

International Journal of

Sciences:

Basic and Applied

Research

STABILISATION ET APPROXIMATION NUMERIQUE D'UNE
CLASSE DE PROBLEMES INVERSES EN EDP

By *Ousmane Samba COULIBALY*

Volume 32, 2017
ISSN (Print & Online): 2307-4531

© IJSBAR THESIS PUBLICATION
www.gssrr.org

Published by:



Visit: www.gssrr.org

ISSN 2307-4531 (Print & Online)

IJSBAR research papers are currently indexed by:



STABILISATION ET APPROXIMATION NUMERIQUE D'UNE CLASSE
DE PROBLEMES INVERSES EN EDP

Copyright © 2017 by Ousmane Samba COULIBALY

All rights reserved. No part of this thesis may be produced or
transmitted in any form or by any
means without written permission of the author.
ISSN(online & Print) 2307-4531

Members of the Editorial Board

Editor in chief

Dr. Mohammad Othman Nassar, Faculty of Computer Science and Informatics, Amman Arab University for Graduate Studies,
Jordan, moanassar@aau.edu.jo , 00962788780593

Editorial Board

Prof. Dr. Felina Panas Espique, Dean at School of Teacher Education, Saint Louis University, Bonifacio St., Baguio City, Philippines.
Prof. Dr. Hye-Kyung Pang, Business Administration Department, Hallym University, Republic Of Korea.
Prof. Dr. Amer Abdulrahman Taqa, basic science Department, College of Dentistry, Mosul University, Iraq.
Prof. Dr. Abdul Haseeb Ansar, International Islamic University, Kuala Lumpur, Malaysia
Dr. kuldeep Narain Mathur, school of quantitative science, Universiti Utara, Malaysia
Dr. Zaira Wahab, Iqra University, Pakistan.
Dr. Daniela Roxana Andron, Lucian Blaga University of Sibiu, Romania.
Dr. Chandan Kumar Sarkar, IUBAT- International University of Business Agriculture and Technology, Bangladesh.
Dr. Azad Ali, Department of Zoology, B.N. College, Dhubri, India.
Dr. Narayan Ramappa Birasal, KLE Society's Gudleppa Hallikeri College Haveri (Permanently affiliated to Karnatak University Dharwad, Reaccredited by NAAC), India.
Dr. Rabindra Prasad Kayastha, Kathmandu University, Nepal.
Dr. Rasmeh Ali AlHuneiti, Brunel University, United Kingdom.
Dr. Florian Marcel Nuta, Faculty of Economics/Danubius University of Galati, Romania.
Dr. Suchismita Satapathy, School of Mechanical Engineering, KIIT University, India.
Dr. Juliana Ajdini, Department of Social Work and Social Policy, Faculty of Social Science, University of Tirana, Albania.
Dr. Arfan Yousaf, Department of Clinical Sciences, Faculty of Veterinary and Animal Sciences, PMAS-Arid Agriculture University Rawalpindi, Pakistan.
Dr. Rajamohan Natarajan, Chemical Engineering, Faculty of Engineering, Sohar university, Oman.
Dr. Tariq Javed, Lahore Pharmacy College (LMDC), University of Health Sciences, Lahore, Pakistan.
Dr. Rogers Andrew, Sokoine University of Agriculture, United Republic Of Tanzania
Dr Feras Fares, Amman Arab University for graduate studies, Jordan.



STABILISATION ET APPROXIMATION NUMERIQUE D'UNE CLASSE DE
PROBLEMES INVERSES EN EDP

Ousmane Samba COULIBALY

Faculté des Sciences et Techniques (FAST)
UNIVERSITE DES SCIENCES, DES TECHNIQUES ET DES TECHNOLOGIES
BAMAKO - MALI

THESE de Doctorat
Présentée pour l'obtention du diplôme de
DOCTORAT en MATHEMATIQUES
Option : Analyse Numérique

Année Universitaire 2016

Résumé

Dans ce travail, nous étudions un problème inverse du type identification de sources pour un problème parabolique modélisant la diffusion dans un rectangle.

En utilisant la décomposition spectrale de l'opérateur différentiel A , on peut expliciter la solution formelle du problème sous la forme d'une série de Fourier avec des coefficients non-bornés (hautes fréquences).

Nous proposons dans notre étude une méthode de régularisation basée sur la troncature spectrale, qui nous permet de construire une solution approchée et stable. Ensuite, la solution stabilisée sera projetée sur un sous espace de Krylov engendré par l'opérateur A en dimension finie. Cet algorithme nous fournit une méthode pratique et simple pour calculer numériquement la solution stabilisée.

Plus précisément, nous faisons une analyse théorique de l'approximation de $\Phi(A)g = (I - e^{-A})^{-1}Ag$ par la méthode de Krylov, et nous démontrons une estimation d'erreur entre la solution discrétisée et sa projection de Krylov.

Ensuite, nous donnons le bilan d'erreur entre la solution exacte et l'approximation de Krylov, pour tester l'efficacité de la méthode et pour permettre de définir un critère d'arrêt selon la taille du sous espace de projection de Krylov.

Mots Clés : Problème mal posé, Problème inverse parabolique, Identification de sources, Sous-espace de Krylov, Fonctions de matrices, Méthodes de projection de Krylov, Régularisation, Filtrage spectral.

**STABILISATION ET APPROXIMATION NUMERIQUE D'UNE CLASSE DE PROBLEMES
INVERSES EN EDP**



U.S.T.T-B

Ousmane Samba COULIBALY

Faculté des Sciences et Techniques (FAST)

UNIVERSITE DES SCIENCES, DES TECHNIQUES ET DES
TECHNOLOGIES DE BAMAKO - MALI

THESE de Doctorat
Présentée pour l'obtention du diplôme de
DOCTORAT en MATHEMATIQUES

Option : Analyse Numérique

Année Universitaire 2016

DÉDICACE

Je dédie ce travail :

- A la mémoire de mon père :

Feu Samba COULIBALY en signe de reconnaissance de l'immense bien qu'il a fait pour moi concernant mon éducation qui aboutit aujourd'hui à la réalisation de cette thèse. Ses conseils et ses bénédictions n'ont jamais fait défaut, que Dieu le tout puissant lui accorde son paradis éternel (amen).

- A ma mère

- A mes frères, sœurs et enfants

- A mon épouse

**A mes parents,
Une profonde marque de reconnaissance.**

REMERCIEMENTS

Les travaux de thèse présentés dans ce mémoire ont été réalisés au Laboratoire de Mathématiques Appliquées d'Annaba. Ils ont été rendus possibles grâce à l'accueil chaleureux et au soutien de madame le Professeur REBBANI Faouzia, et à l'assistance permanente, efficace et amicale de tout le personnel de ce Laboratoire.

J'adresse mes vifs remerciements à mes Directeurs de thèse, Professeur Nadjib BOUSSETILA et Professeur Gaoussou TRAORE. Je leur suis très reconnaissant de la confiance qu'ils m'ont toujours témoigné au cours de ce doctorat. Qu'ils trouvent ici toute ma reconnaissance pour leur suivi attentif, leur soutien, leurs conseils et leur disponibilité.

Je remercie Dr Amadou MAIGA et Dr Diby DIARRA, tous maîtres de conférences à la Faculté des Sciences et Techniques de Bamako (FST), pour leurs encouragements, leur soutien moral et scientifique pour mener ce travail à terme.

Je tiens également à remercier Dr Alain DEZETTER, Dr Jean Emmanuel PATUREL tous chargés de Recherche à l'Institut de Recherche pour le Développement (IRD) et le Professeur Ouaténi DIALLO de la FST qui m'ont initié à la recherche scientifique.

Je voudrais ensuite remercier monsieur le chef de Département d'Etudes et de Recherche (DER) de Mathématiques et Informatique de la FST ainsi qu'à tous les collègues.

Je souhaite exprimer mes remerciements au professeur Ouaténi DIALLO, qui m'a fait l'honneur de présider le jury, ainsi qu'au

- Professeur REBBANI Faouzia, Université Badji Mokhtar Annaba (Algérie),
- Docteur RAMOUL Hichem, Maître de conférences, Université Khenchela (Algérie),
- Professeur SAADI Samira, Université Badji Mokhtar (Algérie),
- Docteur Diby DIARRA, Maître de conférences, Université des Sciences, des Techniques et des Technologies de Bamako (Mali),

- Docteur Amadou MAIGA, Maître de conférences, Université des Sciences, des Techniques et des Technologies de Bamako (Mali),
- Docteur Fana TANGARA, Maître de conférences, Université des Sciences, des Techniques et des Technologies de Bamako (Mali),

pour avoir accepté de faire partie du jury et d'y avoir consacré une partie de leur temps. Je les remercie pour leurs remarques constructives.

Je ne saurais trouver les mots nécessaires pour exprimer ma profonde gratitude envers ma famille pour leur réconfort moral et affectif prodigué sans relâche.

Il convient de remercier la FST et l'USTTB (Université des Sciences, des Techniques et des Technologies de Bamako) pour leurs contributions diverses du début jusqu'à la fin de cette thèse.

En fin, merci à toutes les personnes qui ont contribué de près ou de loin à la réalisation de cette thèse.

Table des matières

1	Introduction	1
1.1	Problématique de la thèse	1
1.2	Contenu de la thèse	3
1.3	Note bibliographique	5
2	Résultats préliminaires et notations	7
2.1	Éléments de théorie spectrale	7
2.1.1	Opérateurs linéaires	7
2.1.2	Opérateurs bornés	8
2.1.3	Opérateurs non-bornés	9
2.1.4	Spectre et résolvante d'un opérateur non borné	11
2.2	Théorie de Riesz-Fredholm	12
2.2.1	Diagonalisation des opérateurs auto-adjoints compacts	12
2.2.2	Famille spectrale et résolution de l'identité	13
2.3	Equations opérationnelles et Alternative de Fredholm	15
2.3.1	Décomposition en valeurs singulières	16
2.4	Problèmes mal posés et problèmes inverses	17
2.4.1	Outils d'analyse des problèmes mal posés : cas linéaire	22
2.4.2	Convexité logarithmique	23
2.5	Méthodes de régularisation	24
2.5.1	La méthode de Tikhonov	25
3	Fonctions de matrices	28
3.1	Généralités	28
3.2	Différentes définitions d'une fonction de matrice	31
3.2.1	Définition par la forme Canonique de Jordan	31
3.2.2	Définition de $f(A)$ par l'interpolation polynomiale	37
3.2.3	Définition de $f(A)$ par l'intégrale de Cauchy	40
3.2.4	Propriétés élémentaires de $f(A)$	44
3.3	Quelques particularités sur les matrices	45
3.3.1	Décomposition en valeurs singulières d'une matrice	45

3.3.2	Matrice de Heisenberg	49
3.4	Evaluation numérique de $f(A)v$	50
3.4.1	Introduction	50
3.4.2	Méthodes de projection de Krylov	50
3.4.3	Principe de fonctionnement	54
3.5	Approximation de $f(A)v$ par la méthode des sous espaces de Krylov	55
3.5.1	Procédure d'Arnoldi : Algorithme	56
3.5.2	Approximation d'Arnoldi de $f(A)v$	59
3.5.3	Valeurs de Ritz	64
4	Etude d'un problème inverse d'identification de sources du type parabolique	65
4.1	Problème de Dirichlet	65
4.1.1	Position du problème	65
4.1.2	Stabilisation et approximation	67
4.1.3	Approximation numérique du problème par La méthode de Krylov	69
4.2	Problème de Neumann	70
4.2.1	Position du problème	71
4.2.2	Exemple d'application	73
4.3	Estimation d'erreur en dimension finie	76
4.3.1	Calcul de l'estimation d'erreur de $\ f - f_m\ $	78
5	Tests numériques	95
5.1	Discrétisation de l'opérateur $A = -\frac{d^2}{dx^2}$	95
5.2	Données exactes	100
Annexe		111
Annexe		112
.1	Programme MATLAB	112
Bibliographie		115

1

Introduction

1.1 Problématique de la thèse

La problématique abordée dans cette thèse s'inscrit dans le cadre des problèmes mal-posés en EDP au sens d'Hadamard.

En 1923, le mathématicien français J. Hadamard a écrit son livre célèbre sur les équations aux dérivées partielles et leur signification physique¹. Cet ouvrage fût le point de départ au développement du concept de problème bien posé en physique mathématique. Il s'agit d'un problème dont la solution existe, est unique et dépend continûment des données (stabilité). Bien entendu, ces notions doivent être précisées par le choix des espaces (et des topologies) dans lesquels les données et la solution sont considérées. Dans ce même livre Hadamard laissait entendre (et c'était aussi une opinion partagée avec I.G. Petrovsky²) que seul un problème bien posé pouvait modéliser correctement un phénomène physique.

La physique mathématique a longtemps ignoré les problèmes mal posés, les considérant, soit dénués de sens physique, soit reflétant une modélisation inadéquate. La réalité actuelle est toute autre : le caractère fondamentalement mal posé de certains

1. J. HADAMARD, *Lectures on Cauchy's problem in linear PDEs*, Yale University Press, New Haven (1923).
2. I.G. PTEROVSKY, *Lecture on Partial Differential Equations*, New York : International Publishers (1954).

problèmes pratiques est reconnu et motive de nombreuses recherches en mathématiques [65, 46, 44, 32, 94, 53, 55, 17, 10, 1, 11].

L'étude des phénomènes dans la nature nous permet de calculer des quantités ou des propriétés physiques d'un modèle donné. On distingue alors deux types de problèmes : les problèmes directs et les problèmes inverses.

- De manière schématique, un problème inverse peut être formulé comme étant une relation fonctionnelle (Input, Système, Output), où l'objectif d'étude est d'identifier des causes connaissant des effets. D'après J.B. Keller³, deux problèmes sont dits inverses l'un de l'autre si la formulation de l'un met l'autre en cause.

- La causalité et l'irréversibilité donnent une dichotomie entre les phénomènes physiques, qui peuvent être quantifiés mathématiquement en deux classes de problèmes : les problèmes bien posés et les problèmes mal posés. En se référant à cette dichotomie, le mot problèmes inverses désigne tous les problèmes qui partagent le caractère mal posé par opposition aux problèmes dits directs.

- La thématique des problèmes inverses et en particulier leurs méthodes de traitement numérique sont d'une grande importance pour les sciences appliquées et la technologie ; elles sont cruciales pour le développement des techniques de mesure et de diagnostique pour des systèmes complexes et partiellement accessibles.

- L'analyse mathématique des problèmes inverses et les problèmes mal-posés est un champ de recherche en plein essor. Même si de grandes avancées ont été effectuées ces dernières années, de nombreuses questions restent encore posées et méritent un effort sérieux pour aboutir à des réponses même partielles. Les questions qui nous intéressent dans cette thèse sont centrées autour de la stabilisation et la régularisation de certains problèmes inverses du type "identification de sources et de complétion de données" et leurs méthodes d'approximation numérique.

- Beaucoup de systèmes en physique, mécanique et biologie, ou en sciences médicales sont décrits par des EDP. Le contrôle des EDP concerne l'étude des variables importantes caractéristiques de ces systèmes, leur évolution dans le temps, et les moyens

3. J. B. KELLER, *Inverse problems*, Amer. Math. Monthly, **83** (1976), 107-118.

d'influencer leur avenir en les agitant à proximité de cibles souhaitées ou reconstruire leur passé, en tenant compte des contraintes et des coûts de contrôle.

Le contrôle des EDP est de plus en plus impliqué dans de nombreuses applications de la technologie. Cette demande en croissance rapide de notre société crée une forte dynamique d'innovation et de nouveaux problèmes mathématiques dans les aspects théoriques ainsi que les méthodes du calcul numériques, et nécessite des processus efficaces intégrés de contrôle pour applications.

Par ailleurs, les problèmes inverses pourraient être décrits comme des problèmes où la réponse est connue, mais pas la question ; ou lorsque les résultats, ou les conséquences sont connus, mais pas la cause. Comme simple exemple, quand on connaît la capacité thermique spécifique d'un liquide, on peut calculer combien de temps il faudra pour faire bouillir une quantité donnée de ce liquide. Supposons maintenant qu'on sait combien de temps prend un liquide à l'ébullition, peut-on savoir de quel liquide il s'agit ? Ils se présentent comme la recherche de l'entrée à partir de la sortie.

Donc la problématique de l'inversion apparaît dans de nombreux domaines tels que les domaines cités ci-dessus et surtout du traitement de données en général.

1.2 Contenu de la thèse

■ Dans cette thèse, notre effort sera consacré à l'aspect numérique d'une classe de problèmes inverses paraboliques. La stratégie numérique est basée sur la troncature (filtrage) spectrale et les méthodes de projection de Krylov. L'étude est complétée aussi par un bilan d'erreur et une série d'expérimentations numériques justifiant l'effet régularisant et la robustesse de la méthode du calcul utilisée dans notre investigation.

■ Quelques problèmes inverses en EDP se ramènent à des problèmes d'approximation matricielle (fonction matricielle $u = \Phi(A)v$) intervenant des matrices carrées de grande taille.

Parmi les méthodes de calcul de l'expression $u = \Phi(A)v$ il y a les méthodes de projection de Krylov, qui sont bien étudiées et expérimentées numériquement pour

certaines classes de fonctions (les fonction décroissantes, les fonctions de type Markov, les fonctions complètement monotones [23, 57, 25, 99, 9, 71, 86, 62, 43]).

Les fonctions intervenant dans les problèmes inverses sont des fonctions croissantes, donc engendrent plus de difficultés, et peu de résultats sont disponibles.

Comme nous traitons un problème mal posé, une procédure de régularisation est nécessaire pour neutraliser le comportement instable. Lorsqu'il s'agit d'identifier ou de calculer une grandeur physique à partir des observation (mesures), on est amené souvent à inverser un opérateur (la résolvante qui donne la solution du problème direct). Cette inversion, qui est souvent mal posée, nécessite un traitement particulier des instabilités, par des techniques dites de régularisation qui consistent à perturber légèrement le problème, ou d'éliminer les hautes fréquences responsables de cette instabilité, de manière à rendre le problème en question bien posé et numériquement réalisable.

Ce Travail s'inscrit dans le cadre des problèmes inverses en EDP, dont les techniques mathématiques utilisées relèvent de l'analyse fonctionnelle et de la théorie spectrale. Il est composé de quatre chapitres, un chapitre rappel et trois chapitres essentiels, une conclusion et quelques annexes.

- Dans le Chapitre 1, nous rappelons quelques résultats d'analyse fonctionnelle, ainsi que les outils d'analyse des problèmes mal posés. Un aperçu non exhaustif est donné pour faciliter la lecture et comprendre les questions liées à cette thématique à travers des exemples académiques.
- Le Chapitre 2 est consacré à l'évaluation numérique de $\Phi(A)v$ (l'action de $\Phi(A)$ sur v) conduisant à l'étude du système de la forme : $\Psi(A)u = v \Leftrightarrow u = \Phi(A)v$, où u est la solution recherchée, v les données du problème, A une matrice symétrique positive (resp. définie positive) de grande taille et Φ une fonction suffisamment régulière (ou analytique) sur le spectre $\Lambda(A)$ de A . Nous rappelons quelques

définitions d'une fonction de matrice et certaines propriétés élémentaires tout en mettant l'accent sur quelques matrices particulières.

- Le troisième Chapitre est dédié à l'application de la méthode de Krylov à un cas concret : l'équation de la chaleur. L'étude est achevée par une démonstration théorique de l'estimation d'erreur entre la solution exacte et la solution de Krylov.
- le chapitre 4 porte sur la discrétisation et la mise en œuvre de quelques exemples académiques. Nous proposons une résolution analytique du problème inverse en question par la méthode de Fourier. Ce qui nous permet d'avoir une idée claire sur l'instabilité du problème et son degré de complexité. Pour tester l'efficacité de la méthode et son effet régularisant, nous l'appliquons au problème de la chaleur avec conditions aux limites de Neumann.

1.3 Note bibliographique

□ Dans la littérature mathématique, on remarque que ces dernières années, un nombre de travaux a été consacré à la méthode de projection de Krylov. Cet effort scientifique est un témoignage sur la popularité et l'efficacité de ces outils mathématiques dans le traitement numérique de l'analyse spectrale de grande échelle.

Notre choix de cette stratégie numérique est basé sur deux motivations :

- Le fondement théorique solide développé par des grands spécialistes reconnus dans le domaine du développement de logiciels industriels. Pour plus de détails, nous renvoyons le lecteur à [62, 43] pour les aspects théoriques, et pour le côté informatique, on peut consulter (<http://www.guettel.com>).

- L'effet régularisant constaté dans le traitement numérique de certains problèmes inverses très complexes, ainsi que le type de convergence qui est super-linéaire [66, 80, 81, 54, 72, 34, 50].

□ Pour les problèmes inverses d'identification de sources, beaucoup d'auteurs se sont intéressés pour cette catégorie de problèmes, où l'effort était comment neutraliser le caractère d'instabilité. On peut citer ici quelques références résumant les différentes approches, comme par exemple : la méthode de Tikhonov, la méthode de troncature

spectrale, la méthode variationnelle de l'état adjoint et d'autres méthodes numérique [41, 79, 26, 27, 5, 7, 94, 84, 85, 95, 12, 13, 14, 58, 59, 60, 97].

□ On note ici que l'étude des problèmes inverses en EDP du type identification de sources/données traités par la méthode de Krylov, se distinguent par la rareté des résultats.

La présente thèse est une continuation dans la même direction développée par les auteurs [49, 75, 2].

2

Résultats préliminaires et notations

L'objectif de ce chapitre est de rappeler quelques notions et résultats qui seront utilisés tout au long de ce travail. Pour plus de détails, des références à la littérature seront systématiquement données.

On se place dans un un cadre hilbertien $(H_1 \longrightarrow H_2)$, où H_i est un espace de Hilbert sur $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} , muni de la norme $|\cdot|_i$ et le produit scalaire $(\cdot, \cdot)_i$, ($i = 1, 2$).

2.1 Éléments de théorie spectrale

2.1.1 Opérateurs linéaires

De manière générale, un opérateur linéaire est une application $A : \mathcal{D}(A) \subseteq H_1 \longrightarrow H_2$ linéaire, où $\mathcal{D}(A)$ est le domaine de définition de l'application linéaire A , qui est un sous-espace vectoriel de H_1 , que l'on suppose en général dense dans H_1 . L'opérateur $A : \mathcal{D}(A) \subseteq H_1 \longrightarrow H_2$ est dit borné si la quantité

$$\|A\| = \sup \{ |Au|_{H_2}, u \in \mathcal{D}(A), |u|_{H_1} = 1 \}$$

est finie. Dans ce cas A est une application linéaire continue sur $\mathcal{D}(A)$, et lorsque $\mathcal{D}(A)$ est dense dans H_1 , A s'étend de manière unique à un opérateur borné sur H_1 .

- Tout opérateur A est complètement défini par son graphe $\mathbf{G}(A)$ qui est un sous-espace vectoriel de $H_1 \times H_2$ défini par $\mathbf{G}(A) = \{ (v, Av), v \in \mathcal{D}(A) \}$.

Pour tout opérateur linéaire $A : \mathcal{D}(A) \subseteq H_1 \longrightarrow H_2$, on note par :

$$\mathbf{N}(A) = \{h \in \mathcal{D}(A), Ah = 0\} \text{ (noyau de } A),$$

$$\mathbf{R}(A) = \{h_2 = Ah_1, h_1 \in \mathcal{D}(A)\} \text{ (image de } A).$$

2.1.2 Opérateurs bornés

On note $\mathcal{L}(H_1, H_2)$ (resp. $\mathcal{L}(H_1)$) l'espace vectoriel des *opérateurs linéaires continus* de H_1 dans H_2 (resp. des *endomorphismes continus* de H_1) muni de la topologie de la convergence uniforme :

$$B \in \mathcal{L}(H_1, H_2), \quad \|B\|_{\mathcal{L}(H_1, H_2)} = \sup_{u \in H_1 \setminus \{0\}} \frac{|Bu|_2}{|u|_1}.$$

Définition 2.1.1. On dit qu'une application linéaire continue $S \in \mathcal{L}(H_1, H_2)$ est *inversible* ssi il existe une application $S' \in \mathcal{L}(H_2, H_1)$ telle que

$$S' \circ S = I_{H_1} \quad \text{et} \quad S \circ S' = I_{H_2}.$$

L'application S' si elle existe est unique. On notera $S' = S^{-1}$ et

$$\mathbf{Inv}(H_1, H_2) := \{S \in \mathcal{L}(H_1, H_2), S \text{ inversible}\}.$$

Théorème 2.1.1. [Théorème des isomorphismes de Banach]

Toute bijection linéaire continue $S \in \mathcal{L}(H_1, H_2)$ est inversible.

Définition 2.1.2. Soit $A \in \mathcal{L}(H)$. On appelle *ensemble résolvant* de A , l'ensemble

$$\rho(A) := \left\{ \lambda \in \mathbb{C}; A_\lambda = (\lambda I - A) \text{ est inversible } (\iff \text{bijectif}) \right\}.$$

Son complémentaire dans le plan complexe s'appelle le *spectre* de A et sera noté $\sigma(A) := \mathbb{C} \setminus \rho(A)$.

On appelle *rayon spectral* (noté $\text{spr}(A)$) la borne supérieure du spectre en module, i.e.,

$$\text{spr}(A) := \sup_{\lambda \in \sigma(A)} |\lambda|.$$

► Le spectre d'un opérateur borné est un compact non vide.

Le spectre ponctuel de A (noté $\sigma_p(A)$) est l'ensemble des $\lambda \in \mathbb{C}$ tels que A_λ soit non injectif :

$$\lambda \in \sigma_p(A) \iff \mathbf{N}(A_\lambda) \neq \{0\}.$$

Un élément $\lambda \in \sigma_p(A)$ est dit valeur propre de A , s'il lui correspond un $0 \neq h \in H$ tel que $Ah = \lambda h$ que l'on appelle vecteur propre correspondant à λ .

Définition 2.1.3. (et proposition) Soit $S \in \mathcal{L}(H_1, H_2)$. Alors il existe un unique opérateur $S^* \in \mathcal{L}(H_2, H_1)$, appelé adjoint de S , qui vérifie la relation suivante :

$$(Sh_1, h_2)_2 = (h_1, S^*h_2)_1, \quad \forall h_1 \in H_1, \quad \forall h_2 \in H_2.$$

De plus, on a les propriétés suivantes :

$$\|S\| = \|S^*\|, \quad S^{**} = (S^*)^* = S.$$

Si S est bijectif (i.e. inversible), alors S^* l'est aussi, et $(S^*)^{-1} = (S^{-1})^*$.

Définition 2.1.4. Soit H un espace de Hilbert. On dit que $A \in \mathcal{L}(H)$ est *auto-adjoint* si $A = A^*$.

$$A = A^* \iff (Ax, y) = (x, Ay), \quad \forall x, y \in H.$$

2.1.3 Opérateurs non-bornés

Définition 2.1.5. On dit qu'un opérateur A est fermé si son graphe $G(A)$ est fermé dans $H_1 \times H_2$, i.e., pour toute suite $(u_n) \subset \mathcal{D}(A)$ telle que $u_n \rightarrow u$ dans H_1 et $Au_n \rightarrow v$ dans H_2 , alors $u \in \mathcal{D}(A)$ et $v = Au$.

► L'opérateur fermé A peut être considéré comme un opérateur borné de son domaine de définition $\mathcal{D}(A)$ muni de la norme du graphe ($\|u\|_G := \|u\|_{H_1} + \|Au\|_{H_2}$) dans H_1 .

Théorème 2.1.2. [*Théorème du graphe fermé*] Si l'opérateur fermé A est défini sur tout l'espace H_1 , alors A est borné

$$(A \text{ fermé et } \mathcal{D}(A) = H_1 \implies A \text{ borné}).$$

Définition 2.1.6. (et proposition) Soit $A : \mathcal{D}(A) \subset H_1 \longrightarrow H_2$ un opérateur non-borné à domaine dense. On peut définir l'opérateur non-borné A^* adjoint de l'opérateur A , comme suit :

$$A^* : \mathcal{D}(A^*) \subset H_2 \longrightarrow H_1$$

$$\mathcal{D}(A^*) = \left\{ v \in H_2 : \exists c > 0 \text{ tel que } |\langle v, Au \rangle| \leq c|u|_{H_1}, \quad \forall u \in \mathcal{D}(A) \right\}.$$

Dans ce cas la fonctionnelle $u \mapsto g(u) = \langle v, Au \rangle$ se prolonge de façon unique en une fonctionnelle linéaire $f : H_1 \longrightarrow \mathbb{K}$ telle que $|f(u)| \leq c|u|_{H_1}, \quad \forall u \in H_1$. Par suite $f \in H_1' \simeq H_1$. On a par conséquent la relation fondamentale qui lie A et A^*

$$\langle v, Au \rangle_{H_2} = \langle A^*v, u \rangle_{H_1}, \quad \forall u \in \mathcal{D}(A), \forall v \in \mathcal{D}(A^*).$$

► Si $A : \mathcal{D}(A) \subset H_1 \longrightarrow H_2$ est un opérateur non-borné à domaine dense, alors A^* est fermé.

Définition 2.1.7. On dit qu'un opérateur $A : \mathcal{D}(A) \subset H \longrightarrow H$ est symétrique lorsque

$$\forall (u, v) \in \mathcal{D}(A) \times \mathcal{D}(A), \quad (Au, v) = (u, Av)$$

Définition 2.1.8. L'opérateur $A : \mathcal{D}(A) \subset H \longrightarrow H$ est dit auto-adjoint si $A = A^*$, i.e.,

$$\mathcal{D}(A) = \mathcal{D}(A^*) \text{ et } (v, Au) = (Av, u), \quad \forall (u, v) \in \mathcal{D}(A) \times \mathcal{D}(A).$$

Théorème 2.1.3. Soit $A : \mathcal{D}(A) \subset H \longrightarrow H$ un opérateur fermé symétrique. A est auto-adjoint si et seulement si $\sigma(A) \subseteq \mathbb{R}$.

Théorème 2.1.4. [Caractérisation des opérateurs à image fermée]

Soit $A : \mathcal{D}(A) \subset H_1 \longrightarrow H_2$ un opérateur non-borné, fermé, avec $\overline{\mathcal{D}(A)} = H_1$.

Les propriétés suivantes sont équivalentes :

(i) $\mathbf{R}(A)$ est fermé, (ii) $\mathbf{R}(A^*)$ est fermé, (iii) $\mathbf{R}(A) = \mathbf{N}(A^*)^\perp$, (iv) $\mathbf{R}(A^*) = \mathbf{N}(A)^\perp$.

Le résultat qui suit est une caractérisation utile des opérateurs surjectifs.

Théorème 2.1.5. Soit $A : \mathcal{D}(A) \subset H_1 \longrightarrow H_2$ un opérateur non-borné, fermé, avec $\overline{\mathcal{D}(A)} = H_1$.

Les propriétés suivantes sont équivalentes :

- (a) A est surjectif, i.e., $\mathbf{R}(A) = H_2$,
 (b) il existe une constante $k > 0$ telle que

$$|v| \leq k|A^*v|, \quad \forall v \in \mathcal{D}(A^*),$$

- (c) $\mathbf{N}(A^*) = \{0\}$ et $\mathbf{R}(A^*)$ est fermé.

Corollaire 2.1.1. Soit $A : \mathcal{D}(A) \subset H_1 \longrightarrow H_2$ un opérateur non-borné, fermé, avec $\overline{\mathcal{D}(A)} = H_2$. L'opérateur A admet un inverse borné A^{-1} sur H_2 si et seulement s'il existe deux constantes m_1 et m_2 telles que

$$|u| \leq m_1 |Au|, \quad \forall u \in \mathcal{D}(A),$$

$$|v| \leq m_2 |A^*v|, \quad \forall v \in \mathcal{D}(A^*).$$

2.1.4 Spectre et résolvante d'un opérateur non borné

Soit $A : \mathcal{D}(A) \subset H \longrightarrow H$ un opérateur non borné que l'on suppose fermé^{1 2 3} et à domaine dense.

Définition 2.1.9. On appelle *ensemble résolvant* de A , l'ensemble

$$\rho(A) = \{\lambda \in \mathbb{C} : A_\lambda = \lambda I - A \text{ est bijectif}\}.$$

Son complémentaire dans le plan complexe s'appelle le *spectre* de A et sera noté $\sigma(A) = \mathbb{C} \setminus \rho(A)$.

- On note que si $\lambda \in \rho(A)$, l'inverse $R(\lambda; A) = A_\lambda^{-1}$ est défini sur tout l'espace et est fermé. Par le théorème du graphe fermé, il est borné, i.e., $A_\lambda^{-1} \in \mathcal{L}(H)$. Cet opérateur est appelé la *résolvante* de A .

- L'ensemble résolvant $\rho(A)$ est un ouvert du plan complexe et l'application $\rho(A) \ni \lambda \longmapsto R(\lambda; A)$ est analytique sur chaque composante connexe de $\rho(A)$. La résolvante satisfait à l'équation fonctionnelle suivante dite *identité de la résolvante* :

$$R(\lambda_1; A) - R(\lambda_2; A) = (\lambda_2 - \lambda_1)R(\lambda_1; A)R(\lambda_2; A).$$

1. L'hypothèse de fermeture est nécessaire pour faire une théorie spectrale raisonnable.
 2. Si A n'est pas fermé, alors $\rho(A) = \emptyset$.
 3. Si $A = A^*$, alors $\sigma(A) \neq \emptyset$ et $\sigma(A) \subseteq \mathbb{R}$.

• Le spectre de A est donc un fermé de \mathbb{C} , et si de plus l'opérateur A est borné, alors $\sigma(A)$ est un compact non vide.

Examinons à présent de plus près la structure du spectre.

• Le premier sous-ensemble important du spectre est le *spectre ponctuel* :

$$\sigma_p(A) = \{\lambda \in \mathbb{C} : A_\lambda \text{ n'est pas injectif}\}.$$

Un élément λ de $\sigma_p(A)$ est dit *valeur propre* de A , il lui correspond un $0 \neq \vartheta \in \mathcal{D}(A)$ tel que $A_\lambda \vartheta = 0$, que l'on appelle *vecteur propre* (fonction propre quand H est un espace de fonctions) correspondant à λ .

- Si $\lambda \in \sigma(A) \setminus \sigma_p(A)$ donc A_λ est injectif mais non surjectif. Deux cas se présentent :
 - Si $\mathbf{R}(A_\lambda)$ n'est pas dense, on dit alors que $\lambda \in \sigma_r(A)$ le spectre *résiduel* de A .
 - Si $\mathbf{R}(A_\lambda)$ est dense, on dit alors que $\lambda \in \sigma_c(A)$ le spectre *continu* de A .

2.2 Théorie de Riesz-Fredholm

2.2.1 Diagonalisation des opérateurs auto-adjoints compacts

Définition 2.2.1. On dit qu'un opérateur $K \in \mathcal{L}(H_1, H_2)$ est **compact** si $K(B_{H_1}(0, 1))$ est relativement compact pour la topologie forte. On désigne par $\mathcal{K}(H_1, H_2)$ l'ensemble des opérateurs compacts de H_1 dans H_2 et on pose $\mathcal{K}(H_1, H_1) = \mathcal{K}(H_1)$.

► La compacité d'un opérateur $T \in \mathcal{L}(H_1, H_2)$ est caractérisée comme suit :

$$T \in \mathcal{K}(H_1, H_2) \iff \forall (x_n) \subset H_1, x_n \rightharpoonup 0 \text{ (faiblement)} \implies Tx_n \longrightarrow 0 \text{ (fortement)}.$$

► Soient E, F et G trois espaces de Banach. Si $S_1 \in \mathcal{L}(E, F)$ et $S_2 \in \mathcal{K}(F, G)$ (resp. $S_1 \in \mathcal{K}(E, F)$ et $S_2 \in \mathcal{L}(F, G)$), alors $S_2 S_1 \in \mathcal{K}(E, G)$.

► [Théorème de Shauder] Si K est compact, alors K^* est compact. Et réciproquement.

Théorème 2.2.1. Soit $K \in \mathcal{K}(H)$ avec $\dim(H) = \infty$. Alors on a :

(a) $0 \in \sigma(K)$,

(b) $\sigma(K) \setminus \{0\} = \sigma_p(K) \setminus \{0\}$,

(c) l'une des situations suivantes :

- ou bien $\sigma(K) = \{0\}$,
- ou bien $\sigma(K) \setminus \{0\}$ est fini,
- ou bien $\sigma(K) \setminus \{0\}$ est une suite qui tend vers 0.

