi$ . Si cette différence est nulle, les angles sont égaux et l'équation correspondant à la ligne où se trouve ce zéro se résume à  $0 = 0$ . En revanche, si les deux angles diffèrent de  $\pi$ , la matrice est non inversible mais les deux angles ne respectent pas la condition d'égalité. Nous proposons alors de contourner le problème en effectuant un pas d'intégration temporelle supplémentaire de façon à utiliser non plus  $T_0$  mais  $\tilde{T}_0$  légèrement supérieure, ce qui modifie normalement la valeur de l'angle posant problème de  $\theta_x(T_0, x_0)$  en  $\theta_x(\tilde{T}_0, x_0)$ . Étant donné que  $T_0$  n'était vraisemblablement pas solution du problème (les angles diffèrent) et n'est de toute façon qu'une supposition, cette façon de procéder semble raisonnable.

### Application au premier point de bifurcation de Hopf

Le résultat de la continuation menée à partir du cycle limite exhibé sur la figure 7.25 est représenté en figure 7.28. On peut voir, par l'analyse des coefficients de Floquet, que la solution périodique est bien instable au voisinage du point de bifurcation. Nous reportons en figure 7.29, les minima et maxima des amplitudes au cours de l'évolution de la solution.

En ce point de bifurcation, il y a donc destruction d'une solution stable sans naissance d'une autre solution stable. On est donc en présence d'une bifurcation catastrophique (cf paragraphe 2.3.2). L'analyse de l'évolution du système nécessite une étude globale et non plus locale pour savoir vers quelle solution le système va tendre lorsqu'il quittera le point fixe devenu instable.

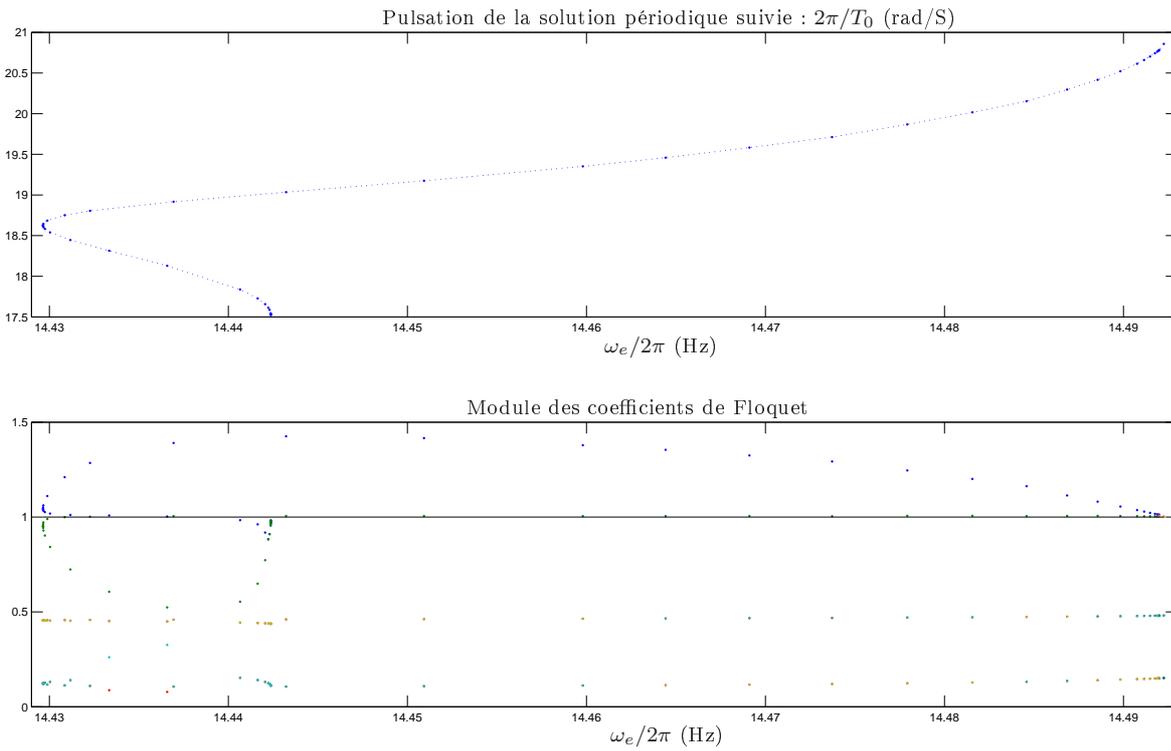


FIG. 7.28 – Contact rotor-stator rigide - Continuation de la solution périodique du premier point de bifurcation de Hopf - Pulsation et stabilité

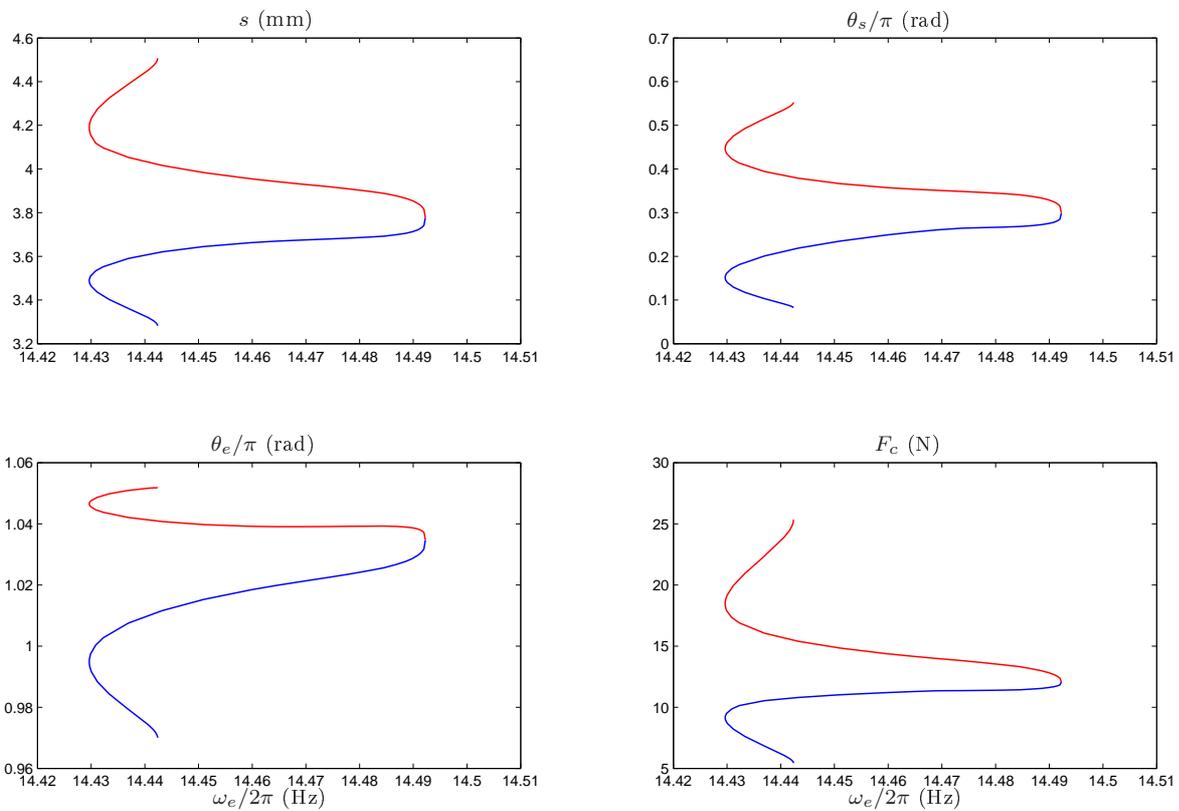


FIG. 7.29 – Contact rotor-stator rigide - Continuation de la solution périodique du premier point de bifurcation de Hopf - Amplitudes

### Étude de la bifurcation de Hopf sur-critique en 16.16 Hz

Nous avons appliqué le même algorithme en partant cette fois du cycle limite exhibé au voisinage de 16.16 Hz. Le résultat de la continuation est représenté en figure 7.30. Pour plus de lisibilité, nous avons reporté les zones de stabilité sur le schéma. Enfin, pour donner une idée des amplitudes de ces solutions, nous avons tracé les minima et maxima des paramètres sur la figure 7.31.

Le suivi du module des coefficients de Floquet nous indique une bifurcation en doublement de période pour  $\omega_e/2\pi \approx 18.11$  Hz (figure 7.32). Nous avons cherché et continué la solution de période double qui naît en ce point de bifurcation. Cette solution, stable à la naissance, devient rapidement instable, subissant une bifurcation de Hopf comme le montre la figure 7.33.

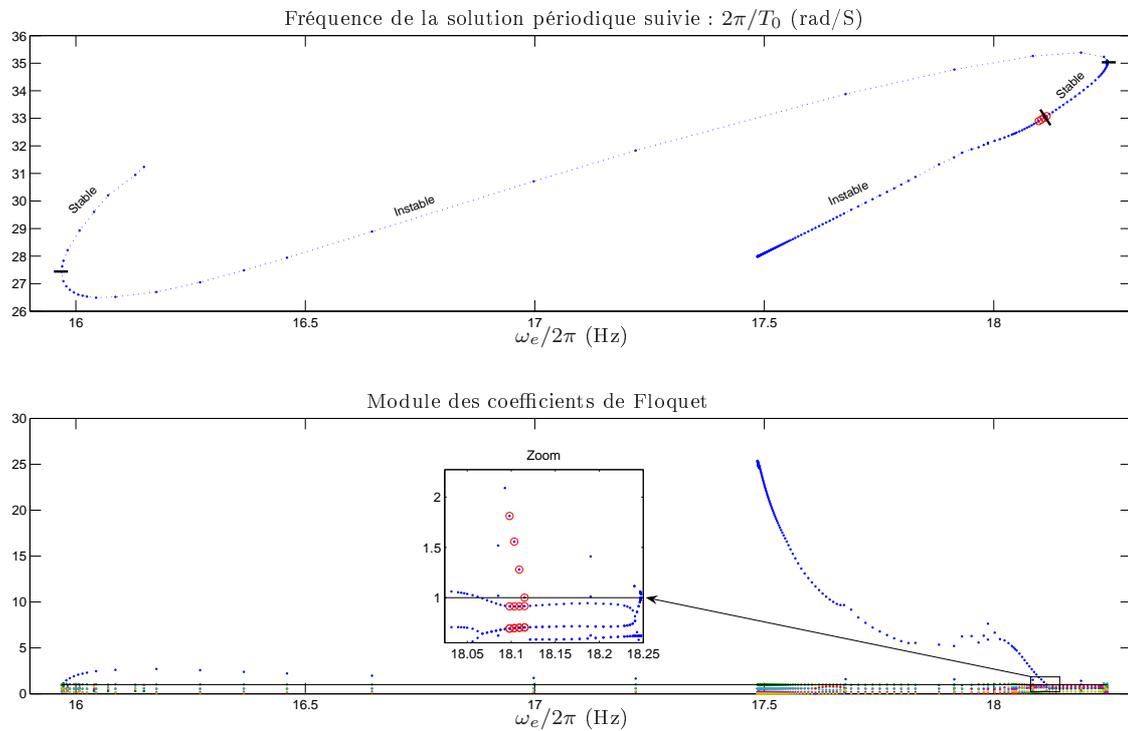


FIG. 7.30 – Contact rotor-stator rigide - Continuation de la solution périodique du second point de bifurcation de Hopf - Pulsation et stabilité

#### 7.4.4 Analyse globale

Comme montré au paragraphe 7.4.3, il existe des solutions déconnectées des solutions précédemment étudiées vers lesquelles le système va se diriger lorsque l'on entrera dans la zone [14.49 16.16] Hz. C'est elles que nous recherchons et étudions dans ce dernier paragraphe. Nous souhaiterions bien évidemment appliquer la méthode globale par homotopie polynomiale ; cependant, ne connaissant pas la période des solutions recherchées, il nous est impossible de l'utiliser. Nous mettons là en évidence la nécessité de faire évoluer cet outil afin qu'il permette de traiter ce cas de figure.

En l'absence d'un outil dédié, nous commençons par effectuer quelques essais à partir de conditions initiales diverses en choisissant la pulsation d'excitation sur la plage de fréquences

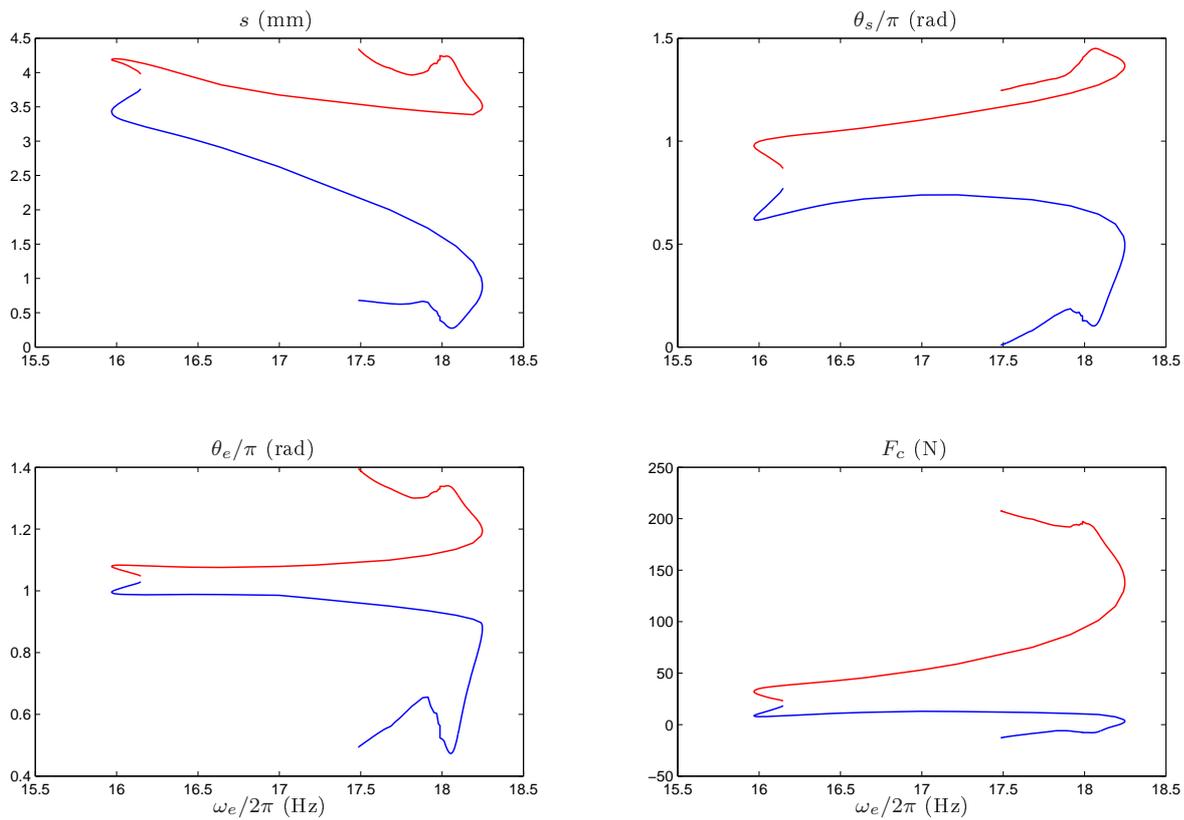


FIG. 7.31 – Contact rotor-stator rigide - Continuation de la solution périodique du second point de bifurcation de Hopf - Amplitudes

