HAL
open science

Méthodes numériques de calcul des valeurs propres et vecteurs propres d'un opérateur linéaire

Françoise Chatelin-Laborde

To cite this version:

Françoise Chatelin-Laborde. Méthodes numériques de calcul des valeurs propres et vecteurs propres d'un opérateur linéaire. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1971. tel-00282822

HAL Id: tel-00282822

https://theses.hal.science/tel-00282822v1

Submitted on 28 May 2008

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

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

n° d'ordre

T H E S E

présentée à

L'UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

pour obtenir

LE GRADE DE DOCTEUR ES SCIENCES MATHEMATIQUES

par

FRANCOISE CHATELIN-LABORDE

METHODES NUMERIQUES DE CALCUL DES VALEURS PROPRES ET VECTEURS PROPRES D'UN OPERATEUR LINEAIRE

Thèse soutenue le 12 mars 1971, devant la Commission d'Examen :

M.	*J. KUNTZMANN*	*Président*
MM.	*J. CEA*	
	N. GASTINEL	*Examinateurs*
	P. MARONI	

LISTE DES PROFESSEURS

Doyen Honoraire : Monsieur M.MORET
Doyen : Monsieur E.BONNIER

PROFESSEURS TITULAIRES

MM.	NEEL Louis	Physique Expérimentale
	KRAVTCHENKO Julien	Mécanique Rationnelle
	CHABAUTY Claude	Calcul différentiel et intégral
	BENOIT Jean	Radioélectricité
	CHENE Marcel	Chimie Papetière
	FELICI Noël	Electrostatique
	KUNTZMANN Jean	Mathématiques Appliquées
	BARBIER Reynold	Géologie Appliquée
	SANTON Lucien	Mécanique des Fluides
	OZENDA Paul	Botanique
	FALLOT Maurice	Physique Industrielle
	KOSZUL Jean-Louis	Mathématiques
	GALVANI Octave	Mathématiques
	MOUSSA André	Chimie Nucléaire
	TRAYNARD Philippe	Chimie Générale
	SOUTIF Michel	Physique Générale
	CRAYA Antoine	Hydrodynamique
	REULOS René	Théorie des Champs
	BESSON Jean	Chimie Minérale
	AYANT Yves	Physique Approfondie
	GALLISSOT François	Mathématiques
Mlle	LUTZ Elisabeth	Mathématiques
MM.	BLANBERT Maurice	Mathématiques
	BOUCHEZ Robert	Physique Nucléaire
	LLIBOUTRY	Géophysique
	MICHEL Robert	Minéralogie et pétrographie
	BONNIER Etienne	Electrochimie et Electrométallurgie
	DESSAUX Georges	Physiologie Animale
	PILLET Emile	Physique Industrielle-Electrotechnique
	YOCCOZ Jean	Physique Nucléaire théorique
	DEBELMAS Jacques	Géologie Générale
	BERBER Robert	Mathématiques
	PAUTENET René	Electrotechnique
	MALGRANGE Bernard	Mathématiques Pures
	VAUQUOIS Bernard	Calcul Electronique
	BARJON Robert	Physique Nucléaire
	BARBIER Jean-Claude	Physique
	SILBERT Robert	Mécanique des Fluides
	BUYLE-BODIN Maurice	Electronique
	DREYFUS Bernard	Termodynamique

MM.	KLEIN Joseph	Mathématiques
	VAILLANT François	Zoologie et Hydrobiologie
	ARNAUD Paul	Chimie
	SENGEL Philippe	Zoologie
	BARNOUD Fernand	Biosynthèse de la Cellulose
	BRISSONNEAU Pierre	Physique
	GAGNAIRE Didier	Chimie Physique
Mme	KOFLER Lucie	Botanique
MM.	DEGRANGE Charles	Zoologie
	PEBAY-PEROULA Jean Claude	Physique
	RASSAT André	Chimie Systématique
	DUCROS Pierre	Cristallographie Physique
	DODU Jacques	Mécanique Appliquée I.U.T.
	ANGLES D'AURIAC Paul	Mécanique des Fluides
	LACAZE Albert	Thermodynamique
	GASTINEL Noël	Analyse Numérique
	GIRAUD Pierre	Géologie
	PERRET René	Servo-mécanisme
	PAYAN Jean Jacques	Mathématiques Pures