Théorème 2.2.2. *On suppose que H est séparable. Soit $T \in \mathcal{K}(H)$ un opérateur auto-adjoint compact. Alors H admet une base Hilbertienne formée de vecteurs propres de T :*

$$\forall x \in H, \quad x = x_0 + \sum_{k \geq 1} (x, e_k) e_k, \quad x_0 \in \mathbf{N}(T), \quad Tx = \sum_{k \geq 1} \lambda_k e_k.$$

2.2.2 Famille spectrale et résolution de l'identité

• Version discrète

Définition 2.2.2. Soit $A : \mathcal{D}(A) \subset H \rightarrow H$ un opérateur non borné. Alors A est dit à *résolvante compacte* si

$$\forall \lambda \in \rho(A), \quad R(\lambda; A) \in \mathcal{K}(H).$$

On a le résultat suivant :

Théorème 2.2.3. *Un opérateur $A : \mathcal{D}(A) \subset H \rightarrow H$ est à résolvante compacte ssi il existe $\mu \in \rho(A)$ tel que $R(\mu; A) \in \mathcal{K}(H)$.*

Théorème 2.2.4. *Soit $A : \mathcal{D}(A) \subset H \rightarrow H$ un opérateur auto-adjoint. Alors*

- (1) $\sigma_r(A) = \emptyset$,
- (2) $\sigma(A) = \sigma_p(A) \cup \sigma_c(A) \subseteq \mathbb{R}$,
- (3) $A \geq \theta \iff \sigma(A) \subset [\theta, \infty[$.

Théorème 2.2.5. *Soit $A : \mathcal{D}(A) \subset H \rightarrow H$ un opérateur auto-adjoint borné inférieurement et à résolvante compacte. Alors A est diagonalisable, i.e., il existe une base hilbertienne dans H , $(e_m)_{m \geq 1} \subset \mathcal{D}(A)$, et une suite de réels $(\lambda_m)_{m \geq 1}$ telles que*

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m \rightarrow +\infty, \quad Ae_m = \lambda_m e_m, \quad m = 1, 2, \dots$$

Remarque 2.2.1. Si $A : \mathcal{D}(A) \subset H \longrightarrow H$ est un opérateur auto-adjoint avec $A \geq \theta > 0 \implies 0 \in \rho(A)$, et l'injection $H^1 := (D(A), |\cdot|_G) \hookrightarrow H$ est compacte, alors A est à résolvante compacte et donc diagonalisable.

- **Version continue**

Définition 2.2.3. Une famille $\{E_\lambda\}_{\lambda \in \mathbb{R}}$ de projections orthogonales dans H est appelée *famille spectrale* ou encore *résolution de l'identité* si elle satisfait aux conditions :

$$\left\{ \begin{array}{l} (i) \quad E_\lambda E_\mu = E_{\inf(\lambda, \mu)}, \quad \lambda, \mu \in \mathbb{R}, \\ (ii) \quad E_{-\infty} = 0, \quad E_{+\infty} = I, \\ \text{où } E_{-\infty} h = \lim_{\lambda \rightarrow -\infty} E_\lambda h, \quad E_{+\infty} h = \lim_{\lambda \rightarrow +\infty} E_\lambda h, \quad h \in H, \\ (iii) \quad E_{\lambda+0} = E_\lambda \text{ où } E_{\lambda+0} h = \lim_{\varepsilon > 0, \varepsilon \rightarrow 0} E_{\lambda+\varepsilon} h, \quad h \in H. \end{array} \right.$$

Les limites sont prises au sens de la norme de H .

Théorème 2.2.6. Soient H un espace de Hilbert et A un opérateur auto-adjoint dans H . Alors il existe une famille spectrale $\{E_\lambda\}_{\lambda \in \mathbb{R}}$ telle que

$$(Ax, y) = \int_{\mathbb{R}} \lambda d(E_\lambda x, y), \quad Ax = \int_{\mathbb{R}} \lambda dE_\lambda x.$$

On note symboliquement $A = \int_{\mathbb{R}} \lambda dE_\lambda$.

Théorème 2.2.7. Soit $\lambda \mapsto f(\lambda)$ une fonction continue à valeurs réelles. Soit $\mathcal{D} \subset H$ définie par :

$$\mathcal{D} = \left\{ h \in H : \int_{\mathbb{R}} |f(\lambda)|^2 d|E_\lambda h|^2 < \infty \right\}.$$

Alors \mathcal{D} est dense dans H et on définit un opérateur auto-adjoint S dans H par :

$$(Sx, y) = \int_{\mathbb{R}} f(\lambda) d(E_\lambda x, y), \quad x \in \mathcal{D}, y \in H,$$

de domaine $\mathcal{D}(S) = \mathcal{D}$.

Fonctions d'un opérateur auto-adjoint

Soit A un opérateur auto-adjoint dans l'espace de Hilbert H , $A = \int_{\lambda_0}^{+\infty} \lambda dE_\lambda$,

$\lambda_0 = \inf \sigma(A) > 0$, sa décomposition spectrale.

Définition 2.2.4. On définit :

- Les puissances de A .

$$A^r = \int_{\lambda_0}^{+\infty} \lambda^r dE_\lambda, \quad r \in \mathbb{R}, \quad h \in \mathcal{D}(A^r) \iff \int_{\lambda_0}^{+\infty} \lambda^{2r} d|E_\lambda h|^2 < \infty.$$

On note ici, que pour tout $r \leq 0$, $A^r \in \mathcal{L}(H)$, et si $r = 0$, $A^0 = I$.

Pour tout $r \geq 0$ et pour tout $h \in \mathcal{D}(A^r)$, on a $(A^r h, h) \geq \lambda_0^r |h|^2$.

Pour tout $r \geq 0$, $\mathcal{D}(A^r)$ muni de la norme $|h|_r^2 = |A^r h|^2$, $h \in \mathcal{D}(A^r)$, est un espace de Hilbert.

Si $0 \leq r_1 \leq r_2$, $\mathcal{D}(A^{r_2}) \hookrightarrow \mathcal{D}(A^{r_1})$ et $\mathcal{D}(A^{r_2})$ est dense dans $\mathcal{D}(A^{r_1})$.

- $f(A)$ pour une fonction f continue sur \mathbb{R} .

$$f(A) = \int_{\lambda_0}^{+\infty} f(\lambda) dE_\lambda, \quad h \in \mathcal{D}(f(A)) \iff \int_{\lambda_0}^{+\infty} |f(\lambda)|^2 d|E_\lambda h|^2 < \infty.$$

2.3 Equations opérationnelles et Alternative de Fredholm

On suppose que H est séparable. Soit $T \in \mathcal{K}(H)$ un opérateur auto-adjoint compact donnée par sa décomposition spectrale :

$$\forall h \in H, \quad h = h_0 + \sum_{k \geq 1} (h, e_k) e_k, \quad h_0 \in \mathbf{N}(A), \quad Th = \sum_{k \geq 1} \lambda_k e_k.$$

Considérons l'équation

$$(T - \lambda I)f = g, \tag{2.3.1}$$

où $f = f_0 + \sum_{n=1}^{\infty} f_n e_n$, $g = g_0 + \sum_{n=1}^{\infty} g_n e_n$ sont deux vecteurs de H donnés.

- Si $\lambda \notin \sigma(A)$, la solution de l'équation (2.3.1) est donnée par :

$$f = \sum_{n=1}^{\infty} \left(\frac{g_n}{\lambda_n - \lambda} \right) e_n - \frac{g_0}{\lambda}.$$

- Si $\lambda = \lambda_s \neq 0$, l'équation (2.3.1) n'a de solution que si $g_s = 0$ ($\iff g \in \ker(T - \lambda_s I)^\perp$) et dans ce cas les solutions sont données par :

$$f = \left[\sum_{n \neq s}^{\infty} \left(\frac{g_n}{\lambda_n - \lambda} \right) e_n - \frac{g_0}{\lambda_s} \right] + G_s,$$

où G_s est un élément arbitraire de $\ker(T - \lambda_s I)$.

- Si $\lambda = 0$, pour que l'équation $Tf = g$ ait une solution il faut et il suffit que

$$f_0 = 0 \iff f \in \mathbf{N}(T)^\perp = \overline{\mathbf{R}(T)}$$

et que la série

$$\sum_{n=1}^{\infty} \frac{|g_n|^2}{\lambda_n^2}$$

soit convergente. Dans ce cas les solutions sont données par :

$$f = \left[\sum_{n=1}^{\infty} \left(\frac{g_n}{\lambda_n - \lambda} \right) e_n \right] + G_0,$$

où G_0 est un élément arbitraire de $\ker(T)$.

2.3.1 Décomposition en valeurs singulières

Considérons maintenant un opérateur compact $T \in \mathcal{K}(H_1, H_2)$, où H_1, H_2 sont deux espaces de Hilbert séparables. L'une des approches les plus pratiques pour étudier le problème inverse $Th_1 = h_2$, consiste à utiliser la décomposition en valeurs singulières (SVD)⁴ de l'opérateur T . Cette décomposition propose des bases pour les espaces de Hilbert H_1 et H_2 permettant d'exprimer et de résoudre simplement le problème.

Définition 2.3.1. Valeurs singulières. Soient H_1, H_2 deux espaces de Hilbert séparables et $T \in \mathcal{K}(H_1, H_2)$. On appelle valeur singulière de l'opérateur T , le réel positif $s = \sqrt{\lambda}$, où λ est une valeur propre de l'opérateur $K = T^*T : H_1 \rightarrow H_1$.

4. La notion de valeurs singulières généralise la notion de valeurs propres liée aux opérateurs auto-adjoints.

Théorème 2.3.1 (Décomposition en valeurs singulières (SVD)). Soit $T \in \mathcal{K}(H_1, H_2)$ et Pr_0 la projection orthogonale sur $\mathbf{N}(T)$. Alors il existe une suite de valeurs singulières (s_n) et deux systèmes orthonormés $\{\varphi_1, \varphi_2, \dots\} \subset H_1, \{\psi_1, \psi_2, \dots\} \subset H_2$ tels que :

1. (s_n) est décroissante, $s_n \rightarrow 0, n \rightarrow \infty$.

2. $T\varphi_k = s_k\psi_k, T^*\psi_k = s_k\varphi_k$.

3. $\forall h \in H_1, h = \sum_{k \geq 1} (h, \varphi_k) \varphi_k + Pr_0 h$.

4. $\forall h \in H_1, Th = \sum_{k \geq 1} s_k (h, \varphi_k) \psi_k$.

5. $\forall \tilde{h} \in H_2, T^*\tilde{h} = \sum_{k \geq 1} s_k (h, \psi_k) \varphi_k$.

Le système $\{(s_k; \varphi_k, \psi_k)\}_{k \geq 1}$ est appelé système singulier de T .

La famille (φ_n) est une base hilbertienne de $\mathbf{N}(T)^\perp$, la famille (ψ_n) est une base hilbertienne de $\overline{\mathbf{R}(T)}$.

Remarque 2.3.1. Le calcul des valeurs singulières et l'étude de leur vitesse de décroissance peut donc fournir des renseignements sur le caractère mal posé d'un problème inverse donné.

2.4 Problèmes mal posés et problèmes inverses

Problèmes directs. Si on note par \mathbf{P} l'espace des paramètres, \mathbf{E} l'espace des excitations et \mathbf{R} l'espace des états (réponses), alors le problème direct $L : \mathbf{P} \times \mathbf{E} \rightarrow \mathbf{R}$, consiste à calculer la réponse d à partir de la donnée des sollicitations x et des paramètres p . Les équations de la physique donnent en général la réponse d comme fonction de x et $p : L(x, p) = d$, la notation L symbolise les équations de la physique du problème considéré; on parle parfois du modèle physique.

Problèmes inverses. D'un point de vue "physique" ou "expérimental", on dit qu'on a un problème inverse toute situation où l'on souhaite évaluer une certaine grandeur physique p inaccessible à l'expérience à partir de la mesure d'une autre grandeur d directement accessible à l'expérience, connaissant un modèle mathématique du problème direct qui donne explicitement d à partir de p (ce que l'on note symboliquement

$d = G(p)$). Réf 5

► **Problèmes directs et Problèmes inverses en EDP**

□ Dans le cas de problèmes directs, étant donné un domaine $\Omega \subset \mathbb{R}^N$, on s'intéresse aux solutions $u : \Omega \times [0, \infty[\ni (x, t) \longrightarrow u(x, t) \in E$ de

$$\begin{cases} u_t + F(t, x, \partial_{x_1}^{\alpha_1} u, \dots, \partial_{x_p}^{\alpha_p} u) = f, & \text{dans } \Omega \\ \{B_i\}_{i=1}^q u = g_i, & \text{sur } \partial\Omega \times [0, \infty[, \\ u(x, 0) = u_0(x) & \text{dans } \Omega \end{cases}$$

□ Dans le cas de problèmes inverses; à partir d'une connaissance partielle de la solution u de l'EDP (mesures internes, mesures frontières), on doit retrouver par exemple :

- $f, g_1, \dots, g_q \longrightarrow$ problème d'identification de sources.
- $u_0 \longrightarrow$ problème d'identification de données initiales.
- $F \longrightarrow$ problème d'identification de coefficients.
- $\Omega \longrightarrow$ problème d'identification géométrique.

La difficulté principale des problèmes inverses est leur caractère généralement mal posé⁶

Définition 2.4.1. [40], Soient X, Y deux espaces de Banach, et $A : X \supseteq D(A) \longrightarrow Y$ un opérateur (linéaire ou non-linéaire). Le problème inverse $Ax = y$ est **bien posé** au sens de HADAMARD si

Existence :	Pour tout $y \in Y$ il existe $x \in X$ tel que $Ax = y$.
Unicité :	Pour tout $y \in Y$, il y a au plus une solution $x \in X$.
Stabilité :	La solution x dépend continûment de la donnée y .

Si au moins une de ces trois conditions n'est pas vérifiée, alors le problème est dit **mal posé**. En pratique, cela veut souvent dire qu'il n'existe pas de solution unique ou que, si elle existe, une légère modification des données conduit à des solutions très différentes.

5. Marc Bonnet, Problèmes inverses : Cours de DEA Dynamique des Structures et Couplages (2004).

6. Alors que les mêmes causes provoquent les mêmes effets, des effets identiques peuvent avoir de multiples causes : les problèmes inverses sont mal posés.

Remarque 2.4.1. Le choix des espaces de départ et d'arrivée X et Y est bien sûr très important dans cette définition. La stabilité est une condition primordiale. En effet, s'il y a un problème de stabilité, le calcul numérique de la solution peut devenir impossible à cause des erreurs de mesures ou d'arrondis.

Remarque 2.4.2. La définition donnée par Hadamard est très contraignante dans la pratique. Il faut donc relaxer la définition d'un problème bien posé.

Définition 2.4.2 (Lavrentiev 1959). (Stabilité conditionnelle) Soit $A : X \supseteq D(A) \rightarrow Y$ un opérateur fermé, densément défini. On dit que le problème $Ax = y$ est conditionnellement stable (ou correct au sens de TIKHONOV) sur $M \subset D(A)$ s'il existe une fonction

$$\omega : \mathbb{R}_+ \rightarrow \mathbb{R}_+, \text{ continue en } 0 \text{ avec } \omega(0) = 0,$$

telle que l'on ait

$$\|x_2 - x_1\| \leq \omega(\|Ax_2 - Ax_1\|), \forall x_2, x_1 \in M.$$

L'ensemble M est appelée ensemble des contraintes (ou ensemble des informations a priori). L'appartenance de u à M signifie une certaine régularité ou une certaine bornitude.

On donne ici quelques exemples de problèmes mal posés.

Exemple 2.4.1. Problème de Cauchy pour l'équation de Laplace. Considérons le problème suivant :

$$\begin{cases} \Delta u = 0, & (x, y) \in \mathbb{R} \times (0, \infty), \\ u(x, 0) = 0, & x \in \mathbb{R}, \\ \partial_y u(x, 0) = \varphi_\varepsilon(x), & x \in \mathbb{R}, \end{cases} \quad (2.4.2)$$

où $\varphi_\varepsilon(x) = \varepsilon \sin\left(\frac{x}{\varepsilon}\right)$, $\varepsilon > 0$. On vérifie aisément que $u_\varepsilon(x, y) = \varepsilon^2 \sinh\left(\frac{y}{\varepsilon}\right) \sin\left(\frac{x}{\varepsilon}\right)$ est une solution du problème (2.4.2). On remarque que $(\varphi_\varepsilon \rightarrow 0, \varepsilon \rightarrow 0)$ mais $(u_\varepsilon(x, y) \rightarrow \infty, \varepsilon \rightarrow 0)$ pour tout $x > 0$ fixé. Ce qui prouve que les solutions de (2.4.2) ne dépendent pas continûment des données initiales, d'où le problème est mal posé.

Exemple 2.4.2. Problème rétrograde pour l'équation de la chaleur. Ce problème consiste à déterminer $u(x, 0) = u_0(x)$ (condition initiale inconnue), sachant que le champ de température $u(x, t)$ vérifié :

$$\begin{cases} u_t - u_{xx} = 0, & x \in (0, \pi), t \in (0, T), \\ u(x, T) = \psi(x), & 0 \leq x \leq \pi, \\ u(0, t) = u(\pi, t) = 0, & 0 \leq t \leq T, \end{cases} \quad (2.4.3)$$

où $\psi \in L_2(0, \pi)$ est une fonction donnée. Par la méthode de Fourier, on peut expliciter la solution du problème (2.4.3) sous la forme :

$$u(x, t) = \sum_{n=1}^{\infty} e^{(T-t)n^2} \psi_n e_n(x),$$

où ψ_n est le coefficient de Fourier d'ordre n de ψ :

$$\psi_n = \langle \psi, e_n \rangle = \sqrt{\frac{2}{\pi}} \int_0^{\pi} \psi(x) \sin(nx) dx, \quad e_n(x) = \sqrt{\frac{2}{\pi}} \sin(nx).$$

Soit $\varphi(x) = u_0(x, 0)$ la température initiale. Alors d'après l'égalité de Parseval, on a :

$$\|\varphi\|^2 = \sum_{n=1}^{\infty} e^{2n^2 T} |\psi_n|^2.$$

On considère maintenant le problème (2.4.3) avec des données bruitées :

$$\psi_k = \psi + \frac{1}{k} e_k(x).$$

On remarque que $\|\psi_k - \psi\| = \frac{1}{k} \rightarrow 0, k \rightarrow +\infty$ mais $\|u(\psi_k; 0) - u(\psi; 0)\| = \frac{1}{k} e^{k^2 T} \rightarrow +\infty, k \rightarrow +\infty$. On voit très clairement que le problème (2.4.3) est instable donc mal posé. C'est pour cela, qu'on dit que les phénomènes de la chaleur sont **irréversibles**.

La solution de l'équation de la chaleur avec la condition initiale $u(x, 0) = \varphi(x)$, telle que $\varphi(x) \in L_2(0, \pi)$ est donnée par la formule :

$$u(x, t) = \sum_{n=1}^{\infty} e^{-n^2 t} \varphi_n e_n(x) = \int_0^{\pi} \left\{ \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-n^2 t} \sin(nx) \sin(n\xi) \right\} \varphi(\xi) d\xi.$$

Ainsi, u est solution du problème (2.4.3) si et seulement si φ satisfait l'équation de Fredholm de première espèce :

$$\mathcal{K}\varphi = \psi, \quad u(x, T) = \int_0^{\pi} K(x, \xi) \varphi(\xi) d\xi = \psi(x), \quad 0 \leq x \leq \pi,$$

où $K(x, \xi) = \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-n^2 T} \sin(nx) \sin(n\xi)$.

L'opérateur intégral K est du type Hilbert-Schmidt (donc compact), d'où K^{-1} n'est pas borné. Ce qui montre le caractère mal posé du problème (2.4.3).

Exemple 2.4.3. Equation hyperbolique avec conditions de Dirichlet. Considérons le problème suivant :

$$\begin{cases} u_{tt}(t) + Au(t) = 0, & 0 < t < T, \\ u(0) = \varphi, \quad u(T) = \psi, \end{cases} \quad (2.4.4)$$

où φ, ψ sont des fonctions données dans H , et $A : \mathcal{D}(A) : H \rightarrow H$ tel que $A = A^*$ et $A \geq \delta > 0$. Si $\lambda_k = \frac{(k\pi)^2}{T^2}, k = 1, 2, \dots$, ne sont pas des valeurs propres de A , alors l'opérateur $(\sin(T\sqrt{A}))$ est injectif, et la solution formelle du problème (2.4.4) est donnée par :

$$u(t) = \sin((T-t)\sqrt{A})(\sin(T\sqrt{A}))^{-1} \psi + \sin(t\sqrt{A})(\sin(T\sqrt{A}))^{-1} \varphi.$$

Inversement, si $\left\{ \lambda_k = \frac{(k\pi)^2}{T^2}, k = 1, 2, \dots \right\} \cap \sigma_p(A) \neq \emptyset$, alors la solution du problème (2.4.4) n'est pas unique. Le problème (2.4.4) est mal posé au sens d'HADAMARD dans les deux cas : les valeurs $\lambda_k = \frac{(k\pi)^2}{T^2}, k = 1, 2, \dots$, peuvent être proches des valeurs propres de A :

$$[\delta, +\infty[\ni \lambda \mapsto \frac{1}{\sin(T\sqrt{\lambda})} \text{ n'est pas bornée au voisinage des } \lambda_k.$$

► On remarque d'après les exemples donnés, qu'il y a deux questions sérieuses liées à cette catégorie de problèmes :

1. **La non unicité :** Pour cette question, il nous faut des informations supplémentaires sur la solution et une bonne connaissance de la nature physique du problème, pour récupérer l'unicité (conditions a priori).
2. **L'instabilité :** Ce caractère est le plus problématique, surtout dans l'implémentation numérique.

Cela veut dire qu'il est impossible de donner un schéma numérique convergent et stable quel que soit la performance de la méthode proposée. Pour traiter ce

caractère d'instabilité, on approxime le problème original par un problème proche (dans un certain sens) qui est stable. Les méthodes de régularisation sont variées. Chaque problème nécessite un traitement spécifique selon sa complexité et son degré de position incorrecte (voir [32]).

2.4.1 Outils d'analyse des problèmes mal posés : cas linéaire

Dans l'étude des équations de la forme :

$$B : D(B) \subseteq H_1 \longrightarrow H_2, u \longmapsto Bu = v,$$

la fermeture de $\mathbf{R}(B)$ est une propriété cruciale, pour que l'inverse de B soit borné. Le Théorème de Banach nous fournit une caractérisation topologique de cette propriété :

Théorème 2.4.1 (Théorème de Banach sur l'inversion bornée). *On suppose que B est injectif. Alors $B^{-1} : \mathbf{R}(B) \longrightarrow H_1$ est borné si et seulement si $\mathbf{R}(B)$ est fermée.*

Théorème 2.4.2 (Théorème de PICARD). *(voir [73]) Soit $K \in \mathcal{K}(H_1, H_2)$ un opérateur compact, et $\{(\sigma_n, \varphi_n, \psi_n), n \in \mathbb{N}\}$ son système singulier. Alors le problème :*

$$Kf = g$$

est résoluble si et seulement si

$$g \in \mathbf{N}(K^*)^\perp = \overline{\mathbf{R}(K)} \text{ et } \sum_{n \in \mathbb{N}} \frac{1}{\sigma_n^2} |\langle g, \psi_n \rangle|^2 < +\infty.$$

Dans ce cas, la solution est donnée par la formule :

$$f = \sum_{n \in \mathbb{N}} \frac{1}{\sigma_n} \langle g, \psi_n \rangle \varphi_n + f_0, f_0 \in \mathbf{N}(K).$$

- Afin de proposer une stratégie de régularisation efficace, on doit mesurer tout d'abord la complexité du problème posé. En général, on ne dispose pas d'un cadre théorique permettant de donner des réponses à ce type de questions, mais dans des cas particuliers, on a des critères qui caractérisent que tels problèmes sont fortement ou faiblement mal posés.

► Pour les opérateurs compacts, on utilise le critère suivant :

Soient H^1, H^2 deux espaces de Hilbert séparables, $T \in \mathcal{K}(H^1, H^2)$, et soit le problème inverse :

$$T : H_1 \longrightarrow H_2, u \longrightarrow Tu = v. \quad (2.4.5)$$

Définition 2.4.3. (cf. [32]) On dit que le problème (2.4.5) est faiblement mal posé (resp. fortement mal posé), si les valeurs singulières s_n de $K = T^*T$ sont équivalentes à $\frac{C}{n^p}$ (resp. Ce^{-n^p}), où C et p sont des constantes positives.

2.4.2 Convexité logarithmique

Parmi les méthodes d'analyse des problèmes mal posés en EDP, on cite ici la méthode de la convexité logarithmique (cf. [46, 31]), qui trouve sa place dans la stabilisation des problèmes mal posés, conditionnellement stables.

Définition 2.4.4. Soit $f : I \subset \mathbb{R} \longrightarrow \mathbb{R}_+$ une fonction positive. On dit que f est log-convexe si la fonction $\hat{f}(t) = \log(f(t))$ est convexe.

Si la fonction f est strictement positive ($f > 0$) et log-convexe sur l'intervalle $[a, b]$, alors on a l'inégalité suivante :

$$\log(f(t)) \leq (1 - \theta(t))\log(f(a)) + \theta(t)\log(f(b)), \quad t \in [a, b], \quad (2.4.6)$$

où $\theta(t) = \frac{b-t}{b-a}$. Cette dernière inégalité nous donne l'inégalité d'interpolation :

$$f(t) \leq f(a)^{1-\theta(t)} f(b)^{\theta(t)}, \quad t \in [a, b]. \quad (2.4.7)$$

► Une fonction $f \in \mathcal{C}^2([a, b]; \mathbb{R}_+^*)$ est log-convexe ssi $ff'' - (f')^2 \geq 0$. Cette propriété découle du fait qu'une fonction est convexe si sa dérivée seconde est positive :

$$\hat{f}'' = \log(f(t))'' \geq 0 \iff \frac{ff'' - (f')^2}{f^2} \geq 0 \iff ff'' - (f')^2 \geq 0 \text{ et } f^2 > 0.$$

Théorème 2.4.3. Soit f une fonction continue, positive et log-convexe sur un intervalle I . Alors $f(t) > 0, \forall t \in I$ ou $f(t) = 0, \forall t \in I$

Théorème 2.4.4. Soit $A : \mathcal{D}(A) \subset H \longrightarrow H$ un opérateur symétrique, et $u(t)$ une solution du problème $u'(t) = Au(t)$, $0 \leq t \leq T$. Alors $\log(|u(t)|)$ est convexe sur $[0, T]$. En conséquence

$$|u(t)| \leq |u(0)|^{1-t/T} |u(T)|^{t/T}. \quad (2.4.8)$$

Cette inégalité montre qu'on peut récupérer la dépendance continue si on impose une contrainte de bornitude sur la solution.

2.5 Méthodes de régularisation

La régularisation des problèmes mal posés, due initialement à TIKHONOV [11], cherche à redéfinir les notions d'inversion et de solution (quasi-solution, solution approchée, ...), de façon que la « solution régularisée » obtenue par « inversion régularisée » dépende continûment des données et soit proche de la solution exacte (supposant que celle-ci existe pour des données proches des valeurs effectivement obtenues par la mesure). En d'autres termes, on remplace le problème initial mal posé par un autre « proche dans un certain sens » du premier et qui est bien posé.

Considérons le problème inverse $Kh_1 = h_2$ où $K : H_1 \longrightarrow H_2$ est un opérateur compact injectif.⁷ On suppose que $h_2 \in \mathbf{R}(K)$, i.e., le problème inverse possède une solution unique.⁸

Définition 2.5.1. Une famille d'opérateurs linéaires bornés $R_\alpha : H_1 \longrightarrow H_2, (\alpha > 0)$ est dite "famille régularisante" pour l'opérateur K si

$$\forall h_1 \in H^1, \lim_{\alpha \rightarrow 0} (R_\alpha K) h_1 = h_1, \text{ i.e., } R_\alpha K \longrightarrow I \text{ simplement.}$$

Remarque 2.5.1. Si R_α est une famille régularisante pour l'opérateur $K : H^1 \longrightarrow H^2$, où H^1 est de dimension infinie, alors les opérateurs R_α ne sont pas uniformément bornés, i.e., il existe une suite $(\alpha_n) \subset \mathbb{R}_+$ telle que $\lim_{n \rightarrow \infty} \|R_{\alpha_n}\| = +\infty$.

7. Le fait de choisir K injectif n'est pas très contraignant car on peut toujours restreindre l'espace H^1 au complément orthogonal de $\mathbf{N}(K)$, où \mathbf{N} désigne le noyau.

8. Il faut noter que notre problème inverse $Kh_1 = h_2$ est toujours mal posé à cause de la non continuité de K^{-1} .

La donnée initiale $h_2 \in H^2$ n'est jamais connue exactement : il y a toujours un bruit qui vient la perturber. Notons h_2^η la donnée perturbée où le nombre $\eta > 0$ est le niveau du bruit, i.e., $|h_2 - h_2^\eta| \leq \eta$.

Notons $h_1^{\alpha,\eta} = R_\alpha h_2^\eta$ l'approximation de la solution du problème inverse $Kh_1 = h_2$ obtenue avec l'opérateur de régularisation et la donnée perturbée. En utilisant l'inégalité triangulaire sur $|h_1 - h_1^{\alpha,\eta}|$, on obtient

$$|h_1 - h_1^{\alpha,\eta}| = |(h_1 - R_\alpha h_2) + (R_\alpha h_2 - h_1^{\alpha,\eta})| \leq \eta \|R_\alpha\| + |(h_1 - R_\alpha h_2)|. \quad (2.5.9)$$

Le premier terme de droite de l'équation (2.5.9) représente la majoration de l'erreur due au niveau de bruit. Par la Remarque 2.5.1, nous avons vu que $\|R_{\alpha_n}\| \rightarrow +\infty$ quand $\alpha \rightarrow 0$. Donc il ne faut pas choisir α trop petit sinon l'erreur peut devenir très grande. Par contre le second terme de droite de (2.5.9) tend vers 0 quand α tend vers 0 par définition de R_α . Nous allons faire tendre le niveau de bruit η vers 0 et nous allons choisir une stratégie de régularisation de manière à ne pas commettre une trop grande erreur sur la vraie solution h_1 .

Définition 2.5.2. Une stratégie de régularisation $\eta \rightarrow \alpha(\eta)$ est admissible si pour tout $h_1 \in H_1$

$$\lim_{\eta \rightarrow 0} \alpha(\eta) = 0 \text{ et } \lim_{\eta \rightarrow 0} \left(\sup_{h_2^\eta \in H^2} \left\{ |R_{\alpha(\eta)} h_2^\eta - h_1| \text{ tel que } |Kh_1 - h_2^\eta| \leq \eta \right\} \right) = 0 \quad (2.5.10)$$

Les stratégies de régularisation sont variées. Chaque problème nécessite un traitement spécifique selon son degré de complexité. (Voir. [32]). Parmi les méthodes les plus connues en problèmes inverses et en calcul matriciel mal conditionné, on a la méthode de Tikhonov et la méthode de la troncature spectrale.

2.5.1 La méthode de Tikhonov

- Le principe de la régularisation de Tikhonov pour stabiliser le problème inverse mal posé $Kf = g$ est de choisir comme solution l'élément f_α qui minimise la fonctionnelle

$$|Kf - g|^2 + \alpha |f|^2, \alpha > 0. \quad (2.5.11)$$

L'existence et l'unicité du minimum sont assurées par la coercivité et la stricte convexité de $f \mapsto |f|^2$. Le paramètre α est appelé paramètre de régularisation et le terme $|f|^2$ est appelé terme de correction. Le choix du paramètre α est basé sur un critère d'équilibre entre l'erreur due au terme de correction et le gain de la stabilité.

On a le Théorème suivant :

Théorème 2.5.1. *Soit $K \in \mathcal{L}(H_1, H_2)$. Alors la fonctionnelle de Tikhonov admet un unique minimum f_α . L'élément f_α est la solution de l'équation normale*

$$S_\alpha f_\alpha = (\alpha I + K^* K) f_\alpha = K^* g. \quad (2.5.12)$$

La famille d'opérateurs $R_\alpha = (\alpha I + K^* K)^{-1} K^* : H^2 \rightarrow H^1$ est appelée famille régularisante de Tikhonov. On a $\|R_\alpha\| \leq \frac{1}{2\sqrt{\alpha}}$ et tout choix de $\alpha(\eta) \rightarrow 0$ avec $\eta^2 \alpha(\eta) \rightarrow 0$ est admissible. Pour les résultats de la vitesse de convergence, on peut consulter les références [32].

- Le paramètre de régularisation $\alpha > 0$ est choisi via le principe d'écart (en anglais : discrepancy principle) de MOROZOV. Ce principe consiste à fixer le paramètre tel que la solution correspondante ait une erreur égale au niveau de bruit.

Le choix optimal est extrêmement difficile et les critères qui existent sont d'application délicate, et nécessitent des méthodes itératives pour être mis en oeuvre.^{10, 11}

- Dans la pratique nous supposerons qu'un paramètre α est valable si l'erreur appartient à un petit intervalle contenant la valeur du niveau de bruit $\eta > 0$.

9. $(S_\alpha = S_\alpha^*, \langle S_\alpha h, h \rangle = |Sh|^2 + \alpha |h|^2 \geq \alpha |h|^2, \forall h \in H_1) \implies (\sigma(S_\alpha) \subset [\alpha, \|S_\alpha\|] \implies 0 \in \rho(S_\alpha))$, i.e., S_α^{-1} existe et $S_\alpha^{-1} \in \mathcal{L}(H_1)$.

10. **Méthodes a priori** : utilisation d'informations sur le niveau d'erreur et sur l'opérateur K .

11. **Méthodes a posteriori** : utilisent aussi les données g_η . $\alpha_{opt} := \max\{\alpha : |Kf_\alpha - g_\eta| \leq \eta\}$,

où $f_\alpha = \inf_f \{|Kf - g_\eta|^2 + \alpha |f|^2\}$.

Pour calculer numériquement le paramètre de régularisation $\alpha(\delta)$, on résout l'équation

$$\varphi(\delta) = \|Kx_{\alpha(\delta)} - f_{\delta}\|^2 - \delta^2 = 0 \quad (2.5.13)$$

par la méthode de Newton

$$\alpha_{j+1} = \alpha_j - \frac{\varphi(\alpha_j)}{\varphi'(\alpha_j)}, \quad \alpha_0 = \text{initialisation}, \quad j = 1, 2, \dots, \quad (2.5.14)$$

où

$$\varphi'(\alpha) = 2\langle -K(\alpha I + K^*K)^{-2}K^*f_{\delta}, K(\alpha I + K^*K)^{-1}K^*f_{\delta} - f_{\delta} \rangle.$$

3

Fonctions de matrices

3.1 Généralités

Nous donnons ici un répertoire de résultats sur la théorie des fonctions de matrices. Pour plus de détails, nous renvoyons le lecteur à l'excellente référence [62], ainsi que les références très intéressantes [99, 43, 28]. Pour les aspects numériques et les récents développements de cette thématique, le lecteur peut visiter la page web du Professeur Stefan Güttel (<http://www.guettel.com>).

Soit

$$f : M_{k,l}(\mathbb{C}) \longrightarrow M_{m,n}(\mathbb{C})$$

une fonction de l'ensemble des matrices de k lignes et l colonnes à éléments dans \mathbb{C} vers l'ensemble des matrices de m lignes et n colonnes à éléments dans \mathbb{C} . Le tableau suivant permet une classification des fonctions de matrices.

- $k = l = 1, m = n = 1, f(A)$: fonction numérique d'une variable scalaire.
- $k = l = 1, m = 1, n > 1$ ou $m > 1, n = 1, f(A)$: fonction vectorielle d'une variable scalaire.
- $k = l = 1, m > 1, n > 1, f(A)$: fonction matricielle d'une variable scalaire.
- $k = 1, l > 1$ ou $k > 1, l = 1$ et $m = n = 1, f(A)$: fonction scalaire d'une variable vectorielle.

- $k = 1, l > 1$ ou $k > 1, l = 1$ et $m = 1, n > 1$, $f(A)$: fonction champ de vecteurs.
- $k = 1, l > 1$ ou $k > 1, l = 1$ et $m > 1, n = 1$, $f(A)$: fonction champ de vecteurs.
- $k = 1, l > 1$ ou $k > 1, l = 1$ et $m > 1, n > 1$, $f(A)$: fonction matricielle d'une variable vectorielle.

- $k = 1, l > 1$ et $m = n = 1$, $f(A)$: fonction scalaire d'une variable matricielle.
- $k = 1, l > 1$ et $m = 1, n > 1$, $f(A)$: fonction vectorielle d'une variable matricielle.
- $k = 1, l > 1$ et $m > 1, n = 1$, $f(A)$: fonction vectorielle d'une variable matricielle.
- $k = 1, l > 1$ et $m > 1, n > 1$, $f(A)$: fonction matricielle d'une variable matricielle.