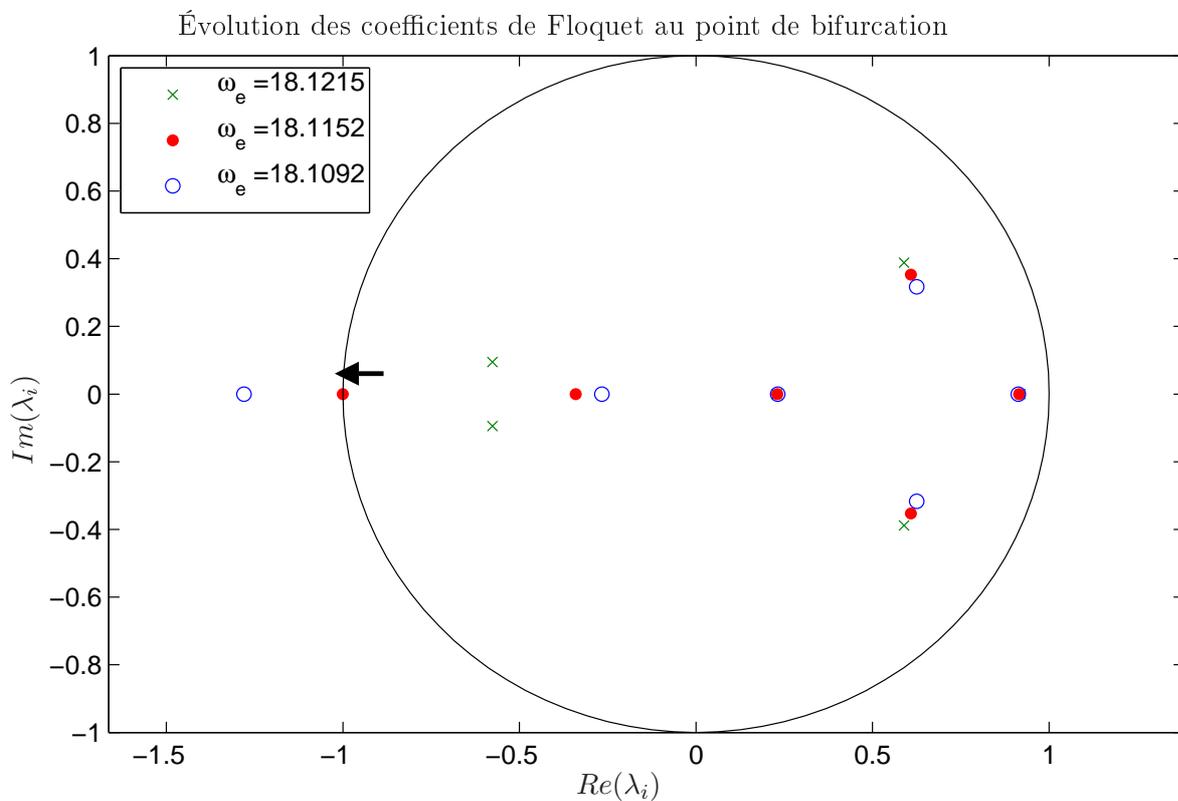


FIG. 7.32 – Contact rotor-stator rigide - Branche périodique issue de la seconde bifurcation de Hopf - Bifurcation de doublement de période en 18.11 Hz

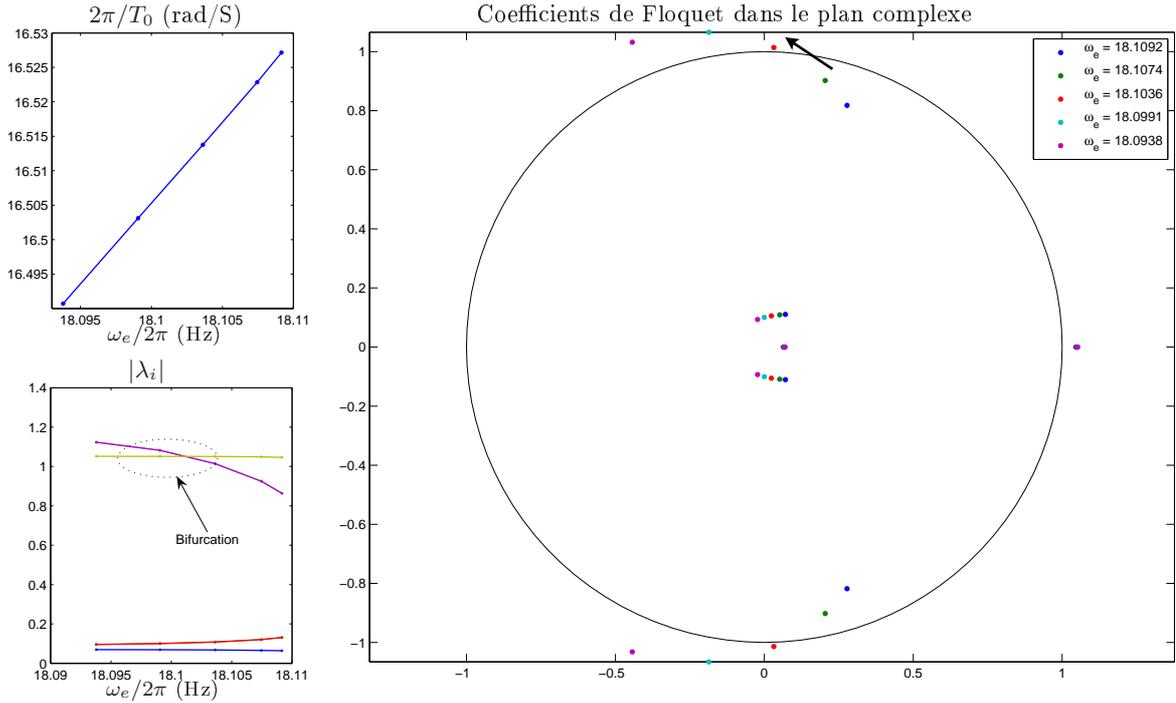


FIG. 7.33 – Contact rotor-stator rigide - Étude de la solution de période double

concernée. Ceci nous permet d’obtenir des solutions sur lesquelles on effectue une continuation puis une étude de stabilité et de bifurcation.

### Essais

On réalise quelques intégrations temporelles de façon à avoir une idée des solutions recherchées. Il s’avère que l’on trouve des solutions périodiques. La figure 7.34 donne l’évolution de leur pulsation  $\omega_s$  sur la gamme de fréquences d’excitation [14.5 14.7] Hz. Il semble qu’on ait donc un doublement de période au voisinage de 14.55 Hz et inversement un doublement de fréquence au voisinage de 14.65 Hz.

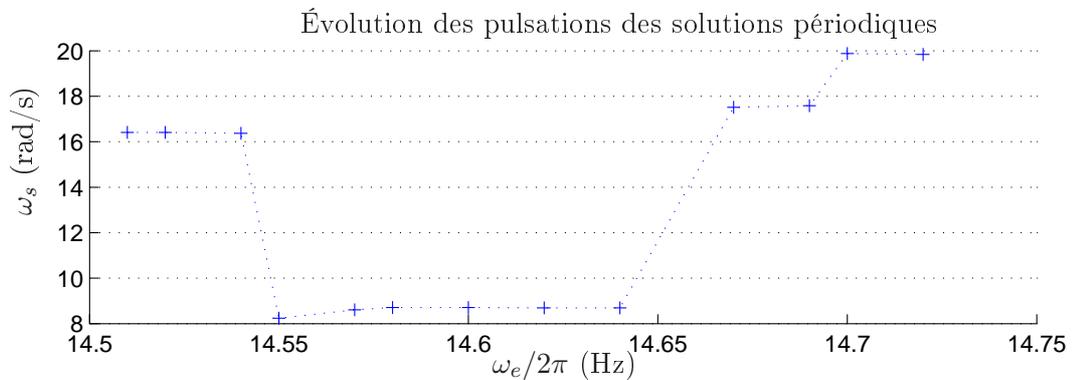


FIG. 7.34 – Contact rotor-stator rigide - Nouvelles solutions périodiques : tirages sur la bande [14.5 14.7] Hz

### Continuation et analyse

Afin d'avoir un portrait plus global de la solution, nous choisissons de réaliser une continuation du cycle limite accroché pour  $\omega_e/2\pi = 14.52$  Hz. On obtient alors les figures 7.35 et 7.36.

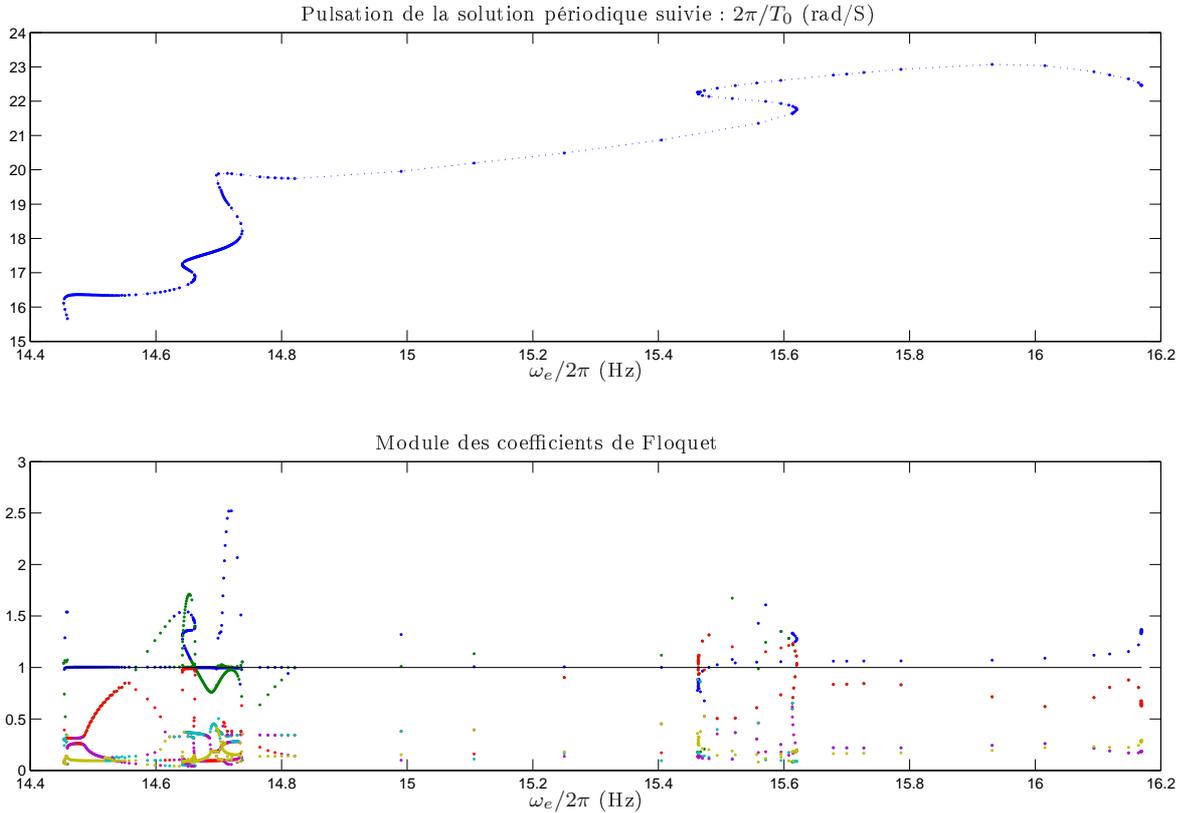


FIG. 7.35 – Contact rotor-stator rigide - Solution périodique déconnectée - Pulsation et stabilité

Au vu de l'évolution des coefficients de Floquet, cette solution subit de nombreuses bifurcations. Dans un souci de respect des échéances, nous n'avons pas poursuivi chacune de ces bifurcations comme nous avons pu le faire pour les autres branches. L'intérêt de cette solution est en effet d'exhiber l'existence d'une solution stable mais complètement déconnectée de celle poursuivie initialement (point fixe) accrochée à l'issue d'une bifurcation catastrophique (au premier point de bifurcation de Hopf).

Nous donnons enfin une représentation graphique de l'ensemble des solutions étudiées par le biais des amplitudes de l'excentricité du stator ( $s$ ) en figure 7.37. Nous proposons ainsi une nouvelle représentation de ce que l'on pourrait nommer diagramme de bifurcation. Elle permet non seulement d'observer les bifurcation mais aussi de donner les amplitudes des solutions périodiques ainsi que leurs fréquences de réponse. Moins synthétique que les diagrammes traditionnels, cette forme de représentation permet de fournir toutes les informations nécessaires à la compréhension de la dynamique du système.