PROFESSEURS SANS CHAIRE

MM.	GIDON Paul	Géologie
Mme	BARBIER Marie-Jeanne	Electrochimie
Mme	SOUTIF Jeanne	Physique
MM.	COHEN Joseph	Electrotechnique
	DEPASSEL R.	Mécanique des Fluides
	GLENAT René	Chimie
	BARRA Jean	Mathématiques Appliquées
	COUMES André	Electronique
	PERRIAUX Jacques	Géologie et Minéralogie
	ROBERT André	Chimie Papetière
	BIARREZ Jean	Mécanique Physique
	BONNET Georges	Electronique
	CAUQUIS Georges	Chimie Générale
	BONNETAIN Lucien	Chimie Minérale
	DEPOMMIER Pierre	Physique Nucléaire-Génie Atomique
	HACQUES Gérard	Calcul Numérique
	POLOUJADOFF Michel	Electrotechnique
Mme	KAHANE Josette	Physique
Mme	BONNIER Jane	Chimie
MM.	VALENTIN Jacques	Physique
	REBECQ Jacques	Biologie
	DEPORTES Charles	Chimie
	SARROT-REYNAULD Jean	Géologie
	BERTRANDIAS Jean Paul	Mathématiques
	AUBERT Guy	Physique

PROFESSEURS ASSOCIES

MM.	RODRIGUES Alexandre	Mathématiques Pures
	MORITA Susumu	Physique Nucléaire
	RADHAKRISHNA	Thermodynamique

MAITRES DE CONFERENCES

MM.	LANCIA Roland	Physique Atomique
Mme	BOUCHE Liane	Mathématiques
MM.	KAHANE André	Physique Générale
	DOLIQUE Jean Michel	Electronique
	BRIERE Georges	Pyysique
	DESRE Georges	Chimie
	LAJZEHOWICZ Joseph	Physique
	LAURENT Pierre	Mathématiques Appliquées
Mme	BERTRANDIAS Françoise	Mathématiques Pures
MM.	LONGEQUEUE Jean Pierre	Physique
	SOHM Jean Claude	Electrochimie
	ZADWORNY François	Electronique
	DURAND Françis	Chimie Physique
	CARLIER Georges	Biologie Végétale
	PFISTER Jean Claude	Physique
	CHIBON Pierre	Biologie Animale
	IDELMAN Simon	Physiologie animale
	BLOCH Daniel	Electrotechnique I.P.
	MARTIN-BOUYER Michel	Chimie (C.S.U.Chambery)
	SIBILLE Robert	Construction mécanique (I.U.T.)
	BRUGEL Lucien	Energétique I.U.T.
	BOUVARD Maurice	Hydrologie
	RICHARD Lucien	Botanique
	PELMONT Jean	Physiologie animale
	BOUSSARD Jean Claude	Mathématiques Appliquées(I.P.G.)
	MOREAU René	Hydraulique I.P.G.
	ARMAND Yves	Chimie I.U.T.
	BOLLIET Louis	Informatique I.U.T.
	KUHN Gérard	Energétique I.U.T.
	PEFFEN René	Chimie I.U.T.
	GERMAIN Jean Pierre	Mécanique
	JOLY Jean René	Mathématiques Pures

Mlle	PIERY Yvette	Biologie Animale
MM.	BERNARD Alain	Mathématiques Pures
	MOHSEN Tahsin	Biologie (C.S.U.Chambery)
	CONTE René	Mesures Physiques I.U.T.
	LE JUNTER Noël	Génie Electrique Electronique I.U.T.
	LE ROY Philippe	Génie Mécanique I.U.T.
	ROMIER Guy	Technique Statistiques Quantitatives I.U.T
	VIALON Pierre	Géologie
	BENZAKEN Claude	Mathématiques Appliquées
	MAYNARD Roger	Physique
	DUSSAUD René	Mathématiques (C.S.U.Chambery)
	BELORIZKY Elie	Physique (C.S.U Chambery)
Mme	LAJZEROWICZ Jeanine	Physique (C.S.U Chambery)
M.	JULLIEN Pierre	Mathématiques Pures
Mme	RINAUDO Marguerite	Chimie
MM.	BLIMAN Samuel	E.I.E.
	BEGUIN Claude	Chimie Organique
	NEGRE Robert	I.U.T.

MAITRES DE CONFERENCES ASSOCIES

MM.	YAMADA Osamu	Physique du Solide
	NAGAO Makoto	Mathématiques Appliquées
	MAREZIO Massimo	Physique du Solide
	CHEECKE John	Thermodynamique
	BOUDOURIS Georges	Radioélectricité
	ROZMARIN Georges	Chimie Papetière