Dans ce travail nous nous intéressons à la définition d'une fonction de matrice pour une fonction scalaire f et une matrice $A \in M_n(\mathbb{C})$, c'est-à-dire, A appartient à l'ensemble des matrices de n lignes et n colonnes à éléments dans \mathbb{C} , telle que $f(A)$ soit une matrice de même dimension que A . De façon générale, on utilise $f(z), z \in \mathbb{C}$.

Soit $A \in M_n(\mathbb{C})$ une matrice carrée d'ordre n à éléments dans \mathbb{C} . Considérons

$$f : \mathbb{C} \longrightarrow \mathbb{C}$$

$$z \longrightarrow f(z)$$

une fonction scalaire (à valeur dans \mathbb{C}).

Si f est une fonction polynomiale ou une fonction rationnelle ou bien une représentation d'une série entière convergente, alors on peut définir :

$$f : M_n(\mathbb{C}) \longrightarrow M_n(\mathbb{C})$$

$$A \longrightarrow f(A).$$

Exemple 3.1.1. ► $f(t) = 1 + t \Rightarrow f(A) = I + A$.

► $f(t) = \frac{1+t^2}{1-t}, t \neq 1 \Rightarrow f(A) = (I + A^2)(I - A)^{-1}$ si $1 \notin Sp(A)$ ($Sp(A)$ désigne l'ensemble des valeurs propres de la matrice A).

$$\text{► } \ln(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} - \frac{t^4}{4} + \dots = \sum_{m=1}^{+\infty} (-1)^{m+1} \frac{t^m}{m}, |t| < 1 \Rightarrow$$

$$\ln(I+A) = A - \frac{A^2}{2} + \frac{A^3}{3} - \frac{A^4}{4} + \dots = \sum_{m=1}^{+\infty} (-1)^{m+1} \frac{A^m}{m} \text{ si } \rho(A) < 1.$$

Notre objectif ici est de donner les différentes définitions pour définir $f(A)$ où f peut être n'importe quelle fonction scalaire.

Proposition 3.1.1. Si p est un polynôme de degré m , $A \in M_n(\mathbb{C})$ s'écrivant sous la forme

$$A = XJX^{-1} \text{ où } J = \begin{pmatrix} J_1 & & & & \\ & J_2 & & & \\ & & J_3 & & \\ & & & \ddots & \\ & & & & J_r \end{pmatrix} = \text{diag}(J_1, J_2, \dots, J_r)$$

est une matrice diagonale par blocs et X une matrice carrée inversible, alors on a :

i) $p(A) = Xp(J)X^{-1}$,

ii) $p(J) = \text{diag}(p(J_1), p(J_2), \dots, p(J_r))$,

iii) Si $AV = \lambda V$ alors $p(A)V = p(\lambda)V$, $V \in \mathbb{C}^n$ un vecteur.

Preuve. Soit p un polynôme de degré m , $p(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m$ et $A = XJX^{-1}$.

i)

$$\begin{aligned} p(A) &= a_0I + a_1A + a_2A^2 + \dots + a_mA^m \\ &= a_0I + a_1(XJX^{-1}) + a_2(XJX^{-1})^2 + \dots + a_m(XJX^{-1})^m \\ &= a_0XIX^{-1} + a_1XJX^{-1} + a_2XJ^2X^{-1} + \dots + a_mXJ^mX^{-1} \\ &= X(a_0I + a_1J + a_2J^2 + \dots + a_mJ^m)X^{-1} = Xp(J)X^{-1}. \end{aligned}$$

D'où $p(A) = p(XJX^{-1}) = Xp(J)X^{-1}$

ii) $J = \text{diag}(J_1, J_2, \dots, J_r)$. Montrons que $p(J) = \text{diag}(p(J_1), p(J_2), \dots, p(J_r))$.

On a $p(J) = a_0I + a_1J + a_2J^2 + \dots + a_mJ^m$, et le lemme suivant permet le passage de J à J^k

Lemme 3.1.1. $\forall k \in \mathbb{N}$, si $J = \text{diag}(J_1, J_2, \dots, J_r)$ alors $J^k = \text{diag}(J_1^k, J_2^k, \dots, J_r^k)$.

Preuve. Raisonnons par récurrence.

Pour $k = 0$, $J^0 = I_n$.

D'autre part $diag(J_1^0, J_2^0, \dots, J_r^0) = diag(I_{n_1}, I_{n_2}, \dots, I_{n_r}) = I_n$ avec

$$n_1 + n_2 + n_3 + \dots + n_r = n.$$

Supposons la relation vraie à l'ordre $k \geq 0$ et démontrons à l'ordre $k + 1$.

$$J^{k+1} = J^k \times J = diag(J_1^k, J_2^k, \dots, J_r^k) \times diag(J_1, J_2, \dots, J_r) = diag(J_1^{k+1}, J_2^{k+1}, \dots, J_r^{k+1})$$

D'où $\forall k \in \mathbb{N}$, si $J = diag(J_1, J_2, \dots, J_r)$ alors $J^k = diag(J_1^k, J_2^k, \dots, J_r^k)$. CQFD

$$\begin{aligned} \text{Donc } p(J) &= \sum_{k=0}^m a_k J^k = \sum_{k=0}^m a_k diag(J_1^k, J_2^k, \dots, J_r^k) = \sum_{k=0}^m diag(a_k J_1^k, a_k J_2^k, \dots, a_k J_r^k) \\ &= diag\left(\sum_{k=0}^m a_k J_1^k, \sum_{k=0}^m a_k J_2^k, \dots, \sum_{k=0}^m a_k J_r^k\right). \end{aligned}$$

D'où

$$p(J) = diag(p(J_1), p(J_2), \dots, p(J_r)).$$

iii) $AV = \lambda V$, $p(A)V = a_0 IV + a_1 AV + a_2 A^2 V + \dots + a_m A^m V = a_0 V + a_1 \lambda V + a_2 \lambda^2 V + \dots + a_m \lambda^m V$,

car si λ est valeur propre de A alors λ^m est valeur propre de A^m d'où

$$p(A)V = (a_0 + a_1 \lambda + a_2 \lambda^2 + \dots + a_m \lambda^m)V = p(\lambda)V.$$

3.2 Différentes définitions d'une fonction de matrice

3.2.1 Définition par la forme Canonique de Jordan

Forme canonique de Jordan d'une matrice A Soit $A \in M_n(\mathbb{C})$, alors il existe une matrice J et une matrice inversible X telle que $A = XJX^{-1}$ où

$$J = X^{-1}AX = diag(J_1(\lambda_1), J_2(\lambda_2), \dots, J_r(\lambda_r)) = diag(J_k(\lambda_k)) \quad (3.2.1)$$

$$\text{avec } k = 1 \text{ à } r \text{ où } J_k = J_k(\lambda_k) = \begin{pmatrix} \lambda_k & 1 & 0 & 0 & \dots & \dots & 0 \\ 0 & \lambda_k & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & \lambda_k & 1 & 0 & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & 1 \\ 0 & \dots & \dots & \dots & \dots & 0 & \lambda_k \end{pmatrix} \in M_{m_k}(\mathbb{C})$$

$m_1 + m_2 + m_3 + \dots + m_r = n$ (la somme des dimensions des blocs de la matrice de J est égale à la multiplicité algébrique). r est le nombre de blocs dans J , i.e., la dimension de

l'espace propre (le nombre de vecteurs linéairement indépendants), r est la multiplicité géométrique de J .

J est appelé matrice de Jordan par blocs.

$$J_k = J_k(\lambda_k) = \begin{pmatrix} \lambda_k & 0 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & \lambda_k & 0 & 0 & \cdot & \cdot & 0 \\ 0 & & \lambda_k & 0 & & & \cdot \\ 0 & & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & & \cdot & \cdot & \cdot \\ \cdot & & & & & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \lambda_k \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & 1 & 0 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}$$

$J_k = \lambda_k I + N_k \in M_{m_k}$ où N_k est appelé bloc nilpotent de Jordan ou matrice sur-diagonale d'ordre m_k , i.e., $N_k^{m_k} = 0$.

Exemple 3.2.1. Pour $m_k = 4$, on a :

$$J_k = \begin{pmatrix} \lambda_k & 1 & 0 & 0 \\ 0 & \lambda_k & 1 & 0 \\ 0 & 0 & \lambda_k & 1 \\ 0 & 0 & 0 & \lambda_k \end{pmatrix} = \begin{pmatrix} \lambda_k & 0 & 0 & 0 \\ 0 & \lambda_k & 0 & 0 \\ 0 & 0 & \lambda_k & 0 \\ 0 & 0 & 0 & \lambda_k \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, N_k = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$N_k^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, N_k^3 = N_k^2 \cdot N_k = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, N_k^4 = N_k^3 \cdot N_k = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$N_k^4 = 0$. On remarque d'une façon générale qu'élever la matrice N_k à la puissance i revient à décaler la sur-diagonale de i rang.

Calculons J_k^ℓ avec $\ell \geq m_k$:

$$J_k^\ell = (\lambda_k I + N_k)^\ell = \sum_{i=0}^{m_k-1} C_\ell^i \lambda_k^{\ell-i} N_k^i$$

$$N_k = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & 1 & 0 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}, N_k^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 1 & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & 0 & 1 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}, N_k^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdot & 0 \\ 0 & \cdot & 0 & 0 & 0 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}$$

$$\text{Ainsi de suite on a : } N_k^{m_k-2} = \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 1 & 0 \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 1 \\ 0 & \cdot & 0 & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix} \text{ et } N_k^{m_k-1} = \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & 0 & 0 & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}$$

$$\begin{aligned} \text{Donc } J_k^\ell &= \sum_{i=0}^{m_k-1} C_\ell^i \lambda_k^{\ell-i} N_k^i = C_\ell^0 \lambda_k^\ell I + C_\ell^1 \lambda_k^{\ell-1} N_k + C_\ell^2 \lambda_k^{\ell-2} N_k^2 + \dots + C_\ell^{m_k-1} \lambda_k^{\ell-m_k+1} N_k^{m_k-1} \\ &= \lambda_k^\ell I + \ell \lambda_k^{\ell-1} N_k + \frac{\ell(\ell-1)}{2!} \lambda_k^{\ell-2} N_k^2 + \dots + \frac{\ell(\ell-1)(\ell-2)\dots(\ell-m_k+2)}{(m_k-1)!} \lambda_k^{\ell-m_k+1} N_k^{m_k-1} \end{aligned}$$

$$= \begin{pmatrix} \lambda_k^\ell & \ell \lambda_k^{\ell-1} & \frac{\ell(\ell-1)}{2!} \lambda_k^{\ell-2} & \cdot & \cdot & \cdot & \frac{\ell(\ell-1)(\ell-2)\dots(\ell-m_k+2)}{(m_k-1)!} \lambda_k^{\ell-m_k+1} \\ 0 & \lambda_k^\ell & \ell \lambda_k^{\ell-1} & \frac{\ell(\ell-1)}{2!} \lambda_k^{\ell-2} & \cdot & \cdot & \frac{\ell(\ell-1)(\ell-2)\dots(\ell-m_k+3)}{(m_k-2)!} \lambda_k^{\ell-m_k+2} \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\ell(\ell-1)}{2!} \lambda_k^{\ell-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \ell \lambda_k^{\ell-1} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \lambda_k^\ell \end{pmatrix}$$

— Si on pose $p(z) = z^\ell$ avec $\ell \geq m_k$, on constate que :

$$J_k^\ell = \begin{pmatrix} p(\lambda_k) & p'(\lambda_k) & \frac{p''(\lambda_k)}{2!} & \cdot & \cdot & \cdot & \frac{p^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ 0 & p(\lambda_k) & p'(\lambda_k) & \frac{p''(\lambda_k)}{2!} & \cdot & \cdot & \frac{p^{(m_k-2)}(\lambda_k)}{(m_k-2)!} \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{p''(\lambda_k)}{2!} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & p'(\lambda_k) \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & p(\lambda_k) \end{pmatrix}$$

$$\text{D'où } p(J_k) = J_k^\ell = \sum_{i=0}^{m_k-1} \frac{p^{(i)}(\lambda_k)}{i!} N_k^i = p(\lambda_k)I + p'(\lambda_k)N_k + \frac{p''(\lambda_k)}{2!} N_k^2 + \dots + \frac{p^{(m_k-1)}(\lambda_k)}{(m_k-1)!} N_k^{m_k-1}.$$

— Si on pose d'une façon générale $p(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_\ell z^\ell$ avec

$\ell \geq m_k$, alors on a :

$$\begin{aligned}
p(z) &= \sum_{j=0}^{\ell} a_j q_j(z) \text{ avec } q_j(z) = z^j \\
p(J_k) &= \sum_{j=0}^{\ell} a_j J_k^j = \sum_{j=0}^{\ell} a_j \left(\sum_{i=0}^{m_k-1} \frac{q_j^{(i)}(\lambda_k)}{i!} N_k^i \right) = \sum_{j=0}^{\ell} a_j \left(\sum_{i=0}^{m_k-1} \frac{q_j^{(i)}(\lambda_k)}{i!} N_k^i \right) = \sum_{i=0}^{m_k-1} \left(\sum_{j=0}^{\ell} a_j \frac{q_j^{(i)}(\lambda_k)}{i!} \right) N_k^i \\
&= \sum_{i=0}^{m_k-1} \frac{p^{(i)}(\lambda_k)}{i!} N_k^i.
\end{aligned}$$

Maintenant, on remplace p par une fonction $f : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$, où $f(J_k)$ est définie sur les valeurs spectrales, i.e., si $f(\lambda_k), f'(\lambda_k), f''(\lambda_k), \dots, f^{m_k-1}(\lambda_k)$ existent.

Domaine de Définition de $f(A)$.

Définition 3.2.1. Soient $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_s$ les valeurs propres distinctes de A et n_i l'indice de λ_i , i.e., la dimension du plus grand bloc de la matrice de Jordan dans lequel λ_i apparaît (l'ordre de multiplicité de λ_i).

f est dite définie sur un voisinage du spectre de A si les dérivées $f^{(j)}(\lambda_i)$ existent pour $j = 0$ à $n_i - 1$, $i = 1$ à s . Les valeurs $f^{(j)}(\lambda_i)$ sont appelées valeurs de la fonction f sur le spectre de A .

Définition de $f(A)$ Soient $f : \mathbb{C} \rightarrow \mathbb{C}$ et $A \in M_n(\mathbb{C})$ avec $J = X^{-1}AX$, où $J = \text{diag}(J_1, J_2, \dots, J_s)$ et $J_i = J_i(\lambda_i) \in M_{n_i}(\mathbb{C})$.

Si f est définie sur un voisinage du spectre de A , on pose alors

$$f(A) = Xf(J)X^{-1} = X \text{diag}\left(f(J_1(\lambda_1)), f(J_2(\lambda_2)), \dots, f(J_s(\lambda_s))\right)X^{-1}, \quad (3.2.2)$$

où

$$f(J_i(\lambda_i)) = \sum_{j=0}^{n_i} \frac{f^{(j)}(\lambda_i)}{j!} N_i^j = \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{f''(\lambda_i)}{2!} & \cdot & \cdot & \cdot & \frac{f^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ 0 & f(\lambda_i) & f'(\lambda_i) & \frac{f''(\lambda_i)}{2!} & \cdot & \cdot & \frac{f^{(n_i-2)}(\lambda_i)}{(n_i-2)!} \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{f''(\lambda_i)}{2!} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & f'(\lambda_i) \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & f(\lambda_i) \end{pmatrix} \quad (3.2.3)$$

Exemple 3.2.2. On donne la matrice de Jordan par bloc $J = \begin{pmatrix} \frac{1}{2} & 1 \\ 0 & \frac{1}{2} \end{pmatrix}$ et $f(x) = x^3$. On donne ici $f(J)$ et on vérifie que $f(J) = J^3$.

On a

$$J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, f(J) = \begin{pmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{pmatrix}$$

$$\lambda = \frac{1}{2}, f\left(\frac{1}{2}\right) = \frac{1}{8}, f'\left(\frac{1}{2}\right) = \frac{3}{4}$$

$$f(J) = \begin{pmatrix} \frac{1}{8} & \frac{3}{4} \\ 0 & \frac{1}{8} \end{pmatrix}.$$

Vérifions que $f(J) = J^3$,

$$J^2 = \begin{pmatrix} \frac{1}{2} & 1 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 1 \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & 1 \\ 0 & \frac{1}{4} \end{pmatrix}$$

$$J^3 = \begin{pmatrix} \frac{1}{4} & 1 \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 1 \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{8} & \frac{3}{4} \\ 0 & \frac{1}{8} \end{pmatrix} = f(J)$$

Exemple 3.2.3. Calcul de e^A pour $A = \begin{pmatrix} 3 & -6 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$.

Cherchons d'abord J , pour cela, calculons l'équation caractéristique

$$p(\lambda) = 0 \Leftrightarrow \det(A - \lambda I) = 0 \Leftrightarrow$$

$$\begin{vmatrix} 3-\lambda & -6 & 4 \\ \frac{1}{2} & -\lambda & 0 \\ 0 & \frac{1}{2} & -\lambda \end{vmatrix} = (3-\lambda) \begin{vmatrix} -\lambda & 0 \\ \frac{1}{2} & -\lambda \end{vmatrix} - \frac{1}{2} \begin{vmatrix} -6 & 4 \\ \frac{1}{2} & -\lambda \end{vmatrix} = 0 \Leftrightarrow \lambda^3 - 3\lambda^2 + 3\lambda - 1 = (\lambda - 1)^3 = 0$$

Soit $\lambda = 1$, $n_1 = 3$. A n'admet qu'une seule valeur propre qui est 1. Cherchons les vecteurs propres pour pouvoir déterminer le nombre de blocs dans la matrice de J . Soit E_1 le sous espace propre associé au vecteur propre 1, $E_1 = \text{Vect}(1)$ où $\text{Vect}(1)$ est

l'ensemble des vecteurs propres associés à 1.

Soit Z un vecteur propre associé à 1.

$$\text{On a : } AZ = Z \Leftrightarrow (A - I)Z = 0 \Leftrightarrow \begin{pmatrix} 2 & -6 & 4 \\ \frac{1}{2} & -1 & 0 \\ 0 & \frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{Ce qui peut s'écrire } \begin{cases} 2x - 6y + 4z = 0 \\ \frac{1}{2}x - y = 0 \\ \frac{1}{2}y - z = 0 \end{cases} \Leftrightarrow \begin{cases} 2x - 6y + 4z = 0 \\ x = 2y \\ y = 2z \end{cases} .$$

Donc $E_1 = \{(4x, 2x, x), x \in \mathbb{R}\}$ et par suite on a un seul vecteur propre $z = \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}$, et

la forme de Jordan de A ne possède qu'un seul bloc ($s = 1$),

$$J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Cherchons X . On sait que $A = XJX^{-1} \Leftrightarrow AX = XJ$ où $X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$.

$$\text{Après calcul on trouve : } X = \begin{pmatrix} 1 & 2 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & -1 \end{pmatrix} \text{ et } X^{-1} = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 2 & -4 \\ 1 & -4 & 4 \end{pmatrix},$$

$$f(A) = Xf(J)X^{-1}, f(J) = e^J, f(x) = e^x, f'(x) = e^x, f''(x) = e^x$$

$$f(J) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} \\ 0 & f(\lambda) & f'(\lambda) \\ 0 & 0 & f(\lambda) \end{pmatrix} = \begin{pmatrix} e & e & \frac{e}{2} \\ 0 & e & e \\ 0 & 0 & e \end{pmatrix}$$

$$e^A = Xf(J)X^{-1} = \begin{pmatrix} 1 & 2 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & -1 \end{pmatrix} \begin{pmatrix} e & e & \frac{e}{2} \\ 0 & e & e \\ 0 & 0 & e \end{pmatrix} \begin{pmatrix} 0 & 0 & 4 \\ 0 & 2 & -4 \\ 1 & -4 & 4 \end{pmatrix} = \begin{pmatrix} \frac{7e}{2} & -8e & 6e \\ \frac{3e}{4} & -e & e \\ -\frac{7e}{8} & 4e & \frac{-7e}{2} \end{pmatrix}.$$

On obtient

$$e^A = e \begin{pmatrix} \frac{7}{2} & -8 & 6 \\ \frac{3}{4} & -1 & 1 \\ -\frac{7}{8} & 1 & \frac{-7}{2} \end{pmatrix}.$$

Remarque 3.2.1. Si A est diagonalisable, alors la forme canonique de Jordan est réduite à $A = XDX^{-1}$ avec $D = \text{diag}(\lambda_i)$ et $f(A) = Xf(D)X^{-1} = X\text{diag}(f(\lambda_i))X^{-1}$.

Donc pour les matrices diagonales, $f(A)$ a les mêmes vecteurs propres que A et ses valeurs propres sont obtenues en calculant $f(\lambda)$, $\lambda \in \text{Sp}(A)$.

Remarque 3.2.2. $f(J_i)$ peut être obtenue avec la série de Taylor. Supposons que f a un développement en série de Taylor convergente telle que

$$f(t) = f(\lambda_i) + f'(\lambda_i)(t - \lambda_i) + \frac{f''(\lambda_i)(t - \lambda_i)^2}{2!} + \dots + \frac{f^{(j)}(\lambda_i)(t - \lambda_i)^j}{j!} + \dots$$

or $J_i = \lambda_i I + N_i$. Or N_i est le bloc nilpotent de Jordan, en remplaçant t par $J_i \in M_n(\mathbb{C})$, nous obtenons alors une série finie :

$$\begin{aligned} f(J_i) &= f(\lambda_i)I + f'(\lambda_i)(J_i - \lambda_i I) + \frac{f''(\lambda_i)(J_i - \lambda_i I)^2}{2!} + \dots + \frac{f^{(n_i-1)}(\lambda_i)(J_i - \lambda_i I)^{n_i-1}}{(n_i - 1)!} + \dots \\ &= f(\lambda_i)I + f'(\lambda_i)N_i + \frac{f''(\lambda_i)}{2!}N_i^2 + \dots + \frac{f^{(n_i-1)}(\lambda_i)}{(n_i - 1)!}N_i^{n_i-1} \quad \text{car } N_i^{n_i} = 0. \end{aligned}$$

Conclusion Cette définition de $f(A)$ exige à la fonction f de prendre des valeurs bien définies sur $\text{Sp}(A)$ y compris les valeurs liées aux dérivées de f .

3.2.2 Définition de $f(A)$ par l'interpolation polynomiale

Polynôme minimal

Définition 3.2.2. On appelle polynôme minimal de A le polynôme unitaire ψ de plus bas degré tel que $\psi(A) = 0$.

Si f admet un développement en série de Taylor convergente alors d'après la remarque précédente, on a :

$$f(t) = f(\lambda_i) + f'(\lambda_i)(t - \lambda_i) + \frac{f''(\lambda_i)(t - \lambda_i)^2}{2!} + \dots + \frac{f^{(n_i-1)}(\lambda_i)(t - \lambda_i)^{n_i-1}}{(n_i - 1)!}.$$

On pose $\psi(t) = \prod_{i=1}^s (t - \lambda_i)^{n_i}$ où $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_s$ sont des valeurs propres distinctes de A et n_i est l'ordre de multiplicité de λ_i .

Théorème 3.2.1. Soient p et q deux polynômes de degré m et $A \in M_n(\mathbb{C})$. Alors, $p(A) = q(A)$ ssi p et q prennent les mêmes valeurs sur $\text{Sp}(A)$.

Preuve. Supposons $p(A) = q(A)$.

$p(A) = q(A) \Rightarrow A$ est un zéro de $d = p - q$ qui est aussi divisible par le polynôme minimal ψ . Donc $d(A) = \psi(A)t(A)$ et $d(A) = 0 \Rightarrow \psi(A) = 0$, ce qui veut dire que d prend la valeur zéro sur $\text{Sp}(A)$. D'où p et q prennent les mêmes valeurs sur $\text{Sp}(A)$.

Réciproquement, supposons que p et q prennent les mêmes valeurs sur $\text{Sp}(A)$.

On a $d(A) = p(A) - q(A) = 0$ sur $\text{Sp}(A)$, donc divisible par le polynôme minimal ψ . Ainsi, $d(A) = \psi(A)t(A)$ pour un polynôme t et puisque $d(A) = \psi(A)t(A) = 0$, il s'ensuit que $p(A) - q(A) = 0 \Rightarrow p(A) = q(A)$. ■

Ainsi, $p(A)$ est complètement déterminé par les valeurs de p sur $\text{Sp}(A)$. Donc on peut généraliser cette propriété sur une fonction arbitraire $f(A)$ définie de manière qu'elle soit déterminée par les valeurs de f sur $\text{Sp}(A)$.

Définition 3.2.3. (Définition de $f(A)$) Si f est une fonction définie sur un voisinage du spectre de $A \in M_n(\mathbb{C})$ et $\psi(A)$ le polynôme minimal de A alors $f(A) = p(A)$ où p est un polynôme de degré $\deg(p) < \deg(\psi) = \deg(\psi_A) = \sum_{i=1}^s n_i$ et satisfait aux conditions d'interpolation de Hermite :

$$p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i) \quad j = 0 : n_i - 1 \text{ et } i = 1 : s \quad (3.2.4)$$

$$\text{avec } p(t) = \sum_{i=1}^s \left(\left(\sum_{j=0}^{n_i-1} \frac{1}{j!} \phi_i^{(j)}(\lambda_i) (t - \lambda_i)^j \right) \prod_{j \neq i} (t - \lambda_j)^{n_j} \right)$$

où

$$\phi_i(t) = \frac{f(t)}{\prod_{j \neq i} (t - \lambda_j)^{n_j}}$$

Mais pour une matrice ayant des valeurs propres distinctes ($n_i \equiv 1, s = n$) cette formule de $p(t)$ est réduite à la formule de Lagrange :

$$p(t) = \sum_{i=1}^n f(\lambda_i) \ell_i(t), \quad \ell_i(t) = \prod_{j=1, j \neq i}^n \frac{t - \lambda_j}{\lambda_i - \lambda_j} \quad (3.2.5)$$

$p(t)$ peut être obtenu aussi par la formule suivante : $f(J_i(\lambda_i)) = \sum_{j=0}^{n_i} \frac{f^{(j)}(\lambda_i)}{j!} N_i^j$

en remplaçant J_i par t et sachant que $N_i = J_i - \lambda_i I$:

$$p(t) = f(\lambda_i) + f'(\lambda_i)(t - \lambda_i) + \frac{f''(\lambda_i)(t - \lambda_i)^2}{2!} + \dots + \frac{f^{(n_i-1)}(\lambda_i)(t - \lambda_i)^{n_i-1}}{(n_i - 1)!} \quad (3.2.6)$$

Exemple 3.2.4. Soient $f(t) = t^{\frac{1}{2}}$ et $A = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$. Calculer $A^{\frac{1}{2}}$. A-t-on $A^{\frac{1}{2}} = e^{\frac{1}{2}\ln(A)}$?

Calculons $A^{\frac{1}{2}}$. $\det(A - \lambda I) = 0 \Leftrightarrow \begin{vmatrix} 2-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} = (2-\lambda)(3-\lambda) - 2 = 0$ ou $\lambda^2 - 5\lambda + 4 = 0$.

Soit $\lambda_1 = 1, \lambda_2 = 4, i = 1 \text{ à } 2, s = 2$ et $n_1 = n_2 = 1, f(t) = t^{\frac{1}{2}}, f(1) = 1, f(4) = 2$.

$$p(t) = f(\lambda_1) \frac{t - \lambda_2}{\lambda_1 - \lambda_2} + f(\lambda_2) \frac{t - \lambda_1}{\lambda_2 - \lambda_1}, = \frac{t-4}{-3} - 2 \frac{t-1}{3} = \frac{t+2}{3},$$

$$p(t) = \frac{t+2}{3} \Rightarrow p(A) = \frac{1}{3}(A + 2I) = f(A), \text{ ce qui peut s'écrire : } A^{\frac{1}{2}} = \frac{1}{3}(A + 2I) = \frac{1}{3} \begin{pmatrix} 4 & 2 \\ 1 & 5 \end{pmatrix}$$

Le polynôme minimal est $\psi(t) = (t-1)(t-4)$.

Posons $f(t) = e^{\frac{1}{2}\ln(t)}, f(1) = e^{\frac{1}{2}\ln(1)} = 1, f(4) = e^{\frac{1}{2}\ln(4)} = 2$.

$$p(t) = f(\lambda_1) \frac{t - \lambda_2}{\lambda_1 - \lambda_2} + f(\lambda_2) \frac{t - \lambda_1}{\lambda_2 - \lambda_1},$$

$$p(t) = \frac{t+2}{3} \Rightarrow p(A) = \frac{1}{3}(A + 2I) = f(A).$$

Finalement $f(A) = e^{\frac{1}{2}\ln(A)} = \frac{1}{3}(A + 2I)$.

Donc pour que $A^{\frac{1}{2}} = e^{\frac{1}{2}\ln(A)}$ il faut que $Sp(A) \subset \mathbb{R}_+^*$.

Exemple 3.2.5. Soit $A = \begin{pmatrix} 3 & -6 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$. Calculer $A^{\frac{1}{2}}$.

Posons $f(t) = t^{\frac{1}{2}}, f'(t) = \frac{1}{2}t^{-\frac{1}{2}}, f''(t) = -\frac{1}{4}t^{-\frac{3}{2}}$. On a $\det(A - \lambda I) = (\lambda - 1)^3, \lambda = 1, n_1 = 3, j = 0 \text{ à } 2$.

$f(t) = p(t) = f(\lambda) + f'(\lambda)(t - \lambda) + \frac{f''(\lambda)(t - \lambda)^2}{2!} = 1 + \frac{1}{2}(t - 1) - \frac{1}{8}(t - 1)^2$, ce qui implique que $p(A) = f(A) = I + \frac{1}{2}(A - I) - \frac{1}{8}(A - I)^2$.

$$\text{On a : } A - I = \begin{pmatrix} 2 & -6 & 4 \\ \frac{1}{2} & -1 & 0 \\ 0 & \frac{1}{2} & -1 \end{pmatrix}, \quad (A - I)^2 = \begin{pmatrix} 1 & -4 & 4 \\ \frac{1}{4} & -2 & 2 \\ \frac{1}{4} & -1 & 1 \end{pmatrix},$$

$$f(A) = A^{\frac{1}{2}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & -3 & 2 \\ \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} - \begin{pmatrix} \frac{1}{8} & -\frac{1}{2} & \frac{1}{4} \\ \frac{16}{1} & -\frac{4}{1} & \frac{4}{1} \\ \frac{1}{32} & -\frac{1}{8} & \frac{1}{8} \end{pmatrix} = \begin{pmatrix} \frac{15}{8} & -\frac{5}{7} & \frac{3}{2} \\ \frac{3}{8} & \frac{7}{4} & \frac{1}{1} \\ -\frac{1}{32} & \frac{4}{8} & \frac{11}{8} \end{pmatrix}.$$

Remarque 3.2.3. 1 Si $\lambda_1 \neq \lambda_2$, alors $f\left(\begin{pmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{pmatrix}\right) = \begin{pmatrix} f(\lambda_1) & \alpha \frac{f(\lambda_1) - f(\lambda_2)}{\lambda_2 - \lambda_1} \\ 0 & f(\lambda_2) \end{pmatrix}$

2 Si $\lambda_1 = \lambda_2 = \lambda$ alors $f\left(\begin{pmatrix} \lambda & \alpha \\ 0 & \lambda \end{pmatrix}\right) = \begin{pmatrix} f(\lambda) & \alpha f'(\lambda) \\ 0 & f(\lambda) \end{pmatrix}$

Exemple 3.2.6. On a les égalités suivantes :

$$— \cos^2(A) + \sin^2(A) = I,$$

$$— (A^{\frac{1}{p}})^p = A,$$

$$— e^{iA} = \cos A + i \sin A.$$

3.2.3 Définition de $f(A)$ par l'intégrale de Cauchy

Soit $A \in M_n(\mathbb{C})$. On définit $f(A)$ par :

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz, \quad (3.2.7)$$

où f est une fonction analytique dans un domaine D , tel que Γ entoure D avec $\Gamma \cap D = \emptyset$.

Il n'est pas facile de calculer $f(A)$ explicitement avec cette formule, mais elle conduit à la démonstration de certains résultats théoriques.

Exemple 3.2.7. Soit $A \in M_n(\mathbb{C})$, $f(t) = e^t$. Calculer $f(A) = e^A$.

Soit l'intégrale $f(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(t)}{(t-z)} dt$ avec $\left| \frac{z}{t} \right| < 1$ sur Γ . On a

$$\frac{1}{t-z} = \frac{1}{t} \times \frac{1}{1 - \frac{z}{t}} = \frac{1}{t} \sum_{k \geq 0} \frac{z^k}{t^k} = \sum_{k \geq 0} \frac{z^k}{t^{k+1}},$$

$$f(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(t)}{(t-z)} dt = \frac{1}{2\pi i} \int_{\Gamma} f(t) \sum_{k \geq 0} \frac{z^k}{t^{k+1}} dt = \frac{1}{2\pi i} \int_{\Gamma} \sum_{k \geq 0} \frac{f(t)}{t^{k+1}} z^k dt.$$

En posant $a_k = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(t)}{t^{k+1}} dt$, on a : $f(z) = \sum_{k \geq 0} a_k z^k$.

D'autre part $a_k = \frac{f^{(k)}(0)}{k!}$ donc $f(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(t)}{(t-z)} dt = \sum_{k \geq 0} \frac{f^{(k)}(0)}{k!} z^k$,

$$f(t) = e^t, f^{(k)}(t) = e^t, f^{(k)}(0) = 1.$$

Finalemment, $f(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(t)}{t-z} dt = \sum_{k \geq 0} \frac{1}{k!} z^k$ et en remplaçant z par A , il vient

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(t)(tI - A)^{-1} dt = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz = \sum_{k \geq 0} \frac{A^k}{k!}.$$

D'où $f(A) = e^A = \sum_{k \geq 0} \frac{A^k}{k!}$ avec $\rho(A) < 1$.

Remarque 3.2.4. Cette série converge normalement sur toute partie bornée de $M_n(\mathbb{C})$.

En effet, Soit E une partie bornée de $M_n(\mathbb{C})$. Il existe donc $\alpha \geq 0$ tel que $\forall A \in E$, $\|A\| \leq \alpha$.

On a alors pour $k \in \mathbb{N}^*$ et pour tout $A \in E$, $\|\frac{1}{k!} A^k\| \leq \frac{1}{k!} \|A^k\| \leq \frac{1}{k!} \alpha^k$.

Donc en introduisant le symbole \sum , on a :

$$\|\sum_{k \geq 0} \frac{1}{k!} A^k\| \leq \sum_{k \geq 0} \|\frac{1}{k!} A^k\| \leq \sum_{k \geq 0} \frac{1}{k!} \|A^k\| \leq \sum_{k \geq 0} \frac{1}{k!} \alpha^k$$

où $\sum_{k \geq 0} \frac{1}{k!} \alpha^k$ est majorée par e^α qui est fini, ce qui montre la convergence.

Théorème 3.2.2. La définition (3.2.2) et la définition (3.2.4) sont équivalentes. Si f est analytique alors la définition (3.2.7) est équivalente aux deux premières.

Preuve. Démontrons que (3.2.4) \Rightarrow (3.2.2), (3.2.2) \Rightarrow (3.2.7) et (3.2.7) \Rightarrow (3.2.4)
* (3.2.4) \Rightarrow (3.2.2)

Supposons que $f(A) = p(A)$ et $p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i)$, $j = 0 : n_i - 1$, $i = 1 : s$, et montrons que $f(A) = Zf(J)Z^{-1}$.

Si A peut s'écrire sous la forme $A = ZJZ^{-1}$, alors

$$f(A) = p(ZJZ^{-1}) = Zp(J)Z^{-1} = Z \text{diag}(p(J_i(\lambda_i)))Z^{-1} = Z \text{diag}(f(J_i(\lambda_i)))Z^{-1} = Zf(J)Z^{-1}$$

, car $p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i)$ et de plus les valeur de p sur $Sp(J_i)$ constituent un sous ensemble des valeurs de p sur $Sp(A)$. D'où $f(A) = Zf(J)Z^{-1}$.