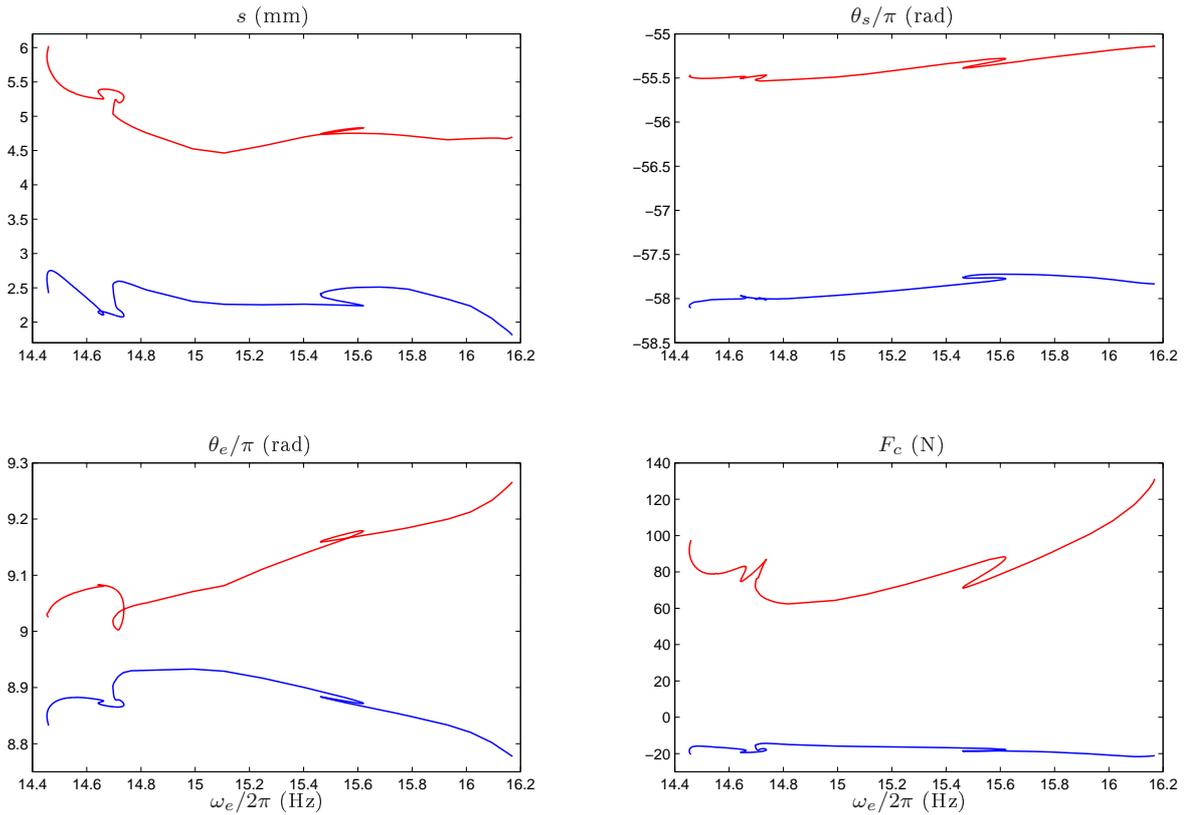


FIG. 7.36 – Contact rotor-stator rigide - Solution périodique déconnectée - Amplitudes

### Considérations numériques

Nous tenons ici à rendre compte de deux problèmes liés à l'étude nécessairement numérique et non purement analytique d'un problème.

**Valeur des coefficients de Floquet :** Les coefficients de Floquet sont issus du calcul des valeurs propres de la matrice de monodromie obtenue elle-même par intégration temporelle. Si le calcul des amplitudes du signal (variable  $x$ ) n'est pas trop sensible au pas en temps, il semble que celui de la matrice de monodromie le soit particulièrement, et ainsi, celui des valeurs propres extraites. Ceci est particulièrement visible sur la figure 7.35 : la théorie donne qu'un des coefficients de Floquet doit toujours être égal à 1 dû à la formulation autonome de notre problème ; or, il est évident sur la gamme de fréquences [15.7 16.2]  $Hz$ , que cette assertion est mise en défaut. Pourtant, les solutions trouvées sont bien périodiques et dans la continuité des précédentes. Les calculs menés sont donc « justes » en un certain sens. Ceci montre qu'il faut être prudent à l'analyse de ces coefficients dont le calcul manque vraisemblablement de précision.

Une solution serait d'augmenter le nombre de points de calcul sur une période. Cependant, ceci augmente simultanément le temps de calcul, ainsi que le cumul d'erreur numérique inhérent à l'utilisation d'une machine limitée dans son seuil de précision. Il faut donc trouver un compromis entre l'exactitude de ces coefficients et le temps consacré à leur calcul.

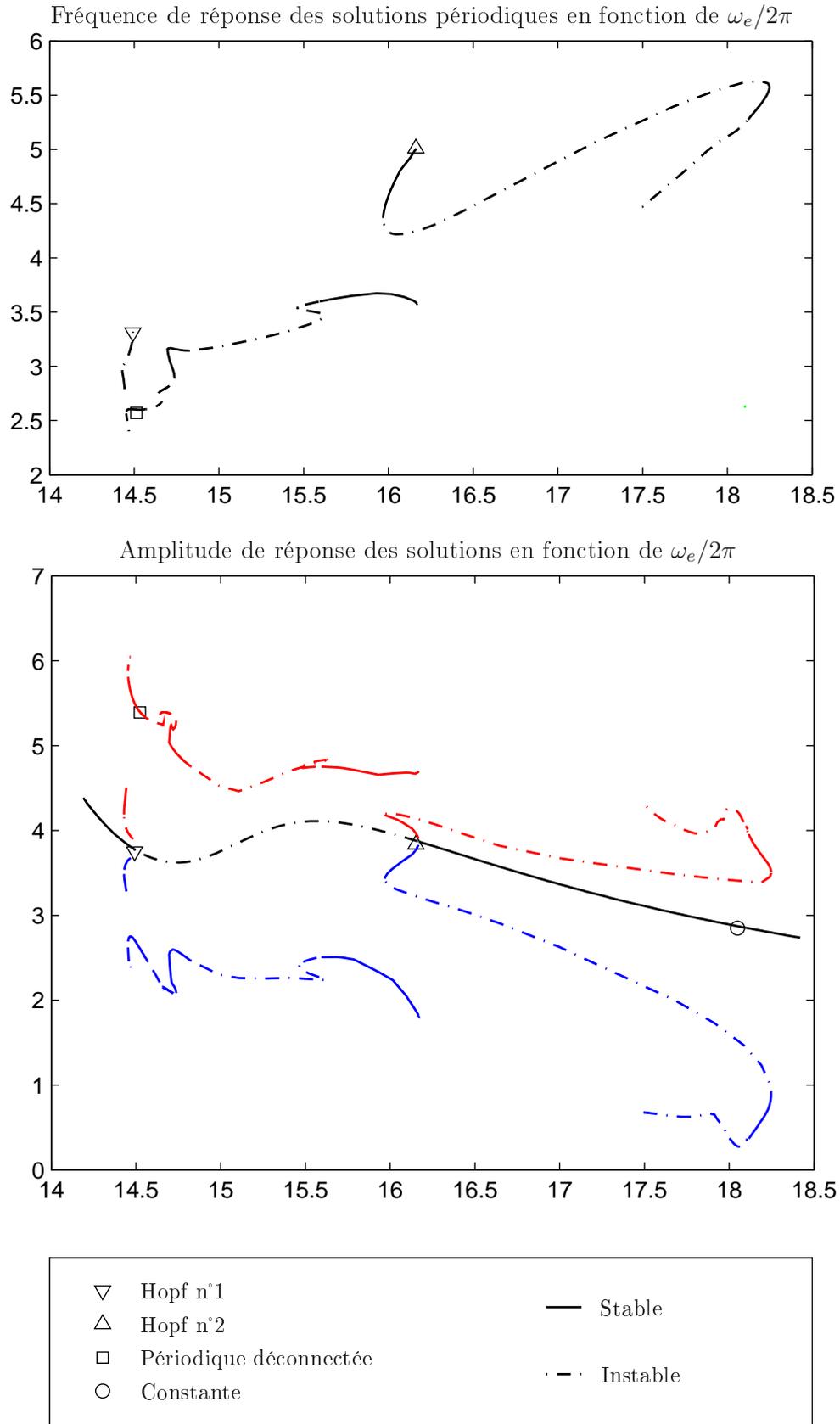


FIG. 7.37 – Contact rotor-stator rigide - Nouveau diagramme de bifurcation

**Comparaison entre méthode de tir et différences finies :** Lors de l'étude de la dernière solution périodique, nous avons tenté de mettre en œuvre une autre méthode de construction des solutions périodiques : celle des différences finies (cf. section 2.3.2). Nous nous sommes alors heurtés à plusieurs problèmes qu'il nous semble intéressant de mentionner.

Tout d'abord, intrinsèquement, cette méthode génère un système de grande taille dû à la discrétisation en temps qu'elle utilise. Elle nécessite donc une attention particulière à la résolution du système, qui exploiterait l'aspect « quasi-bande » de la matrice obtenue. De plus, si on la compare à celle de tir, elle s'avère plus coûteuse : considérons le nombre d'opérations nécessaires pour la réalisation d'une itération de chacune des deux méthodes en utilisant  $N_t$  points de discrétisation d'une période et une résolution de type Newton-Raphson.

a) Méthode des différences finies :  $N_t$  calculs de  $F(x^i, \omega_e)$  et de  $D_x F(x^i, \omega_e)$  puis inversion d'un système de taille  $n \times N_t$ .

b) Méthode de tir : Calcul de  $F(x^i, \omega_e)$  et de  $D_x F(x^i, \omega_e)$   $N_t$  fois puis inversion d'un système de taille  $n$ .

La méthode de tir semble donc moins coûteuse car nécessite le même nombre de calculs élémentaires mais ne nécessite l'inversion que d'un petit système. Elle a de surcroît l'avantage de fournir la matrice de monodromie, et donc de permettre rapidement l'analyse de la stabilité.

De plus, d'une façon purement qualitative, la convergence de la méthode des différences finies semble assez mauvaise : lorsqu'il lui est donné un point initial trop éloigné du cycle recherché, elle ne converge pas.

Enfin, dans le cas particulier de notre problème, étant donné que nous traitons des angles et qu'il est nécessaire d'avoir recours à une fonction  $\varphi$  supplémentaire afin de vérifier la périodicité du signal (cf. paragraphe 7.4.3), elle se révèle inadaptée : elle intègre en effet dans sa formulation la condition de périodicité  $x(T) - x(0) = 0$ , à laquelle il semble peu aisé de substituer  $\varphi(x(T) - x(0)) = 0$ .

### 7.4.5 Conclusion

L'étude de ce problème montre qu'un système apparemment simple peut avoir des comportements complexes que seule une étude complète peut dévoiler. Ceci nous a permis d'appliquer la majorité des méthodes d'étude locale exposées au chapitre 2, puis à proposer une forme de diagramme de bifurcation plus complète que celle rencontrée usuellement.

Par ailleurs, la présence d'une bifurcation de Hopf sous-critique a souligné l'importance d'un outil d'analyse globale. Ne connaissant pas a priori la période des solutions recherchées, nous n'avons pas pu mettre en œuvre l'outil que nous avons développé qui requiert que ce paramètre soit une donnée et non une inconnue. Ceci permet de dégager l'intérêt d'un travail ultérieur visant à rendre ce type de recherche possible.

## 7.5 Conclusion

Un rapide bilan sur les résultats obtenus par les quatre méthodes d'analyse globale appliquées à l'oscillateur de Duffing a permis d'affirmer sans peine que la méthode donnant le plus de satisfaction était celle fondée sur l'homotopie polynomiale. Cette conclusion a ensuite été

réaffirmée sur l'exemple d'un contact entre un rotor et un stator non-linéaire.

L'application avec succès de cette dernière méthode à différents systèmes illustrant des cas typiques d'étude dans le domaine des machines tournantes permet de démontrer son efficacité, son intérêt, et d'envisager sereinement son utilisation sur des systèmes plus importants correspondant aux besoins industriels.

Enfin, le dernier exemple de contact permanent entre un rotor et son stator, traité à l'aide des outils d'analyse locale cités dans le second chapitre de ce mémoire, a exhibé le besoin fort de poursuivre encore le travail pour être capable de libérer la pulsation de réponse du système comme une inconnue supplémentaire ; ceci permettra d'effectuer une analyse globale efficace à l'issue de bifurcations catastrophiques par exemple.

# Conclusion

L'objectif principal de ce travail de thèse était de proposer un outil permettant l'analyse globale de systèmes dynamiques non-linéaires. La principale difficulté dans cette recherche de toutes les solutions d'un système non-linéaire était de trouver une méthode qui ne soit pas trop consommatrice en termes de ressources numériques. Pour parvenir à élaborer un outil satisfaisant, nous avons exploré quatre familles de méthodes permettant de trouver toutes les solutions d'un système non-linéaire.

La première méthode, dite de cell-mapping et reposant sur une double discrétisation spatiale et temporelle, est attrayante par sa proximité à la formulation dynamique du problème et par la richesse des informations qu'elle fournit : solutions de toutes natures, bassins d'attraction et possibilité d'analyse de stabilité. Cependant, elle nécessite un temps de calcul ainsi qu'une allocation d'espace mémoire prohibitifs dès que l'on souhaite traiter des systèmes à plus de deux degrés de liberté.

Nous avons donc ensuite mis en œuvre des méthodes ne traitant pas directement le système différentiel mais un système algébrique obtenu en faisant une hypothèse sur la forme du déplacement. Elles ne peuvent donc pas rendre compte directement de toutes les solutions du système dynamique mais de toutes les solutions d'une catégorie donnée. La contrepartie de cette restriction est un gain en temps considérable. Les deux premières méthodes traitant des équations algébriques partitionnent l'espace de variation des inconnues algébriques en cellules qu'elles essaient d'exclure à l'aide d'un test d'exclusion pour la première et d'exclure ou à défaut de réduire pour la méthode travaillant sur l'arithmétique des intervalles. On observe une amélioration des performances par rapport à la méthode de cell-mapping, mais les estimations et les mises en œuvre montrent qu'il est difficile de traiter des systèmes à plus de trois degrés de liberté avec un temps de calcul inférieur à la journée.

L'absence d'hypothèse sur la forme de la non-linéarité nous a alors semblé être un obstacle majeur à l'efficacité des algorithmes testés. En effet, nous étions partis du principe que la méthode proposée ne devait pas être dédiée à un type de non-linéarité particulier ; ceci a pour inconvénient l'absence de cadre mathématique permettant de faire un travail analytique en amont qui réduirait en aval le coût numérique. Nous avons donc eu l'idée de nous placer dans le cadre richement étudié des polynômes en choisissant d'approcher l'expression de la non-linéarité algébrique par une famille d'éléments polynomiaux. On ne cherche donc plus toutes les solutions d'un système algébrique mais toutes les solutions d'un système multi-polynomial à plusieurs variables. La résolution de ce système se fait alors par une méthode homotopique : on construit un système multi-polynomial de même structure que celui que l'on souhaite résoudre dont on connaît les solutions ; puis, on injecte petit à petit le système à résoudre tandis qu'on diminue la contribution du second système dans les équations. En faisant une continuation des solutions de ce système au cours de son évolution, on a la garantie d'obtenir

toutes les solutions du système visé. Cette dernière méthode présente des temps de calcul très intéressants. Son application à différents systèmes du champ de la dynamique des rotors a permis de montrer d'une part son efficacité et d'autre part son intérêt en tant que méthode d'analyse globale, permettant de mettre à jour des branches de solutions déconnectées les unes des autres.