** (3.2.2) \Rightarrow (3.2.7).

Supposons que $f(A) = Zf(J)Z^{-1} = Z \text{diag}(f(J_i(\lambda_i)))Z^{-1}$ et montrons que

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz.$$

On a $J_k = \lambda_k I + N_k \in M_{m_k}$ où $N_k^{n_j} = 0$, $n_1 + n_2 + n_3 + \dots + n_s = N$, $j = 1 : s$,

$$f(J_k) = f(J_k(\lambda_k)) = \sum_{j=0}^{n_i-1} \frac{f^{(j)}(\lambda_k)}{j!} N_k^j.$$

On a vu dans l'exemple précédent que $a_k = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(t)}{t^{k+1}} dt = \frac{f^{(k)}(0)}{k!}$ (f étant analytique). Donc au point λ_k on a :

$$a_j = \frac{f^{(j)}(\lambda_k)}{j!} = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(z)}{(z - \lambda_k)^{j+1}} dz.$$

D'autre part,

$$(zI - J_k)^{-1} = (zI - \lambda_k I - N_k)^{-1},$$

$$\begin{aligned} ((zI - \lambda_k I) - N_k)^{-1} &= \frac{1}{z - \lambda_k} I + \frac{1}{(z - \lambda_k)^2} N_k + \frac{1}{(z - \lambda_k)^3} N_k^2 + \dots + \frac{1}{(z - \lambda_k)^{n_i}} N_k^{n_i-1} \\ &= \frac{1}{z - \lambda_k} \sum_{j=0}^{n_i-1} \left(\frac{N_k}{z - \lambda_k} \right)^j = \sum_{j=0}^{n_i-1} \frac{N_k^j}{(z - \lambda_k)^{j+1}}. \end{aligned}$$

Donc

$$\frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - J_k)^{-1} dz = \sum_{j=0}^{n_i-1} \frac{1}{2\pi i} \int_{\Gamma} \frac{f(z)}{(z - \lambda_k)^{j+1}} N_k^j dz = \sum_{j=0}^{n_i-1} \frac{f^{(j)}(\lambda_k)}{j!} N_k^j = f(J_k).$$

D'où en remplaçant J_k par A , on a :

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz$$

(3.2.7) \Rightarrow (3.2.4).

Supposons que $f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz$ où f est analytique et montrons que $f(A) = p(A)$ où p est un polynôme de degré $\deg(p) < \deg(\psi) = \sum_{i=1}^s n_i$ et satisfait aux conditions d'interpolation de Hermites :

$$p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i), \quad j = 0 : n_i - 1 \text{ et } i = 1 : s.$$

Lemme 3.2.1. $f(A) = p(A)$ si et seulement si

$$p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i) \quad j = 0 : n_i - 1 \text{ et } i = 1 : s$$

pour tout $\lambda_i \in Sp(A)$.

Preuve. Soit $J = Z^{-1}AZ$ avec $J = \text{diag}(J_1, J_2, \dots, J_s) = \text{diag}(J_i) \quad i = 1 : s$.

Donc il est clair $f(A) = p(A) \Leftrightarrow f(J_i) = p(J_i)$. Par définition,

$$\begin{aligned} f(J_i) = f(J_i(\lambda_i)) &= \sum_{j=0}^{n_i} \frac{f^{(j)}(\lambda_i)}{j!} N_i^j = \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{f''(\lambda_i)}{2!} & \cdot & \cdot & \cdot & \frac{f^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ 0 & f(\lambda_i) & f'(\lambda_i) & \frac{f''(\lambda_i)}{2!} & \cdot & \cdot & \frac{f^{(n_i-2)}(\lambda_i)}{(n_i-2)!} \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{f''(\lambda_i)}{2!} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & f'(\lambda_i) \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & f(\lambda_i) \end{pmatrix} \\ &= \begin{pmatrix} p(\lambda_i) & p'(\lambda_i) & \frac{p''(\lambda_i)}{2!} & \cdot & \cdot & \cdot & \frac{p^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ 0 & p(\lambda_i) & p'(\lambda_i) & \frac{p''(\lambda_i)}{2!} & \cdot & \cdot & \frac{p^{(n_i-2)}(\lambda_i)}{(n_i-2)!} \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{p''(\lambda_i)}{2!} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & p'(\lambda_i) \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & p(\lambda_i) \end{pmatrix} \\ &= p(J_i). \end{aligned}$$

Donc $f(J_i) = p(J_i)$ pour tout $i = 1 : s$ si et seulement si $p^{(j)}(\lambda_i) = f^{(j)}(\lambda_i)$. (CQFD)

Ainsi

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz = \frac{1}{2\pi i} \int_{\Gamma} p(z)(zI - A)^{-1} dz = p(A).$$

D'où l'équivalence des trois définitions de $f(A)$. ■

Remarque 3.2.5. — $f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz$

$$\text{On a} \quad \frac{d^m f}{dz^m}(A) = \frac{m!}{2\pi i} \int_{\Gamma} f(z) \left((zI - A)^{-1} \right)^{(m+1)} dz$$

- La première définition de $f(A)$ facilite la résolution des équations matricielles de la forme $X^2 = A$ et $e^X = A$ et les équations différentielles contenant des dérivées de très grand ordre.
- La deuxième définition de $f(A)$ fournit la manière la plus simple de l'obtenir.
- La troisième définition peut être généralisée aux opérateurs.

3.2.4 Propriétés élémentaires de $f(A)$

Soient $A \in M_n(\mathbb{C})$ et f une fonction définie sur un voisinage de $Sp(A)$. On a les propriétés suivantes :

1. $f(A)A = Af(A)$: , $f(A)$ permute avec A ,
2. $f(A^t) = f(A)^t$,
3. pour Z inversible, $f(ZAZ^{-1}) = Zf(A)Z^{-1}$,
4. $f(\text{diag}(A_1, A_2, A_3, \dots, A_m)) = \text{diag}(f(A_1), f(A_2), f(A_3), \dots, f(A_m))$.
5. Soient f et g définies sur un voisinage de $Sp(A)$
 - (a) Si $h(t) = f(t) + g(t)$ alors $h(A) = f(A) + g(A)$,
 - (b) Si $h(t) = f(t)g(t)$ alors $h(A) = f(A)g(A)$.
6. Supposons que f admet un développement en série de Taylor

$$f(z) = \sum_{k=0}^{+\infty} \frac{f^{(k)}(\alpha)}{k!} (z - \alpha)^k = \sum_{k=0}^{+\infty} a_k (z - \alpha)^k$$

de rayon de convergence R . Si $A \in M_n(\mathbb{C})$ alors $f(A)$ est définie par

$f(A) = \sum_{k=0}^{+\infty} a_k (A - \alpha I)^k$ si et seulement si les λ_i , ($i = 1 : s$) satisfont aux conditions suivantes :

- (a) $|\lambda_i - \alpha| < R$,
- (b) $|\lambda_i - \alpha| = R$ et la série $f^{n_i-1}(\lambda)$ où n_i est l'indice de λ_i est convergente au point $\lambda = \lambda_i$, ($i = 1 : s$)

7. Si f est analytique sur un ouvert $\Omega \subseteq \mathbb{C}$, $f(A) \in M_n(\mathbb{C})$ et $\mathcal{D} = \{A \in M_n(\mathbb{C}), Sp(A) \subseteq \Omega\}$ alors

$$(a) f(A^*) = f(A)^* \quad \text{pour tout } A \in \mathcal{D}$$

$$(b) f(\bar{A}) = \overline{f(A)} \quad \text{pour tout } A \in \mathcal{D}$$

(cf.[62])

3.3 Quelques particularités sur les matrices

3.3.1 Décomposition en valeurs singulières d'une matrice

Nous donnons la définition avec $A \in M_{m,n}(\mathbb{R})$ (resp. $A \in M_{m,n}(\mathbb{C})$).

La décomposition en valeurs singulières ("Singular value decomposition" ou SVD) repose sur un théorème d'Algèbre linéaire qui dit qu'une matrice rectangulaire peut être exprimée comme le produit de 3 matrices : une matrice orthogonale U (unitaire), une matrice diagonale Σ et la transposée d'une matrice orthogonale V^t (unitaire V^*)

$$A = U\Sigma V^t. \quad (3.3.8)$$

$U \in M_{m,m}(\mathbb{R})$, $\Sigma \in M_{m,n}(\mathbb{R})$, $V^t \in M_{n,n}(\mathbb{R})$ (resp. $U \in M_{m,m}(\mathbb{C})$, $\Sigma \in M_{m,n}(\mathbb{C})$, $V^* \in M_{n,n}(\mathbb{C})$), où

$$UU^t = I, \text{ et } V^tV = I \text{ respectivement } (UU^* = I, \text{ et } V^*V = I).$$

Les colonnes de U sont des vecteurs orthogonaux de AA^t (resp. de AA^*).

Les colonnes de V sont des vecteurs orthogonaux de A^tA (resp. de A^*A) et Σ est une matrice diagonale qui contient les racine carrées des valeurs propres de U ou de V en ordre décroissant.

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_p) \text{ avec } p = \min(m, n), \sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_p \geq 0.$$

Exemple 3.3.1. Décomposons en valeurs singulières la matrice suivante :

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}$$

Trouvons d'abord U en calculant AA^t

$$AA^t = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix}$$

Cherchons maintenant les valeurs propres et les vecteurs propres correspondants de AA^t . L'équation aux valeurs propres s'écrit :

$$\begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Leftrightarrow \begin{cases} (11 - \lambda)x_1 + x_2 = 0 \\ x_1 + (11 - \lambda)x_2 = 0 \end{cases} .$$

Cherchons les valeurs propres en égalisant le déterminant de la matrice des coefficients à 0 :

$$\begin{vmatrix} 11 - \lambda & 1 \\ 1 & 11 - \lambda \end{vmatrix} = 0 \Leftrightarrow (11 - \lambda)^2 - 1 = 0.$$

Ce qui peut s'écrire : $\lambda^2 - 22\lambda + 120 = 0$ ou encore $(\lambda - 12)(\lambda - 10) = 0 \Rightarrow \lambda = 10$ ou $\lambda = 12$.

Les valeurs propres de A sont $\lambda = 10$ et $\lambda = 12$.

En remplaçant les valeurs propres de A dans les équations de départ, nous pouvons calculer les vecteurs propres associés. Pour la valeur propre $\lambda = 10$, on peut écrire

$$(11 - 10)x_1 + x_2 = 0 \Leftrightarrow x_1 = -x_2$$

Ce qui est vraie pour un grand nombre de valeurs. On choisit $x_1 = 1$ et alors $x_2 = -1$.

On a donc $\lambda_1 = 10$ et $v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Pour la valeur propre $\lambda = 12$ on a :

$$(11 - 12)x_1 + x_2 = 0 \Leftrightarrow x_1 = x_2$$

donc en prenant $x_1 = 1$, on a $x_2 = 1$ et

$$\lambda_2 = 12 \text{ et } v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Les vecteurs v_1 et v_2 deviennent les colonnes d'une matrice ordonnée en fonction des valeurs propres décroissantes. On a donc

$$(v_2, v_1) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Il faut donc convertir cette matrice en matrice orthogonale. On le fait par une procédure d'orthogonalisation de Gram-Schmidt. Cette approche consiste à orthonormaliser (rendre orthogonal + normée) le premier vecteur et d'orthonormaliser ensuite les autres vecteurs en termes d'eux mêmes auxquels on soustrait une multiplication avec les vecteurs déjà normalisés. Plus précisément, si on a un ensemble de vecteurs $u_1, u_2, u_3, \dots, u_{\theta-1}$, alors on peut écrire :

$$w_\theta = v_\theta - \sum_{i=1}^{\theta-1} u_i \cdot v_\theta * u_i \quad \text{et} \quad u_\theta = \frac{w_\theta}{\|w_\theta\|},$$

où "*" est le produit d'un scalaire avec un vecteur et "." est le produit scalaire.

Par le procédé de Gram-Schmidt, on prend

$$u_1 = w_1, \quad v_1 = (1, 1)^t, \quad v_2 = (1, -1)^t$$

$$u_1 = \frac{w_1}{\|w_1\|} = \frac{v_1}{\|v_1\|} = \frac{(1, 1)^t}{\sqrt{1^2 + 1^2}} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)^t.$$

Dans les calculs, on peut être amené à prendre souvent $u_1^t = u_1$ pour raison de commodité.

$$w_2 = v_2 - u_1 \cdot v_2 * u_1 = (1, -1)^t - \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) (1, -1)^t * \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$$

$$= (1, -1)^t - 0 * \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) = (1, -1)^t$$

$$u_2 = \frac{w_2}{\|w_2\|} = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)^t$$

On a donc pour

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Le calcul de V est similaire et se fait sur $A^t A$

$$A^t A = \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix}.$$

Les valeurs propres de $A^t A$ se calculent avec

$$\begin{cases} 10x_1 + 0x_2 + 2x_3 = \lambda x_1 \\ 0x_1 + 10x_2 + 4x_3 = \lambda x_2 \\ 2x_1 + 4x_2 + 2x_3 = \lambda x_3 \end{cases} \Leftrightarrow \begin{cases} (10 - \lambda)x_1 + 0x_2 + 2x_3 = 0 \\ 0x_1 + (10 - \lambda)x_2 + 4x_3 = 0 \\ 2x_1 + 4x_2 + (2 - \lambda)x_3 = 0 \end{cases}$$

Si on fixe le déterminant on a :

$$\lambda(\lambda - 10)(\lambda - 12) = 0 \Rightarrow \lambda = 0 \text{ ou } \lambda = 10 \text{ ou } \lambda = 12.$$

Donc les valeurs propres de $A^t A$ sont : $\lambda_1 = 12$, $\lambda_2 = 10$ et $\lambda_3 = 0$. Cherchons les vecteurs propres associés.

Pour $\lambda_1 = 12$ on peut écrire $(10 - 12)x_1 + 2x_3 = 0 \Leftrightarrow -2x_1 + 2x_3 = 0$

En choisissant $x_1 = 1$ et $x_3 = 1$ on trouve $x_2 = 2$ et donc

$$\lambda_1 = 12, \quad v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

On procède de même sur λ_2 et λ_3 et on trouve :

$$\lambda_2 = 10, \quad v_2 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \text{ et } \lambda_3 = 0, \quad v_3 = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}.$$

On utilise les v_i comme les vecteurs colonnes d'une matrice ordonnés par λ_i décroissants :

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & -5 \end{pmatrix}$$

On orthonormalise par Gram-Schmidt,

$$u_1 = \frac{w_1}{\|w_1\|} = \frac{v_1}{\|v_1\|} = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right)^t$$

et

$$w_2 = v_2 - u_1 \cdot v_2 * u_1 = (2, -1, 0)^t \Rightarrow u_2 = \left(\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}, 0 \right)^t$$

$$w_3 = v_3 - u_1 \cdot v_3 * u_1 - u_2 \cdot v_3 * u_2 = (1, 2, -5)^t \Rightarrow u_3 = \left(\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, -\frac{5}{\sqrt{30}} \right)^t$$

et donc

$$V = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{5}} & \frac{\sqrt{30}}{2} \\ \frac{\sqrt{6}}{1} & -\sqrt{5} & \frac{\sqrt{30}}{5} \\ \frac{1}{\sqrt{6}} & 0 & -\frac{\sqrt{30}}{\sqrt{30}} \end{pmatrix} \Rightarrow V^t = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{\sqrt{6}}{2} & -\frac{\sqrt{6}}{1} & 0 \\ \frac{\sqrt{5}}{1} & \frac{\sqrt{5}}{2} & \frac{5}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & -\frac{5}{\sqrt{30}} \end{pmatrix}.$$

Pour la matrice Σ , on prend la racine carrée des valeurs propres non-nulles et on les place sur la diagonale en ordre

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_p \geq 0.$$

Les valeurs propres de U et V sont toujours les mêmes et peu importe de quelle matrice U ou V on les prend.

$$\Sigma = \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix}$$

D'où

$$A = U\Sigma V^t = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{\sqrt{6}}{2} & -\frac{\sqrt{6}}{1} & 0 \\ \frac{\sqrt{5}}{1} & \frac{\sqrt{5}}{2} & \frac{5}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & -\frac{5}{\sqrt{30}} \end{pmatrix} = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}.$$

3.3.2 Matrice de Heisenberg

Définition 3.3.1. Une matrice $H_m = [h_{i,j}, 1 \leq i, j \leq m]$ est dite matrice de Heisenberg supérieure si $h_{i,j} = 0$ pour $j + 1 < i$, i.e.,

$$H_m = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdot & \cdot & \cdot & h_{1,m-1} & h_{1,m} \\ h_{2,1} & h_{2,2} & \cdot & \cdot & \cdot & h_{2,m-1} & h_{2,m} \\ 0 & h_{3,2} & \cdot & \cdot & \cdot & h_{3,m-1} & h_{3,m} \\ 0 & 0 & h_{4,3} & \cdot & \cdot & h_{4,m-1} & h_{4,m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & h_{m,m-1} & h_{m,m} \end{pmatrix}$$

Si $h_{j+1,j} \neq 0$ avec $j = 1, 2, 3, \dots, m-1$ alors H_m est dite irréductible.

3.4 Evaluation numérique de $f(A)v$

3.4.1 Introduction

Etant donné une matrice $A \in M_n(\mathbb{C})$ et un vecteur $v \in \mathbb{C}^n$, nous menons l'étude sur des systèmes de la forme :

$$u = f(A)v, \quad (3.4.9)$$

c'est-à-dire, l'action de $f(A)$ sur un vecteur v , A une matrice symétrique positive (ou définie positive) de grande taille et f une fonction suffisamment régulière (ou analytique) sur le spectre de A . Nous voulons calculer le produit $f(A)v$ où f est une fonction de matrice définie sur le spectre de A . Il est possible de calculer $f(A)$ avec les trois définitions vues dans le chapitre précédent :

- la forme canonique de Jordan;
- l'interpolation polynômiale;
- l'intégrale de Cauchy.

Mais si A est de très grande taille il n'est pas toujours évident de pouvoir calculer $f(A)$, puis le produit $f(A)v$. Nous proposons donc une méthode d'approximation numérique pour l'approximation de $f(A)$ basée sur les sous espaces de Krylov de A et de v . L'avantage de cette méthode est qu'elle nécessite seulement la connaissance de A pour le calcul $f(A)v$. Pour certaines fonctions telles que la fonction exponentielle et la fonction inverse, elle converge simplement.

3.4.2 Méthodes de projection de Krylov

La résolution de systèmes linéaires repose, dans la plupart des codes de calcul, sur l'utilisation de méthodes de projection sur des sous-espaces de Krylov dont les représentants les plus connus sont la méthode de Lanczos pour les matrices symétriques définies positives et la méthode d'Arnoldi. Ceci est dû à leur efficacité sur une grande variété de problèmes.

Définition 3.4.1. (Sous espaces de Krylov) Le m -sous espace de Krylov de $A \in M_n(\mathbb{C})$ et $v \in \mathbb{C}^n$, $v \neq 0$ est défini par :

$$K_m = K_m(A, v) = \text{vect}\{v, Av, A^2v, \dots, A^{m-1}v\} = \{p(A)v, p \in P_{m-1}\}, \quad (3.4.10)$$

où $\text{vect}\{\cdot\}$ désigne l'ensemble des combinaisons linéaires des éléments qui se trouvent entre accolade et P_{m-1} l'ensemble des polynômes de degré inférieur ou égal à $m-1$.

Lemme 3.4.1. Le polynôme minimal de A par rapport à $v \in \mathbb{C}^n$ est donné par :

$$\psi_{A,v}(t) = \prod_{i=1}^s (t - \lambda_i)^{l_i} \in P_L, \text{ où } L = \text{deg}(\psi_{A,v}) = \sum_{i=1}^s l_i, \quad 0 \leq l_i \leq n_i,$$

l_i dépend de v aussi bien que de A et $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_s$ les valeurs propres distinctes de A , n_i étant la dimension du plus grand bloc de la matrice de Jordan dans le quel λ_i apparaît (l'ordre de multiplicité de λ_i).

Preuve. Montrons d'abord l'unicité de $\psi_{A,v}$. Supposons qu'il existe un autre polynôme minimal $\chi_{A,v}$ de A par rapport à v .

Par définition on a : $\text{deg}(\psi_{A,v}) = \text{deg}(\chi_{A,v})$ et par conséquent on peut poser $p = \alpha(\chi_{A,v} - \psi_{A,v})$ qui est polynôme unitaire avec $\alpha \in \mathbb{C}^*$. Donc $p(A)v = 0$ et $\text{deg}(p) < \text{deg}(\psi_{A,v})$, ce qui est contradictoire car $\psi_{A,v}$ a le plus bas degré.

Supposons maintenant que $\psi_{A,v}$ contienne un facteur de la forme $(t - \tilde{\lambda}_i)$, où $\tilde{\lambda}_i \notin \text{Sp}(A)$. Donc $(A - \tilde{\lambda}_i I)$ est inversible et par définition de $\psi_{A,v}$, on a :

$$(A - \tilde{\lambda}_i I)^{-1} \psi_{A,v}(A)v = 0.$$

D'autre part, $\text{deg}((t - \tilde{\lambda}_i)^{-1} \psi_{A,v}) < \text{deg}(\psi_{A,v})$. Ce qui est contradictoire avec $\psi_{A,v}$ minimal.

On a donc :

$$\psi_{A,v}(t) = \prod_{i=1}^s (t - \lambda_i)^{l_i}.$$

■

Remarque 3.4.1. On a $\text{deg}(\psi_{A,v}) \leq \text{deg}(\psi_A)$

Théorème 3.4.1. (cf.[77]) Pour tous polynômes p et q et une matrice $A \in M_n(\mathbb{C})$,
 $p(A)v = q(A)v$ si et seulement si

$$p^{(j)}(\lambda_i) = q^{(j)}(\lambda_i), \quad j = 0 : l_i - 1, \quad i = 1 : s \text{ avec } 0 \leq l_i \leq n_i.$$

Théorème 3.4.2. (cf. [62]) Soient $A \in M_n(\mathbb{C})$, $\psi_{A,v}$ le polynôme minimal de A par rapport à $v \in \mathbb{C}^n$. Si f est une fonction définie sur le $Sp(A)$, alors $f(A)v = q(A)v$, où q est l'unique polynôme d'interpolation d'Hermite de degré

$$\deg(q) \leq \deg(\psi_{A,v}) = \sum_{i=1}^s l_i$$

et satisfait aux conditions d'interpolation de Hermite :

$$q^{(j)}(\lambda_i) = f^{(j)}(\lambda_i), \quad j = 0 : l_i - 1, \quad i = 1 : s.$$

D'après le théorème 3.4.2, on a : $f(A)v \in K_L(A, v)$ où l'indice

$$L = L(A, v) = \deg(\psi_{A,v}) \in \mathbb{N}$$

est le plus petit nombre pour lequel $K_L(A, v) = K_{L+1}(A, v)$.

Proposition 3.4.1. Si $A^m v \in K_m$ alors $A^{m+r} v \in K_m$, pour tout $r > 0$.

Preuve. Supposons $A^m v \in K_m$. Raisonnons par récurrence.

Par définition, $K_m = K_m(A, v) = \text{vect}\{v, Av, A^2v, \dots, A^{m-1}v\}$ et

$$A^m v \in K_m \Rightarrow A^m v = \sum_{k=0}^{m-1} \alpha_k A^k v, \text{ avec } A^0 = I.$$

Pour $r = 1$, on a

$$\begin{aligned}
A^{m+1}v &= A(A^m v) \\
&= A\left(\sum_{k=0}^{m-1} \alpha_k A^k v\right) = \sum_{k=0}^{m-1} \alpha_k A^{k+1} v \\
&= \sum_{k=0}^{m-2} \alpha_k A^{k+1} v + \alpha_{m-1} A^m v \\
&= \sum_{k=0}^{m-2} \alpha_k A^{k+1} v + \alpha_{m-1} \sum_{k=0}^{m-1} \alpha_k A^k v \\
&= \sum_{k=0}^{m-2} \alpha_k A^{k+1} v + \alpha_{m-1} \alpha_0 v + \alpha_{m-1} \sum_{k=1}^{m-1} \alpha_k A^k v \\
&= \sum_{k=0}^{m-2} \alpha_k A^{k+1} v + \alpha_{m-1} \alpha_0 v + \alpha_{m-1} \sum_{k=0}^{m-2} \alpha_{k+1} A^{k+1} v \\
&= \alpha_{m-1} \alpha_0 v + \sum_{k=0}^{m-2} (\alpha_k + \alpha_{m-1} \alpha_{k+1}) A^{k+1} v \\
&= \alpha_{m-1} \alpha_0 v + \sum_{k=1}^{m-1} (\alpha_{k-1} + \alpha_{m-1} \alpha_k) A^k v.
\end{aligned}$$

Posons $\beta_0 = \alpha_{m-1} \alpha_0$, $\beta_k = \alpha_{k-1} + \alpha_{m-1} \alpha_k$, $k = 1 : m-1$.

On a :

$$A^{m+1}v = \sum_{k=0}^{m-1} \beta_k A^k v \in K_m.$$

Donc pour $r = 1$, la relation est vraie.

Supposons que pour $r \geq 1$, $A^{m+r}v \in K_m$ et montrons que $A^{m+r+1}v \in K_m$.

$$A^{m+r}v \in K_m \Rightarrow A^{m+r}v = \sum_{k=0}^{m-1} \gamma_k A^k v$$

et donc

$$\begin{aligned}
A^{m+r+1}v &= A(A^{m+r}v) \\
&= \sum_{k=0}^{m-1} \gamma_k A^{k+1}v \\
&= \sum_{k=0}^{m-2} \gamma_k A^{k+1}v + \gamma_{m-1} A^m v \\
&= \sum_{k=0}^{m-2} \gamma_k A^{k+1}v + \gamma_{m-1} \sum_{k=0}^{m-1} \gamma_k A^k v \\
&= \sum_{k=0}^{m-2} \gamma_k A^{k+1}v + \gamma_{m-1} \gamma_0 v + \gamma_{m-1} \sum_{k=1}^{m-1} \gamma_k A^k v \\
&= \sum_{k=0}^{m-2} \gamma_k A^{k+1}v + \gamma_{m-1} \gamma_0 v + \gamma_{m-1} \sum_{k=0}^{m-2} \gamma_{k+1} A^{k+1}v \\
&= \gamma_{m-1} \gamma_0 v + \sum_{k=0}^{m-2} (\gamma_k + \gamma_{m-1} \gamma_{k+1}) A^{k+1}v \\
&= \gamma_{m-1} \gamma_0 v + \sum_{k=1}^{m-1} (\gamma_{k-1} + \gamma_{m-1} \gamma_k) A^k v.
\end{aligned}$$

En posant $\sigma_0 = \gamma_{m-1} \gamma_0$ et $\sigma_k = \gamma_{k-1} + \gamma_{m-1} \gamma_k$, On obtient :

$$A^{m+r+1}v = \sum_{k=0}^{m-1} \sigma_k A^k v, \quad \text{et par suite} \quad A^{m+r+1}v \in K_m.$$

En conclusion, $A^{m+r}v \in K_m$, pour tout $r > 0$. ■

3.4.3 Principe de fonctionnement

Soit A la matrice (n, n) du système à résoudre, v un vecteur quelconque, on note : $K_m = K_m(A, v) = \text{vect}\{v, Av, A^2v, \dots, A^{m-1}v\}$ le sous-espace de Krylov engendré par A et v .

V_m est la matrice des vecteurs d'une base de $K_m(A, v)$, $\{v_1, \dots, v_m\}$.

Construction de V_m

Pour la construction de la base $\{v_1, \dots, v_m\}$ de $K_m(A, v)$, on utilise l'algorithme d'Arnoldi ou l'algorithme de Lanczos si la matrice A est symétrique.

Soit $A = VHV^*$, où V est orthogonale. Alors $V^*AV = H$ (matrice Heisenberg). Posons $V = [v_1, v_2, \dots, v_m]$, on a : $AV = VH$ alors

$$Av_j = \sum_{i=1}^{j+1} h_{i,j} v_i \quad j = 1 : m-1.$$

Ce qui peut s'écrire

$$Av_j = \sum_{i=1}^j h_{i,j} v_i + h_{j+1,j} v_{j+1} \Leftrightarrow h_{j+1,j} v_{j+1} = Av_j - \sum_{i=1}^j h_{i,j} v_i = w_j.$$

Comme V est unitaire, $h_{i,j} = v_i^* Av_j$, $i = 1 : j$.

Supposons que $w_j \neq 0$, $v_{j+1} = w_j / h_{j+1,j}$ avec $h_{j+1,j} = \|w_j\|_2$.

Comme $Av_j = \sum_{i=1}^{j+1} h_{i,j} v_i$ on a :

$$\text{vect}\{v_1, v_2, \dots, v_j\} = \text{vect}\{v_1, Av_1, \dots, A^{j-1} v_1\}.$$

Donc cette procédure produit une séquence de vecteurs orthonormaux v_1, v_2, \dots, v_m tels que pour chaque j , les vecteurs v_1, v_2, \dots, v_j constituent des sous-espaces de Krylov. Ces vecteurs constituent la base orthonormale d'Arnoldi $\{v_i\}_{i=1}^j$ du sous espace de Krylov $K_m(A, v_1)$.

D'après la méthode d'Arnoldi, on a, à l'étape m , la factorisation suivante :

$$AV_m = V_{m+1} H_{m+1,m} = V_m H_m + h_{m+1,m} v_{m+1} e_m^T.$$

3.5 Approximation de $f(A)v$ par la méthode des sous espaces de Krylov

Considérons $f_m = p(A)v \in K_m$ l'approximation de $f(A)v$. Pour chercher cette approximation, on utilise la méthode d'Arnoldi, basée sur la décomposition de K_m .

Soit

$$K_m = K_m(A, v) = \text{vect}\{v, Av, A^2v, \dots, A^{m-1}v\}.$$

Les vecteurs colonnes de ces matrices ne sont pas orthogonaux mais nous pouvons extraire une base orthogonale par la méthode d'orthogonalisation de Gram-Schmidt.

Mais il se trouve que cette méthode n'est pas toujours stable. D'où l'utilisation de l'algorithme d'Arnoldi qui permet de garantir la stabilité de la méthode de Gram-Schmidt et qui consiste à calculer récursivement la mise de A sous la représentation unitairement semblable H , H étant de type Heisenberg.

Nous nous plaçons dans le cas où $\dim(K_m(A, v)) = m$.

3.5.1 Procédure d'Arnoldi : Algorithme

La méthode d'Arnoldi est une méthode de projection orthogonale sur un sous-espace de Krylov permettant de construire, pour toute matrice A , une base orthonormale du sous-espace et une matrice de Heisenberg. La construction est basée sur la multiplication successive d'un vecteur initial v par la matrice A et l'orthonormalisation par une procédure de Gram-Schmidt modifiée appliquée aux vecteurs obtenus, selon le schéma suivant :

$$\begin{aligned} v_1 = \frac{v}{\|v\|}, & \longrightarrow v_2 = Av_1 \quad \rightsquigarrow v_2 = \frac{v_2}{\|v_2\|}, \\ \dots & \longrightarrow v_m = Av_{m-1} \quad \rightsquigarrow v_m = \frac{v_m}{\|v_m\|}. \end{aligned} \quad (3.5.11)$$

Soient $A \in M_n(\mathbb{C})$, $v_1 \in \mathbb{C}^n$. Cet Algorithme calcule la factorisation de A sous la forme $AV = VH$, où $V \in M_{n,m}(\mathbb{C})$ ($m = \deg(\psi_{A,v_1}) \leq n$) de colonnes orthonormales et $H \in M_m(\mathbb{C})$ Heisenberg supérieur. C'est-à-dire il calcule une base orthogonale d'un espace de Krylov engendré par $(v, Av, A^2v, \dots, A^{m-1}v)$ par la méthode d'Arnoldi et renvoie cette base stockée en vecteurs colonnes dans V et la matrice Heisenberg des coefficients d'orthogonalisation dans H . On Choisit un vecteur v_1

Algorithme d'Arnoldi

- 1: Données A, v, m
- 2: Résultats V_m, H_m
- 3: $v_1 = \frac{v}{\|v\|_2} \longrightarrow$ initialisation
- 4: **for** $j = 1 : m$ **do**
- 5: $w := Av_j$

```

6:  for  $i = 1 : j$  do
7:     $h_{i,j} := v_i^* w$ 
8:     $w := w - h_{i,j} v_i$      $\rightarrow$  orthogonalisation de  $w$ ,  $h_{i,j} = (Av_j, v_i)$ ,  $i \leq j$ 
9:  end for
10:  $h_{j+1,j} := \|w\|_2$      $\rightarrow$      $h_{j+1,j} = 0$  si  $j = m$ 
11: if  $h_{j+1,j} > 0$  then
12:    $v_{j+1} := \frac{w}{h_{j+1,j}}$ 
13: else
14:    $v_{j+1} := 0$ 
15: end if
16: end for

```

A la sortie on a $V_m = [v_1, \dots, v_m]$, v_{m+1} , $H_m = (h_{i,j})$ $1 \leq i, j \leq m$, $h_{m+1,m}$, où les $(v_i)_{i=1,2,\dots,j}$ sont les vecteurs du système orthonomé trouvé à l'étape $m-1$. Les vecteurs sont calculés récursivement.

La matrice V_m est alors une base d'un sous espace invariant associé à A . On obtient donc une base orthonormée de $K_m(A, v)$.

Les $h_{i,j}$ sont calculées par l'Algorithme et donne une matrice de type Heisenberg supérieure.

On a ensuite $H_m = V_m^* A V_m$. La matrice H_m est donc la représentation de la base formée par les vecteurs d'Arnoldi de la projection orthogonale de A sur les sous-espaces de Krylov.

$$H_m = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdot & \cdot & \cdot & h_{1,m-1} & h_{1,m} \\ h_{2,1} & h_{2,2} & \cdot & \cdot & \cdot & h_{2,m-1} & h_{2,m} \\ 0 & h_{3,2} & \cdot & \cdot & \cdot & h_{3,m-1} & h_{3,m} \\ 0 & 0 & h_{4,3} & \cdot & \cdot & h_{4,m-1} & h_{4,m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & h_{m,m-1} & h_{m,m} \end{pmatrix}. \quad (3.5.12)$$

Le problème a une solution si seulement si l'itération d'Arnoldi s'arrête. L'Algorithme d'Arnoldi s'arrête lorsque $h_{m+1,m} = 0$.

L'intérêt de la méthode d'Arnoldi pour les grandes matrices ($n > 10^3$) est de pouvoir

stopper le processus à l'étape $m \times n$ et de travailler avec H de taille $m \times m$ ou $(m+1) \times m$.
L'écriture matricielle du procédé d'Arnoldi conduit à la relation :

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T, \quad (3.5.13)$$

ou encore

$$AV_m = V_{m+1} \tilde{H}_{m+1,m}. \quad (3.5.14)$$

Soit

$$A[v_1, v_2, \dots, v_m] = [v_1, v_2, \dots, v_m, v_{m+1}] \tilde{H}_{m+1,m}.$$

Avec

$$\tilde{H}_{m+1,m} = \begin{pmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,m-1} & h_{1,m} \\ h_{2,1} & h_{2,2} & \dots & h_{2,m-1} & h_{2,m} \\ 0 & h_{3,2} & \dots & h_{3,m-1} & h_{3,m} \\ 0 & 0 & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & h_{m,m-1} & h_{m,m} \\ 0 & 0 & \dots & \cdot & h_{m+1,m} \end{pmatrix} \quad (3.5.15)$$

$$= \begin{pmatrix} (Av_1, v_1) & (Av_2, v_1) & \dots & \dots & (Av_m, v_1) \\ (Av_1, v_2) & (Av_2, v_2) & \dots & \dots & (Av_m, v_2) \\ 0 & (Av_2, v_3) & \dots & \dots & (Av_m, v_3) \\ 0 & 0 & \dots & \dots & \cdot \\ \cdot & \cdot & \dots & \dots & \cdot \\ \cdot & \cdot & \dots & \dots & (Av_m, v_m) \\ 0 & 0 & \dots & \dots & (Av_m, v_{m+1}) \end{pmatrix} \quad (3.5.16)$$

Nous définissons $H_m = [h_{i,j}, 1 \leq i, j \leq m]$ en délogant le dernier rang de $\tilde{H}_{m+1,m}$ une matrice Heisenberg supérieure. Donc

$$A[v_1, v_2, \dots, v_m] = [v_1, v_2, \dots, v_m, v_{m+1}] \tilde{H}_{m+1,m}$$

peut s'écrire comme

$$A[v_1, v_2, \dots, v_m] = [v_1, v_2, \dots, v_m, v_{m+1}] H_m + h_{m+1,m} v_{m+1} e_m^T.$$

En remplaçant $V_m = [v_1, \dots, v_m]$ on a :

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T$$

où $H_m = (h_{i,j})$ et $e_m = [0, 0, \dots, 0, 1]^T$, $e_m \in \mathbb{R}^m$ vecteur dont la $m^{i\text{eme}}$ composante est 1.

Corollaire 3.5.1. Si $m = n$, alors $AV_n = V_n H_n$ et H_n est unitairement semblable à A .

Remarque 3.5.1. On dit que H_m est la **compression (ou projection)** de A sur K_m . Elle représente le calcul de A sur K_m par rapport à la base d'Arnoldi $\{v_i\}_{i=1}^m$ avec $j = 1, 2, 3, \dots, m$. On peut poser

$$\text{vect}\{v_1, v_2, \dots, v_j\} = \text{vect}\{v_1, Av_1, \dots, A^{j-1}v_1\},$$

car la procédure d'Arnoldi produit une séquence de vecteurs orthonormaux v_1, v_2, \dots, v_m tels que pour chaque n , les vecteurs v_1, v_2, \dots, v_n constituent les sous-espaces de Krylov. Ces vecteurs constituent la base orthonormale d'Arnoldi $\{v_i\}_{i=1}^m$ du sous espace de Krylov $K_m(A, v_1)$.