La solution proposée fournit donc un début de réponse à cette vaste question qu'est l'analyse globale de systèmes non-linéaires de grande taille. Il reste maintenant à étendre son champ d'action à d'autres types de solutions que les points fixes et les cycles limites. On imagine aisément sa mise en œuvre dans le cas des solutions quasi-périodiques en utilisant une transformation des équations dynamiques non par balance harmonique simple mais par balance multi-harmonique. L'enjeu majeur nous semble, pour ce type de solutions, de pouvoir étendre la recherche à des solutions de période inconnue. Le domaine de recherche de solutions chaotiques reste, lui, totalement vierge étant donnée la difficulté de trouver des descripteurs fiables des attracteurs.

Une autre piste de recherche qui nous semble judicieuse est l'utilisation de calculs parallèles : chaque solution étant continuée indépendamment des autres, cette méthode se prête particulièrement bien à la parallélisation. Sa mise en œuvre dans le contexte industriel où l'on rencontrera un parc informatique important mais dont les unités seront peu puissantes profiterait pleinement, à notre sens, de ce type d'architecture.

Enfin mentionnons le travail, secondaire dans le temps consacré et la difficulté rencontrée, de mise en œuvre de méthodes de continuation, d'analyse aux bifurcations et de stabilité. Ces méthodes, relativement bien représentées dans la littérature, sont des outils complémentaires à la conduite d'une analyse globale, permettant d'une part de suivre l'évolution des solutions éventuellement multiples exhibées par analyse globale et d'autre part de mettre en évidence les moments clés où effectuer une telle analyse en cas de détection d'une bifurcation catastrophique ou de perte de stabilité.

# Bibliographie

- [AG03] E. L. Allgower et K. Georg: *Introduction to Numerical Continuation Methods*. Springer-Verlag, 2003.
- [AM00] G. Alefeld et G. Mayer: *Interval analysis : theory and application*. Journal of Computational and Applied Mathematics, 121 :421–464, 2000.
- [AR03] F. Assis Rodriguez: *Étude de la dynamique non-linéaire d’amortisseurs du type squeeze film*. Thèse de doctorat, École Centrale de Lyon, 2003.
- [BBBE06] G. Baldo, A. Bonelli, O. Bursi et S. Erlicher: *The accuracy of the generalized- $\alpha$  method in the time integration of non-linear single- and two-DOF forced systems*. Computational Mechanics, 38 :15–31, 2006.
- [CH93] J. Chung et G. M. Hulbert: *A time integration algorithm for structural dynamics with improved numerical dissipation : the generalized- $\alpha$  method*. Journal of Applied Mechanics, 60 :371–374, June 1993.
- [Chi93] D. W. Childs: *Turbomachinery rotordynamics - Phenomena, modeling and analysis*. Wiley, 1993.
- [DCL05] Q. Ding, J. E. Cooper et A. Y. T. Leung: *Application of an improved cell mapping method to bilinear stiffness aeroelastic systems*. Journal of Fluids and Structures, 20 :35–49, 2005.
- [DJKP07] I. Danaila, P. Joly, S.M. Kaber et M. Postel: *An Introduction to Scientific Computing*. Springer, 2007.
- [EBB02] S. Erlicher, L. Bonaventura et O. S. Bursi: *The analysis of the generalized- $\alpha$  method for non-linear dynamic problems*. Computational Mechanics, 28 :83–104, 2002.
- [Ehr99] F. F. Ehrich: *Handbook of Rotordynamics*. Krieger Publishing Co, 1999.
- [Geo01] K. Georg: *Improving the efficiency of exclusion algorithms*. Advances in Geometry, 1 :193–210, 2001.
- [Geo03] K. Georg: *A new exclusion test*. Journal of Computational and Applied Mathematics, 152 :147–160, 2003.
- [GK99] R. Guder et E. Kreuzer: *Control of an Adaptive Refinement Technique of Generalized Cell Mapping by System Dynamics*. Nonlinear Dynamics, 20 :21–32, 1999.
- [Gon00] O. Gonzalez: *Exact energy and momentum conserving algorithms for general models in nonlinear elasticity*. Computer Methods in Applied Mechanics and Engineering, 190 :1763–1783, 2000.
- [GR93] M. Géradin et D. Rixen: *Théorie des Vibrations, Application à la dynamique des structures*. Masson, 1993.

- [Hsu87] C.S. Hsu: *Cell-to-Cell Mapping, a method of global analysis for non linear systems*. Springer-Verlag, 1987.
- [HW04] E. Hansen et G. W. Walster: *Global Optimization Using Interval Analysis*. M. Dekker, 2004.
- [KC99] D. Kuhl et A. Crisfield: *Energy-conserving and decaying algorithms in non-linear structural dynamics*. International Journal for Numerical Methods in Engineering, 45 :569–599, 1999.
- [Kuz04] Y. A. Kuznetsov: *Elements of Applied Bifurcation Theory, 3rd ed*. Springer, 2004.
- [Les07] N. Lesaffre: *Stabilité et Analyse Linéaire du Contact Rotor-Stator*. Thèse de doctorat, École Centrale de Lyon, 2007.
- [LF98] M. Lalanne et G. Ferraris: *Rotordynamics Prediction in Engineering*. Wiley, 1998. 2nd ed.
- [Li97] T. Y. Li: *Numerical solution of multivariate polynomial systems by homotopy continuation methods*. Acta Numerica, 6 :399–436, 1997.
- [LST07] N. Lesaffre, J. J. Sinou et F. Thouverez: *Contact analysis of a flexible bladed-rotor*. Journal of Mechanics A/Solids, 26(3) :541–557, 2007.
- [LSW97] J. Levitas, J. Singer et T. Weller: *Global dynamic stability of a shallow arch by Poincaré like simple cell mapping*. International Journal of Non-Linear Mechanics, 32(2) :411–424, 1997.
- [LSY87] T. Y. Li, T. Sauer et J. A. Yorke: *The random product homotopy and deficient polynomial systems*. Numerische Mathematik, 51(5) :481–500, 1987.
- [MBC03] R.O. Medrano, M.S. Baptista et I.L. Caldas: *Homoclinic orbits in a piecewise system and their relation with invariant sets*. Physica D : Nonlinear Phenomena, 186(3-4) :133–147, 2003.
- [MS87a] A. Morgan et A. Sommese: *Computing all solutions to polynomial systems using homotopy continuation*. Applied Mathematics and Computation, 24(2) :115–38, 1987.
- [MS87b] A. Morgan et A. Sommese: *A homotopy for solving general polynomial systems that respects m-homogeneous structures*. Applied Mathematics and Computation, 24(2) :101–113, 1987.
- [NB95] A. H. Nayfeh et B. Balachandran: *Applied Nonlinear Dynamics*. Wiley Series in Nonlinear Science, 1995.
- [NSP04] L. Noels, L. Stainier et J. P. Ponthot: *An energy-momentum conserving algorithm for non-linear hypoelastic constitutive models*. International Journal for Numerical Methods in Engineering, 59 :83–114, 2004.
- [PSP01] E. Pesheck, S. W. Shaw et C. Pierre: *Accurate Reduced-Order Models for a Simple Rotor Blade Model Using Nonlinear Normal Modes*. Mathematical and Computer Modelling, 33 :1085–1097, 2001.
- [PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling et B. P. Flannery: *Numerical Recipes in C, The Art of Scientific Computing, 2nd ed*. Cambridge University Press, 1992.

- [Sey88] R. Seydel: *From Equilibrium to Chaos, Practical Bifurcation and Stability Analysis*. Elsevier, 1988.
- [SN97] P. Sundararajan et S. T. Noah: *Dynamics of forced nonlinear systems using shooting/arclength continuation method - Application to rotor systems*. Journal of Vibration and Acoustics, 119 :9–20, 1997.
- [SP94] S. W. Shaw et C Pierre: *Normal Modes of Vibration for Non-Linear Continuous Systems*. Journal of Sound and Vibration, 169(3) :319–347, 1994.
- [ST92] J.C. Simo et N. Tarnow: *The discrete energy-momentum method. Conserving algorithm for nonlinear elastodynamics*. ZAMP, 43 :757–792, 1992.
- [SV00] A. J. Sommese et J. Verschelde: *Numerical Homotopies to Compute Generic Points on Positive Dimensional Algebraic Sets*. Journal of Complexity, 16 :572–602, 2000.
- [SVW03] A. J. Sommese, J. Verschelde et C. W. Wampler: *Introduction to numerical algebraic geometry*. Dans *Graduate School on Systems of Polynomial Equations : From Algebraic Geometry to Industrial Applications*, pages 229–247, 2003.
- [SW05] A. J. Sommese et C. W. Wampler: *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*. World Scientific Publishing Co. Pte. Ltd., 2005.
- [TS02] J. M. T. Thompson et H. B. Stewart: *Nonlinear Dynamics and Chaos, 2nd ed*. Wiley, 2002.
- [Van88] J. M. Vance: *Rotordynamics of Turbomachinery*. Wiley, 1988.
- [vGE01] G. von Groll et D. J. Ewins: *The harmonic balance method with arclength continuation in rotor/stator contact problems*. Journal of Sound and Vibration, 241(2) :223–233, 2001.
- [Wam92] C. W. Wampler: *Bézout number calculations for multi-homogeneous polynomial systems*. Applied Mathematics and Computation, 157(2-3) :143–157, 1992.
- [Wat02] L. T. Watson: *Probability-one homotopies in computational science*. Journal of Computational and Applied Mathematics, 140(1-2) :785–807, 2002.
- [Wri85] A. H. Wright: *Finding all solutions to a system of polynomial equations*. Mathematics of Computation, 44(169) :125–133, 1985.
- [XZL97] Z B. Xu, J S. Zhang et Y W. Leung: *A general CDC Formulation for Specializing the Cell Exclusion Algorithm of Finding All Zeros of Vector Functions*. Applied Mathematics and Computation, 86 :235–259, 1997.
- [XZW96] Z B. Xu, J S. Zhang et W. Wang: *A cell exclusion algorithm for determining all the solutions of a nonlinear system of equations*. Applied Mathematics and Computation, 80 :181–208, 1996.
- [YKT98] K. Yamamura, P. Kawata et A. Tokue: *Interval solution of nonlinear equations using linear programming*. BIT, 38(1) :186–199, 1998.
- [YT02] K. Yamamura et S. Tanaka: *Finding all solutions of systems of nonlinear equations using the dual simplex method*. BIT, 42(1) :214–230, 2002.

- [ZH95] J. Y. Zhao et E. J. Hahn: *Eccentric Operation and Blade-Loss Simulation of a Rigid Rotor Supported by an Improved Squeeze Film Damper*. Journal of Tribology, 117 :490–497, 1995.
- [Zul88] W. Zulehner: *A simple homotopy method for determining all isolated solutions to polynomial systems*. Mathematics of Computation, 50(181) :167–177, 1988.

# Table des figures

1.1	Illustration des temps direct et indirect . . . . .	7
1.2	Bassin d'attraction schématique . . . . .	8
1.3	Portrait de phase schématique . . . . .	9
1.4	Sections de Poincaré schématiques . . . . .	10
1.5	Intérêt du spectre fréquentiel . . . . .	12
1.6	Portraits de phase types des différentes classes de points fixes hyperboliques . . . . .	13
1.7	Bifurcations des solutions types points fixes . . . . .	16
1.8	Bifurcations des solutions périodiques . . . . .	17
1.9	Diagramme de bifurcation schématique . . . . .	18
2.1	Intégration temporelle - Algorithme de Newmark - Cas linéaire . . . . .	24
2.2	Intégration temporelle - Algorithme de Newmark - Cas non-linéaire . . . . .	26
2.3	Algorithme général d'une méthode de continuation . . . . .	36
2.4	Continuation : Principaux prédicteurs . . . . .	38
2.5	Continuation : Principaux correcteurs . . . . .	39
2.6	Intérêt de l'adaptation du pas de continuation . . . . .	40
2.7	Destruction du point d'embranchement par perturbation . . . . .	41
3.1	Modes de construction d'un mapping . . . . .	46
3.2	Illustration d'un simple cell-mapping . . . . .	48
3.3	Algorithme d'analyse d'un cell-mapping . . . . .	49
3.4	Cell-mapping - Illustration de la méthode d'analyse . . . . .	51
3.5	Cell-mapping - Duffing : Bassin d'attraction pour $\omega = 1.2 \text{ rad.s}^{-1}$ . . . . .	53
3.6	Cell-mapping - Duffing : Bassin d'attraction pour $\omega = 1.5 \text{ rad.s}^{-1}$ . . . . .	53
3.7	Cell-mapping - Duffing : raffinement autour de la première solution . . . . .	55
3.8	Cell-mapping - Duffing : raffinement autour de la deuxième solution . . . . .	55
3.9	Cell-mapping - Duffing : raffinement autour de la troisième solution . . . . .	56
3.10	Cell-mapping - Duffing : portraits de phase des trois solutions . . . . .	56
3.11	Cell-mapping - Extrapolation des temps de calcul et des espaces de stockage . . . . .	58
4.1	Méthodes d'exclusion - Algorithme général . . . . .	60
4.2	Méthodes d'exclusion - Méthode du simplexe - Tableaux de la Phase II . . . . .	71
4.3	Méthode du simplexe - Cas 1H . . . . .	74

4.4	Méthode du simplexe - Cas C1H . . . . .	75
4.5	Méthode du simplexe - Cas 1H3H . . . . .	76
4.6	Méthode du simplexe - Extrapolation du nombre de cellules . . . . .	77
4.7	Méthode du simplexe - Extrapolation du temps par cellule . . . . .	77
4.8	Méthode du simplexe - Extrapolation du temps de calcul . . . . .	78
5.1	Arithmétique des intervalles - Exemple d'ensembles de départ . . . . .	81
5.2	Arithmétique des intervalles - Algorithme Général . . . . .	82
5.3	Arithmétique des intervalles - Illustration du hull . . . . .	84
5.4	Arithmétique des intervalles - Algorithme de Box Consistency . . . . .	86
5.5	Arithmétique des intervalles - Cas 1H . . . . .	93
5.6	Arithmétique des intervalles - Cas C1H . . . . .	93
5.7	Arithmétique des intervalles - Cas 1H3H . . . . .	94
5.8	Arithmétique des intervalles - Extrapolation du nombre de cellules . . . . .	95
5.9	Arithmétique des intervalles - Extrapolation du temps par cellule . . . . .	96
5.10	Arithmétique des intervalles - Extrapolation du temps de calcul . . . . .	96
6.1	Illustration du principe de la méthode homotopique . . . . .	103
6.2	Graphique des différentes méthodes permettant d'élaborer $\mathbf{Q}(\tilde{x})$ . . . . .	107
6.3	Système étudié . . . . .	114
6.4	Continuation du système non-linéaire cubique à 3 ddl : amplitude de $r_3$ . . . . .	116
6.5	Solutions réelles multiples obtenues par homotopie . . . . .	118
6.6	Homotopies polynomiales - Cas 1H - Solution 1 . . . . .	120
6.7	Homotopies polynomiales - Cas 1H - Solution 2 . . . . .	120
6.8	Homotopies polynomiales - Cas 1H - Solution 3 . . . . .	121
6.9	Homotopies polynomiales - Cas 1H - Chemin divergent . . . . .	121
6.10	Homotopies polynomiales - Extrapolation du nombre de cellules . . . . .	122
6.11	Homotopies polynomiales - Extrapolation du temps par chemin . . . . .	123
6.12	Homotopies polynomiales - Extrapolation du temps de calcul . . . . .	123
7.1	Contact permanent non-frottant aube-carter - Modèle . . . . .	127
7.2	Contact rotor-stator non frottant - Analyse linéaire . . . . .	128
7.3	Contact rotor-stator non frottant - Solutions . . . . .	132
7.4	Squeeze-film - Schéma simplifié d'un palier . . . . .	133
7.5	Squeeze-film - Modèle dynamique étudié . . . . .	134
7.6	Squeeze-film - Influence de $\beta$ sur $\epsilon$ . . . . .	136
7.7	Squeeze-film - Influence de $\beta$ sur $\epsilon_1$ et $\epsilon_2$ . . . . .	136
7.8	Squeeze-film - Erreur de l'approximation polynomiale . . . . .	138
7.9	Squeeze-film - Erreur relative moyenne par couronne (%) - Degré 6 . . . . .	139
7.10	Squeeze-film - Erreur relative moyenne par couronne (%) - Degré 12 . . . . .	139

7.11 Squeeze-film - Découverte d'une nouvelle branche de solutions . . . . .	141
7.12 Squeeze-film - Stabilité des branches . . . . .	142
7.13 Squeeze-film - Intégrations temporelles . . . . .	143
7.14 Contact rotor-stator rigide - Modèle . . . . .	145
7.15 Contact rotor-stator rigide - Notations . . . . .	145
7.16 Contact rotor-stator rigide - Amortissement hystérétique - $s$ . . . . .	150
7.17 Contact rotor-stator rigide - Amortissement hystérétique - $F_c$ . . . . .	150
7.18 Contact rotor-stator souple - Amortissement hystérétique - $s$ . . . . .	151
7.19 Contact rotor-stator souple - Amortissement hystérétique - $F_c$ . . . . .	151
7.20 Contact rotor-stator rigide - Amortissement visqueux . . . . .	152
7.21 Contact rotor-stator rigide - Stabilité des points fixes . . . . .	153
7.22 Contact rotor-stator rigide - Intégration temporelle pour $\omega_e/2\pi = 12.63 \text{ Hz}$ . . . . .	155
7.23 Contact rotor-stator rigide - 12.63 Hz - Conditions initiales perturbées . . . . .	156
7.24 Contact rotor-stator rigide - 14.60 Hz - Instabilité de la solution constante . . . . .	157
7.25 Contact rotor-stator rigide - Intégration temporelle en $\omega_e \approx \omega_{e1}$ . . . . .	157
7.26 Contact rotor-stator rigide - Intégration temporelle en $\omega_e \approx \omega_{e2}$ . . . . .	158
7.27 Algorithmes utilisés pour la continuation périodique . . . . .	159
7.28 Contact rotor-stator rigide - Bifurcation de Hopf n°1 Pulsation et stabilité . . . . .	162
7.29 Contact rotor-stator rigide - Bifurcation de Hopf n°1 - Amplitudes . . . . .	162
7.30 Contact rotor-stator rigide - Bifurcation de Hopf n°2 - Pulsation et stabilité . . . . .	163
7.31 Contact rotor-stator rigide - Bifurcation de Hopf n°2 - Amplitudes . . . . .	164
7.32 Contact rotor-stator rigide - Bifurcation de Hopf n°2 - Seconde bifurcation . . . . .	164
7.33 Contact rotor-stator rigide - Étude de la solution de période double . . . . .	165
7.34 Contact rotor-stator rigide - Solution déconnectée - Périodes . . . . .	165
7.35 Contact rotor-stator rigide - Solution déconnectée - Pulsation et stabilité . . . . .	166
7.36 Contact rotor-stator rigide - Solution déconnectée - Amplitudes . . . . .	167
7.37 Contact rotor-stator rigide - Nouveau diagramme de bifurcation . . . . .	168
A.1 Schéma d'un oscillateur de Duffing . . . . .	191
A.2 Oscillateur de Duffing - Fonction de réponse en fréquence . . . . .	192
A.3 Oscillateur de Duffing étudié - Fonction de réponse en fréquence . . . . .	197
A.4 Oscillateur de Duffing étudié - Evolution des composantes harmoniques . . . . .	198
B.1 Arithmétique des intervalles - Algorithme d'Hansen & Walster détaillé . . . . .	202
B.2 Arithmétique des intervalles - Algorithme d'Hansen & Walster détaillé (suite) . . . . .	203



# Liste des tableaux

2.1	Synthèse des différents algorithmes d'intégration temporelle en linéaire . . . . .	27
3.1	Cell-mapping - Illustration de la méthode d'analyse . . . . .	51
3.2	Cell-mapping - Oscillateur de Duffing : caractéristiques numériques . . . . .	54
4.1	Méthode du simplexe - Cas d'application à un oscillateur de Duffing . . . . .	73
4.2	Méthode du simplexe - Résultats . . . . .	73
5.1	Arithmétique des intervalles - Résultats . . . . .	92
6.1	Homotopies polynomiales - Comparaison des méthodes . . . . .	117
6.2	Homotopies polynomiales - Autres cas d'application à un oscillateur de Duffing .	119
6.3	Homotopies polynomiales - Résultats . . . . .	119
7.1	Contact rotor-stator non frottant - Résultats . . . . .	130
7.2	Contact rotor-stator non frottant - Solutions . . . . .	131
7.3	Squeeze-film - Résultats des différentes méthodes homotopiques . . . . .	140



# Liste des définitions

1.1	Système autonome ou non-autonome . . . . .	4
1.2	Stabilité au sens de Lyapunov ou uniforme . . . . .	5
1.3	Stabilité asymptotique . . . . .	6
1.4	Application de Poincaré . . . . .	9
1.5	Transformée de Fourier discrète (DFT) . . . . .	11
2.1	Schéma d'intégration temporelle explicite ou implicite . . . . .	20
2.2	Consistance d'une méthode d'intégration . . . . .	20
2.3	Stabilité d'une méthode d'intégration . . . . .	20
2.4	Tangente induite par une matrice . . . . .	36
2.5	Matrice Pseudo-Inverse de Moore-Penrose . . . . .	39
2.6	Point d'embranchement (Branching Point) . . . . .	40
2.7	Point d'embranchement simple (Simple branching point) . . . . .	42
6.1	Polynôme à plusieurs variables . . . . .	100
6.2	Espace projectif $\mathbb{P}^P$ . . . . .	104



# Liste des lemmes et théorèmes

2.1	Théorème de Floquet . . . . .	34
2.2	Lemme relatif aux points d'embranchement . . . . .	43
4.1	Domination par une fonction décomposable en série entière . . . . .	64
6.1	Homotopie de degré total . . . . .	108
6.2	Homotopie de multi-homogénéisation . . . . .	110
6.3	Homotopie par produit d'éléments linéaires . . . . .	113



# Nomenclature

**Équation dynamique (ordre 2) :**  $M.\ddot{q} + C.\dot{q} + K.q + \hat{f}(q, \dot{q}) = f_e(t)$

- $q$  : vecteur d'inconnues issues du paramétrage du problème,  $q \in \mathbb{R}^N$
- $\dot{q}$ ,  $\ddot{q}$  : respectivement vecteurs vitesse et accélération,  $\dot{q} = \frac{dq}{dt}$ ,  $\ddot{q} = \frac{d^2q}{dt^2}$
- $M$ ,  $C$ ,  $K$  : matrices respectives dites de masse, de dissipation et de raideur
- $\hat{f}(q, \dot{q})$  : vecteur des efforts non-linéaires
- $f_e(t)$  : vecteur des efforts extérieurs

**Équation augmentée (ordre 1) :**  $A.\dot{x} + B.x + \hat{F}(x) = F_e(t)$

- $x = \begin{Bmatrix} q \\ \dot{q} \end{Bmatrix}$  : vecteur d'inconnues augmenté,  $x \in \mathbb{R}^{2N}$
- $A = \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix}$ ,  $B = \begin{bmatrix} 0 & -K \\ K & C \end{bmatrix}$
- $\hat{F}(x, t) = \begin{Bmatrix} 0 \\ \hat{f}(q, \dot{q}) \end{Bmatrix}$  : vecteur augmenté des efforts non-linéaires
- $F_e = \begin{Bmatrix} 0 \\ f_e(t) \end{Bmatrix}$  : vecteur augmenté des efforts extérieurs

**Équation algébrisée :**  $\bar{H}.\tilde{x} + \hat{H}(\tilde{x}) = H_c$  ou  $H(\tilde{x}) = 0$

- $\tilde{x}$  : vecteur d'inconnues algébriques,  $\tilde{x} \in \mathbb{R}^P$
- $\bar{H}$  : partie linéaire
- $\hat{H}(\tilde{x})$  : partie non-linéaire
- $H_c$  : terme constant (contribution des efforts extérieurs)

**Dérivation :**

- d'une fonction  $F$  par rapport à la variable  $x$  :  $D_x F = \frac{\partial F}{\partial x}$
- d'une variable  $x$  par rapport au temps « vrai »  $t$  :  $\dot{x} = \frac{\partial x}{\partial t}$
- d'une variable  $x$  par rapport au temps adimensionnalisé  $\tau$  :  $x' = \frac{\partial x}{\partial \tau}$

**Cellules et intervalles :**

- Pavé de  $\mathbb{R}^Q$  :  $\tilde{X} = \{\tilde{X}_1, \dots, \tilde{X}_Q\}$  ou  $\tilde{X} = \tilde{X}_1 \times \dots \times \tilde{X}_Q$
- Bornes inférieure et supérieure d'un intervalle  $\tilde{X}_i$  : respectivement  $\underline{\tilde{X}}_i$  et  $\overline{\tilde{X}}_i$
- Centre d'un pavé  $\tilde{X}$  :  $\text{mid}(\tilde{X})$ ,  $\forall i \in \llbracket 1, Q \rrbracket$ ,  $\text{mid}(\tilde{X}_i) = \frac{\underline{\tilde{X}}_i + \overline{\tilde{X}}_i}{2}$
- Largeur d'un pavé  $\tilde{X}$  :  $w(\tilde{X}_i) = \overline{\tilde{X}}_i - \underline{\tilde{X}}_i$  et  $w(\tilde{X}) = \max_{1 \leq i \leq Q} (w(\tilde{X}_i))$
- Magnitude d'une composante  $\tilde{X}_i$  :  $\text{mag}(\tilde{X}_i) = \max(|\underline{\tilde{X}}_i|, |\overline{\tilde{X}}_i|)$

**Schémas d'algorithmes :**

- Stockage de l'élément  $x$  dans l'élément  $y$  :  $y \leftarrow x$
- Utilisation de l'élément  $x$  pour le calcul de  $y$  :  $x \mapsto y$
- Ajout de l'élément  $x$  à une liste  $xList$  :  $xList \stackrel{+}{\leftarrow} x$

**Symboles divers :**

- Ensemble des conditions initiales considérées lors de l'analyse globale d'un système mis sous la forme d'une équation augmentée :  $\mathcal{E}_{IC} \subset \mathbb{R}^{2N}$
- Produit scalaire hermitien :  $\langle x, y \rangle = \bar{x}^t \cdot y$

**Abréviations :**

- ddl : degré de liberté
- CDC : Cell Discarded Condition, test d'exclusion

# Annexes



## Annexe A

# Oscillateur de Duffing

---

Un oscillateur de Duffing est un oscillateur classique dans le cadre des systèmes dynamiques. On peut le représenter mécaniquement par un oscillateur amorti classique auquel on adjoint une raideur non-linéaire de forme cubique comme le montre la figure A.1. L'équation

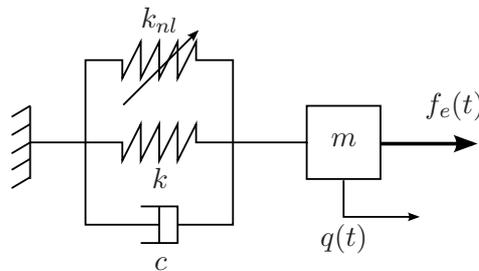


FIG. A.1 – Schéma d'un oscillateur de Duffing

différentielle s'écrit très simplement :

$$m\ddot{q} + c\dot{q} + kq + k_{nl}q^3 = f_e \cos(\omega t) \quad (\text{A.1})$$

ou encore

$$\ddot{q} + 2\xi\omega_0\dot{q} + \omega_0^2q + \beta q^3 = \Gamma \cos(\omega t) \quad (\text{A.2})$$

en posant  $\omega_0^2 = \frac{k}{m}$ ,  $\xi = \frac{c}{2m\omega_0}$ ,  $\beta = \frac{k_{nl}}{m}$  et  $\Gamma = \frac{f_e}{m}$ .

La présence d'une telle non-linéarité peut avoir un effet dit mollissant ou raidissant comme on peut le voir sur les courbes typiques de la figure A.2. La courbe (a) est obtenue pour une valeur négative de  $k_{nl}$  et représente un effet mollissant : le pic de résonance penche vers la

gauche ; la résonance arrive donc plus « tôt » ou du moins se prolonge pour des valeurs de fréquence d'excitation inférieures à celle où le pic prend naissance. Inversement, la figure (c) représente un effet raidissant pour une valeur de  $k_{nl}$  positive.

Outre ces effets remarquables, cette non-linéarité présente l'avantage d'être simple dans son expression analytique, continue et d'être bien tabulée dans la littérature. En mécanique, elle est utilisée avec un coefficient  $k_{nl}$  positif pour la modélisation de grands déplacements.

Nous proposons ici d'appliquer en détail deux méthodes non-linéaires permettant de transformer l'équation différentielle (A.1) en un système d'équations algébriques. Nous présenterons donc dans un premier temps la méthode des échelles multiples qui fait partie des méthodes de perturbations puis la méthode de balance harmonique, méthode fréquentielle, qui nous donnera des expressions réutilisées au fil des chapitres 3 à 6.