3.5.2 Approximation d'Arnoldi de $f(A)v$

Trouvons l'approximation de $f(A)v$ par le procédé d'Arnoldi en prenant

$$v_1 = \frac{v}{\|v\|_2}.$$

Lemme 3.5.1. Pour $m < n$ fixé et

$$p(z) = a_m z^m + a_{m-1} z^{m-1} + \dots + a_1 z + a_0 \in P_m,$$

avec $AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T$ on a :

$$p(A)v = \|v\| V_m p(H_m) e_1 + \|v\| a_m \alpha_m v_{m+1}, \quad \text{avec } \alpha_m = \prod_{j=1}^m h_{j+1,j}. \quad (3.5.17)$$

En particulier, pour tout $p \in P_{m-1}$ on a :

$$p(A)v = \|v\| V_m p(H_m) e_1. \quad (3.5.18)$$

Preuve. Il est suffisant de montrer ces résultats pour $p(z) = z^j$, i.e.,

$$A^j v = \|v\| V_m H_m^j e_1 \quad \text{pour } j < m$$

et

$$A^m v = \|v\| V_m H_m^m e_1 + \|v\| \alpha_m v_{m+1}.$$

Par construction de la base d'Arnoldi on a : $v = \|v\|V_m e_1$.

Montrons que $A^j v = \|v\|V_m H_m^j e_1$. Raisonnons par récurrence.

Pour $j = 0$, on a $A^0 v = \|v\|V_m H_m^0 e_1 \Leftrightarrow v = \|v\|V_m e_1$ vraie.

Supposons la relation est vraie à l'ordre $k - 1$ et montrons qu'elle est aussi vraie à l'ordre k .

$$\begin{aligned} A^k v &= A(A^{k-1} v) = A(\|v\|V_m H_m^{k-1} e_1) = \|v\|(AV_m)H_m^{k-1} e_1 \\ &= \|v\|(V_m H_m + h_{m+1,m} v_{m+1} e_m^t) H_m^{k-1} e_1 \\ A^k v &= \|v\|V_m H_m^k e_1 + \|v\|h_{m+1,m} v_{m+1} e_m^t H_m^{k-1} e_1 \\ &= \|v\|V_m H_m^k e_1 + \|v\|h_{m+1,m} (e_m^t H_m^{k-1} e_1) v_{m+1}. \end{aligned}$$

En posant $\alpha_m = h_{m+1,m} (e_m^t H_m^{k-1} e_1)$ on a :

$$A^k v = \|v\|V_m H_m^k e_1 + \|v\|\alpha_m v_{m+1}.$$

Or

$$e_m^t H_m^{k-1} e_1 = \begin{cases} 0, & k < m; \\ \prod_{j=1}^{m-1} h_{j+1,j}, & k = m. \end{cases}$$

(cf.[77])

Pour $k < m$, $A^k v = \|v\|V_m H_m^k e_1$.

Pour $k = m$, $A^m v = \|v\|V_m H_m^m e_1 + \|v\|\alpha_m v_{m+1}$. ■

Définition 3.5.1. (Approximation d'Arnoldi) Soit f une fonction définie pour H_m . L'approximation d'Arnoldi d'ordre m de $f(A)v$ est définie comme

$$\mathbf{f}_m = \|v\|V_m \mathbf{f}(H_m) e_1. \quad (3.5.19)$$

L'Algorithme d'obtention de cette approximation f_m peut être résumé par :

Algorithme. Données : $A \in M_n(\mathbb{C})$, $v \in \mathbb{C}^n$ et $m < n$.

Résultat : L'approximation d'Arnoldi f_m .

1. On détermine V_m, H_m utilisées dans la procédure d'Arnoldi.

2. On pose $f_m = \|v\|V_m f(H_m)e_1$.

Remarque 3.5.2. 1. La dimension du problème qui était de n est réduite à m ,

$f(A)v \simeq \|v\|V_m f(H_m)e_1 = f_m$, V_m est la base orthonormale de Krylov et
 $v = \|v\|V_m e_1$.

2. On n'a besoin que de la première colonne de $f(H_m)$ car on prend le produit $f(H_m)e_1$ sinon le calcul de $f(H_m)$ demeure toujours un petit problème si on ne connaît pas les valeurs propres de H_m .

3. Même si nous supposons que f est définie sur le $Sp(A)$, H_m peut avoir des valeurs propres en des points où f n'est pas définie. Dans ce cas $f(H_m)$ n'est pas défini et par conséquent on ne pourra pas déterminer l'approximation d'Arnoldi d'ordre m .

Conclusion Dans le cas de données exactes on pose l'approximation de $u = f(A)v$ par la projection sur les sous-espaces de Krylov, on a :

— Pour $m = n$, La méthode d'Arnoldi nous fournit une formule très pratique pour son calcul en posant $v_1 = \frac{v}{\|v\|}$ et en calculant

$$f(A)v = V_n f(H_n) V_n^* v = \|v\| V_n f(H_n) e_1 \quad (3.5.20)$$

avec $v = \|v\|V_n e_1$.

En effet :

D'après la définition de $f(A)$ par la formule d'intégrale de Cauchy on a :

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz$$

où f est une fonction analytique dans un domaine D , tel que Γ entoure D avec $\Gamma \cap D = \emptyset$.

On a

$$f(A)v = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} v dz.$$

D'autre part

$$zI - H_m = zI - V_m^t A V_m = z(V_m^t V_m) - V_m^t A V_m = V_m^t (zV_m - A V_m) = V_m^t (zI - A) V_m,$$

et

$$(zI - H_m)^{-1} = V_m^{-1}(zI - A)^{-1}(V_m^t)^{-1} = V_m^t(zI - A)^{-1}V_m.$$

Donc

$$f(H_m) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - H_m)^{-1} dz = \frac{1}{2\pi i} \int_{\Gamma} f(z)(V_m^t(zI - A)^{-1}V_m) dz.$$

Par suite

$$\begin{aligned} \|v\|V_m f(H_m)e_1 &= \frac{1}{2\pi i} \int_{\Gamma} f(z)\|v\|(V_m V_m^t(zI - A)^{-1}V_m^t e_1) dz \\ &= \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1}(\|v\|V_m^t e_1) dz = f(A)v \end{aligned}$$

— Pour $m \ll n$, on le projette sur un sous espace de Krylov de dimension $m \ll n$, pour obtenir $f(A)v \simeq \|v\|V_m f(H_m)e_1$ car

$$f(A)v = f(A)V_k V_k^t v \simeq V_k V_k^t f(A)V_k V_k^t v \simeq V_k f(V_k^t A V_k)V_k^t v = V_k \|v\| f(H_k)e_k.$$

L'Approximation de Krylov est de la forme

$$f_k = \|v\|V_k f(H_k)e_1. \quad (3.5.21)$$

Exemple 3.5.1. Déterminons la matrice V_m , H_m et f_m pour la matrice

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 6 & 7 & 3 \\ 12 & 20 & 13 & 11 & 24 & 0 \\ 1 & 0 & 0 & 4 & 6 & 8 \\ 6 & 3 & 2 & 1 & 9 & 4 \\ 8 & 7 & 6 & 5 & 4 & 1 \end{pmatrix}, \quad v = \begin{pmatrix} 2 \\ 1 \\ 3 \\ 4 \\ 6 \\ 7 \end{pmatrix}$$

et $f(x) = e^x$. On prend $n = 6$ et $m = 1 : 5$.

On utilise l'Algorithme d'Arnoldi avec

$$f_m = \|v\|V_m f(H_m)e_1 = \beta V_m f(H_m)e_1$$

On a :

$$V_5 = \begin{pmatrix} 0.1865 & 0.2168 & -0.3203 & 0.8897 & 0.1351 & -0.0759 \\ 0.0933 & 0.3421 & -0.1803 & -0.0251 & -0.5124 & 0.7607 \\ 0.2798 & 0.8283 & 0.3631 & -0.1734 & 0.1679 & -0.2133 \\ 0.3730 & 0.0462 & -0.6561 & -0.3133 & -0.3412 & -0.4622 \\ 0.5595 & -0.1749 & -0.2034 & -0.2128 & 0.6462 & 0.3900 \\ 0.6528 & -0.3423 & 0.5110 & 0.1851 & -0.3963 & -0.0658 \end{pmatrix}$$

$$H_5 = \begin{pmatrix} 24.7565 & 9.7575 & -9.1742 & 3.7422 & 2.4835 & 0 \\ 22.1476 & 11.5986 & -11.3695 & -3.1973 & 4.1546 & 0 \\ 0 & 11.5242 & -6.8678 & 1.6609 & 1.2044 & 0 \\ 0 & 0 & 2.0322 & -0.6344 & -2.5060 & 0 \\ 0 & 0 & 0 & 1.5132 & 3.2580 & 0 \\ 0 & 0 & 0 & 0 & 1.1345 & 0 \end{pmatrix}$$

$$f(H_5) = \begin{pmatrix} 3.8448 \times 10^{13} & 1.1635 \times 10^{13} & -1.2502 \times 10^{13} & 0.2880 \times 10^{13} & 0.4310 \times 10^{13} & 0 \\ 3.6476 \times 10^{13} & 1.1038 \times 10^{13} & -1.1861 \times 10^{13} & 0.2733 \times 10^{13} & 0.4089 \times 10^{13} & 0 \\ 1.0999 \times 10^{13} & 0.3329 \times 10^{13} & -0.3576 \times 10^{13} & 0.0824 \times 10^{13} & 0.1233 \times 10^{13} & 0 \\ 0.0694 \times 10^{13} & 0.0210 \times 10^{13} & -0.0226 \times 10^{13} & 0.0052 \times 10^{13} & 0.0078 \times 10^{13} & 0 \\ 0.0037 \times 10^{13} & 0.0011 \times 10^{13} & -0.0012 \times 10^{13} & 0.0003 \times 10^{13} & 0.0004 \times 10^{13} & 0 \\ 0.0001 \times 10^{13} & 0.000041 \times 10^{13} & -0.000044 \times 10^{13} & 0.000010 \times 10^{13} & 0.000015 \times 10^{13} & 1 \end{pmatrix},$$

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad f_5 = \begin{pmatrix} 1.3058 \times 10^{14} \\ 1.5060 \times 10^{14} \\ 4.8095 \times 10^{14} \\ 0.9200 \times 10^{14} \\ 1.3696 \times 10^{14} \\ 1.9673 \times 10^{14} \end{pmatrix}$$

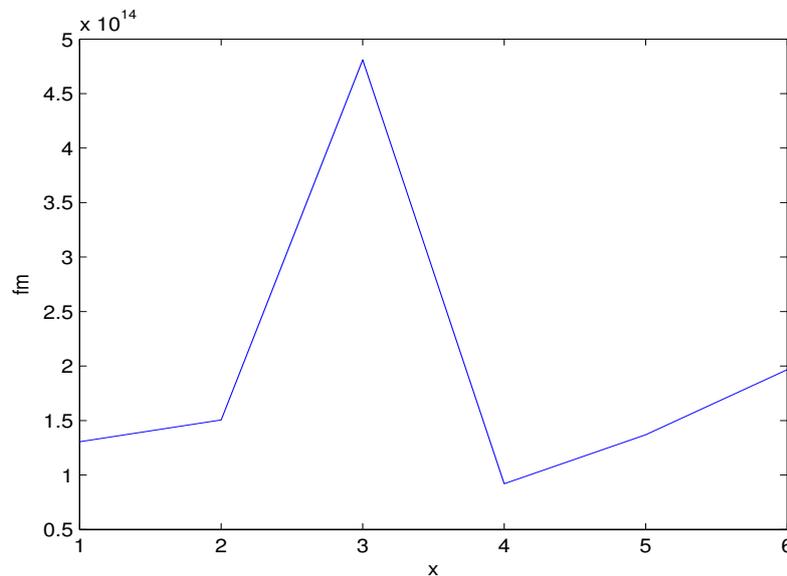


FIGURE 3.1: Représentation graphique de $f_m \simeq f(A)v = e^A v$.

La figure 2.1 montre l'Approximation de $f(A)v$ par l'Algorithme d'Arnoldi avec $f(x) = e^x$.

Tentons de trouver un moyen pour pouvoir déterminer $f(H_m)$ d'une façon plus générale.

3.5.3 Valeurs de Ritz

Soit $1 \leq m \leq n$. Les valeurs propres de la matrice H_m sont appelées valeurs *Ritz*(m) de A .

En pratique, le calcul des couples propres d'une matrice carrée A est un travail souvent fastidieux, c'est pour cela, il est commode de calculer les couples de Ritz qui peuvent être considérés comme des bonnes approximations à une certaine précision. Cette approche est bien justifiée théoriquement et numériquement.

Soient $(\lambda_i)_{i=1}^n$ les valeurs propres de la matrice A et $(\beta_i)_{i=1}^m$ les valeurs propres de H_m si $A = A^t$ d'après Golub et Van Loan [29] (voir aussi, [77]) on a la localisation suivante :

$$\lambda_{\min} \leq \beta_1 \leq \beta_2 \leq \dots \leq \beta_m \leq \lambda_{\max}$$

Chaque intervalle $(-\infty, \beta_1]$, $[\beta_1, \beta_2]$, $[\beta_2, \beta_3]$, \dots , $[\beta_{m-1}, \beta_m]$, $[\beta_m, +\infty)$ contient au moins une valeur propre de A . Ceci montre que les valeurs de Ritz $(\beta_i)_{i=1}^m$ sont des approximations des valeurs propres de A qui dépendent explicitement du choix du vecteur initial v dans la procédure d'Arnoldi (Lanczòs dans le cas symétrique). L'Approximation de $f(A)v$ avec la procédure d'Arnoldi en prenant $v_1 = \frac{v}{\|v\|_2}$ et pour $(m = \deg(\psi_{A,v_1}) \leq n)$ est : $f_m = \|v\|V_m f(H_m)e_1 = f(V_m H_m V_m^*)v = V_m f(H_m)V_m^*v$ qui est l'évaluation de $f(A)v$ sur le plus petit sous-espace de Krylov $K_m(A, v_1)$. Donc on peut l'étendre sur tout l'espace de Krylov $K_m(A, v_n)$ avec $n = 1 : m$.

Théorème 3.5.1. (cf. [96]) Soient les matrices V_m et H_m obtenues par la procédure d'Arnoldi sur $A \in M_n(\mathbb{C})$ et $v \in \mathbb{C}^n$. On a :

$$\|v\|V_m f(H_m)e_1 = \tilde{p}_{m-1}(A)v, \quad (3.5.22)$$

où \tilde{p}_{m-1} est l'unique polynôme d'interpolation de Hermite degré $m - 1$ qui interpole f sur $Sp(H_m)$.

4

Etude d'un problème inverse d'identification de sources du type parabolique

4.1 Problème de Dirichlet

4.1.1 Position du problème

Soient $(H, \langle \cdot, \cdot \rangle, \|\cdot\|)$ un espace de Hilbert séparable de dimension infinie, et $A : D(A) \subset H \rightarrow H$ un opérateur non borné à domaine dense dans H .

Considérons le problème parabolique suivant :

$$\begin{cases} u_t(t) + Au(t) = f, & 0 < t < T, \\ u(0) = 0, \end{cases} \quad (4.1.1)$$

où la source $f \in H$ est une fonction donnée indépendante de t .

On fait les hypothèses suivantes :

H1 L'opérateur A est auto-adjoint, i.e., $A = A^*$.

H2 L'opérateur A est strictement positif, i.e., $A \geq \gamma I$ où $0 < \gamma = \inf \sigma(A)$.

H3 L'injection $H^1 := (D(A), \|\cdot\|_G) \hookrightarrow H$ est compacte, où $\|\cdot\|_G$ est la norme du graphe.

• Sous les hypothèses (H1, H2 et H3) l'opérateur A est diagonalisable.^{1 2} Soit alors $(\xi_k, \lambda_k)_{k \geq 1}$ les couples propres (valeur propres, vecteurs propres) de l'opérateur A :

$$A\xi_k = \lambda_k \xi_k, \quad k = 1, 2, \dots, \quad \langle \xi_m, \xi_n \rangle = \delta_{mn},$$

$$0 < \gamma \leq \lambda_1 \leq \lambda_2 \leq \dots \lambda_k \rightarrow \infty, \quad k \rightarrow \infty,$$

$$\forall h \in H, \quad h = \sum_{k=1}^{\infty} h_k \xi_k, \quad h_k = \langle h, \xi_k \rangle \text{ le coefficient de Fourier d'ordre } k \text{ de } h.$$

Théorème 4.1.1. Pour tout $f \in H$, le problème 4.1.1 admet une solution unique $u \in C([0, T]; H)$ donnée par l'expression

$$u(t) = R(t)f = \int_0^t e^{-(t-s)A} f ds = (I - e^{-tA})A^{-1}f = \sum_{k=1}^{\infty} \left(\frac{1 - e^{-t\lambda_k}}{\lambda_k} \right) \langle f, \xi_k \rangle \xi_k. \quad (4.1.2)$$

Comme exemple d'application, on prend $H = L^2(0, 1)$ et

$$A = -\frac{d}{dx^2}, \quad D(A) = H_0^1(0, 1) \cap H^2(0, 1),$$

i.e., l'équation de la chaleur avec conditions de Dirichlet :

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = f(x), & 0 < x < 1, 0 < t < 1, \\ u(x, 0) = 0, & x \in [0, 1], \\ u(0, t) = u(1, t) = 0, & t \in [0, 1]. \end{cases} \quad (4.1.3)$$

Ici $f \in L^2(0, 1)$ est la source recherchée à partir de la condition finale $u(x, 1) = g(x) \in L^2(0, 1)$.

Il est facile de vérifier que A est auto-adjoint, strictement positif et à résolvante compacte, où les couples propres sont donnés par

$$\lambda_n = (n\pi)^2, \quad e_n(x) = \sqrt{2} \sin(n\pi x), \quad n = 1, 2, \dots$$

Notre problème inverse se formule comme suit :

-
1. H2 $\implies 0 \in \rho(A)$, H3 $\implies A^{-1}$ est compact.
 2. (H1, H2, H3) $\implies A^{-1}$ est auto-adjoint compact, et donc diagonalisable d'après le Théorème de la décomposition spectrale des opérateurs auto-adjoints compacts. En conséquence A est diagonalisable.

Soit $u(\tau; f) = g$ une observation finale de la solution $u(t; f)$ au point $t = \tau$, $0 < \tau \leq T$. Notre objectif est de déterminer la donnée f à partir de la condition g .

On peut schématiser notre problème inverse comme suit :

$$f \longrightarrow u(t) = R(t)f \longrightarrow \Theta(u) = u(\tau) = g, \quad (4.1.4)$$

on obtient donc une équation de Fredholm de première espèce

$$Kf = g, \quad K = R(\tau). \quad (4.1.5)$$

Il est clair que l'opérateur K est auto-adjoint, compact et injectif, et donc d'après le théorème de Picard, la solution recherchée f sera donnée par la formule

$$f = K^{-1}g = \left(\frac{A}{I - e^{-\tau A}} \right) g = \sum_{k=1}^{\infty} \left(\frac{\lambda_k}{1 - e^{-\tau \lambda_k}} \right) \langle g, \xi_k \rangle \xi_k, \quad (4.1.6)$$

sous la condition que

$$\sum_{k=1}^{\infty} \left(\frac{\lambda_k}{1 - e^{-\tau \lambda_k}} \right)^2 |\langle g, \xi_k \rangle|^2 < +\infty. \quad (4.1.7)$$

De l'inégalité

$$\lambda_k^2 \leq \left(\frac{\lambda_k}{1 - e^{-\tau \sqrt{\lambda_k}}} \right)^2 \leq \left(\frac{\lambda_k}{1 - e^{-\tau \gamma}} \right)^2,$$

on déduit que la condition (4.1.7) est équivalente à $g \in D(A)$.

On note ici que notre problème inverse est instable, ceci découle du comportement des hautes fréquences $\omega_k = \left(\frac{\lambda_k}{1 - e^{-\tau \lambda_k}} \right) \approx \lambda_k \longrightarrow \infty$ quand $k \longrightarrow \infty$.

4.1.2 Stabilisation et approximation

Nous voulons maintenant aborder la question de stabilisation dans le cas de données entachées de bruit, parce que les données dont on dispose viennent de l'expérimentation, ce qui implique l'existence d'erreurs de mesure. Cette cause d'incertitude induit une image floue à cause de la sensibilité des problèmes inverses aux incertitudes, et l'interprétation des réponses devient presque impossible et engendre un grand risque.

On suppose maintenant que la donnée $u(\tau) = g$ est entachée de bruit (inexacte), i.e., on dispose d'une approximation g_δ de g : $\|g_\delta - g\| \leq \delta$, δ est le niveau de bruit.

Régularisation par troncature spectrale

Définition 4.1.1. Pour $N > 0$, on définit la solution régularisée du problème (4.1.5) pour des données exactes (resp. inexactes) comme suit :

$$f_N = \sum_{k=1}^N \left(\frac{\lambda_k}{1 - e^{-\tau\lambda_k}} \right) \langle g, \xi_k \rangle \xi_k, \quad (4.1.8)$$

$$f_N^\delta = \sum_{k=1}^N \left(\frac{\lambda_k}{1 - e^{-\tau\lambda_k}} \right) \langle g_\delta, \xi_k \rangle \xi_k, \quad (4.1.9)$$

Cette méthode est connue sous le nom *troncature spectrale* qui élimine les hautes fréquences responsables de l'instabilité.

Théorème 4.1.2. On suppose que $f \in B(p, E) = \{f \in D(A^p) : \|A^p f\| \leq E\}$, $p > 0$, et soit $\lambda_{N+1} \approx \left(\frac{E}{\delta}\right)^{1/(2+p)}$, alors on a l'estimation d'erreur suivante

$$\|f - f_N^\delta\| \leq K \delta^{\frac{p}{p+2}} E^{\frac{2}{p+2}},$$

$$\text{où } K = (1 + M) = 1 + \frac{1}{1 - e^{-\tau\lambda_1}}.$$

Preuve. Notons

$$\omega_k = \frac{\lambda_k}{1 - e^{-\tau\lambda_k}} \leq \frac{\lambda_k}{1 - e^{-\tau\lambda_1}} = M \lambda_k,$$

$$g_k = \langle g, \xi_k \rangle, \quad g_k^\delta = \langle g_\delta, \xi_k \rangle.$$

En utilisant l'inégalité triangulaire, on peut écrire

$$\|f - f_N^\delta\| = \|f - f_N + f_N - f_N^\delta\| \leq \|f - f_N\| + \|f_N - f_N^\delta\| = \Delta_1 + \Delta_2. \quad (4.1.10)$$

$$\Delta_1^2 = \|f - f_N\|^2 = \left\| \sum_{k=1}^{\infty} f_k \xi_k - \sum_{k=1}^N f_k \xi_k \right\|^2 = \sum_{k=N+1}^{\infty} |f_k|^2. \quad (4.1.11)$$

$$\Delta_2^2 = \|f_N - f_N^\delta\|^2 = \left\| \sum_{k=1}^N \omega_k g_k \xi_k - \sum_{k=1}^N \omega_k g_k^\delta \xi_k \right\|^2 = \sum_{k=1}^{N+1} \omega_k^2 |g_k - g_k^\delta|^2. \quad (4.1.12)$$

$$\Delta_1^2 = \sum_{k=N+1}^{\infty} \lambda_k^{-2p} \lambda_k^{2p} |f_k|^2 \leq \lambda_{N+1}^{-2p} \sum_{k=N+1}^{\infty} \lambda_k^{2p} |f_k|^2 \leq \lambda_{N+1}^{-2p} E^2. \quad (4.1.13)$$

$$\Delta_2^2 = \sum_{k=1}^{N+1} \omega_k^2 |g_k - g_k^\delta|^2 \leq \lambda_{N+1}^2 M^2 \sum_{k=1}^{N+1} |g_k - g_k^\delta|^2 \leq \lambda_{N+1}^2 M^2 \delta^2. \quad (4.1.14)$$

Ce qui implique que

$$\Delta_1 + \Delta_2 \leq \lambda_{N+1}^{-2p} E + \lambda_{N+1}^2 M \delta \approx \left(\left(\frac{E}{\delta} \right)^{\frac{1}{2+p}} \right)^{-p} E + M \delta \left(\frac{E}{\delta} \right)^{\frac{2}{2+p}} = (1 + M) E^{\frac{2}{2+p}} \delta^{\frac{2p}{2+p}}.$$

4.1.3 Approximation numérique du problème par La méthode de Krylov

Discrétisation et projection de la solution

La discrétisation par éléments finis ou différences finies d'une EDP intervenant dans un problème inverse est une source d'erreur et son influence sur la convergence est certaine. Pour remédier à ce phénomène, on propose une **semi-discrétisation** du problème, ce qui va nous permettre d'une part de diminuer les erreurs, et d'autre part de transformer le problème semi-discrétisé en une EDO, dont l'étude repose en général sur des méthodes spectrales en dimension finie.

Par un calcul opérationnel, on transforme notre problème en un problème de calcul matriciel de la forme : $u = \Phi(A)v$, où $u :=$ est la solution recherchée, $v :=$ les données du problème, $A \in \mathbb{R}^{m \times m}$ est une matrice symétrique définie positive, et $\Phi :=$ une fonction définie et continue sur une partie contenant le spectre de A .

Notons $A_m \in \mathbb{R}^{m \times m}$ l'opérateur discret de A , qui est une matrice symétrique, définie positive. On suppose que les erreurs de la discrétisation sont petites par rapport à l'incertitude δ , et que A_m est de taille suffisamment grande de façon que A_m soit une bonne approximation de l'opérateur différentiel A .

Dans la suite, nous utilisons les notations suivantes :

$$E_m := \mathbb{R}^m, \quad m \geq 1.$$

$$G \in E_m := \text{l'analogue discret de } g.$$

$$G^\delta \in E_m := \text{l'analogue discret de } g_\delta.$$

$$\sigma(A_m) = \{\alpha_k\}_{k=1}^m := \text{l'ensemble des valeurs propres de } A_m \quad (0 < \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m).$$

$$\{\zeta_k\}_{k=1}^m := \text{l'ensemble des vecteurs propres normalisés de } A_m.$$

Les analogues discrets de (4.1.6) et (4.1.8) sont notés respectivement par :

$$F = \sum_{k=1}^m \frac{\alpha_k}{1 - e^{-\tau\alpha_k}} \langle G, \zeta_k \rangle \zeta_k, \quad (4.1.15)$$

$$F_n = \sum_{k=1}^n \frac{\alpha_k}{1 - e^{-\tau\alpha_k}} \langle G, \zeta_k \rangle \zeta_k, \quad (n \leq m). \quad (4.1.16)$$

$$F_n^\delta = \sum_{k=1}^n \frac{\alpha_k}{1 - e^{-\tau\alpha_k}} \langle G_\delta, \zeta_k \rangle \zeta_k, \quad (n \leq m). \quad (4.1.17)$$

La méthode de projection de Krylov consiste à générer l'espace de Krylov

$$\mathcal{K}_l(A_m, G) = \text{vect} \{G, A_m G, \dots, A_m^{l-1} G\}$$

par un algorithme itératif.

Soient $(q_i)_{i=1}^l$ la base orthonormée de $\mathcal{K}_l(A_m, G)$ avec $q_1 = \frac{G}{\|G\|}$, $Q_l = (q_1, q_2, \dots, q_l)$ et $T_l = Q_l^\top A_m Q_l \in \mathbb{R}^{l \times l}$ la représentation symétrique de A_m sur l'espace \mathcal{K}_l . L'approximation de (4.1.15) dans le triplet de Krylov $\mathbf{K}_l = (\mathcal{K}_l(A_m, G), Q_l, T_l)$ est donnée par :

$$[F]_l = \|G\| Q_l \left(\frac{T_l}{I - \exp(-\tau T_l)} \right) e_1 \in \mathbb{R}^m, \quad (4.1.18)$$

où $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^l$.

Remarque 4.1.1. L'approximation (4.1.18) converge vers la solution (4.1.15), et le type de convergence est super-linéaire.

Soit $\left(\left(\theta_j^{(l)} \right) \right)_{j=1, \dots, l}$ les valeurs propres de T_l , et $\left(y_j^{(l)} \right)_{j=1, \dots, l}$ les vecteurs propres de T_l . Puisque $\left(\theta_j^{(l)}, y_j^{(l)} \right)_{j=1, \dots, l}$ approchent les couples propres de A_m , alors l'approximation du vecteur F_n^δ dans le triplet de Krylov $\mathbf{K}(A_m, G_\delta)$ prend la forme :

$$[F]_l^\delta = \|G_\delta\| Q_l \sum_{\theta_j \leq \lambda_n} \left(\frac{\theta_j^{(l)}}{1 - \exp(-\tau \theta_j^{(l)})} \right) \left[\left(y_j^{(l)} \right)^\top e_1 \right] y_j^{(l)}. \quad (4.1.19)$$

4.2 Problème de Neumann

Dans cette section, on considère le problème inverse (4.1.3) mais avec les conditions de Neumann.

4.2.1 Position du problème

Soient $(H, \langle \cdot, \cdot \rangle, \|\cdot\|)$ un espace de Hilbert séparable de dimension infinie, et $A : D(A) \subset H \rightarrow H$ un opérateur non borné à domaine dense dans H .

Considérons le problème parabolique (direct) suivant :

$$\begin{cases} u_t(t) + Au(t) = f, & 0 < t < T, \\ u(0) = 0, \end{cases} \quad (4.2.20)$$

où la source $f \in H$ est une fonction donnée indépendante de t .

► Notre problème inverse est d'identifier la source f à partir de la condition finale $u(T) = g$.

On fait les hypothèses suivantes :

H1 L'opérateur A est auto-adjoint, i.e., $A = A^*$.

H2 L'opérateur A est positif, i.e., $A \geq 0$ où $0 \in \sigma(A)$.

H3 L'opérateur A est diagonalisable, i.e., il existe une famille dénombrable de couples propres $(\xi_k, \lambda_k)_{k \geq 0}$ telle que :

$$A\xi_k = \lambda_k \xi_k, \quad k = 0 = 1, 2, \dots, \quad \langle \xi_m, \xi_n \rangle = \delta_{mn},$$

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \lambda_k \rightarrow \infty, \quad k \rightarrow \infty,$$

$$\forall h \in H, \quad h = \sum_{k=1}^{\infty} h_k \xi_k, \quad h_k = \langle h, \xi_k \rangle \text{ le coefficient de Fourier d'ordre } k \text{ de } h.$$

On note par $S(t) = e^{-tA}$ le semi-groupe engendré par $(-A)$, i.e.,

$$S(t)h = \sum_{k=0}^{\infty} e^{-t\lambda_k} \langle h, \xi_k \rangle \xi_k, \quad h = \sum_{k=0}^{\infty} \langle h, \xi_k \rangle \xi_k \in H.$$

Notons $S(T) = e^{-TA}$. Il est clair que $S(T)$ est auto-adjoint, compact³ et positif.

3. Un opérateur discret $M : H \rightarrow H$ défini par :

$$Mh = \sum_{k=0}^{\infty} \alpha_k \langle h, \xi_k \rangle \xi_k$$

est compact SSI $\lim_{k \rightarrow \infty} \alpha_k = 0$.

Notons $R(t) = \int_0^t e^{-(t-s)A} ds$ l'opérateur borné défini par

$$R(t)f = \sum_{k=0}^{\infty} \int_0^t e^{-(t-s)\lambda_k} ds \langle f, \xi_k \rangle \xi_k = t \langle f, \xi_0 \rangle \xi_0 + \sum_{k=1}^{\infty} \left\{ \frac{1 - e^{-t\lambda_k}}{\lambda_k} \right\} \langle f, \xi_k \rangle \xi_k.$$

On remarque que l'opérateur discret $R(t)$ est compact ($\alpha_k = \frac{1 - e^{-\lambda_k}}{\lambda_k} \rightarrow 0$), de plus il est auto-adjoint et positif.

En utilisant la méthode de Fourier et la représentation de la solution du problème direct (4.2.20) :

$$u(t) = R(t)f$$

et la condition finale

$$u(T) = T \langle f, \xi_0 \rangle \xi_0 + \sum_{k=1}^{\infty} \left\{ \frac{1 - e^{-T\lambda_k}}{\lambda_k} \right\} \langle f, \xi_k \rangle \xi_k = g = \langle g, \xi_0 \rangle \xi_0 + \sum_{k=1}^{\infty} \langle g, \xi_k \rangle \xi_k,$$

on tire que notre problème inverse admet une solution unique donnée par :

$$f = \frac{1}{T} \langle g, \xi_0 \rangle \xi_0 + \sum_{k=1}^{\infty} \left\{ \frac{\lambda_k}{1 - e^{-T\lambda_k}} \right\} \langle g, \xi_k \rangle \xi_k. \quad (4.2.21)$$

Remarque 4.2.1. On note ici que notre problème inverse est instable, ceci découle du comportement des hautes fréquences $\omega_k = \left(\frac{\lambda_k}{1 - e^{-T\lambda_k}} \right) \approx \lambda_k \rightarrow \infty$ quand $k \rightarrow \infty$.

Remarque 4.2.2. Pour construire une approximation stable pour (4.2.21), on tronque la série de Fourier et on ne prend que la partie finie.

Cas particulier. On suppose ici que

$$g \in [\xi_0]^\perp = \widehat{H} = \{g \in H : g_0 = \langle g, \xi_0 \rangle = 0\}.$$

Dans ce cas, la solution (4.2.21) devient

$$f = \sum_{k=1}^{\infty} \left\{ \frac{\lambda_k}{1 - e^{-T\lambda_k}} \right\} \langle f, \xi_k \rangle \xi_k. \quad (4.2.22)$$

Autrement dit,

$$f = \widehat{A} \left(I - e^{-T\widehat{A}} \right) g = \Phi(\widehat{A})f, \quad \Phi(s) = \frac{s}{1 - e^{-Ts}}.$$

Donc, on peut définir l'opérateur $\widehat{\Phi}(\widehat{A}) = \Phi(\widehat{A})$, où l'opérateur $\widehat{A} = A$ avec le domaine de définition $D(\widehat{A}) = u \in D(A) \cap \widehat{H}$.

Notons

$$\widehat{A}: D(\widehat{A}) \subset \widehat{H} \longrightarrow \widehat{H}, \quad \widehat{A}u = Au, \quad \forall u \in D(\widehat{A}).$$

• Sous cette notation, l'opérateur \widehat{A} est auto-adjoint et strictement positif ($\inf(\sigma(\widehat{A})) = \lambda_1 > 0$), de plus, la fonction $\Phi(s) = \frac{s}{1 - e^{-Ts}}$, $s \in [\lambda_1, +\infty[$ se comporte bien lorsque on remplace \widehat{A} par une approximation matricielle. Cette procédure est bien connue dans l'analyse numérique des problèmes mal conditionnés (en anglais, shift-scaling method).

4.2.2 Exemple d'application

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = f(x), & 0 < x < 1, 0 < t < 1, \\ u(x, 0) = 0, & x \in [0, 1], \\ u_x(0, t) = u_x(1, t) = 0, & t \in [0, 1]. \end{cases} \quad (4.2.23)$$

Il est important de noter ici, que ce problème du point de vue numérique est plus difficile en comparant avec le problème de Dirichlet. En effet, l'opérateur A est auto-adjoint, positif et $\lambda = 0$ est une valeur propre. Ce qui induit une difficulté technique quand on passe au problème discret.