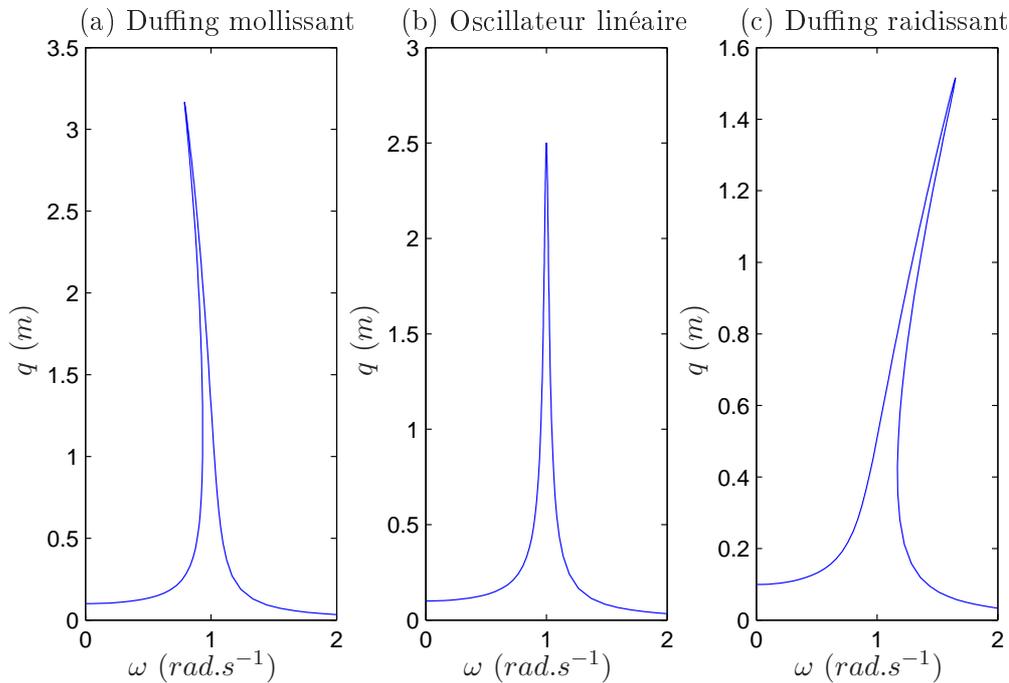


FIG. A.2 – Oscillateur de Duffing - Fonction de réponse en fréquence

## A.1 Application de la méthode des échelles multiples

Afin d'appliquer cette méthode, il convient de considérer les termes :

- d'amortissement :  $2\xi\omega_0\dot{q}$
- non linéaire :  $\beta q^3$
- d'excitation :  $\Gamma \cos(\omega t)$

comme des termes de perturbation du système linéaire (A.3).

$$\ddot{q} + \omega_0^2 q = 0 \quad (\text{A.3})$$

Il est alors légitime de les supposer petits et de les affecter d'un coefficient de pondération adimensionné  $0 < \epsilon < 1$ . On obtient ainsi la réécriture suivante de l'équation (A.2) :

$$\ddot{q} + \omega_0^2 q = \epsilon [-2\xi\omega_0\dot{q} - \beta q^3 + \Gamma \cos(\omega t)] \quad (\text{A.4})$$

La solution  $q(t, \epsilon)$  de cette équation peut s'écrire sous la forme du développement selon  $\epsilon$  suivant :

$$q(t, \epsilon) = q_0(t, \epsilon = 0) + \sum_{k=1}^n \epsilon^k \frac{\partial^k q}{\partial \epsilon^k}(t, \epsilon = 0) + o(\epsilon^n) \quad (\text{A.5})$$

En notant  $q_k(t) = \frac{\partial^k q}{\partial \epsilon^k}(t, \epsilon = 0)$ , on obtient l'écriture de l'approximation de  $q(t, \epsilon)$  :

$$q(t, \epsilon) \approx q_0(t) + \epsilon q_1(t) + \epsilon^2 q_2(t) + \dots + \epsilon^n q_n(t) \quad (\text{A.6})$$

$q_0(t)$  est la solution à (A.3) et les termes en  $\epsilon^k$ , la réponse à la perturbation. En réinjectant le développement (A.6) dans (A.4), et en ne gardant que les termes d'ordre inférieur ou égal à 1 en  $\epsilon$ , on obtient l'équation :

$$(\ddot{q}_0 + \epsilon \ddot{q}_1) + \omega_0^2(q_0 + \epsilon q_1) = \epsilon [-2\xi\omega_0 q_0 - \beta q_0^3 + \Gamma \cos(\omega t)] \quad (\text{A.7})$$

On introduit alors les échelles de temps suivantes :  $T_i = \epsilon^i t$ ,  $i = 1, \dots, n$ . Ainsi, si on considère les  $q_k(t)$  comme des fonctions de  $T_0, T_1, \dots, T_n$ , on a :

$$\frac{dq_k(T_0, T_1, \dots, T_n)}{dt} = \frac{dT_0}{dt} \frac{\partial q_k}{\partial T_0} + \frac{dT_1}{dt} \frac{\partial q_k}{\partial T_1} + \dots + \frac{dT_n}{dt} \frac{\partial q_k}{\partial T_n} = \frac{\partial q_k}{\partial T_0} + \epsilon \frac{\partial q_k}{\partial T_1} + \dots + \epsilon^n \frac{\partial q_k}{\partial T_n}$$

Étant donné que nous avons fait le choix d'un développement au premier ordre en  $\epsilon$ , il nous faut retenir  $T_0$  et  $T_1$  comme échelles de temps afin d'être cohérent dans les ordres du développement. On écrira donc pour une quantité  $x(T_0, T_1)$  :

$$\frac{dx(T_0, T_1)}{dt} = \frac{\partial x}{\partial T_0} + \epsilon \frac{\partial x}{\partial T_1} = D_0 x + \epsilon D_1 x \quad (\text{A.8})$$

$$\frac{d^2 x(T_0, T_1)}{dt^2} = \frac{\partial^2 x}{\partial T_0^2} + 2\epsilon \frac{\partial^2 x}{\partial T_0 \partial T_1} + \epsilon^2 \frac{\partial^2 x}{\partial T_1^2} = D_0^2 x + 2\epsilon D_0 D_1 x \quad (\text{A.9})$$

avec la notation  $D_i = \frac{\partial}{\partial T_i}$

Enfin, souhaitant étudier le comportement de l'oscillateur proche de sa pulsation de résonance ( $\omega = \omega_0 + \epsilon\sigma$ ), on pose :

$$\omega t = (\omega_0 + \epsilon\sigma)t = (\omega_0 + \epsilon\sigma)t = \omega_0 T_0 + \sigma T_1 \quad (\text{A.10})$$

$\sigma$  est couramment appelé *detuning parameter* et permet de gérer l'écart de la pulsation d'excitation à celle propre de l'oscillateur non amorti.

En réinjectant l'utilisation des échelles de temps  $T_0$  et  $T_1$  dans l'équation (A.2) puis en égalant les coefficients de même degré en  $\epsilon$ , on obtient le système d'équations différentielles suivant :

$$D_0^2 q_0 + \omega_0^2 q_0 = 0 \quad (\text{A.11a})$$

$$D_0^2 q_1 + \omega_0 q_1 = -2D_0 D_1 q_0 - 2\xi\omega_0 D_0 q_0 - \beta q_0^3 + \Gamma \cos(\omega_0 T_0 + \sigma T_1) \quad (\text{A.11b})$$

L'équation (A.11a) donne la forme suivante pour  $q_0$  :

$$q_0(T_0, T_1) = A(T_1) e^{i\omega_0 T_0} + \overline{A(T_1)} e^{-i\omega_0 T_0} \quad (\text{A.12})$$

Puis, en injectant cette forme de  $q_0$  dans (A.11b) :

$$D_0^2 q_1 + q_1 = -[2i\omega_0(A' + \xi\omega_0 A) + \beta(3A^2\bar{A})]e^{i\omega_0 T_0} - \beta A^3 e^{i3\omega_0 T_0} + \frac{\Gamma}{2} e^{i\omega_0 T_0} e^{i\sigma T_1} + [cc]$$

où  $A' = \frac{dA}{dT_1}$  et  $[cc]$  désigne la partie complexe conjuguée du terme écrit à droite.

Afin d'éliminer les termes séculiers induits par la présence de  $e^{i\sigma T_1}$ , il convient d'imposer :

$$-[2i\omega_0(A' + \xi\omega_0 A) + \beta(3A^2\bar{A})] + \frac{\Gamma}{2} e^{i\sigma T_1} = 0 \quad (\text{A.13})$$

Il reste donc l'équation suivante servant à trouver  $q_1$  :

$$D_0^2 q_1 + q_1 = -\beta A^3 e^{i3\omega_0 T_0} \quad (\text{A.14})$$

En revanche, nous nous intéressons à  $q_0$ , donc à l'expression de  $A(T_1)$ , régie par (A.13). En cherchant  $A(T_1)$  sous sa forme polaire (A.15), on obtient l'équation différentielle complexe (A.16) que l'on sépare en un système de deux équations différentielles réelles (A.17).

$$A(T_1) = \frac{a(T_1)}{2} e^{i\theta(T_1)}, \quad a(T_1) \in \mathbb{R}, \quad \theta(T_1) \in \mathbb{R} \quad (\text{A.15})$$

$$i\omega_0(a' + ia\theta' + \xi\omega_0 a) + \frac{3}{8}\beta a^3 - \frac{\Gamma}{2} e^{i(\sigma T_1 - \theta)} = 0 \quad (\text{A.16})$$

$$\begin{cases} a' + \xi\omega_0 a &= \frac{\Gamma}{2\omega_0} \sin(\sigma T_1 - \theta) \\ a\theta' - \frac{3}{8\omega_0}\beta a^3 &= \frac{\Gamma}{2\omega_0} \cos(\sigma T_1 - \theta) \end{cases} \quad (\text{A.17})$$

Afin de rendre le système autonome (i.e. faire disparaître  $T_1$ ), on pose :

$$\gamma = \sigma T_1 - \theta(T_1) \quad (\text{A.18})$$

Le système d'équations différentielles à traiter devient :

$$\begin{cases} a' + \xi\omega_0 a &= \frac{\Gamma}{2\omega_0} \sin \gamma \\ -a\gamma' + \sigma a - \frac{3}{8\omega_0}\beta a^3 &= \frac{\Gamma}{2\omega_0} \cos \gamma \end{cases}, \text{ en gardant la convention } x' = \frac{dx}{dT_1} \quad (\text{A.19})$$

Nous ne nous intéressons qu'aux solutions stationnaires :  $a' = 0$  et  $\theta' = 0$ . En prenant en compte ces simplifications et en sommant les équations de (A.19) élevées au carré, on obtient la relation suivante entre l'amplitude  $a$  de la réponse stationnaire, le décalage de la pulsation d'excitation  $\sigma$  par rapport à la pulsation propre du système ( $\omega_0$ ) et l'amplitude de l'excitation  $\Gamma$  :

$$\left[ \xi^2 \omega_0^2 + \left( \sigma - \frac{3}{8} \frac{\beta}{\omega_0} a^2 \right)^2 \right] a^2 = \frac{\Gamma^2}{4\omega_0^2} \quad (\text{A.20})$$

Grâce à cette équation, et en appliquant une méthode d'analyse globale sur l'amplitude  $a$ , on peut ainsi pour un jeu de paramètres d'excitation  $(\sigma, \Gamma)$ , calculer toutes les solution pour  $a$  puis  $\gamma$ . L'expression de  $q(t, \epsilon)$  est alors :

$$q(t, \epsilon) = a \cos(\omega t - \gamma) + O(\epsilon)$$

## A.2 Application de la méthode de la balance harmonique

Étant donné que nous avons restreint nos recherches à des solutions stationnaires qui s'avèrent être synchrones avec l'excitation, donc, de pulsation connue, nous nous proposons de retrouver le résultat de l'équation (A.20) par une méthode simple de balance harmonique qui n'est pas une méthode perturbative. On recherche donc  $q(t)$  solution de (A.2) sous la forme :

$$q(t) = a_0 + a_1 \cos(\omega t) + b_1 \sin(\omega t) \quad (\text{A.21})$$

La seule difficulté est ici le terme en  $q^3$  qu'il faut développer sous forme d'harmoniques. Ceci se fait en utilisant les formules suivantes :

$$\begin{aligned} (a_0 + a_1 \cos \omega t + b_1 \sin \omega t)^3 &= a_0^3 + a_1^3 \cos^3 \omega t + b_1^3 \sin^3 \omega t + 3(a_0^2 a_1 \cos \omega t + a_0^2 b_1 \sin \omega t \\ &\quad + a_0 a_1^2 \cos^2 \omega t + a_0 b_1^2 \sin^2 \omega t + a_0 b_1^2 \sin^2 \omega t + a_1 b_1^2 \cos \omega t \sin^2 \omega t) \\ &\quad + 6a_0 a_1 b_1 \cos \omega t \sin \omega t \\ \cos^2 \omega t &= \frac{1}{2} + \frac{1}{2} \cos 2\omega t \\ \sin^2 \omega t &= \frac{1}{2} - \frac{1}{2} \cos 2\omega t & \cos \omega t \sin \omega t &= \frac{1}{2} \sin 2\omega t \\ \cos^3 \omega t &= \frac{3}{4} \cos \omega t + \frac{1}{4} \cos 3\omega t & \cos^2 \omega t \sin \omega t &= \frac{1}{4} \sin \omega t + \frac{1}{4} \sin 3\omega t \\ \sin^3 \omega t &= \frac{3}{4} \sin \omega t - \frac{1}{4} \sin 3\omega t & \cos \omega t \sin^2 \omega t &= \frac{1}{4} \cos \omega t - \frac{1}{4} \cos 3\omega t \end{aligned}$$

En appliquant successivement le produit scalaire  $(.,.)_\omega$  (cf. (2.19)) à l'équation (A.2) dans laquelle l'expression harmonique (A.21) de  $q$  a été injectée, pour  $g(t) = 1$ ,  $g(t) = \cos \omega t$ , et  $g(t) = \sin \omega t$ , on obtient respectivement les équations (A.22a), (A.22b) et (A.22c) qui constituent le système algébrisé par balance harmonique à un harmonique correspondant à l'équation différentielle (A.2) :

$$a_0[\omega_0^2 + \beta(a_0^2 + \frac{3}{2}(a_1^2 + b_1^2))] = 0 \quad (\text{A.22a})$$

$$(\omega_0^2 - \omega^2)a_1 + 2\xi\omega_0\omega b_1 + 3\beta a_1(a_0^2 + \frac{1}{4}(a_1^2 + b_1^2)) = \Gamma \quad (\text{A.22b})$$

$$(\omega_0^2 - \omega^2)b_1 - 2\xi\omega_0\omega a_1 + 3\beta b_1(a_0^2 + \frac{1}{4}(a_1^2 + b_1^2)) = 0 \quad (\text{A.22c})$$