Notons $H = L^2[0, 1]$ et A l'opérateur non borné défini par :

$$\mathcal{D}(A) = \{u \in H^2(0, 1) : u'(1) = u'(0) = 0\}, \quad Au = -\frac{d^2u}{dx^2}. \quad (4.2.24)$$

Proposition 4.2.1. *L'opérateur A est auto-adjoint.*

Preuve. L'opérateur A est symétrique. En effet, $\forall (u, v) \in \mathcal{D}(A) \times \mathcal{D}(A)$, on a :

$$(Au, v) = \int_0^1 -u''v dx.$$

Par une intégration par parties et sachant que $u'(1) = u'(0) = 0$ et $v'(1) = v'(0) = 0$ on a :

$$(Au, v) = \int_0^1 -u''v dx = [-u'v]_0^1 + \int_0^1 u'v' dx = \int_0^1 u'v' dx = [uv']_0^1 - \int_0^1 uv'' dx = (u, Av).$$

Il est également auto-adjoint. En effet, nous allons montrer que l'opérateur $(A + I)$ est inversible, i.e., $(A + I)$ est injectif ($N(A + I) = \{0\}$) et son image égale à H ($R(A + I) = H$).⁴

$$R(I + A) = H \iff \forall z \in H, \exists u \in \mathcal{D}(A) \text{ tq } Au + u = z.$$

Or d'après le théorème de Lax-Milgram, pour toute fonction $f \in L^2[0, 1]$, il existe une unique solution $u \in H^1(0, 1)$ telle que

$$a(u, v) = \int_0^1 (u'v' + uv) dx = \int_0^1 f v dx = L(v).$$

On a de plus $u \in \mathcal{D}(A)$. D'où le résultat. ■

On calcule l'ensemble des couples propres de A :

$$-u''(x) = \lambda u(x), \quad \lambda \geq 0, u \neq 0,$$

avec les conditions

$$u'(0) = u'(1) = 0,$$

on trouve

$$\lambda_0 = 0, e_0(x) = 1, \lambda_n = (n\pi)^2, e_n(x) = \cos(n\pi x), \quad n = 1, 2, \dots,$$

Montrons que l'ensemble $\mathbb{B} = \{e_n\}_{n=0}^\infty$ est une base hilbertienne de H . La démonstration de cette propriété repose sur le théorème de caractérisation des bases hilbertiennes suivant :

4. Si un opérateur symétrique est tel que $R(I + A) = H$, alors le domaine de A est dense dans H ($\mathcal{D}(A)^\perp = \{0\} \iff \overline{\mathcal{D}(A)} = H$) et A est auto-adjoint.

Théorème 4.2.1. [39, Critère de Vitali-Dalzell, p. 39]. Soit $\mathbb{B} = (\zeta)_{n=0}^{\infty}$ une suite orthonormale de l'espace de Hilbert $H = L^2((a, b), \mathbb{R})$, où (a, b) est un intervalle borné de \mathbb{R} . Alors, la suite \mathbb{B} est totale (et donc une base hilbertienne) ssi :

$$\frac{2}{(b-a)^2} \sum_{n=0}^{\infty} \int_a^b \left| \int_a^x \zeta_n(t) dt \right|^2 dx = 1. \quad (4.2.25)$$

Théorème 4.2.2. Les éléments de l'ensemble \mathbb{B} forment une base orthonormée dans l'espace de Hilbert H .

Preuve. On a :

$$\int_0^1 e_i(x)e_j(x) dx = \delta_{i,j}, \quad \forall (i, j) \in \mathbb{N} \times \mathbb{N}. \quad (4.2.26)$$

Il reste à vérifier le critère de de Vitali-Dalzell pour conclure que \mathbb{B} est une base hilbertienne de l'espace de Hilbert $H = L^2(-1, 1), \mathbb{R}$. En effet,

$$\int_0^1 \left| \int_0^x e_0(t) dt \right|^2 dx = \frac{1}{3},$$

$$\int_0^1 \left| \int_0^x e_n(t) dt \right|^2 dx = \frac{1}{(n\pi)^2},$$

d'où

$$\frac{2}{(1-0)^2} \left\{ a_0 + \sum_{n=1}^{\infty} a_n \right\} = 2 \left(\frac{1}{3} + \frac{1}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \right) = 2 \left(\frac{1}{3} + \frac{1}{\pi^2} \times \frac{\pi^2}{6} \right) = 1.$$

■.

On revient, maintenant à notre problème inverse (4.2.23). Puisque l'opérateur A est diagonalisable, alors en utilisant le technique de Fourier et la solution du problème direct, on peut écrire :

$$f = f_0 e_0 + \sum_{n=1}^{\infty} \langle f_n e_n, g \rangle = g_0 e_0 + \sum_{n=1}^{\infty} g_n e_n,$$

où les scalaires f_i et g_i ($i=0,1,\dots$) sont les coefficients de Fourier donnés par

$$f_0 = \int_0^1 f(x)dx, \quad g_0 = \int_0^1 g(x)dx, \quad f_n = \sqrt{2} \int_0^1 f(x) \cos(n\pi x)dx, \quad g_n = \sqrt{2} \int_0^1 g(x) \cos(n\pi x)dx.$$

Ce qui donne

$$u(x, t) = \sum_{n=0}^{\infty} \left\{ \int_0^t e^{-(t-s)\lambda_n} ds \right\} e_n(x),$$

$$u(x, t) = t f_0 e_0(x) + \sum_{n=1}^{\infty} \left\{ \frac{1 - e^{-t\lambda_n}}{\lambda_n} \right\} f_n e_n(x). \quad (4.2.27)$$

En utilisant la condition $u(x, 1) = g(x)$ on peut identifier les coefficient de la source f . L'égalité vectorielle

$$f_0 e_0 + \sum_{n=1}^{\infty} \left\{ \frac{1 - e^{-\lambda_n}}{\lambda_n} \right\} f_n e_n = g = g_0 e_0 + \sum_{n=1}^{\infty} g_n e_n,$$

nous donne

$$f = g_0 e_0 + \sum_{n=1}^{\infty} \left\{ \frac{\lambda_n}{1 - e^{-\lambda_n}} \right\} g_n e_n. \quad (4.2.28)$$

4.3 Estimation d'erreur en dimension finie

Soit $A \in M_n(\mathbb{R})$ une matrice symétrique définie positive. On pose

$$f = \varphi(A)g = A(I - e^{-A})^{-1}g = \sum_{i=1}^n \frac{\lambda_i}{1 - e^{-\lambda_i}} g_i e_i = \sum_{i=1}^n \frac{\lambda_i}{1 - e^{-\lambda_i}} (g, e_i) e_i = \sum_{i=1}^n \varphi(\lambda_i) (g, e_i) e_i,$$

où $g \in \mathbb{R}^n$, où (λ_i, e_i) sont les couples propres de la matrice A :

$$Ae_i = \lambda_i e_i, \quad (e_i, e_j) = \delta_{ij}, \quad \mathbb{R}^n = \bigoplus_{i=1}^n [e_i].$$

Notons

$$f_m = \|g\|_2 Q_m \varphi(H_m) e_1 = \|g\|_2 [Q_m H_m (I - e^{-H_m})^{-1}] e_1$$

l'approximation de Krylov de la fonction f .

Notre objectif ici est d'estimer la quantité $\Delta_m = \|f - f_m\|$.

Pour réaliser cette tâche (qui n'est pas facile), nous utilisons les séries de Tchebychev qui sont importantes en théorie de l'approximation.

Il est en effet connu que le développement tronqué d'une fonction en série de Tchebychev est un polynôme proche du meilleur approximant au sens de la norme uniforme sur le segment $[-1, 1]$.

Soient $\lambda_1, \lambda_2, \dots, \lambda_n$ et $\beta_1, \beta_2, \dots, \beta_m$ avec $m \ll n$ les valeurs propres respectives de A et de $H = Q^t A Q$ où $H \in M_m(\mathbb{R})$ une matrice de Heisenberg, $Q \in M_{n,m}(\mathbb{R})$, $Q^t Q = I$, $\|Q\| < 1$.

On a la relation suivante entre les valeurs propres de A et celles de H (valeurs de Ritz) :

$$\lambda_1 \leq \beta_i \leq \lambda_n, \quad i = 1, 2, \dots, m \text{ avec } \lambda_{i+1} \geq \lambda_i, \quad \beta_{i+1} \geq \beta_i.$$

Si $g = \sum_{i=1}^n g_i e_i = \sum_{i=1}^n (g, e_i) e_i$ où (e_i) est une base orthogonale de \mathbb{R}^n , alors

$$f = \sum_{i=1}^n g \varphi(\lambda_i) e_i.$$

Les (g, e_i) sont appelés coefficients de Fourier.

Comme vecteur approché de f nous prenons le vecteur $f_m = p_m(A)g$ où p_m est un polynôme de degré $\deg(p_m) \leq m - 1$.

Les λ_i et β_i sont connues. Soient λ_1 et λ_n respectivement les bornes inférieure et supérieure de l'intervalle spectral de A .

Soit $B = \frac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1} I_n - \frac{2}{\lambda_n - \lambda_1} A$ avec $\|B\| \leq 1$ et $y = \frac{1}{\lambda_n - \lambda_1} (\lambda_n + \lambda_1 - 2x)$ une fonction linéaire définie sur l'intervalle spectral de A et I_n la matrice unité d'ordre n .

Définissons une fonction h sur le segment $[-1, 1]$ de façon que $h(B) = \varphi(A)$:

$$h(x) = \varphi\left(\frac{\lambda_n + \lambda_1 - (\lambda_n - \lambda_1)x}{2}\right).$$

Considérons la série de Tchebychev de la fonction h , $h(x) = \sum_{k=0}^{+\infty} h_k T_k(x)$ où les h_k sont les coefficients de Tchebychev avec

$$h_k = \frac{\min(2, k+1)}{\pi} \int_{-1}^1 h(x) T_k(x) (1-x^2)^{-\frac{1}{2}} dx$$

et les T_k sont les polynômes de Tchebychev de première espèce et satisfont à la relation de récurrence : $T_0(x) = 1$, $T_1(x) = x$, $T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$, $k \geq 1$.

La famille (T_k) forme une base de l'espace de Hilbert $L^2_\omega(-1, 1)$, où $\omega = \frac{1}{\sqrt{1-x^2}}$ est la mesure du produit scalaire (appelée aussi fonction poids).

Cette famille de polynômes est orthogonale relativement à la fonction poids ω :

$$\int_{-1}^1 T_n(x) T_m(x) \omega(x) dx = 0 \quad \text{pour } n \neq m.$$

Comme $f_m = p_m(A)g = \|g\| Q p(H_m) e_1$

D'après Saad [96], on peut poser aussi de la même façon :

$$V = \frac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1} I_n - \frac{2}{\lambda_n - \lambda_1} H, \quad \|V\| \leq 1.$$

Théorème 4.3.1. (cf.[88]) Si la série $h(x) = \sum_{k=0}^{+\infty} h_k T_k(x)$ est absolument convergente sur l'intervalle $[-1, 1]$, alors on a les égalités suivantes :

$$f - f_m = \sum_{k=m}^{+\infty} \left(h_k T_k(B)g - h_k Q \|g\| T_k(V) e_1 \right) = \sum_{k=m}^{+\infty} h_k \left(T_k(B)g - \|g\| Q T_k(V) e_1 \right)$$

et l'erreur commise satisfait à :

$$\|f - f_m\| \leq 2 \|g\| \sum_{k=m}^{+\infty} |h_k| < +\infty.$$

4.3.1 Calcul de l'estimation d'erreur de $\|f - f_m\|$

Pour la fonction $f = \varphi(A)g = A(I - e^{-A})^{-1}g$, nous essayons d'estimer les coefficients de Fourier-Tchebychev de la fonction linéaire sur l'intervalle spectral de la matrice A

(symétrique définie positive) en estimant les valeurs de cette fonction sur des ellipses dont ses pôles⁵ sont les extrémités de l'intervalle spectral. Ces ellipses ne doivent pas contenir les singularités de la fonction. On suppose que la somme des demi-axes de l'ellipse peut être optimisée par un paramètre en fonction duquel on cherchera une majoration de la fonction.

Soit E_R une ellipse de Foyers $F(1, 0)$ et $F'(-1, 0)$ et R la somme des demi-axes, $R >$

1. Cette ellipse est déterminée par

$$E_R = \left\{ \frac{Re^{i\theta} + R^{-1}e^{-i\theta}}{2}, 0 \leq \theta \leq 2\pi \right\}.$$

Posons

$$a = \frac{\lambda_n - \lambda_1}{2} > 0, \quad c = \frac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1} > 1.$$

$$B = \frac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1} I_n - \frac{2}{\lambda_n - \lambda_1} A, \quad \|B\| \leq 1$$

On a :

$$[I - e^{-A}]^{-1} A = [I - e^{-a(cI_n - B)}]^{-1} a(cI_n - B).$$

Comme $h(B) = \varphi(A)$ on a $h(\omega) = (1 - e^{-a(c-\omega)})^{-1} a(c - \omega)$, $\omega \in E_R$.

Nous admettons le Théorème suivant :

Théorème 4.3.2. (cf.[76]) Si une fonction f est analytique dans le domaine délimité par l'ellipse E_R et continue sur E_R de sorte que

$$|f(z)| \leq M(R), \quad z \in E_R, \quad M \in \mathbb{R},$$

alors les coefficients de Tchebychev a_k de f satisfont à l'inégalité

$$|a_k| \leq 2M(R)R^{-k},$$

et pour tout $r \in [1, R)$ les séries de Tchebychev de la fonction f

$$\frac{a_0}{2} T_0(z) + \sum_{k=1}^{\infty} a_k T_k(z)$$

convergent uniformément vers f sur l'ensemble fermé borné par l'ellipse E_r .

5. Si par un point fixe on mène des droites rencontrant une ellipse, et que par les points d'intersection de ces droites avec la courbe, on lui mène des couples de tangentes, les points d'intersection de ces tangentes sont en ligne droite; cette droite est la "polaire" du point fixe, qui est le pôle.

En appliquant ce théorème, on peut écrire :

Soit $h(\omega) = (1 - e^{-a(c-\omega)})^{-1} a(c-\omega)$ avec $\omega \in E_R$. Elle peut être prolongée par continuité :

$$\hat{h}(\omega) = \begin{cases} \frac{a(c-\omega)}{1 - e^{-a(c-\omega)}}, c \neq \omega \\ \hat{h}(c) = 1, \end{cases}$$

\hat{h} est définie en c . Donc le point c est une singularité amovible de la fonction h . Comme l'ellipse ne doit pas contenir les singularités de la fonction \hat{h} , cherchons les autres singularités dans le plan complexe proches de l'intervalle spectral.

On sait d'une manière générale que la fonction

$$e^{x+iy} = e^z$$

est une fonction entière qui a un point isolé à l'infini. C'est une fonction périodique de y de période 2π donc une fonction de z de période $2\pi i$.

$$1 - e^{-a(c-\omega)} = 0 \Leftrightarrow e^{-a(c-\omega)} = 1 = e^{2\pi ni}, n \in \mathbb{Z}.$$

De ce fait, la fonction \hat{h} a pôle d'ordre 1 en

$$-a(c-\omega) = 2\pi ni \text{ pour } n = \dots, -1, 0, 1, \dots.$$

Les plus proches des singularités de l'intervalle spectral sont $n = -1, 0, 1$ et par suite on a :

$$-a(c-\omega) = 2\pi ni \Leftrightarrow \omega = c + \frac{2\pi ni}{a}.$$

Pour $n = -1$, on a $\omega = c - \frac{2\pi i}{a}$.

Pour $n = 0$, on a $\omega = c$

Pour $n = 1$, on a $\omega = c + \frac{2\pi i}{a}$.

Donc $c + \frac{2\pi i}{a}$ et $c - \frac{2\pi i}{a}$ sont les singularités non-amovibles de \hat{h} .

Pour les systèmes symétriques définis positifs comme dans notre cas, l'ellipse enveloppe le spectre de A qui dégénère en intervalle spectral $[\lambda_{min}, \lambda_{max}]$ sur la partie positive de l'axe des abscisses.

$$\omega \in E_R \Rightarrow \omega = \frac{Re^{i\theta} + R^{-1}e^{-i\theta}}{2}.$$

Et sur l'ellipse si on pose le grand axe $\rho_1 = \frac{1}{2}(R + \frac{1}{R})$ et $\rho_2 = \frac{1}{2}(R - \frac{1}{R})$ le petit axe, on a

$$\frac{1}{2}(R - \frac{1}{R}) \leq |\omega| \leq \frac{1}{2}(R + \frac{1}{R}).$$

Introduisons la fonction de Zhukovsky $\psi(z) = \frac{1}{2}(z + \frac{1}{z})$, $z \neq 0$ et son inverse est $\phi(\omega) = \omega + \sqrt{\omega^2 - 1}$, $\omega \in \mathbb{C} \setminus [-1, 1]$ où \mathbb{C} désigne l'ensemble des nombres complexes.

($\phi(\omega)$) est définie pour tout nombre complexe ω sauf les éléments appartenant au segment $[-1, 1]$.

Selon une définition

$$\Gamma_R = \left\{ \psi(z) : |z| = R \right\}, R > 1, \text{ où } \psi \text{ est la fonction de Zhukovsky.}$$

Puisque ϕ est la fonction inverse de ψ , on peut fixer R par

$$\Gamma_R = \left\{ z : |\phi(z)| = R \right\}, R > 1.$$

Les singularités les plus proches de \hat{h} se trouvent sur l'ellipse E_R avec

$$R = \left| \phi\left(c \pm \frac{2\pi i}{a}\right) \right|. \quad (4.3.29)$$

La valeur de R est indépendante du signe dans (4.3.29) à cause de la symétrie. Fixons cette valeur de R et notons que les deux singularités non-amovibles sont des pôles simples. Ensuite, décomposons \hat{h} par

$$\hat{h} = \hat{h}_1 + \hat{h}_2 + \hat{h}_3,$$

où

$$\hat{h}_1 = \hat{h} - \frac{\text{Res}(h, c + \frac{2\pi i}{a})}{\omega - c - \frac{2\pi i}{a}} - \frac{\text{Res}(h, c - \frac{2\pi i}{a})}{\omega - c + \frac{2\pi i}{a}}, \quad \hat{h}_2 = \frac{\text{Res}(h, c + \frac{2\pi i}{a})}{\omega - c - \frac{2\pi i}{a}}, \quad \hat{h}_3 = \frac{\text{Res}(h, c - \frac{2\pi i}{a})}{\omega - c + \frac{2\pi i}{a}},$$

et Res désigne le résidu d'une fonction analytique à un pôle.

Calculons R .

$$R = \left| \phi\left(c + \frac{2\pi i}{a}\right) \right| = \left| c + \frac{2\pi i}{a} + \sqrt{\left(c + \frac{2\pi i}{a}\right)^2 - 1} \right| = \left| c + \frac{2\pi i}{a} + \sqrt{c^2 - \frac{4\pi^2}{a^2} + \frac{4i\pi c}{a} - 1} \right|.$$

Posons $z = x + iy$ tel que $c^2 - \frac{4\pi^2}{a^2} - 1 + \frac{4i\pi c}{a} = x^2 - y^2 + 2ixy$, on a :

$$\begin{cases} x^2 - y^2 = c^2 - \frac{4\pi^2}{a^2} - 1 & (i) \\ xy = \frac{2\pi c}{a} & (ii) \\ x^2 + y^2 = \sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}} & (iii) \end{cases}$$

$$(i) + (iii) \Rightarrow x^2 = \frac{1}{2}\left(c^2 - \frac{4\pi^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}}$$

Prenons x positif pour raison de commodité, on a :

$$x = \left[\frac{1}{2}\left(c^2 - \frac{4\pi^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}} \right]^{\frac{1}{2}}.$$

$$(ii) \Rightarrow y = \frac{2\pi c}{a} \times \frac{1}{x} = \frac{2\pi c}{a} \left[\frac{1}{2}\left(c^2 - \frac{4\pi^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}} \right]^{-\frac{1}{2}}.$$

Donc

$$\begin{aligned} z = x + iy &= \left[\frac{1}{2}\left(c^2 - \frac{4\pi^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}} \right]^{\frac{1}{2}} \\ &+ i \frac{2\pi c}{a} \left[\frac{1}{2}\left(c^2 - \frac{4\pi^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}} \right]^{-\frac{1}{2}}. \end{aligned}$$

Posons

$$\alpha = \left[\frac{1}{2}\left(c^2 - \frac{4\pi^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}} \right]^{\frac{1}{2}},$$

et

$$\alpha^{-1} = \left[\frac{1}{2}\left(c^2 - \frac{4\pi^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(c^2 - \frac{4\pi^2}{a^2} - 1\right)^2 + \frac{16\pi^2 c^2}{a^2}} \right]^{-\frac{1}{2}}.$$

Par suite

$$R = \left| \phi\left(c + \frac{2\pi i}{a}\right) \right| = \left| c + \frac{2\pi i}{a} + \alpha + i \frac{2\pi c}{a} \alpha^{-1} \right| = \left| c + \alpha + i \frac{2\pi}{a} (1 + c\alpha^{-1}) \right| = \sqrt{(c + \alpha)^2 + \frac{4\pi^2}{a^2} (1 + c\alpha^{-1})^2}.$$

Soit

$$R = \sqrt{(c + \alpha)^2 \left(1 + \frac{4\pi^2}{a^2 \alpha^2}\right)}.$$

Calculons

$$\hat{h} = \hat{h}_1 + \hat{h}_2 + \hat{h}_3.$$

Les pôles sont simples :

$$\omega_n = c + \frac{2\pi ni}{a}, n \in \mathbb{Z}.$$

Posons $\omega = \omega_n + t$ avec t un nombre très petit.

Le Résidu de h en ces points :

$$h(\omega_n + t) = \frac{a\left(\frac{-2n\pi}{a} - t\right)}{1 - e^{-a\left(\frac{-2n\pi}{a} + t\right)}} = \frac{-at - 2n\pi i}{1 - e^{at}}.$$

En faisant la division on a :

$$h(\omega_n + t) = \frac{-2n\pi i - at}{-at - \frac{1}{2!}a^2 t^2 - \frac{1}{3!}a^3 t^3 \dots} = \frac{2n\pi i}{at} + \dots.$$

Donc $\text{Res}(h, \omega_n) = \frac{2n\pi i}{a}$ et par suite

$$\text{Res}\left(h, c + \frac{2\pi i}{a}\right) = \frac{2\pi i}{a}, \text{ et } \text{Res}\left(h, c - \frac{2\pi i}{a}\right) = \frac{-2\pi i}{a}.$$

Remarque 4.3.1. Comme les pôles sont simples, on pouvait utiliser la formule

$$h(\omega) = \frac{f(\omega)}{g(\omega)}.$$

Si $f(a) \neq 0$, $g(a) = 0$ et $g'(a) \neq 0$ alors

$$\text{Res}\left(\frac{f}{g}, a\right) = a_{-1} = \lim_{\omega \rightarrow a} (\omega - a) \frac{f(\omega)}{g(\omega)} = \frac{f(a)}{g'(a)}.$$

$$\hat{h}_1 = \hat{h} - \frac{\text{Res}\left(h, c + \frac{2\pi i}{a}\right)}{\omega - c - \frac{2\pi i}{a}} - \frac{\text{Res}\left(h, c - \frac{2\pi i}{a}\right)}{\omega - c + \frac{2\pi i}{a}} = \frac{a(c - \omega)}{1 - e^{-a(c - \omega)}} - \frac{\frac{2\pi i}{a}}{\omega - c - \frac{2\pi i}{a}} + \frac{\frac{2\pi i}{a}}{\omega - c + \frac{2\pi i}{a}}.$$

$$\hat{h}_1(\omega) = \frac{a(c - \omega)}{1 - e^{-a(c - \omega)}} - \frac{2\pi i}{a} \times \frac{1}{\omega - c - \frac{2\pi i}{a}} + \frac{2\pi i}{a} \times \frac{1}{\omega - c + \frac{2\pi i}{a}}$$

$$= \frac{a(\omega - c)}{e^{a(\omega - c)} - 1} - \frac{2\pi i}{a(\omega - c - \frac{2\pi i}{a})} + \frac{2\pi i}{a(\omega - c + \frac{2\pi i}{a})}.$$

La fonction \hat{h} a un prolongement holomorphe au point c . En effet :

$$\lim_{\omega \rightarrow 0} \frac{\hat{h}(c + \omega) - \hat{h}(c)}{\omega} = \hat{h}'(c).$$

$$\frac{\hat{h}(c + \omega) - \hat{h}(c)}{\omega} = \frac{\frac{a(c - c - \omega)}{1 - e^{-a(c - c - \omega)}} - 1}{\omega} = \frac{\frac{-a\omega}{1 - e^{a\omega}} - 1}{\omega}.$$

$$e^{a\omega} = 1 + a\omega + \frac{a^2\omega^2}{2} + \omega^2(\varepsilon(\omega)) \text{ donc } 1 - e^{a\omega} = -a\omega - \frac{a^2\omega^2}{2} - \omega^2(\varepsilon(\omega)).$$

Par suite

$$\begin{aligned} \frac{\hat{h}(c + \omega) - \hat{h}(c)}{\omega} &= \frac{\frac{-a\omega}{1 - e^{a\omega}} - 1}{\omega} = \frac{\frac{-a\omega}{-a\omega - \frac{a^2\omega^2}{2} - \omega^2(\varepsilon(\omega))} - 1}{\omega} = \frac{-a\omega + a\omega + \frac{a^2\omega^2}{2} + \omega^2(\varepsilon(\omega))}{\omega(-a\omega - \frac{a^2\omega^2}{2} - \omega^2(\varepsilon(\omega)))} \\ &= \frac{\omega^2(\frac{a^2}{2} + \varepsilon(\omega))}{\omega^2(-a - \frac{a^2\omega}{2} - \omega(\varepsilon(\omega)))} = \frac{(\frac{a^2}{2} + \varepsilon(\omega))}{(-a - \frac{a^2\omega}{2} - \omega(\varepsilon(\omega)))}. \end{aligned}$$

Lorsque $\omega \rightarrow 0$ on a :

$$\hat{h}'(c) = \frac{-a}{2}.$$

Considérons deux Disques $A_R(c - \frac{2\pi i}{a}, \frac{\pi}{a})$ et $B_R(c + \frac{2\pi i}{a}, \frac{\pi}{a})$. Nous cherchons une majoration de $\hat{h}_1(\omega)$ sur l'ellipse tout en entier. Nous l'estimons séparément sur un ensemble autour des pôles et sur son complémentaire dans l'ellipse. Nous le faisons en 3 étapes.

Etape 1

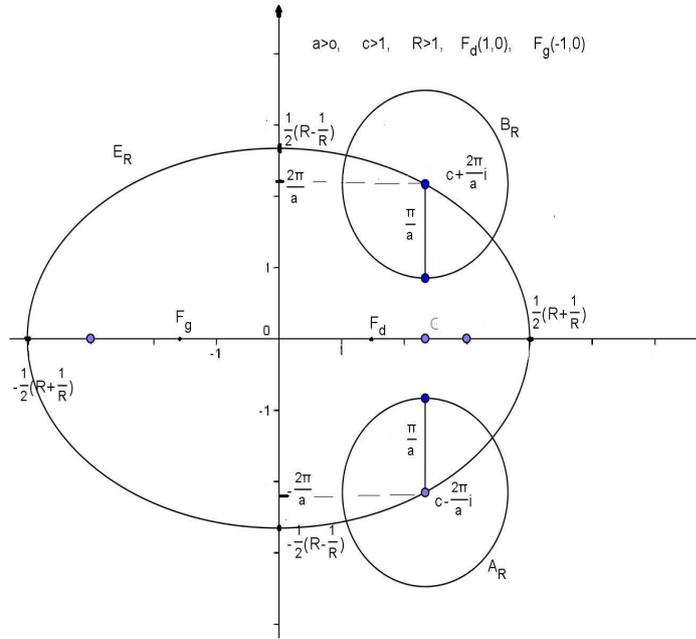


FIGURE 4.1: Estimation d'erreur autour des pôles sur l'ellipse

Considérons le Disque A_R de centre $c - \frac{2\pi i}{a}$ et de rayon $\frac{\pi}{a}$:

$$A_R = \left\{ \omega \in E_R, \left| \omega - c + \frac{2\pi i}{a} \right| < \frac{\pi}{a} \right\}.$$

Posons $X = a(\omega - c + \frac{2\pi i}{a}) = a(\omega - c) + 2\pi i$ donc $\omega = c - \frac{2\pi i}{a} + \frac{X}{a}$ et $\omega \in A_R \Rightarrow |X| < \pi$.

On aura

$$F(X) = \hat{h}_1(\omega) = \hat{h}_1\left(c - \frac{2\pi i}{a} + \frac{X}{a}\right) = \frac{X - 2\pi i}{e^{X-2\pi i} - 1} + \frac{2\pi i}{X} - \frac{2\pi i}{X - 4\pi i} = \frac{X - 2\pi i}{e^X - 1} + \frac{2\pi i}{X} - \frac{2\pi i}{X - 4\pi i}.$$

$$\text{ou } F(X) = \frac{X}{e^X - 1} - \frac{2\pi i X}{X(e^X - 1)} + \frac{(2\pi i)X(e^X - 1)}{X^2(e^X - 1)} + \frac{1}{2} \frac{1}{\left(1 - \frac{X}{4\pi i}\right)}.$$

$$= \frac{X}{e^X - 1} - \frac{2\pi i X}{X(e^X - 1)} + \frac{(2\pi i)X(e^X - 1)}{X^2(e^X - 1)} + \frac{1}{2} \times \frac{1 - \frac{X}{4\pi i} + \frac{X}{4\pi i}}{\left(1 - \frac{X}{4\pi i}\right)}.$$

Ou encore

$$F(X) = \left[1 - \frac{2\pi i}{X} + \frac{(2\pi i)(e^X - 1)}{X^2} \right] \frac{X}{e^X - 1} + \frac{1}{2} + \frac{1}{8\pi i} \frac{X}{1 - \frac{X}{4\pi i}}.$$

On pose $H(X) = \frac{X}{e^X - 1} = 1 - \frac{X}{2} + \sum_{n=1}^{+\infty} \frac{b_{2n}}{(2n)!} X^{2n}$ où les b_{2n} sont les nombres de Bernoulli.

On sait que

$$|b_{2n}| \leq 4 \frac{(2n)!}{(2\pi)^{2n}}, \quad n \geq 1. \quad (I)$$

En posant $\Psi(X) = -\frac{X}{2} + \sum_{n=1}^{+\infty} \frac{b_{2n}}{(2n)!} X^{2n}$ on a $H(X) = 1 + \Psi(X)$ et $F(X)$ devient

$$F(X) = \frac{1}{2} + \left[1 - \frac{2\pi i}{X} + \frac{(2\pi i)(e^X - 1)}{X^2} \right] (1 + \Psi(X)) + \frac{1}{8\pi i} \frac{X}{1 - \frac{X}{4\pi i}}.$$

On sait que

$$\frac{e^X - 1}{X^2} = \frac{1}{X^2} \left(\sum_{n=1}^{+\infty} \frac{X^n}{n!} \right) = \sum_{n=1}^{+\infty} \frac{X^{n-2}}{n!} = \frac{1}{X} + \frac{1}{2} + \sum_{n=3}^{+\infty} \frac{X^{n-2}}{n!} = \frac{1}{X} + \frac{1}{2} + \sum_{n=1}^{+\infty} \frac{X^n}{(n+2)!}.$$

Calculons le terme $1 - \frac{2\pi i}{X} + \frac{(2\pi i)(e^X - 1)}{X^2}$ dans $F(X)$:

$$1 - \frac{2\pi i}{X} + \frac{(2\pi i)(e^X - 1)}{X^2} = 1 - \frac{2\pi i}{X} + 2\pi i \left(\frac{1}{X} + \frac{1}{2} + \sum_{n=1}^{+\infty} \frac{X^n}{(n+2)!} \right) = 1 + \pi i + 2\pi i \sum_{n=1}^{+\infty} \frac{X^n}{(n+2)!}.$$

Ensuite $\frac{1}{1 - \frac{X}{4\pi i}} = \sum_{n=0}^{+\infty} \left(\frac{X}{4\pi i} \right)^n$.

En remplaçant ces termes par leur valeur dans $F(X)$, il vient

$$\begin{aligned} F(X) &= \frac{1}{2} + \left[1 + \pi i + 2\pi i \sum_{n=1}^{+\infty} \frac{X^n}{(n+2)!} \right] [1 + \Psi(X)] + \frac{1}{8\pi i} X \sum_{n=0}^{+\infty} \left(\frac{X}{4\pi i} \right)^n \\ &= \frac{1}{2} + 1 + \pi i + (1 + \pi i)\Psi(X) + 2\pi i(1 + \Psi(X)) \left(\sum_{n=1}^{+\infty} \frac{X^n}{(n+2)!} \right) + \frac{1}{8\pi i} X \sum_{n=0}^{+\infty} \left(\frac{X}{4\pi i} \right)^n \\ &= \frac{3}{2} + \pi i + (1 + \pi i)\Psi(X) + 2\pi i(1 + \Psi(X)) \left(\sum_{n=1}^{+\infty} \frac{X^n}{(n+2)!} \right) + \frac{1}{8\pi i} X \sum_{n=0}^{+\infty} \left(\frac{X}{4\pi i} \right)^n. \end{aligned}$$

$$|F(X)| \leq \frac{\sqrt{9+4\pi^2}}{2} + \sqrt{1+\pi^2}|\Psi(X)| + 2\pi(1+|\Psi(X)|) \left(\sum_{n=1}^{+\infty} \frac{|X|^n}{(n+2)!} \right) + \frac{1}{8\pi}|X| \sum_{n=0}^{+\infty} \left| \frac{X}{4\pi} \right|^n.$$

Pour $\omega \in A_R$ on a $|X| < \pi$ donc

$$\begin{aligned} |\Psi(X)| &\leq \frac{|X|}{2} + \sum_{n=1}^{+\infty} \left| \frac{b_{2n}}{(2n)!} \right| |X|^{2n} \leq \frac{\pi}{2} + 4 \sum_{n=1}^{+\infty} \frac{1}{(2\pi)^{2n}} \pi^{2n} \text{ (suivant la relation (I))} \\ &= \frac{\pi}{2} + 4 \sum_{n=1}^{+\infty} \left(\frac{1}{4} \right)^n \\ &= \frac{\pi}{2} + \frac{4}{3}. \end{aligned}$$

$$\sum_{n=1}^{+\infty} \frac{|X|^n}{(n+2)!} < \sum_{n=1}^{+\infty} \frac{|X|^n}{n!} = e^{|X|} - 1 < e^\pi - 1.$$

$$|X| \sum_{n=0}^{+\infty} \left| \frac{X}{4\pi} \right|^n < \pi \sum_{n=0}^{+\infty} \left(\frac{1}{4} \right)^n = \frac{4\pi}{3}.$$

Et finalement

$$|F(X)| \leq \frac{\sqrt{9+4\pi^2}}{2} + \sqrt{1+\pi^2} \left(\frac{\pi}{2} + \frac{4}{3} \right) + 2\pi \left(\frac{\pi}{2} + \frac{7}{3} \right) (e^\pi - 1) + \frac{1}{6} = M_1.$$

Soit

$$\sup_{\omega \in A_R} |\hat{h}_1(\omega)| \leq M_1$$

Etape 2

Considérons le Disque B_R de centre $c + \frac{2\pi i}{a}$ et de rayon $\frac{\pi}{a}$

$$B_R = \left\{ \omega \in E_R, \left| \omega - c - \frac{2\pi i}{a} \right| < \frac{\pi}{a} \right\}$$

Posons $Y = a(\omega - c - \frac{2\pi i}{a}) = a(\omega - c) - 2\pi i$ donc $\omega = c + \frac{2\pi i}{a} + \frac{Y}{a}$ et $\omega \in B_R \Rightarrow |Y| < \pi$.

On aura

$$\begin{aligned}
G(Y) &= \hat{h}_1(\omega) = \hat{h}_1\left(c + \frac{2\pi i}{a} + \frac{Y}{a}\right) = \frac{Y + 2\pi i}{e^Y - 1} - \frac{2\pi i}{Y} + \frac{2\pi i}{Y + 4\pi i} \\
&= \frac{Y}{e^Y - 1} + \frac{2\pi i Y}{Y(e^Y - 1)} - \frac{(2\pi i)Y(e^Y - 1)}{Y^2(e^Y - 1)} + \frac{1}{2} \frac{1}{\left(1 - \frac{Yi}{4\pi}\right)} \\
&= \frac{Y}{e^Y - 1} + \frac{2\pi i Y}{Y(e^Y - 1)} - \frac{(2\pi i)Y(e^Y - 1)}{Y^2(e^Y - 1)} + \frac{1}{2} \times \frac{1 - \frac{Yi}{4\pi} + \frac{Yi}{4\pi}}{\left(1 - \frac{Yi}{4\pi}\right)} \\
&= \left[1 + \frac{2\pi i}{Y} - \frac{(2\pi i)(e^Y - 1)}{Y^2}\right] \frac{Y}{e^Y - 1} + \frac{1}{2} + \frac{i}{8\pi} \frac{Y}{\left(1 - \frac{Yi}{4\pi}\right)} \\
&= \left[1 + \frac{2\pi i}{Y} - \frac{(2\pi i)(e^Y - 1)}{Y^2}\right] [1 + \Psi(Y)] + \frac{1}{2} + \frac{i}{8\pi} \frac{Y}{\left(1 - \frac{Yi}{4\pi}\right)}.
\end{aligned}$$

Calculons le terme $1 + \frac{2\pi i}{Y} - \frac{(2\pi i)(e^Y - 1)}{Y^2}$ dans $G(Y)$.

$$\begin{aligned}
1 + \frac{2\pi i}{Y} - \frac{(2\pi i)}{Y^2}(e^Y - 1) &= 1 + \frac{2\pi i}{Y} - \frac{(2\pi i)}{Y^2} \left(\sum_{n=1}^{+\infty} \frac{Y^n}{n!}\right) = 1 + \frac{2\pi i}{Y} - (2\pi i) \left(\sum_{n=1}^{+\infty} \frac{Y^{n-2}}{n!}\right) \\
&= 1 + \frac{2\pi i}{Y} - \frac{2\pi i}{Y} - \pi i - 2\pi i \left(\sum_{n=1}^{+\infty} \frac{Y^n}{(n+2)!}\right) \\
&= 1 - \pi i - 2\pi i \left(\sum_{n=1}^{+\infty} \frac{Y^n}{(n+2)!}\right).
\end{aligned}$$

On a :

$$\begin{aligned}
G(Y) &= \frac{1}{2} + \left[1 - \pi i - 2\pi i \sum_{n=1}^{+\infty} \frac{Y^n}{(n+2)!}\right] [1 + \Psi(Y)] + \frac{i}{8\pi} Y \sum_{n=0}^{+\infty} \left(\frac{iY}{4\pi}\right)^n \\
&= \frac{3}{2} - \pi i + (1 - \pi i)\Psi(Y) - 2\pi i(1 + \Psi(Y)) \left(\sum_{n=1}^{+\infty} \frac{Y^n}{(n+2)!}\right) + \frac{i}{8\pi} Y \sum_{n=0}^{+\infty} \left(\frac{iY}{4\pi}\right)^n.
\end{aligned}$$

Et Finalement on a :

$$|G(Y)| \leq \frac{\sqrt{9 + 4\pi^2}}{2} + \sqrt{1 + \pi^2} \left(\frac{\pi}{2} + \frac{4}{3}\right) + 2\pi \left(\frac{\pi}{2} + \frac{7}{3}\right)(e^\pi - 1) + \frac{1}{6} = M_1.$$

Soit

$$\sup_{\omega \in B_R} |\hat{h}_1(\omega)| \leq M_1$$

Remarque 4.3.2. On pouvait remarquer que la fonction étudiée est analytique et à valeurs réelles sur l'axe réel. Ce qui fait que les valeurs de la fonction aux points complexes conjugués sont également complexes conjugués de sorte que le module est le même.

Etape 3

Considérons $\omega \in E_R \setminus (A_R \cup B_R)$

$$\omega \in E_R \Rightarrow \left| \omega - c - \frac{2\pi i}{a} \right| \geq \frac{\pi}{a} \text{ et } \left| \omega - c + \frac{2\pi i}{a} \right| \geq \frac{\pi}{a}.$$

Soit le grand axe $\rho_1 = \frac{1}{2}(R + \frac{1}{R})$ et $\rho_2 = \frac{1}{2}(R - \frac{1}{R})$ le petit axe et considérons le cercle de centre c contenant l'ellipse.

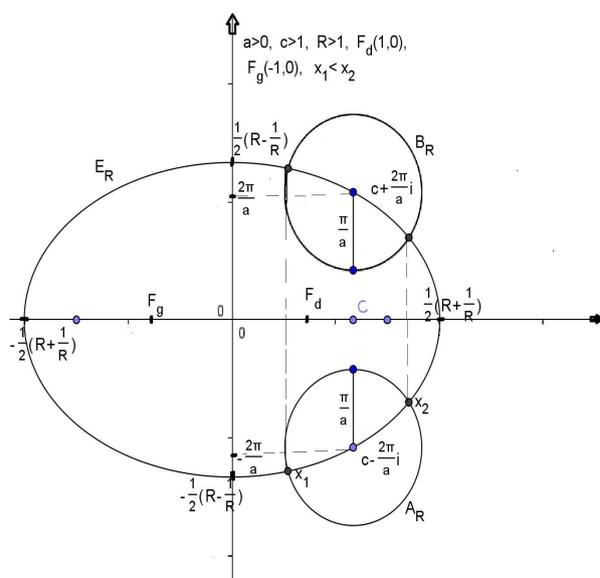


FIGURE 4.2: Estimation d'erreur sur le complément de l'ensemble

On a

$$\forall \omega \in E_R \setminus (A_R \cup B_R), |\omega - c| \leq \frac{1}{2} \left(R + \frac{1}{R} \right) + c = \rho_1 + c$$

$$\hat{h}_1(\omega) = \frac{a(\omega - c)}{e^{a(\omega - c)} - 1} - \frac{2\pi i}{a(\omega - c - \frac{2\pi i}{a})} + \frac{2\pi i}{a(\omega - c + \frac{2\pi i}{a})}.$$

En passant au module

$$\begin{aligned} |\hat{h}_1(\omega)| &= \left| \frac{a(\omega - c)}{(e^{a(\omega - c)} - 1)} - \frac{2\pi}{a(\omega - c - \frac{2\pi i}{a})} + \frac{2\pi}{a(\omega - c + \frac{2\pi i}{a})} \right| \leq \frac{a(\rho_1 + c)}{|(e^{a(\omega - c)} - 1)|} + \frac{2\pi}{a} \times \frac{a}{\pi} + \frac{2\pi}{a} \times \frac{a}{\pi} \\ &\leq 4 + \frac{a(\rho_1 + c)}{|e^{a(\omega - c)} - 1|}. \end{aligned}$$

On sait que $|e^{a(\omega - c)} - 1| \geq ||e^{a(\omega - c)}| - 1|$ donc

$$|\hat{h}_1(\omega)| \leq 4 + \frac{a(\rho_1 + c)}{|e^{a(\omega - c)} - 1|} \leq 4 + \frac{a(\rho_1 + c)}{||e^{a(\omega - c)}| - 1|}.$$

Pour $\omega \in E_R \setminus (A_R \cup B_R)$ prenons $\omega = x + iy$ et on a :

$$e^{a(\omega - c)} = e^{a(x - c + iy)} = e^{a(x - c)} \times e^{ia y}, \quad |e^{a(\omega - c)}| = e^{a(x - c)}.$$

Par suite

$$||e^{a(\omega - c)}| - 1| = |e^{a(x - c)} - 1|.$$

Cherchons $\mu > 0$ tel que $\forall \omega \in E_R \setminus (A_R \cup B_R), ||e^{a(\omega - c)}| - 1| \geq \mu$.

Soit x_1, x_2 les abscisses de la projection des deux points limites des disques A_R et B_R sur le grand axe de l'ellipse.

$$\omega = x + iy \Rightarrow x \leq x_1 \text{ ou } x \geq x_2.$$

Soit $\omega_1 = (x_1 + iy_1)$,

$$\left| \omega_1 - c + \frac{2\pi i}{a} \right|^2 = \frac{\pi^2}{a^2} \Rightarrow (x_1 - c)^2 + \left(y_1 + \frac{2\pi}{a} \right)^2 = \frac{\pi^2}{a^2} \Rightarrow (x_1 - c)^2 \leq \frac{\pi^2}{a^2}$$

$$-\frac{\pi}{a} \leq x_1 - c \leq \frac{\pi}{a} \Rightarrow c - \frac{\pi}{a} \leq x_1 \leq c + \frac{\pi}{a}$$

Ce qui entraîne que

$$c - \frac{\pi}{a} \leq x_1 < c \text{ et } c < x_2 < c + \frac{\pi}{a}.$$

Ensuite

$$\begin{cases} a(x-c) \leq a(x_1-c) < 0 \\ a(x-c) \geq a(x_2-c) > 0 \end{cases} \Rightarrow \begin{cases} e^{a(c-x)} - 1 \leq e^{a(x_1-c)} - 1 = e^{-a(c-x_1)} - 1 \\ e^{a(x_2-c)} - 1 \geq e^{a(x-c)} - 1 \end{cases}$$

$$|e^{a(x-c)} - 1| \geq 1 - e^{-a(c-x_1)} \text{ pour } x < x_1,$$

et

$$|e^{a(x-c)} - 1| \geq \min(1 - e^{-a(c-x_1)}, e^{a(x_2-c)} - 1).$$

Posons $\mu = \min(1 - e^{-a(c-x_1)}, e^{a(x_2-c)} - 1)$, $0 < \mu < 1$. x_1 et x_2 sont solutions de

$$\begin{cases} \frac{x^2}{\rho_1^2} + \frac{y^2}{\rho_2^2} = 1 \\ (x-c)^2 + \left(y + \frac{2\pi}{a}\right)^2 = \frac{\pi^2}{a^2} \end{cases}$$

$$|\hat{h}_1(\omega)| \leq 4 + \frac{a(\rho_1 + c)}{\mu}.$$

Posons $M(R) = \max\left(M_1, 4 + \frac{a(\rho_1 + c)}{\mu}\right)$ on a $\sup_{\omega \in E_R} |\hat{h}_1(\omega)| \leq M(R)$.

Remarque 4.3.3. La Résolution graphique ou résolution algébrique permet de trouver les abscisses x_1 et x_2 des points d'intersections de E_R avec le Disque

$$(x-c)^2 + \left(y + \frac{2\pi}{a}\right)^2 = \frac{\pi^2}{a^2} \quad (4.3.30)$$

où

$$\mu = \min(1 - e^{-a(c-x_1)}, e^{a(x_2-c)} - 1) \quad 0 < \mu < 1. \quad (4.3.31)$$

D'après le théorème (4.3.2),

$$|h_{k,1}| \leq 2M(R)R^{-k} = 2 \max\left(M_1, 4 + \frac{a(\rho_1 + c)}{\mu}\right) \times \left((c + \alpha)^2 \left(1 + \frac{4\pi^2}{a^2\alpha^2}\right)\right)^{\frac{-k}{2}}$$

$$\hat{h}_2(\omega) = \frac{\text{Res}(h, c + \frac{2\pi i}{a})}{\omega - c - \frac{2\pi i}{a}} = \frac{\frac{2\pi i}{a}}{\omega - c - \frac{2\pi i}{a}} = \frac{-2\pi i}{a} \left(\frac{1}{(c + \frac{2\pi i}{a}) - \omega} \right)$$

D'après le théorème 10.4 (cf.[76]) on a la formule suivante :

$$\frac{1}{z_0 - \omega} = \frac{4\phi(z_0)^{-1}}{1 - \phi(z_0)^{-2}} \sum_{k=0}^{\infty} \phi(z_0)^{-k} T_k(\omega), \quad z_0 \in \mathbb{C} \setminus [-1, 1].$$

Appliquons cette formule à \hat{h}_2

$$\hat{h}_2(\omega) = \frac{2\pi i}{a} \left(\frac{1}{(c + \frac{2\pi i}{a}) - \omega} \right) = \frac{2\pi i}{a} \left(\frac{4\phi(c + \frac{2\pi i}{a})^{-1}}{1 - \phi(c + \frac{2\pi i}{a})^{-2}} \sum_{k=0}^{\infty} \phi\left(c + \frac{2\pi i}{a}\right)^{-k} T_k(\omega) \right).$$

Posons $p = \phi(c + \frac{2\pi i}{a})^{-1}$ on a :

$$\hat{h}_2(\omega) = \frac{2\pi i}{a} \left(\frac{4p}{1 - p^2} \sum_{k=0}^{\infty} p^k T_k(\omega) \right) = \sum_{k=0}^{\infty} \frac{2\pi i}{a} \left(\frac{4p}{1 - p^2} p^k \right) T_k(\omega) = \sum_{k=0}^{\infty} \frac{8\pi i}{a} \left(\frac{p^{k+1}}{1 - p^2} \right) T_k(\omega).$$

Or

$$\hat{h}_2(\omega) = \sum_{k=0}^{\infty} h_k T_k(\omega).$$

Donc par identification

$$h_{k,2} = \frac{8\pi i}{a} \left(\frac{p^{k+1}}{1 - p^2} \right).$$

En passant au module

$$|h_{k,2}| = \frac{8\pi}{a} \left| \frac{p^{k+1}}{1 - p^2} \right|.$$

De même

$$\hat{h}_3(\omega) = \frac{-2\pi i}{a} \left(\frac{1}{(c - \frac{2\pi i}{a}) - \omega} \right) = \frac{-2\pi i}{a} \left(\frac{4\phi(c - \frac{2\pi i}{a})^{-1}}{1 - \phi(c - \frac{2\pi i}{a})^{-2}} \sum_{k=0}^{\infty} \phi\left(c - \frac{2\pi i}{a}\right)^{-k} T_k(\omega) \right).$$

Posons $q = \phi(c - \frac{2\pi i}{a})^{-1}$ on a :

$$\hat{h}_3(\omega) = \frac{-2\pi i}{a} \left(\frac{4q}{1-q^2} \sum_{k=0}^{\infty} q^k T_k(\omega) \right) = \sum_{k=0}^{\infty} \frac{-2\pi i}{a} \left(\frac{4q}{1-q^2} q^k \right) T_k(\omega) = \sum_{k=0}^{\infty} \frac{-8\pi i}{a} \left(\frac{q^{k+1}}{1-q^2} \right) T_k(\omega).$$

Or

$$\hat{h}_3(\omega) = \sum_{k=0}^{\infty} h_k T_k(\omega).$$

Donc par identification

$$h_{k,3} = \frac{-8\pi i}{a} \left(\frac{q^{k+1}}{1-q^2} \right).$$

En passant au module

$$|h_{k,3}| = \frac{8\pi}{a} \left| \frac{q^{k+1}}{1-q^2} \right|.$$

Par suite

$$|h_k| \leq |h_{k,1}| + |h_{k,2}| + |h_{k,3}| = 2 \max \left(M_1, 4 + \frac{a(\rho_1 + c)}{\mu} \right) \times \left((c + \alpha)^2 \left(1 + \frac{4\pi^2}{a^2 \alpha^2} \right) \right)^{\frac{-k}{2}} + \frac{8\pi}{a} \left| \frac{p^{k+1}}{1-p^2} \right| + \frac{8\pi}{a} \left| \frac{q^{k+1}}{1-q^2} \right|.$$

$$\sum_{k=m}^{+\infty} |h_k| \leq 2 \max \left(M_1, 4 + \frac{a(\rho_1 + c)}{\mu} \right) \times \sum_{k=m}^{+\infty} R^{-k} + \sum_{k=m}^{+\infty} \frac{8\pi}{a} \left| \frac{p^{k+1}}{1-p^2} \right| + \sum_{k=m}^{+\infty} \frac{8\pi}{a} \left| \frac{q^{k+1}}{1-q^2} \right|.$$

$$\sum_{k=m}^{+\infty} R^{-k} = \frac{R^{-m}}{1-R^{-1}}, \quad \sum_{k=m}^{+\infty} p^{k+1} = \frac{p^{m+1}}{1-p}$$

$$\sum_{k=m}^{+\infty} \left| \frac{p^{k+1}}{1-p^2} \right| = \left| \frac{p^{m+1}}{(1+p)(1-p)^2} \right| = \left| \frac{\phi(c + \frac{2\pi i}{a})^{-m-1}}{(1 + \phi(c + \frac{2\pi i}{a})^{-1})(1 - \phi(c + \frac{2\pi i}{a})^{-1})^2} \right|.$$

De même

$$\sum_{k=m}^{+\infty} \left| \frac{q^{k+1}}{1-q^2} \right| = \left| \frac{q^{m+1}}{(1+q)(1-q)^2} \right| = \left| \frac{\phi(c - \frac{2\pi i}{a})^{-m-1}}{(1 + \phi(c - \frac{2\pi i}{a})^{-1})(1 - \phi(c - \frac{2\pi i}{a})^{-1})^2} \right|.$$

Finalement d'après le Théorème 4.3.1, on a l'estimation d'erreur suivante :

$$\begin{aligned} \|f - f_m\| &\leq 2\|g\| \sum_{k=m}^{+\infty} |h_k| \\ &\leq 2\|g\| \left(2 \max \left(M_1, 4 + \frac{a(\rho_1 + c)}{\mu} \right) \times \frac{\left((c + \alpha)^2 \left(1 + \frac{4\pi^2}{a^2 \alpha^2} \right) \right)^{\frac{-m}{2}}}{1 - \left((c + \alpha)^2 \left(1 + \frac{4\pi^2}{a^2 \alpha^2} \right) \right)^{\frac{-1}{2}}} + \right. \\ &\quad \left. \frac{8\pi}{a} \left| \frac{\phi \left(c + \frac{2\pi i}{a} \right)^{-m-1}}{\left(1 + \phi \left(c + \frac{2\pi i}{a} \right)^{-1} \right) \left(1 - \phi \left(c + \frac{2\pi i}{a} \right)^{-1} \right)^2} \right| + \frac{8\pi}{a} \left| \frac{\phi \left(c - \frac{2\pi i}{a} \right)^{-m-1}}{\left(1 + \phi \left(c - \frac{2\pi i}{a} \right)^{-1} \right) \left(1 - \phi \left(c - \frac{2\pi i}{a} \right)^{-1} \right)^2} \right| \right) \\ &\|f - f_m\| \leq 2\|g\| \left\{ 2 \max \left(M_1, 4 + \frac{a(\rho_1 + c)}{\mu} \right) \times \frac{\left((c + \alpha)^2 \left(1 + \frac{4\pi^2}{a^2 \alpha^2} \right) \right)^{\frac{-m}{2}}}{1 - \left((c + \alpha)^2 \left(1 + \frac{4\pi^2}{a^2 \alpha^2} \right) \right)^{\frac{-1}{2}}} + \right. \\ &\quad \left. \frac{8\pi}{a} \left| \frac{\phi \left(c + \frac{2\pi i}{a} \right)^{-m-1}}{\left(1 + \phi \left(c + \frac{2\pi i}{a} \right)^{-1} \right) \left(1 - \phi \left(c + \frac{2\pi i}{a} \right)^{-1} \right)^2} \right| + \frac{8\pi}{a} \left| \frac{\phi \left(c - \frac{2\pi i}{a} \right)^{-m-1}}{\left(1 + \phi \left(c - \frac{2\pi i}{a} \right)^{-1} \right) \left(1 - \phi \left(c - \frac{2\pi i}{a} \right)^{-1} \right)^2} \right| \right\}. \end{aligned}$$

Les singularités de la fonction \hat{h} les plus proches de l'intervalle spectral de la matrice A déterminent le taux de convergence de la méthode de Lanczos. Il reste entendu qu'une singularité sur un point de l'ellipse produit le même taux de convergence asymptotique qu'un pôle ou un autre point de l'ellipse.

5

Tests numériques

Dans ce chapitre, on donne quelques exemples numériques illustrant le cas de Neumann.

5.1 Discrétisation de l'opérateur $A = -\frac{d^2}{dx^2}$

Soit \sum une partition uniforme de l'intervalle $[0, 1]$ donnée par $0 = x_0 < x_1 < x_2 < \dots < x_{i-1} < x_i < \dots < x_N = 1$ $i = 1, \dots, N$ $h = \frac{1}{N}$ où N est un entier positif, h étant un pas constant de discrétisation et $x_i = ih$.

Le développement de Taylor à l'ordre n d'une fonction f au point x_0 est :

$$f(x_0 + \Delta x) = u(x_0) + \sum_{i=1}^n \frac{\Delta x^i}{i!} f^{(i)}(x_0) + \theta(\Delta x^{n+1})$$

où $f^{(i)}$ désigne la dérivée i^{ieme} de f , et f supposée suffisamment continûment dérivable. Soit l'opérateur :

$$-\frac{d^2 u(x)}{dx^2}.$$

On l'écrit au point x_i : $-\frac{d^2 u(x_i)}{dx^2}$.

Par développement de Taylor, on a :

$$u(x_{i+1}) = u(x_i + h) = u(x_i) + h \left(\frac{du(x_i)}{dx} \right) + \frac{h^2}{2} \left(\frac{d^2u(x_i)}{dx^2} \right) + \frac{h^3}{6} \left(\frac{d^3u(x_i)}{dx^3} \right) + \theta(h^4),$$

$$u(x_{i-1}) = u(x_i - h) = u(x_i) - h \left(\frac{du(x_i)}{dx} \right) + \frac{h^2}{2} \left(\frac{d^2u(x_i)}{dx^2} \right) - \frac{h^3}{6} \left(\frac{d^3u(x_i)}{dx^3} \right) + \theta(h^4).$$

Par sommation, on trouve donc

$$\begin{aligned} u(x_{i+1}) + u(x_{i-1}) &= 2u(x_i) + h^2 \left(\frac{d^2u(x_i)}{dx^2} \right) + \theta(h^4) \\ \Leftrightarrow \left(\frac{d^2u(x_i)}{dx^2} \right) &= \frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1}))}{h^2} + \theta(h^2). \end{aligned}$$

Soit u_i l'approximation de $u(x_i)$, on a :

$$u_i \simeq u(x_i).$$

On obtient le schéma suivant :

$$\left(\frac{d^2u(x_i)}{dx^2} \right) = \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + \theta(h^2).$$

D'autre part il faut discrétiser la condition de Neumann $u'(0) = u'(1) = 0$.

On a

$$u'(x) = \frac{u(x+h) - u(x)}{h} + \theta(h)$$

et

$$u'(x) = \frac{u(x) - u(x-h)}{h} + \theta(h)$$

$$u'(0) = 0 \Rightarrow \frac{u(h) - u(0)}{h} + \theta(h) = 0 \Rightarrow u_1 = u_0,$$

$$u'(1) = 0 \Rightarrow \frac{u(1) - u(1-h)}{h} + \theta(h) = 0 \Rightarrow u_N = u_{N-1}.$$

L'opérateur discrétisé devient alors

$$\begin{cases} A_h u_h = -\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} \\ u(0) = u_0 \\ u_1 = u_0, u_N = u_{N-1} \end{cases}$$

Forme matricielle

On a

$$\frac{1}{h^2}(-u_{i+1} + 2u_i - u_{i-1})$$

$$\text{pour } i = 1, \frac{1}{h^2}(-u_2 + 2u_1 - u_0) = \frac{1}{h^2}(-u_2 + u_1)$$

$$\text{pour } i = 2, \frac{1}{h^2}(-u_3 + 2u_2 - u_1)$$

$$\text{pour } i = 3, \frac{1}{h^2}(-u_4 + 2u_3 - u_2)$$

.....

$$\text{pour } i = N - 2, \frac{1}{h^2}(-u_{N-1} + 2u_{N-2} - u_{N-3})$$

$$\text{pour } i = N - 1, \frac{1}{h^2}(-u_N + 2u_{N-1} - u_{N-2}) = \frac{1}{h^2}(u_{N-1} - u_{N-2})$$

Sous forme matricielle, on obtient :

$$\frac{1}{h^2} \begin{pmatrix} 1 & -1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot & 0 \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \cdot \\ \cdot \\ u_{N-2} \\ u_{N-1} \end{pmatrix}$$

Par identification

$$A_h = \frac{1}{h^2} \begin{pmatrix} 1 & -1 & 0 & 0 & \cdot & \cdot & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot & 0 \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & 0 & 0 & -1 & 1 \end{pmatrix}, u_h = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \cdot \\ \cdot \\ u_{N-2} \\ u_{N-1} \end{pmatrix}$$

La matrice A_h est tridiagonale (par construction), symétrique et positive.

Proposition 5.1.1. *La matrice A_h est symétrique positive.*

Preuve. On applique la définition usuelle, en formant le produit scalaire $h^2(x, A_h x)$, pour un vecteur x de \mathbb{R}^{N-1} quelconque de composantes x_1, x_2, \dots, x_{N-1} .

On a

$$\begin{aligned}
h^2(x, A_h x) &= h^2 \sum_{i=1}^{N-1} x_i (A_h x)_i \\
&= x_1(x_1 - x_2) + \sum_{i=2}^{N-2} x_i(-x_{i-1} + 2x_i - x_{i+1}) + x_N(-x_{N-1} + x_N) \\
&= x_1^2 - x_1 x_2 + (-x_2 x_1 + 2x_2^2 - x_2 x_3 - x_3 x_2 + 2x_3^2 - x_3 x_4 - x_4 x_3 + 2x_4^2 - x_4 x_5 + \dots \\
&\quad - x_{N-1} x_{N-2} + 2x_{N-1}^2 - x_{N-1} x_N) + x_N^2 - x_N x_{N-1} \\
&= x_1^2 + x_N^2 + 2 \sum_{i=2}^{N-2} x_i^2 - x_2 x_1 - 2 \sum_{i=2}^{N-2} (x_i x_{i+1}) - x_1 x_2 \\
&= x_1^2 + x_N^2 + 2 \sum_{i=2}^{N-2} x_i^2 - 2 \sum_{i=2}^{N-2} (x_i x_{i+1}) - 2x_1 x_2 \\
&= x_1^2 + x_N^2 + x_2^2 - x_N^2 + \sum_{i=2}^{N-2} (x_i^2 - 2x_i x_{i+1} + x_{i+1}^2) \\
&= x_1^2 + x_2^2 + \sum_{i=2}^{N-2} (x_i - x_{i+1})^2 - 2x_1 x_2 \\
&= (x_1 - x_2)^2 + \sum_{i=2}^{N-2} (x_i - x_{i+1})^2 \\
&= \sum_{i=1}^{N-2} (x_i - x_{i+1})^2.
\end{aligned}$$

Ainsi, $(x, A_h x) \geq 0$. Il est clair que la forme quadratique $(x, A_h x)$ s'annule pour $x_1 = x_2 = \dots = x_{N-1} = 1$. ■

Remarque 5.1.1. Pour le cas de Dirichlet, la matrice $A_h = \frac{1}{h^2} \text{Tridiag}[-1, 2, -1] \in M_N(\mathbb{R})$ est une matrice tridiagonale, symétrique et définie positive.

On se limite ici au cas de Neumann qui est intéressant du point de vue pratique. On suppose donc que $g \in \widehat{H}$. Cette condition devient en dimension finie $g_h \in [e_0]^\perp = E \subset \mathbb{R}^{N-1}$, où $e_0 = (1, 1, \dots, 1) \in \mathbb{R}^{N-1}$.

La solution du problème discrétisée est notée par


```

11:   Stop
12: end if
13:    $v_{j+1} := \frac{z_j}{\beta_{j+1}}$           → Normalisation
14: end for

```

On obtient ainsi m vecteurs orthonormés v_j et une matrice T tridiagonale symétrique $m \times m$ (matrice de Rayleigh) constituée d'éléments diagonaux $\alpha_j = T_{j,j}$ et extradiagonaux $\beta_j = T_{j,j-1} = T_{j-1,j}$. Les valeurs propres et vecteurs propres de T se calculent par la méthode SVD (Singular Value Decomposition). On obtient ainsi les m plus grandes valeurs propres de A et les composantes des vecteurs propres de A dans la base des v_j . Donc la méthode de Lanczos permet de calculer les valeurs propres d'une matrice réelle en utilisant la notion d'espace de Krylov. L'intérêt d'utilisation de cette méthode est dû au fait de sa capacité de permettre le calcul d'une petite partie du spectre sans avoir à calculer tout le spectre.

5.2 Données exactes

Les figures ci-dessous montrent la solution exacte et l'approximation de Krylov de l'équation de la chaleur avec les conditions de Neumann. Pour l'approximation de Krylov, nous faisons varier N (la taille de la matrice) et m (la dimension du sous-espace de Krylov) pour voir l'efficacité de la méthode. Nous prenons dans chaque cas comme vecteur de départ de l'Algorithme de Lanczos $v = g$ où $g \in \mathbb{R}^N$ qui est la fonction de test. On a $f_m = \|g\| V_m \varphi(T_m) e_1$.

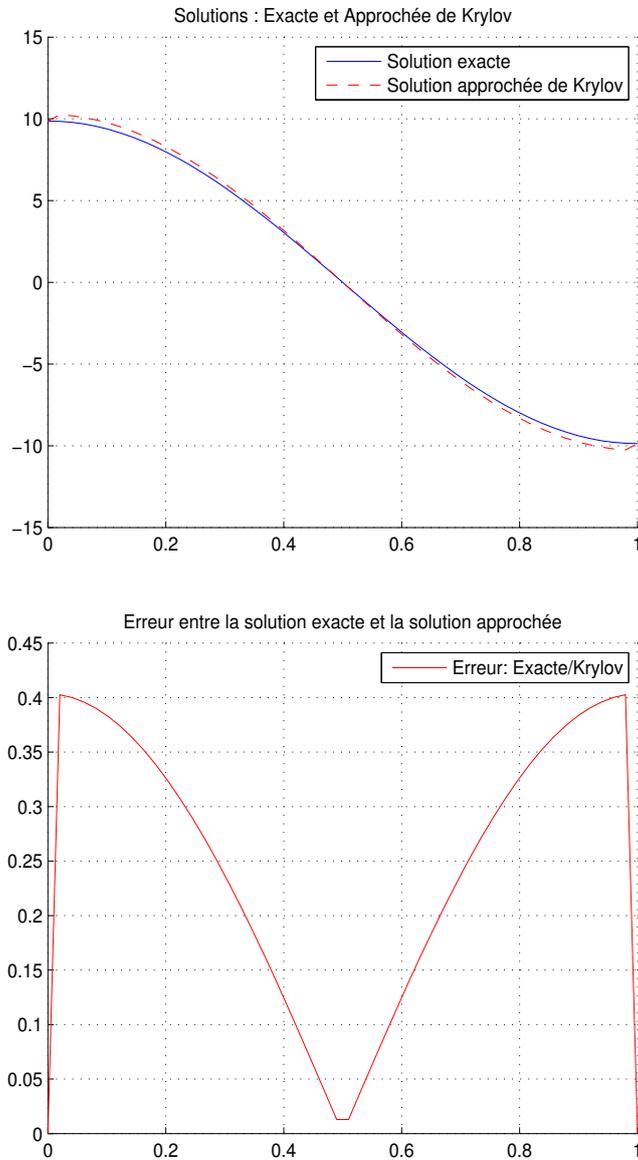


FIGURE 5.1: Comparaison entre la Solution exacte et l'Approximation de Krylov pour $N=50$, $m=50$ et la Courbe d'erreur pour la précision.

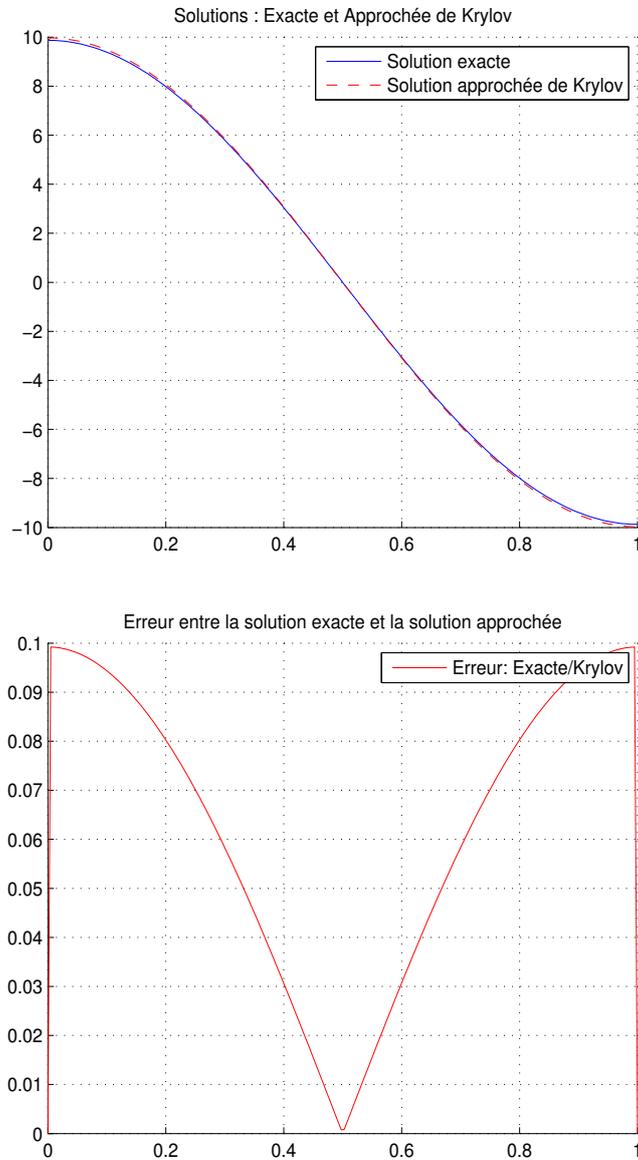


FIGURE 5.2: Comparaison entre la Solution exacte et l'Approximation de Krylov pour $N=200$, $m=150$ et la Courbe d'erreur pour la précision.

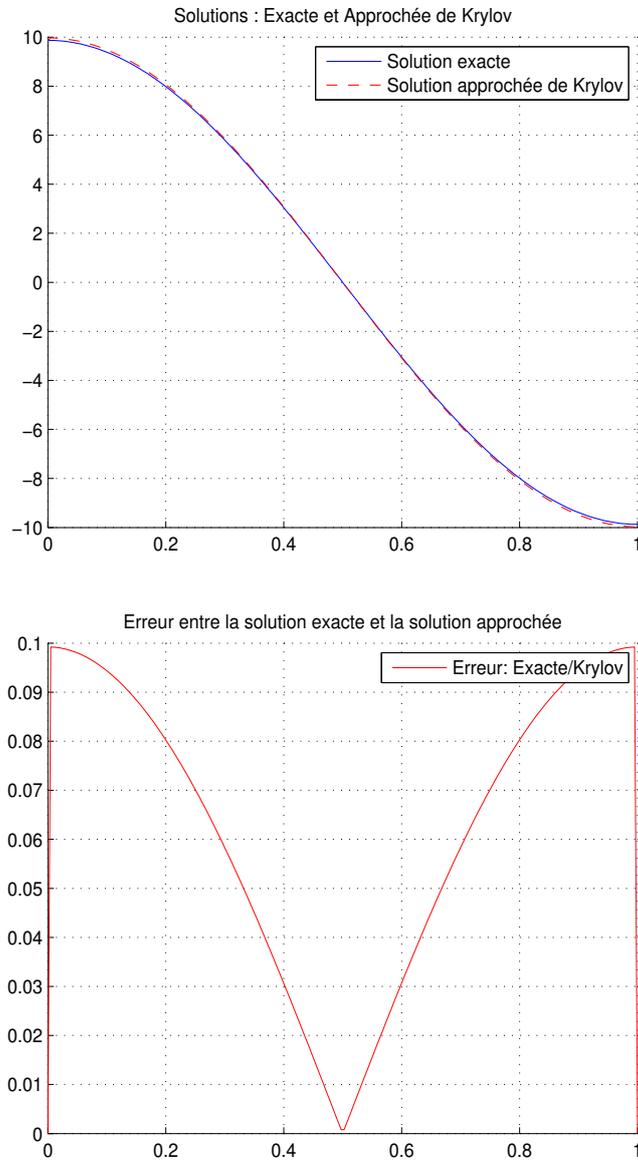


FIGURE 5.3: Comparaison entre la Solution exacte et l'Approximation de Krylov pour $N=200$, $m=100$ et la Courbe d'erreur pour la précision.

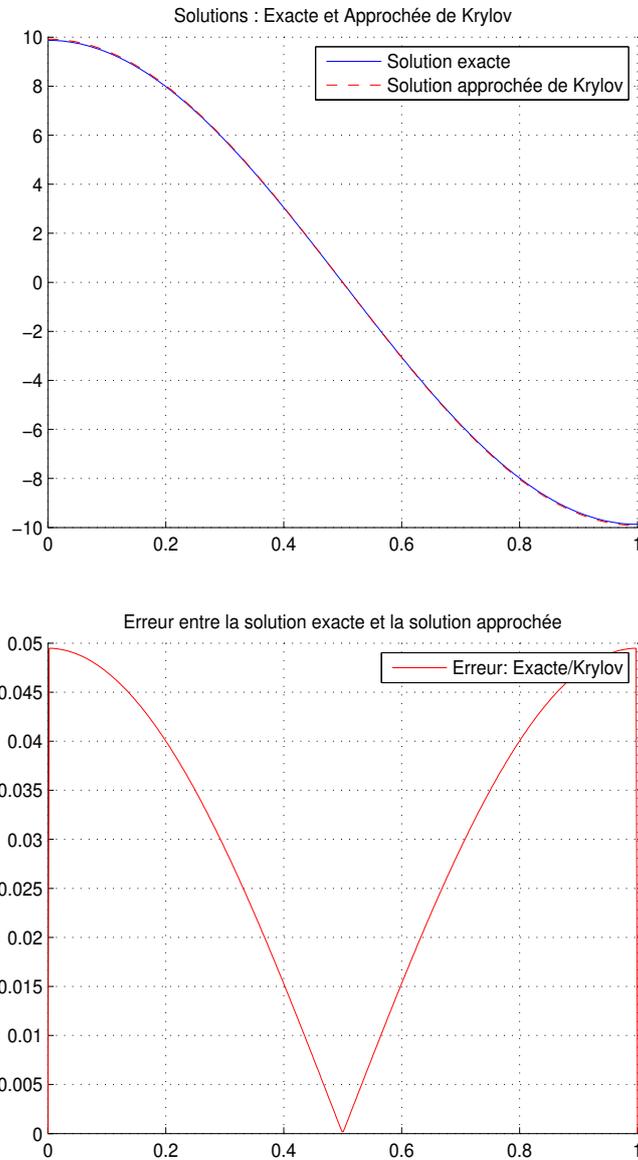


FIGURE 5.4: Comparaison entre la Solution exacte et l'Approximation de Krylov pour $N=400$, $m=100$ et la Courbe d'erreur pour la précision.