Si l'on souhaite retrouver le résultat (A.20) portant sur le module  $a$  de  $q(t)$ , il suffit de constater que l'équation (A.22a) impose la nullité de  $a_0$  puis d'écrire  $a = \sqrt{a_1^2 + b_1^2}$ . On élève alors au carré (A.22b) et (A.22c) et on somme pour obtenir :

$$a^2 \left[ ((\omega_0^2 - \omega^2) + \frac{3}{4}\beta a)^2 + (2\xi\omega_0\omega)^2 \right] = \Gamma^2$$

On peut, par le même procédé trouver les équations pour une troncature réalisée non pas à l'ordre 1 mais à l'ordre 3 en écrivant :

$$q(t) = a_0 + a_1 \cos(\omega t) + b_1 \sin(\omega t) + a_2 \cos(\omega t) + b_2 \sin(\omega t) + a_3 \cos(\omega t) + b_3 \sin(\omega t) \quad (\text{A.23})$$

On obtient alors les expressions suivantes pour les trois composantes linéaire, constante et non-linéaire de la fonction  $H$  :

$$\bar{H} = \begin{bmatrix} \omega_0^2 & & & \\ & \Lambda_1 & & \\ & & \Lambda_2 & \\ & & & \Lambda_3 \end{bmatrix}, \quad \Lambda_n = \begin{bmatrix} \omega_0^2 - (n\omega)^2 & 2\xi\omega_0(n\omega) \\ -2\xi\omega_0(n\omega) & \omega_0^2 - (n\omega)^2 \end{bmatrix} \quad (\text{A.24})$$

$$H_c = \begin{Bmatrix} 0 \\ \Gamma \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \quad \tilde{x} = \begin{Bmatrix} a_0 \\ a_1 \\ b_1 \\ a_2 \\ b_2 \\ a_3 \\ b_3 \end{Bmatrix} \quad (\text{A.25})$$

$$\hat{H}(\tilde{x}) = \beta \frac{3}{2} \left\{ \begin{array}{l} a_0 b_3^2 + a_0 a_2^2 + a_0 a_3^2 + \frac{1}{2} a_1^2 a_2 + a_1 b_2 b_3 + \frac{2}{3} a_0^3 + a_0 b_2^2 + a_0 b_1^2 \dots \\ \dots - b_1 b_2 a_3 + a_0 a_1^2 + a_1 a_2 a_3 - \frac{1}{2} b_1^2 a_2 + b_1 a_2 b_3 + a_1 b_1 b_2 \\ \frac{1}{2} a_1^3 - \frac{1}{2} b_1^2 a_3 + a_1 b_2^2 + 2 a_0 b_2 b_3 - \frac{1}{2} b_2^2 a_3 + \frac{1}{2} a_1 b_1^2 + 2 a_0 a_2 a_3 \dots \\ \dots + a_1 a_3^2 + 2 a_0 a_1 a_2 + 2 a_0^2 a_1 + \frac{1}{2} a_2^2 a_3 + a_1 b_1 b_3 + a_1 a_2^2 \dots \\ \dots + 2 a_0 b_1 b_2 + a_1 b_3^2 + a_2 b_2 b_3 + \frac{1}{2} a_1^2 a_3 \\ a_2 b_2 a_3 - \frac{1}{2} a_2^2 b_3 + 2 a_0 a_2 b_3 + 2 a_0 a_1 b_2 + \frac{1}{2} a_1^2 b_1 + \frac{1}{2} a_1^2 b_3 \dots \\ \dots + b_1 b_2^2 + \frac{1}{2} b_2^2 b_3 - a_1 b_1 a_3 - 2 a_0 b_2 a_3 + b_1 a_3^2 \dots \\ \dots - \frac{1}{2} b_1^2 b_3 - 2 a_0 b_1 a_2 + b_1 a_2^2 + 2 a_0^2 b_1 + \frac{1}{2} b_1^3 + b_1 b_3^2 \\ b_1 b_2 a_3 - a_0 b_1^2 - b_1 a_2 b_3 + 2 a_0 a_1 a_3 + a_0 a_1^2 + b_1^2 a_2 \dots \\ \dots + a_1^2 a_2 + a_2 b_3^2 + 2 a_0^2 a_2 + a_1 a_2 a_3 + \frac{1}{2} a_2 b_2^2 \dots \\ \dots + a_1 b_2 b_3 + 2 a_0 b_1 b_3 + \frac{1}{2} a_2^3 + a_2 a_3^2 \\ \frac{1}{2} b_2^3 + 2 a_0 a_1 b_3 + 2 a_0^2 b_2 - a_1 b_2 a_3 + b_1 b_2 b_3 + b_2 a_3^2 \dots \\ \dots + b_1 a_2 a_3 + a_1^2 b_2 + \frac{1}{2} a_2^2 b_2 + b_2 b_3^2 \dots \\ \dots + 2 a_0 a_1 b_1 + a_1 a_2 b_3 + b_1^2 b_2 - 2 a_0 b_1 a_3 \\ \frac{1}{2} a_1 a_2^2 + 2 a_0^2 a_3 + \frac{1}{2} a_3^3 + a_1^2 a_3 + \frac{1}{2} a_3 b_3^2 - 2 a_0 b_1 b_2 + a_2^2 a_3 + \frac{1}{4} a_1^3 \dots \\ \dots + b_2^2 a_3 - \frac{1}{2} a_1 b_1^2 + b_1 a_2 b_2 + 2 a_0 a_1 a_2 + b_1^2 a_3 - \frac{1}{2} a_1 b_2^2 \\ b_1^2 b_3 + \frac{1}{2} b_3^3 + \frac{1}{2} a_1^2 b_1 + \frac{1}{2} a_3^2 b_3 - \frac{1}{4} b_1^3 - \frac{1}{2} b_1 a_2^2 + \frac{1}{2} b_1 b_2^2 \dots \\ \dots + 2 a_0 a_1 b_2 + a_1 a_2 b_2 + 2 a_0 b_1 a_2 + 2 a_0^2 b_3 + a_2^2 b_3 + a_1^2 b_3 + b_2^2 b_3 \end{array} \right\} \quad (\text{A.26})$$

### A.3 Fonction de réponse en fréquence et stabilité

Comme nous utiliserons cet exemple comme premier fil conducteur entre les différentes méthodes d'analyse globale, nous donnons ici la fonction de réponse en fréquence de l'oscillateur défini par les valeurs numériques suivantes :

$$\begin{aligned} m &= 1 \text{ (kg)}, \quad k = 1 \text{ (N.m}^{-1}\text{)}, \quad c = 2\xi\sqrt{km} \text{ (kg.s}^{-1}\text{)}, \quad \xi = 0.02 \\ k_{nl} &= 1 \text{ (N.m}^{-1}\text{)}, \quad f_e = 0.1, \quad \omega \in [0, 3] \text{ (rad.s}^{-1}\text{)} \end{aligned} \quad (\text{A.27})$$

La figure A.3 représente l'amplitude  $a = \sqrt{a_1^2 + b_1^2}$  du déplacement en fonction de la fréquence d'excitation tandis que la figure A.4 représente l'évolution des deux composantes harmoniques constituant le déplacement  $a_1$  et  $b_1$ . La première figure constitue donc une fonction de réponse en fréquence ; l'inclinaison du pic de résonance est significative de la présence d'une non-linéarité.

Cette figure indique qu'il existe pour certaines fréquences, plusieurs solutions qui co-existent. C'est le cas par exemple pour une excitation de pulsation  $\omega = 1.5 \text{ rad.s}^{-1}$ . On peut aussi la lire ainsi : si l'on applique une rampe montante à la pulsation d'excitation pour parcourir l'intervalle  $[0, 3] \text{ rad.s}^{-1}$ , le système suivra la courbe du point A au point  $M^+$ , puis il

« décrochera » en amplitude pour passer du point  $M^-$  avant d'atteindre le point  $B$  en suivant la courbe. Inversement, si l'on part du point  $B$  et que l'on applique à la pulsation d'excitation une rampe descendante, le système suivra la courbe de  $B$  jusque  $D^-$  d'où il « sautera » à  $D^+$  pour finalement rejoindre le point  $A$ . On voit ainsi premièrement qu'il y aura des sauts en amplitude des solutions stationnaires. Deuxièmement, on s'aperçoit qu'une portion de la courbe n'est jamais parcourue. En effet, les solutions situées entre  $M^+$  et  $D^-$  sont instables comme il est possible de le montrer par une analyse de Floquet.

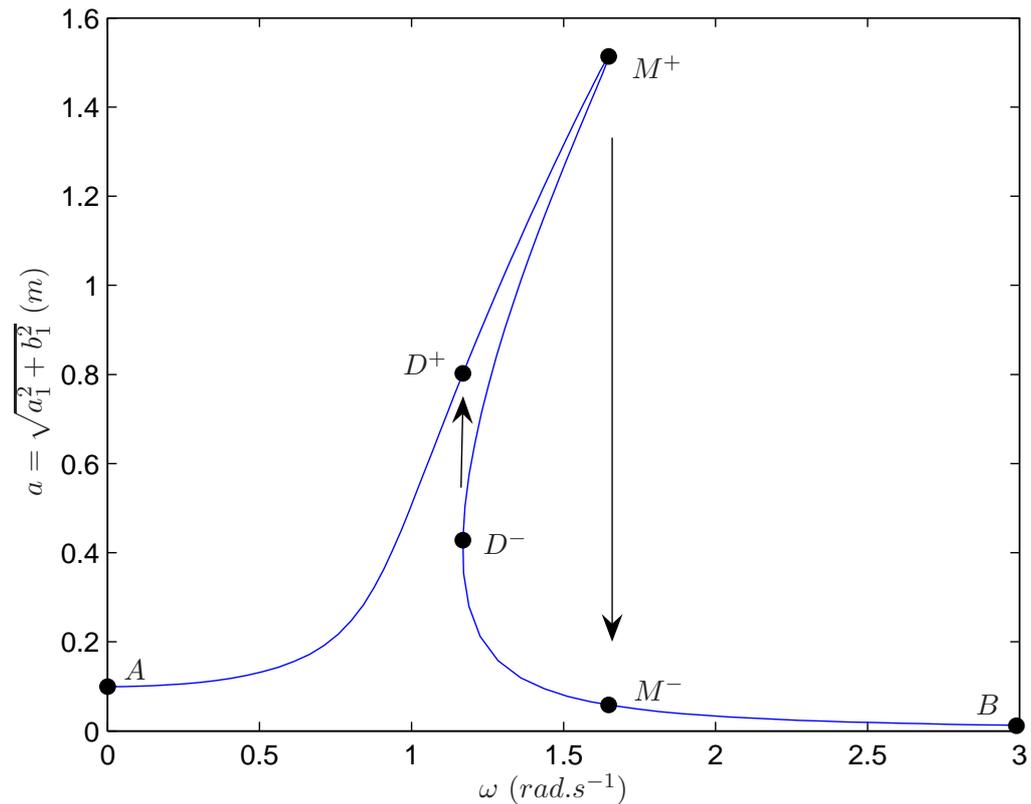


FIG. A.3 – Oscillateur de Duffing étudié - Fonction de réponse en fréquence

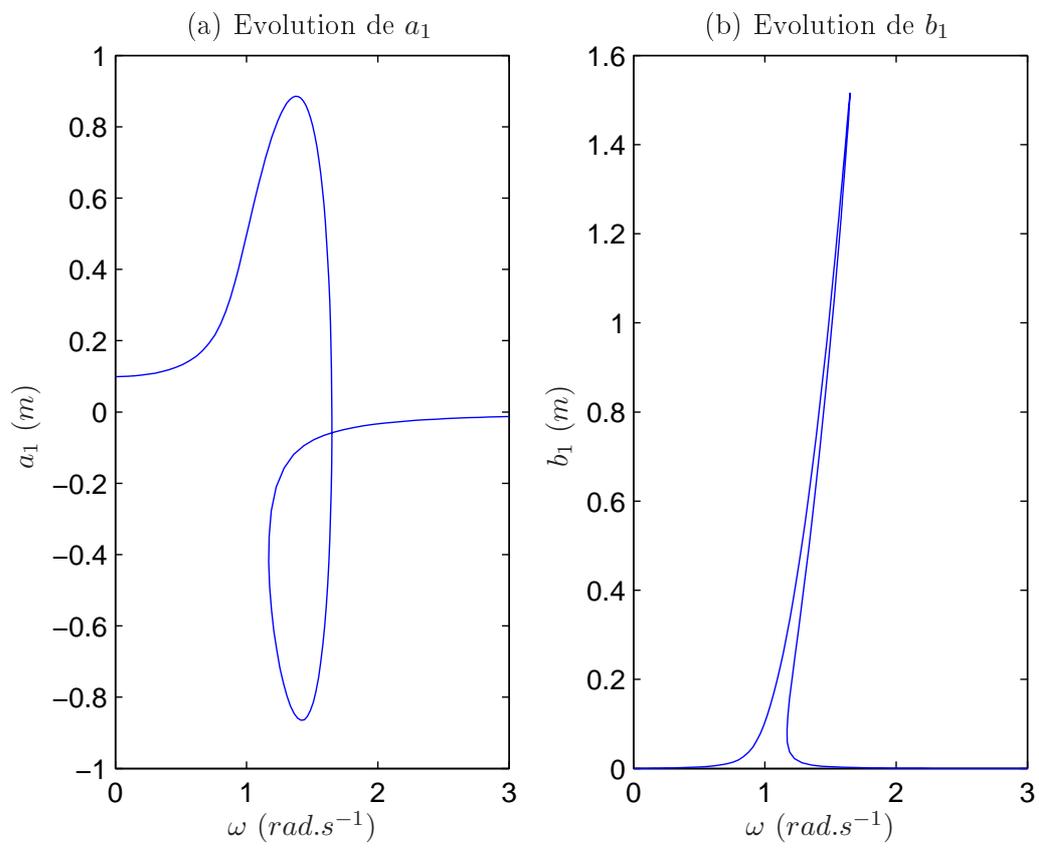


FIG. A.4 – Oscillateur de Duffing étudié - Evolution des composantes harmoniques

## Annexe B

# Arithmétique des intervalles

---

*Nous donnons dans cette annexe les définitions des opérations de base une fois transposées au cas des intervalles ainsi qu'une courte illustration des conséquences de la sous-distributivité. Enfin nous présentons un schéma complet de l'algorithme proposé par Hansen et Walster que nous n'avons présenté que sous forme de blocs résumés dans le chapitre 5.*

### B.1 Définition des opérations élémentaires pour les intervalles

Premièrement, nous décidons, conformément au choix des auteurs de [HW04] de nous placer dans  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ . Nous considérons les deux intervalles de  $\overline{\mathbb{R}}$ ,  $A = [\underline{a}, \overline{a}]$  et  $B = [\underline{b}, \overline{b}]$ . La transposition d'une loi  $\circ$  quelconque sur  $\mathbb{R}$  est prise - sauf mention contraire - au sens suivant :  $A \circ B = \{a \circ b \mid a \in A, b \in B\}$ . Nous détaillons ci-dessous les opérations élémentaires  $+$ ,  $-$ ,  $\times$  et  $/$  :

- Addition :

- $A + B = [\underline{a} + \underline{b}, \overline{a} + \overline{b}]$
- $(+\infty) + (-\infty) = \overline{\mathbb{R}}$

- Soustraction :

- $A - B = [\underline{a} - \overline{b}, \overline{a} - \underline{b}]$
- $(+\infty) - (+\infty) = \overline{\mathbb{R}}$
- $A \ominus B = [\underline{a} - \underline{b}, \overline{a} - \overline{b}]$  (soustraction étendue)

- Multiplication

- $A \times B = [\min(\underline{a} \times \underline{b}, \underline{a} \times \overline{b}, \overline{a} \times \underline{b}, \overline{a} \times \overline{b}), \max(\underline{a} \times \underline{b}, \underline{a} \times \overline{b}, \overline{a} \times \underline{b}, \overline{a} \times \overline{b})]$

- $A^n =$ 

	$\underline{a}, \bar{a} < 0$	$\underline{a} < 0, \bar{a} > 0$	$\underline{a}, \bar{a} > 0$
Si $n = 0$	$[1, 1]$	$[1, 1]$	$[1, 1]$
Si $n$ est pair	$[\bar{a}^n, \underline{a}^n]$	$[0, \max(\underline{a}^n, \bar{a}^n)]$	$[\underline{a}^n, \bar{a}^n]$
Si $n$ est impair	$[\underline{a}^n, \bar{a}^n]$	$[\underline{a}^n, \bar{a}^n]$	$[\underline{a}^n, \bar{a}^n]$
  - $\pm\infty \times 0 = \overline{\mathbb{R}}$
  - Division
    - $\frac{1}{B} = \begin{cases} \text{Si } 0 \notin B, [\min(\frac{1}{\bar{b}}, \frac{1}{\underline{b}}), \max(\frac{1}{\underline{b}}, \frac{1}{\bar{b}})] \\ \text{Si } 0 \in B, [-\infty, \frac{1}{\underline{b}}] \cup [\frac{1}{\bar{b}}, +\infty] \end{cases}$
    - $A/B = A \times \frac{1}{B}$
- Remarquons qu'alors, si  $0 \in A$  et  $0 \in B$ , on arrive au cas très défavorable  $A/B = \overline{\mathbb{R}}$ .

## B.2 Illustration des conséquences de la sous-distributivité

Nous avons conseillé en section 5.1 de factoriser un maximum les expressions utilisées pour le calcul de la fonction  $H$  en remarquant que pour  $A$ ,  $B$ , et  $C$  trois intervalles, on a  $A.(B+C) \subseteq A.B+A.C$ , c'est-à-dire inclusion et non égalité lorsque l'on distribue l'expression. D'une façon générale, il faut donc minimiser les occurrences d'une variable pour ne pas faire grossir inutilement l'intervalle image.

Imaginons par exemple que l'on veuille calculer l'image de  $x.y + x.z = x.(y + z)$  pour  $x \in [-1, 2]$ ,  $y \in [7, 8]$  et  $z \in [-3, 4]$ . L'expression de gauche renvoie l'intervalle  $[-12, 24]$  tandis que l'expression de droite renvoie  $[-14, 24]$ . Faire apparaître  $x$  deux fois, c'est-à-dire une de trop est donc préjudiciable. Cependant, il est bon de remarquer que ce n'est pas la forme factorisée qui est la plus efficace si l'on identifie une puissance supérieure ou égale à 2 : calculons pour les mêmes intervalles de variations  $x^2 + x.y = x.(x + y)$ . La forme de gauche, développée mais qui prend en compte le fait que  $x$  se multiplie avec lui-même donc dont l'image est l'intervalle  $\{x^2, x \in X\}$ , renvoie  $[-8, 20]$  tandis que la forme factorisée de droite renvoie  $[-16, 32]$ . En effet, lors du calcul du membre de droite, on envisage les cas où l'on multiplie entre eux deux éléments  $x_1$  et  $x_2$  distincts de  $X$ , alors qu'on ne l'envisage pas dans le premier cas.

### B.3 Schéma détaillé de l'algorithme proposé par Hansen et Walster

Le schéma complet de l'algorithme décrit en section 5.2 est donné en double page par les figures B.1 et B.2. Nous avons entouré les différents blocs mentionnés sur la figure 5.2 par des pointillés étiquetés :

- HC : Hull Consistency
- BC : Box Consistency
- NG : Méthode de Newton Généralisée
- BCp : Box Consistency Partiel

La question «  $\tilde{X}$  assez réduite comparée à  $\tilde{X}^{(i)}$  ? » revient à la question « Méthode efficace ? » de la figure 5.2 et reçoit une réponse positive si l'équation (5.1) est vérifiée.

La question «  $w(\tilde{X}) > (w_R + w_I)/2$  ? » située en amont de la méthode de Newton Généralisée revient à l'opposé de la question « Cellule suffisamment petite pour une approche locale ? » du schéma 5.2.

Notons enfin que la question en fin du bloc « BCp » qui compare la taille de la cellule après avoir subi les différents procédés de réduction avec celle qu'elle avait initialement lors de son extraction de *intList* permet de tester l'efficacité globale de l'algorithme.

Nous rappelons ci-dessous les définitions des symboles utilisés :

- Stockage de l'élément  $x$  dans l'élément  $y$  :  $y \leftarrow x$
- Utilisation de l'élément  $x$  pour le calcul de  $y$  :  $x \hookrightarrow y$
- Ajout de l'élément  $x$  à une liste  $xList$  :  $xList \stackrel{+}{\leftarrow} x$

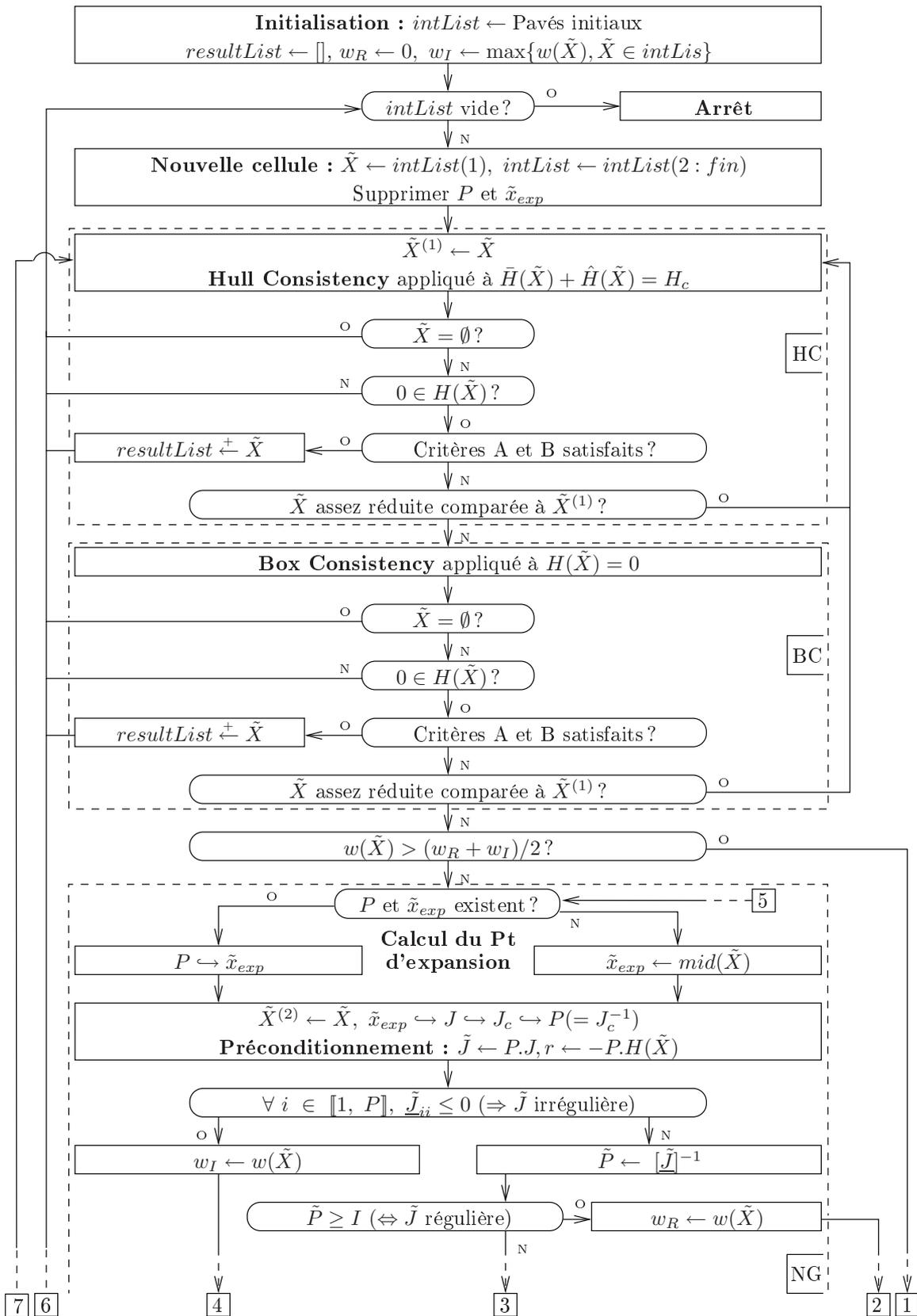


FIG. B.1 – Arithmétique des intervalles - Algorithme d'Hansen &amp; Walster détaillé

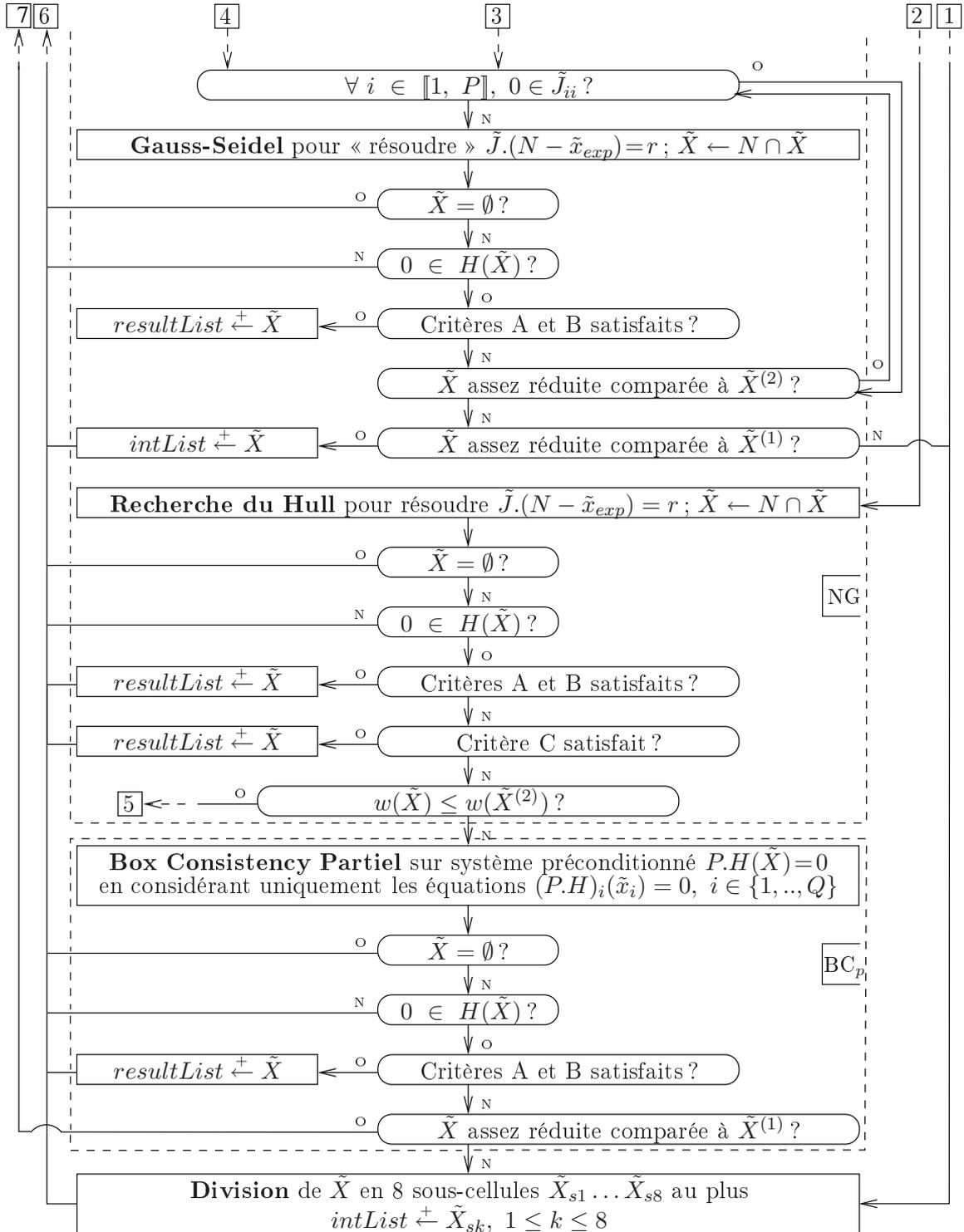


FIG. B.2 – Arithmétique des intervalles - Algorithme d'Hansen & Walster détaillé (suite)



## AUTORISATION DE SOUTENANCE

Vu les dispositions de l'arrêté du 7 août 2006,

Vu la demande du Directeur de Thèse

Monsieur F. THOUVEREZ

et les rapports de

Monsieur J-C. GOLINVAL  
Professeur - Université de Liège - LTAS - Chemin des Chevreuils, 1 - Bât. B52 - B-4000 LIEGE 1  
Belgique

et de

Monsieur B. COCHELIN  
Professeur - Ecole Centrale de Marseille - Plot 6 bureau 107 - Technopôle Château-Gombert  
38 rue Joliot-Curie - 13451 MARSEILLE CEDEX 20

**Mademoiselle SARROUY Emmanuelle**

est autorisé à soutenir une thèse pour l'obtention du grade de **DOCTEUR**

**Ecole doctorale MECANIQUE, ENERGETIQUE, GENIE CIVIL ET ACOUSTIQUE (MEGA)**

Fait à Ecully, le 17 octobre 2008

P/Le Directeur de l'E.C.L.  
La Directrice des Etudes