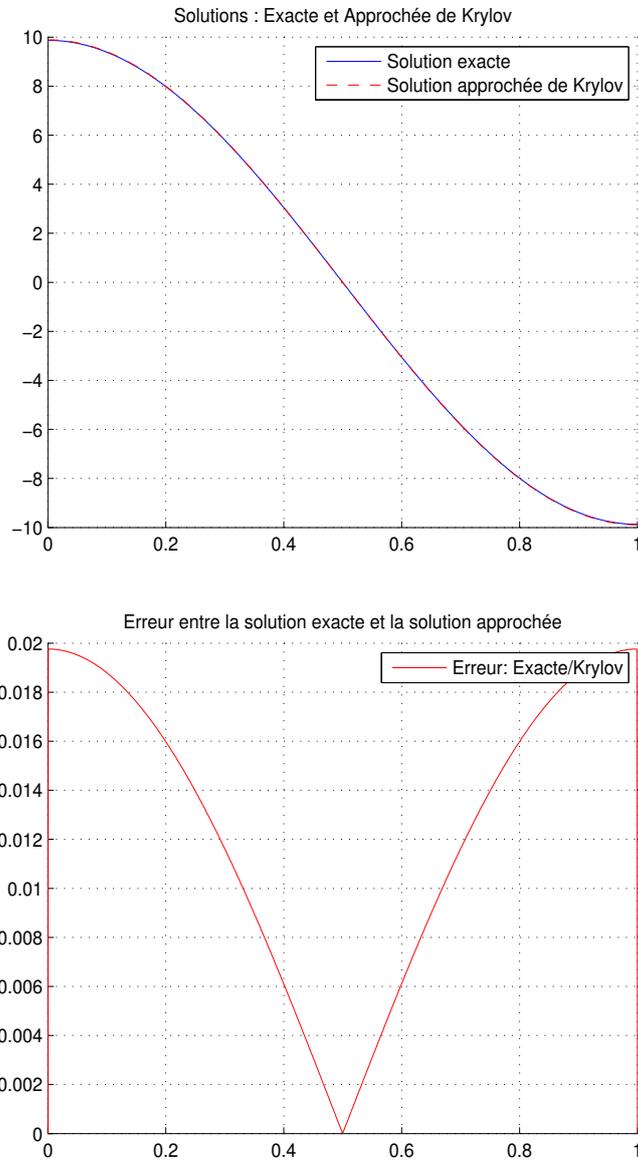


FIGURE 5.5: Comparaison entre la Solution exacte et l'Approximation de Krylov pour $N=1000$, $m=100$ et la Courbe d'erreur pour la précision.

Nous constatons que quand $m \ll N$ la méthode de Krylov donne une approximation beaucoup plus efficace de la solution exacte. Elle est particulièrement utile du fait de la réduction du coup de calcul (nombre d'opérations) pour les matrices de grande taille.

Remarque 5.2.1. Nous pouvons tracer les différentes courbes en fonction de la taille N de la matrice et la dimension de l'espace de Krylov m mais notre fonction f est définie dans $[0, 1]$.

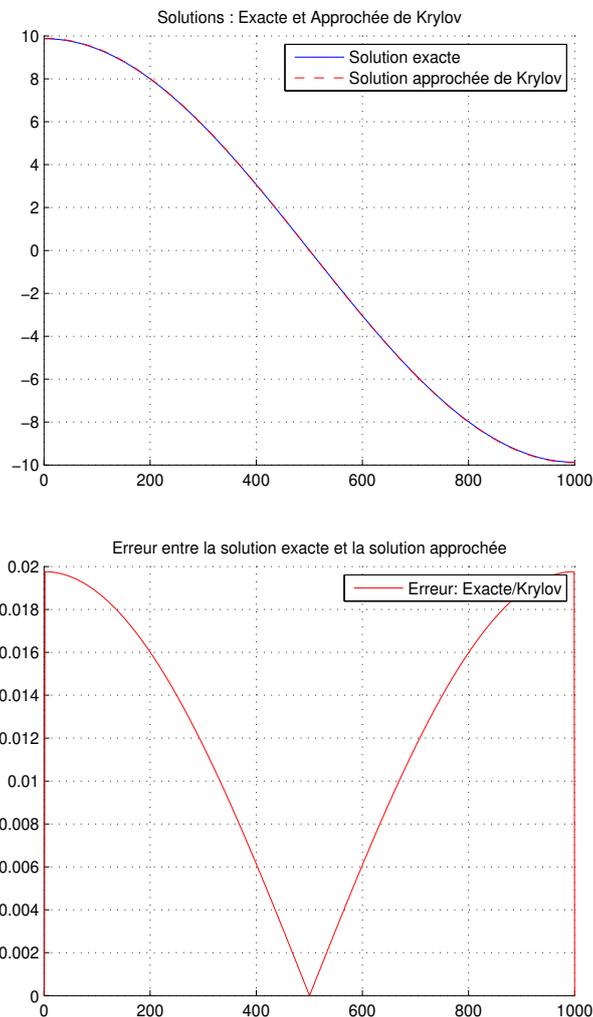


FIGURE 5.6: Comparaison entre la Solution exacte et l'Approximation de Krylov pour $N=1000$, $m=100$ et la Courbe d'erreur pour la précision.

Remarque 5.2.2. Il est possible de calculer certaines fonctions de matrice avec Matlab en utilisant la méthode SVD telles que e^A , $A^{\frac{1}{2}}$, $\cos(A)$, $\sin(A)$, $\ln(A)$. D'une manière

générale si A peut s'écrire sous la forme $A = T^{-1}DT$ où T est une matrice carrée inversible, D une matrice diagonale et T^{-1} matrice de passage et f une fonction de matrice alors

$$f(A) = f(T^{-1}DT) = T^{-1}f(D)T = T^{-1}diag(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n))T.$$

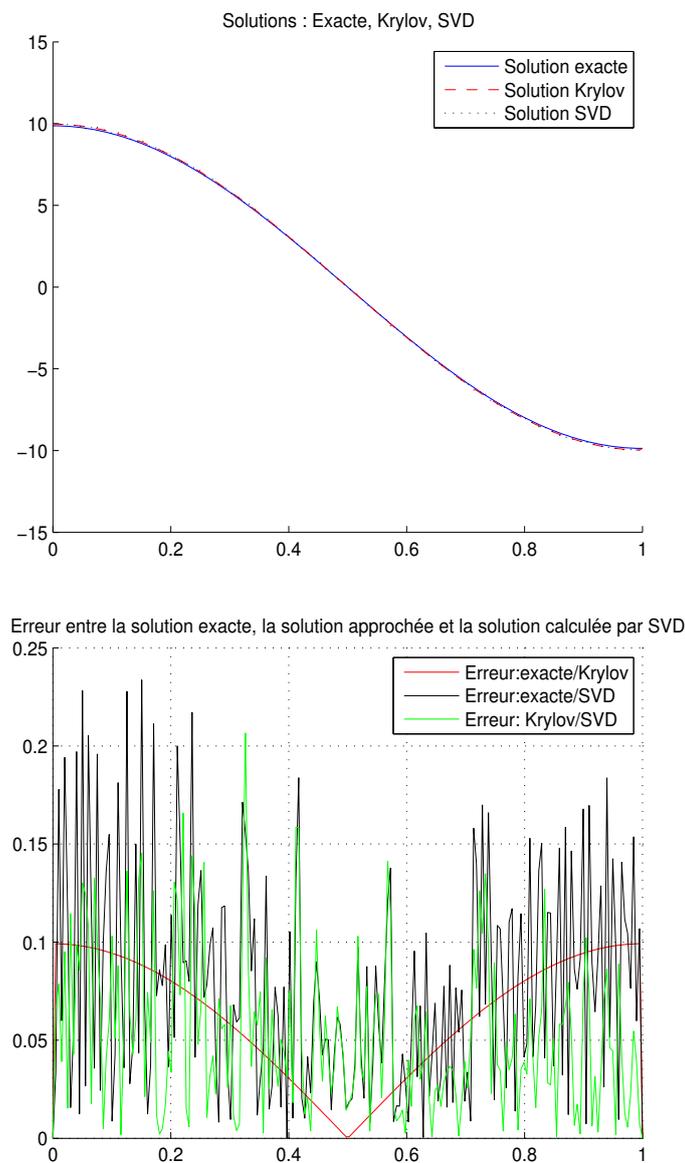


FIGURE 5.7: Comparaison entre les solutions exacte, SVD et Krylov pour $N=200$, $m=100$ et la Courbe d'erreur pour la précision.

Les solutions dans le cas des données perturbées (bruitées)

Pour examiner le cas de données bruitées, on utilise la commande Matlab "randn" (Le rand fonction génère des tableaux de nombres aléatoires dont les éléments sont répartis uniformément dans l'intervalle $[0, 1]$) pour générer des données perturbées g_δ :

$$g_\delta = g + \varepsilon \text{randn}(\text{size}(g)) \quad (5.2.2)$$

où ε est la taille du bruit, $\text{randn}(\cdot)$ est la fonction du bruit aléatoire distribué selon la loi normale avec les paramètres : la moyenne 0, la variance $\sigma^2 = 1$, et l'écart type $\sigma = 1$. "randn(size(g))" renvoie un tableau d'entrées aléatoires qui est de même taille que g .

Pour l'approximation de Krylov, nous fixons N (la taille de la matrice) et m (la dimension du sous-espace de Krylov) et nous faisons varier ε jusqu'à l'obtention d'une solution raffinée. Nous prenons dans chaque cas comme vecteur de départ de l'Algorithme d'Lanczos $v = g_\delta$ où $g_\delta \in \mathbb{R}^N$ qui est la fonction de test.

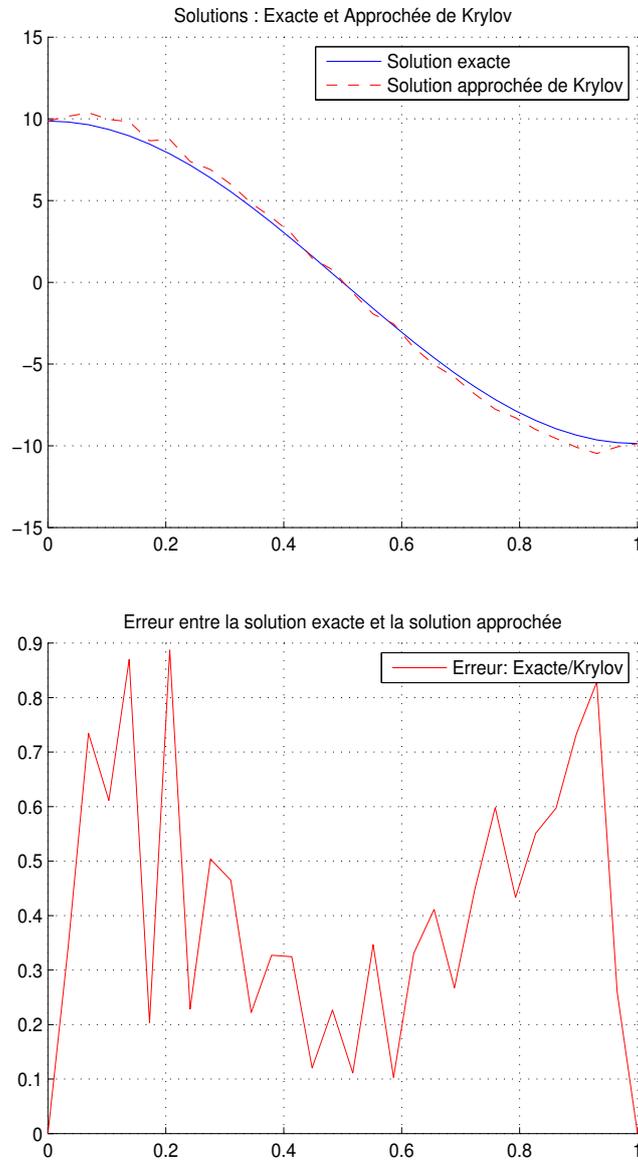


FIGURE 5.8: Comparaison entre la solution exacte et l'approximation de Krylov (resp. la solution exacte et la solution discrétisée) dans le cas perturbé pour $N=30$, $m=12$, $\varepsilon = 10^{-4}$ et la courbe d'erreur pour la précision.

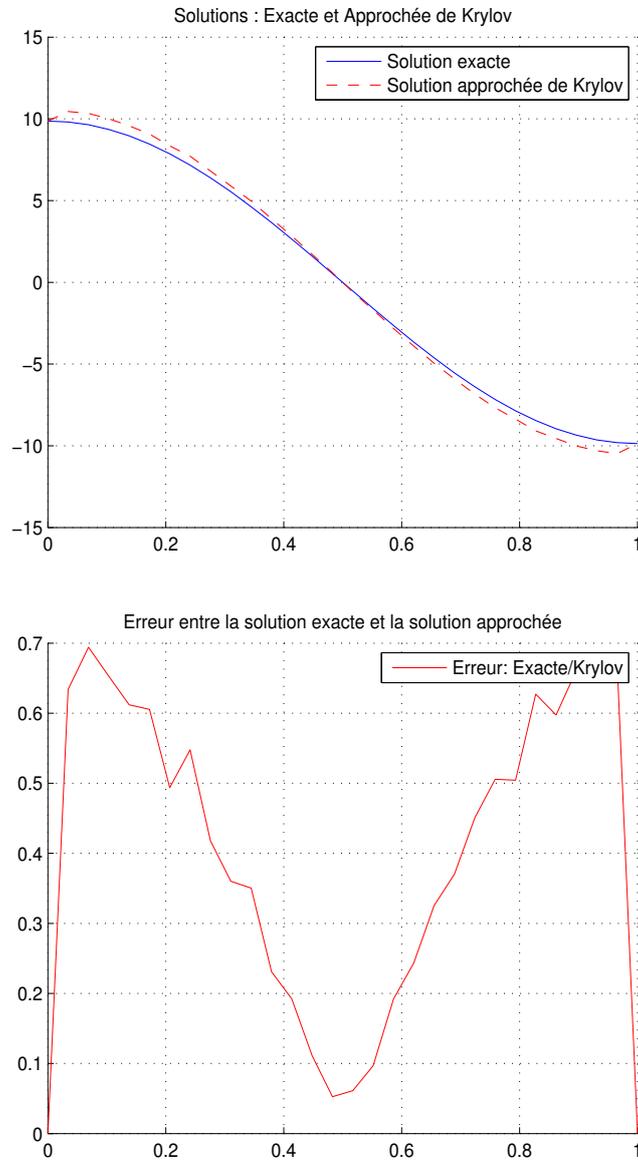


FIGURE 5.9: Comparaison entre la solution exacte et l'approximation de Krylov (resp. la solution exacte et la solution discrétisée) dans le cas perturbé pour $N=30$, $m=12$, $\varepsilon = 10^{-5}$ et la courbe d'erreur pour la précision.

Conclusion et perspectives

L'objectif principal était la mise en œuvre d'une méthode inverse par la technique de projection de Krylov permettant le calcul numérique des systèmes de la forme $u = \Phi(A)v$.

Nous avons étudié la possibilité d'utiliser une technique de calcul matriciel basée sur une méthode de projection de type Krylov. L'analyse de cette technique proposée dans ce travail montre que la méthode de projection peut fournir un outil efficace pour l'approximation d'une fonction de matrice croissante.

Nous avons ainsi illustré l'intérêt de disposer d'outils élaborés de résolution d'équations matricielles.

Pour tester l'efficacité de la méthode, nous l'avons appliqué à un problème inverse d'identification de sources avec des conditions aux limites différentes (Neumann/Dirichlet).

A fin de mieux comprendre l'effet régularisant de la méthode de Krylov nous avons considéré des données entachées du bruit pour voir l'influence des perturbations sur la régularisation de la solution.

Ensuite nous avons développé une nouvelle approche qui permettrait de calculer l'estimation d'erreur entre la solution exacte et l'approximation de Krylov avec une fonction croissante en introduisant les notions de résidu autour des points de singularités de ladite fonction. Cette estimation nous a permis d'étudier théoriquement le comportement de la convergence de la méthode de Krylov ainsi que sa stabilité.

Cette estimation nous permet de donner la taille exacte de l'espace de Krylov selon le test d'arrêt fixé et les précisions souhaitées à établir.

Des résultats tant sur le plan théorique que sur le plan pratique de performance ont été obtenus pour la résolution des systèmes de la forme $u = \Phi(A)v$.

L'une des perspectives que nous envisageons d'entamer dans un futur proche est d'étendre l'estimation d'erreur en dimension 2 (2D) par exemple à un problème inverse elliptique de type identification de conditions aux limites et identification de sources.

.1 Programme MATLAB

```

1  clear ; clc ;
2  n=input('Donner la taille de la matrice carree n=');
3  h=1/n;
4  A=gallery('tridiag',n,-1/h^2,2/h^2,-1/h^2);
5  % A=tridiag(-1/h^2,2/h^2,-1/h^2,n);
6  A(1,1)=1/h^2;
7  A(n,n)=1/h^2;
8  m=input('Donner la dimension de l''espace de krylov associee a
  b_inferieur a n_m=');
9  epsilon=input('Donner la valeur de epsilon pour la perturbation
  de g_epsilon=');
10 if nargin < 3, m=n; end
11 %Calcul de la fonction de test g
12 %
-----
13 x=linspace(0,1,n);
14 g=(1-exp(-pi^2))*cos(pi*x);
15 gn=g+epsilon*randn(size(g));
16 v=gn';
17 %
-----
18 %Calcul de la solution exacte
19 fe=pi^2*cos(pi*x);
20
21 %Construction de la base de Krylov par Lanczos
22 %%
23     n = length(v);
24     T = zeros(m);
25     Q = zeros(n,m);
26
27     v1 = v/norm(v);
28
29     z = A*v1;

```

```

30     alpha = v1' * z;
31     z = z - v1 * alpha;
32
33     Q(:, 1) = v1; T(1, 1) = alpha;
34
35     for j = 2:m,
36
37         beta = norm(z);
38         v0 = v1; v1 = z/beta;
39
40         z = A*v1 - v0*beta;
41         alpha = v1' * z;
42         z = z - v1 * alpha;
43
44         T(j, j-1) = beta; T(j-1, j) = beta; T(j, j) = alpha;
45         Q(:, j) = v1;
46
47     end
48
49     e1=zeros(m,1);
50     e1(1,1)=1;
51     %%
52
53     fm=norm(v)*Q*(T*(eye(size(T))-expm(-T))^-1)*e1;
54     fm(1)=fe(1);
55     fm(n)=fe(n);
56
57     fd=(A*(eye(size(A))-expm(-A))^-1)*v;
58     fd(1)=fe(1);
59     fd(n)=fe(n);
60     figure()
61     hold on,
62     %grid on,
63     plot(fe, 'b'),
64     plot(fm, 'r--'),
65     plot(fd, 'k:'),
66     legend('Exacte', 'Krylov', 'SVD'),
67     title('Solutions : Exacte, Krylov, SVD'),
68     hold off,
69     %%
70     figure()
71     hold on,
72     grid on,

```

```
73     plot(abs(fe'-fm), 'r'),  
74     plot(abs(fe'-fd), 'k'),  
75     legend('Exacte/Krylov', 'Exacte/SVD'),  
76     title('Erreurs'),  
77     hold off,
```

Bibliographie

- [1] A.A. Samarskii, P.N. Vabishchevich, *Numerical Methods for Solving Inverse Problems of Mathematical Physics*, Inverse and Ill-Posed Problems Series, Walter de Gruyter, Berlin. New York (2007).
- [2] A. Bouzitouna, N. Boussetila, F. Rebbani, *Sur Quelques Méthodes Itératives Appliquées à Certains Problèmes Inverses : Méthodes de projection de Krylov, Kozlov-Mazyra et troncature*, Éditions universitaires européennes (2016).
- [3] A. Bouzitouna, *Sur Quelques Méthodes Itératives Appliquées à Certains Problèmes Inverses*, Thèse de Doctora, U.BM-Annaba (2014).
- [4] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Appl. Math. Sci. 44, Springer Verlag, New York (1983).
- [5] A. Frackowiak, N.D. Botkin, M. Cialkowski, K.H. Hoffmann, *Iterative algorithm for solving the inverse heat conduction problems with the unknown source function*, Inverse Problems in Science and Engineering, (2015) 23 :6, 1056-1071.
- [6] A. Hamdia, I. Mahfoudhi, *Inverse source problem in a one-dimensional evolution linear transport equation with spatially varying coefficients : application to surface water pollution*, Inverse Problems in Science and Engineering (2013). Vol. 21, No. 6, 1007-1031.
- [7] A. Hasanov, M. Slodicka, *An analysis of inverse source problems with final time measured output data for the heat conduction equation : A semigroup approach*, Applied Mathematics Letters 26 (2) (2013) 207-214.
- [8] A. Hasanov, O. Baysal, *Identification of unknown temporal and spatial load distributions in a vibrating Euler–Bernoulli beam from Dirichlet boundary measured data*, J. Automatica. (2016) 106-117.
- [9] A. Frommer, M. Schweitzer, *Error bounds and estimates for Krylov subspace approximations of Stieltjes matrix functions*, BIT Numerical Mathematics (2016), Vol. 56, Issue 3, 865–892.
- [10] A.I. Prilepko, D.G. Orlovsky, I.A. Vasin, *Methods for Solving Inverse Problems in Mathematical Physics*, Monographs and textbooks in Pure and Applied Mathematics 222, Marcel Dekker (2000).
- [11] A.N. Tikhonov and V.Y. Arsenin, *Solution of Ill-posed Problems*, Winston & Sons, Washington, DC, (1977).
- [12] A. Qian, *Identifying an unknown source in the Poisson equation by a wavelet dual least square method*, Boundary Value Problems (2013), 2013 :267.
- [13] A. Qian, Y. Li, *Optimal error bound and generalized Tikhonov regularization for identifying an unknown source in the heat equation*, J. Math. Chem (2011), 49 :765-775.
- [14] A. Qian, J.F. Mao, *Optimal error bound and a generalized tikhonov regularization method for identifying an unknown source in the poisson equation*, International Journal of Wavelets, Multiresolution and Information Processing (2014), Vol. 12, No. 1, 1450004, DOI : 10.1142/S0219691314500040.

- [15] A. Shidfar, A. Babaei, A. Molabahrani, *Solving the inverse problem of identifying an unknown source term in a parabolic equation*, Computers and Mathematics with Applications 60 (2010), 1209-1213.
- [16] Bjorck Ake, *Numerical Methods in Matrix Computations*, Springer (2014).
- [17] D.A. Murio, *The mollification method and the numerical solution of ill-posed problems*, Wiley-Interscience Publication (1993).
- [18] D.D. Ang, R. Gorenflo, V.K.Le and D.D. Trong, *Moment Theory and Some Inverse Problems in Potential Theory and Heat Conduction*, Lecture Notes in Mathematics 1792, Springer-Verlag, Berlin (2002).
- [19] D.G. Orlovskii, *Determination of a parameter of a parabolic equation in Hilbert's structure*, Mathematical Notes (1994), Vol. 55, No. 3, 321-317.
- [20] D.G. Orlovskii, *Inverse Dirichlet Problem for an Equation of Elliptic Type*, Differential Equations (2008), Vol. 44, No. 1, 124-134.
- [21] D. Huet; *Décomposition Spectrale et Opérateurs*, PUF (1976).
- [22] E.B. Davies; *Linear Operator and their Spectra*, Cambridge University Press (2007).
- [23] E. Celledoni, I. Moret, *A Krylov projection method for systems of ODES*, Applied Numerical Mathematics 24 (1997), 365-378.
- [24] E.G. Savateev, *On problems of determining the source function in a parabolic equation*, J. Inv. Ill-posed Probl. 3 (1995), 83-102.
- [25] F. Diele, I. Moret, S. Ragni, *Error estimates for polynomial krylov approximations to matrix functions*, SIAM J. MATRIX ANAL. APPL.(2008), Vol. 30, No. 4, pp. 1546-1565.
- [26] F.F. Dou, C.L. Fua, F. Yang, *Identifying an unknown source term in a heat equation*, Inverse Problems in Science and Engineering (2009), Vol. 17, No. 7, 901-913.
- [27] F.F. Dou, C.L. Fu, F.L. Yang, *Optimal error bound and Fourier regularization for identifying an unknown source in the heat equation*, Journal of Computational and Applied Mathematics 230 (2009), 728-737.
- [28] F.R. Gantmacher, *The Theory of Matrices*, Vol.1, Chelsea, New York, (1959).
- [29] G.H. Golub, C.F. Van Loan, *Matrix Computation*, The Johns Hopkins University Press (2013).
- [30] H. Brezis; *Analyse Fonctionnelle, Théorie et Applications*, Masson (1993).
- [31] H. O. Fattorini, *The Cauchy Problem*, Encyclopedia of Mathematics and its Applications, 18, Cambridge University Press 1983.
- [32] H.W. Engel, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic (2000).
- [33] I. Gohberg, S. Goldberg and M.A. Kaashoek; *Basic Classes of Linear Operators*, Birkhäuser (2003).
- [34] I. Hnetynkova, M. Plesinger, Z. Strakos, *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT Numerical Mathematics (2009), Vol., Issue 4, 669-696.
- [35] I. Moret, *Shift-and-Invert Krylov Methods for Time-Fractional Wave Equations*, Numerical Functional Analysis and Optimization (2014), 36 :1, 86-103.

- [36] I. Moret, *A Note on Krylov Methods for Fractional Evolution Problems*, Numerical Functional Analysis and Optimization (2014), 34 :5, 539-556.
- [37] I. Moret, M. Popolizio, *The restarted shift-and-invert Krylov method for matrix functions*, Numer. Linear Algebra Appl. (2014), 21 :68-80.
- [38] I. Moret, *Rational Lanczos approximations to the matrix square root and related functions*, Linear Algebra Appl. (2009), 16 :431-445.
- [39] J. R. Higgins. *Completeness and basis properties of sets of special functions*, Cambridge University Press (1977).
- [40] J. Hadamard, *Lecture note on Cauchy's problem in linear partial differential equations*, Yale Uni Press, New Haven, 1923.
- [41] J. R. Cannon, *Determination of an unknown heat source from overspecified boundary data*, SIAM J. Numer. Anal. 5 (1968), 275-286.
- [42] J.R. Cannon, *One Dimensional Heat Equation*, Addison-Wesley Publishing Company, California (1984).
- [43] J. Liesen, Z. Strakos, *Krylov Subspace Methods : Principles and Analysis*, Oxford University Press (2013).
- [44] J.V. Beck, B. Blackwell, C.R. Clair, *Inverse heat conduction : ill-posed problems*. New York, NY : John Wiley (1985).
- [45] J. Zhaoa, S. Liua, *Central difference regularization method for inverse source problem on the Poisson equation*, Complex Variables and Elliptic Equations : An International Journal, DOI : 10.1080/17476933.2014.940926.
- [46] K.A. Ames, B. Straughan, *Non-Standard and Improperly Posed Problems*, Academic Press (1997).
- [47] K. Schmüdgen, *Unbounded Self-adjoint Operators on Hilbert Space*, Springer (2012).
- [48] K.V. Korshun, *On some Inverse Problem for a Parabolic Equation with a Parameter*, Journal of Siberian Federal University. Mathematics and Physics (2015), 8(3), 281-290.
- [49] L. Eldén and V. Simoncini, *A numerical solution of a Cauchy problem for an elliptic equation by Krylov subspaces*, Inverse Problems 25 (2009) 065002 (22pp).
- [50] L. Reichel, F. Qiang, *Tikhonov regularization based on generalized Krylov subspace methods*, Numerical Analysis and Scientific Computation with Applications (2012), Vol.62, Issue 9, 1215-1228.
- [51] M. Andrie, A. El Badia, *On an inverse source problem for the heat equation : Application to a pollution detection problem*, Inverse Problems in Science and Engineering (2015), 23 :3, 389-412.
- [52] M. Grimmonprez, M. Slodicka, *Reconstruction of an unknown source parameter in a semilinear parabolic problem*, J.C.A.M (2015), <http://dx.doi.org/10.1016/j.cam.2014.12.027>.
- [53] M.M. Lavrentev, V.G. Romanov and G.P. Shishatskii, *Ill-posed Problems in Mathematical Physics and Analysis*, Providence, RI : American Mathematical Society 1986.
- [54] M. Hanke, *On Lanczos based methods for the regularization of discrete ill-posed problems*, BIT Numerical Mathematics (2001), 41 :1008-1018.
- [55] M.M. Lavrentiev and L.Ya. Saveliev, *Operator theory and ill-posed problems*, Inverse and Ill-Posed Problems Series. VSP, Leiden Boston (2006).

- [56] M. Zaslavsky, V. Druskin, L. Knizhnerman, *Solution of 3D time-domain electromagnetic problems using optimal subspace projection : Geophysics* (2011), Vol. 76, 339-351.
- [57] M. Caliari, P. Kandolf, A. Ostermann, S. Rainer, *Comparison of software for computing the action of the matrix exponential*, BIT Numerical Mathematics (2014), Vol.54, Issue 1, 113–128.
- [58] M. Slodicka, *A parabolic inverse source problem with a dynamical boundary condition*, Applied Mathematics and Computation 256 (2015), 529-539.
- [59] M. Slodicka, B. Tomas Johansson, *Uniqueness and counterexamples in some inverse source problems*, Appl. Math. Lett. (2016), <http://dx.doi.org/10.1016/j.aml.2016.02.001>.
- [60] M. Slodicka, *A source identification problem in linear parabolic problems : A semigroup approach*, J. Inverse Ill-Posed Probl. 21 (2013), 579-600.
- [61] N. Boussetila, *Etude de Problèmes Non Locaux et Régularisation de Problèmes Mal Posés en EDP*, Thèse de Doctortat, U. BM-Annaba (2006).
- [62] N.J. Higham, *Functions of Matrices : Theory and Computation*, SIAM (2008).
- [63] N.Trefethen, *Approximation Theory and Approximation Practice*, University of Oxford (2011).
- [64] N. Dunford and J. Schwartz, *Linear Operators, Part II*, John Wiley and Sons, Inc., New York (1967).
- [65] O.M. Alifanov, *Inverse Heat Transfer Problems*. Berlin-Heidelberg : Springer-Verlag (1994).
- [66] P. Brianzi, P. Favati, O. Menchi, F. Romani, *A framework for studying the regularizing properties of Krylov subspace methods*, Inverse Problems (2006), Vol.ume 22, Number 3.
- [67] P. Lévy-Bruhl; *Introduction à la Théorie Spectrale : Cours et Exercices Corrigés*, Dunod (2003).
- [68] P.N. Vabishchevich, V.I. Vasilev, M.V. Vasileva, D.Ya. Nikiforov, *Numerical Solution of an Inverse Filtration Problem*, Lobachevskii Journal of Mathematics (2016), Vol. 37, No. 6, 777-786.
- [69] P. Novati, M. Redivo-Zaglia, M.R. Russo, *Preconditioning linear systems via matrix function evaluation*, Applied Numerical Mathematics 62 (2012), 1804-1818.
- [70] R. Dautray, J.-L. Lions; *Analyse mathématique et calcul numérique. Tome 5 (spectre des opérateurs)*, Edt. Masson, (1988). [§3. page 136-180].
- [71] R. Garrappa, I. Moret, M. Popolizio, *Solving the time-fractional Schrödinger equation by Krylov projection methods*, J. Comput. Phys. (2014), jcp.2014.09.023.
- [72] R. Herzog, E. Sachs, *Superlinear convergence of Krylov subspace methods for self-adjoint problems in Hilbert space*, SIAM J. NUMER. ANAL. (2015), Vol. 53, No. 3, 1304-1324.
- [73] R. Kress, *Linear Integral Equations*, vol. 82 of Applied Mathematical Sciences, Springer (1989).
- [74] R. Lattès, J.-L. Lions, *The method of quasi-reversibility. Applications to partial differential equations*, Elsevier, New York (1969).
- [75] S. Meziani, *Etude d'un Problème de Cauchy Mal Posé avec des Conditions de Dirichlet*, Mémoire de Magister, U. Guelma (2010).
- [76] Stefan Paszkowski, *Computation applications of Chebyshev polynomials and series*, Moscow, Nauka (1983).

- [77] S. Güttel, Convergence estimates of Krylov subspace methods for the approximation of matrix functions using tools from potential theory, M. Thesis, Technische Universität Freiberg (2006).
- [78] S. Güttel, Rational Krylov Methods for operator functions, Ph.D. Thesis, Technischen Universität Bergakademie Freiberg (2010).
- [79] S. D'haeyer, B. Johansson, M. Slodicka, *Reconstruction of a spacewise dependent heat source in a time-dependent heat diffusion process*, IMA Journal of Applied Mathematics 79 (1) (2014) 33-53.
- [80] S. Gazzola, E. Onunwor, L. Reichel, G. Rodriguez, *On the Lanczos and Golub–Kahan reduction methods applied to discrete ill-posed problems*, Numer. Linear Algebra Appl. (2016), 23 :187-204.
- [81] S. Gazzola, P. Novati, M.R. Russo, *On Krylov projection methods and Tikhonov regularization method*, Electronic Transactions on Numerical Analysis (2015), 44 :83-123.
- [82] S. Hamida, Etude de certains Problèmes inverses elliptiques et paraboliques, Thèse de Doctorat, U.BM-Annaba (2015).
- [83] S.I. Kabanikhin and M. Schieck, *Impact of conditional stability : convergence rates for general linear regularization methods*, J. Inverse Ill-Posed Probl., (2008) 16(3) :267-282.
- [84] T. Johansson, D. Lesnic, *A variational method for identifying a spacewise dependent heat source*, IMA J. Appl. Math. 72 (2007), 748-760.
- [85] T. Johansson and D. Lesnic, *Determination of a spacewise dependent heat source*, J. Comput. Appl. Math. 209 (2007), 66-80.
- [86] T. Gockler, V. Grimm, *Convergence analysis of an extended Krylov subspace method for the approximation of operator functions in exponential integrators*, SIAM J. NUMER. ANAL. (2013), Vol. 51, No. 4, pp. 2189-2213.
- [87] V.A. Kozlov, V.G. Maz'ya, *On iterative procedure for solving ill-posed boundary value problems that preserve differential equations*, Leningrad Math J., (1990) 1, 1207-1228.
- [88] V. Druskin, L. Knizhnerman, *Two Polynomial methods of calculating functions of symmetric matrices*. Comput. Maths. Math. Phys., Vol.29, No.6, (1989),112-121.
- [89] V. Druskin, L. Knizhnerman, *Spectral approach to solving three-dimensional Maxwell's diffusion equations in the time and frequency domains*, Radio Science, 29(4),(1994), 937-953.
- [90] V. Druskin, L. Knizhnerman, *Krylov subspace approximation of eigenpairs and matrix functions in exact and computer arithmetic*, Numer Linear Algebra Appl. 2 (1995), 205-217.
- [91] V. Druskin, L. Knizhnerman, T. Tamarchenko, S. Kostek, *Krylov subspace reduction and its extensions for option pricing*, Journal of Computational Finance (1997), DOI : 10.21314/JCF.1997.004.
- [92] V. Druskin, L. Knizhnerman, *Extended Krylov subspaces : approximation of the matrix square root and related functions*, SIAM J. Matrix Anal. Appl. 19 (1998), 755-771.
- [93] V. Druskin, R. Remis, M. Zaslavsky, *An Extended Krylov Subspace Model-Order Reduction Technique to Simulate Wave Propagation in Unbounded Domains*, arXiv :1402.7090v1 [math-ph] 27 Feb 2014.
- [94] V. Isakov, Inverse Source Problems, Math. Surveys Monograph. 34, American Mathematical Society, Providence (1990).

- [95] X.X. Li, H.Z. Guo, S.M. Wan, F. Yang, *Inverse Source Identification by the Modified Regularization Method on Poisson Equation*, Journal of Applied Mathematics (2012), Article ID 971952.
- [96] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM (2003).
- [97] Z. Zhao, O. Xie, Z. Meng, L. You, *Determination of an Unknown Source in the Heat Equation by the Method of Tikhonov Regularization in Hilbert Scales*, Journal of Applied Mathematics and Physics (2014), 2, 10-17.
- [98] Z. Yi, D.A. Murio, *Source Term identification in 1-D IHCP*, Comput. Math. Appl. 47 (2004), 1921-1933.
- [99] Z. Bai, W. Gao, Y. Su, *Matrix Functions and Matrix Equations*, Series in Contemporary Applied Mathematics : Vol 19 (2015).



International Journal of

Sciences: Basic and Applied Research

Print & Online

Published by:



Visit: www.gssrr.org

ISSN 2307-4531 (Print & Online)