TABLE DES MATIERES

INTRODUCTION ---------- I-1

NOTATIONS ---------- I-6

CHAPITRE - I

CALCUL DES VALEURS PROPRES D'UNE MATRICE : ALGORITHME MR

I Factorisation MR relative à N ---------- 1

1 - Notations ---------- 1

2 - Factorisation MR ---------- 1

3 - Forme de Hessenberg ---------- 3

4 - Lemme ---------- 6

II Algorithme MR relatif à une suite $\{P_s\}_{s\in\mathbb{N}}$ ---------- 7

1 - Définition ---------- 7

2 - Propriétés ---------- 7

3 - Convergence lorsque la suite $\{P_s\}$ est donnée ---------- 8

III Etude de la convergence de l'algorithme MR au 1er ordre -------- 11

1 - Convergence de $(M_1 \ldots M_s)$ vers M_X au 1er ordre ---------- 11

2 - Comparaison avec la convergence de LR ---------- 13

IV Accélération de la convergence des algorithmes LR et QR -------- 14

1 - Algorithme LR modifié ---------- 15

2 - Algorithme QR modifié ---------- 16

3 - Expériences numériques et procédures ALGOL ---------- 17

CHAPITRE - II

PERTURBATION D'UNE MATRICE HERMITIENNE - APPLICATION AU CALCUL DE VALEURS ET VECTEURS PROPRES

I Perturbation d'une matrice hermitienne ---------- 25

1 - Préliminaire ---------- 25

2 - Majorant de $|\lambda'-(A'\varphi,\varphi)|$ ---------- 27

3 - Cas particulier où $(H\varphi,\varphi) = 0$ ---------- 33

4 - Valeurs propres multiples ---------- 35

5 - Matrices normales ---------- 41

II Application au calcul des éléments propres. Résultats numériques

1 - Algorithmes QR et Jacobi ---------- 43

2 - Valeurs propres simples séparées ou voisines. Valeurs propres multiples ---------- 51

3 - Application pratique ---------- 58

CHAPITRE - III

PROBLEME INVERSE D'UN PROBLEME DE VALEURS PROPRES

I Une condition nécessaire et suffisante ---------- 64

II Etude de (P) lorsque A est hermitienne ---------- 65

1 - Conditions suffisantes pour que (P) admette une solution ----- 65

2 - Convergence de la méthode des approximations successives ----- 70

3 - Exemple ---------- 73

III A est réelle ---------- 74

IV Généralisation de (P) à une variété linéaire ---------- 75

CHAPITRE - IV

ETUDE DE LA STABILITE DES METHODES D'APPROXIMATION D'ELEMENTS PROPRES D'OPERATEURS LINEAIRES DANS UN ESPACE DE BANACH

I Rappels ---------- 80

II Définitions ---------- 81

III Approximation d'un ensemble fini de valeurs propres isolées ------ 83

1 - Notations ---------- 83

2 - Définition et lemmes ---------- 83

3 - Théorème ---------- 87

4 - Exemples d'instabilité ---------- 88

5 - Continuité de la notion de stabilité ---------- 91

IV Approximation des sous-espaces propres et invariants ---------- 93

1 - Approximation des sous-espaces invariants ---------- 93

2 - Approximation des sous-espaces propres ---------- 95

3 - Cas où dim $P_h X_h \leq$ dim PX : stabilité forte ---------- 96

V Cas particuliers dans un espace de Hilbert ---------- 102

1 - T_h est normal ---------- 103

2 - T et T_h sont autoadjoints ---------- 104

VI Conditions suffisantes de stabilité ---------- 113

1 - Approximation uniforme d'un opérateur borné ---------- 113

2 - Approximation supérieure d'un opérateur autoadjoint fermé ---- 119

VII Exemple simple de calcul de la constante de stabilité ---------- 130

1 - Approximation ---------- 130

2 - Stabilité ---------- 131

Suite CHAPITRE - IV

VIII Méthodes de projection ---------- 132
1 - Définitions ---------- 132
2 - Méthodes classiques ---------- 134
3 - Stabilité ---------- 134

CHAPITRE - V

BORNES D'ERREUR DANS L'APPROXIMATION STABLE DES ELEMENTS PROPRES D'UN OPERATEUR LINEAIRE

I Opérateurs fermés dans un espace de Banach ---------- 137
1 - Résultat général ---------- 137
2 - Cas d'une restriction stable ---------- 138

II Opérateurs autoadjoints dans un Hilbert ---------- 143
1 - Perturbation d'un opérateur autoadjoint ---------- 143
2 - Application aux méthodes d'approximation ---------- 145
3 - Inégalité de Wielandt-Hoffmann pour des opérateurs compacts autoadjoints ---------- 149

CHAPITRE - VI

ALGORITHMES LR ET QR APPLIQUES A CERTAINES CLASSES D'OPERATEURS LINEAIRES DANS UN ESPACE DE HILBERT

I Isomorphismes dans un Hilbert ---------- 155
1 - Définitions et notations ---------- 155
2 - Conditions suffisantes pour que T soit un isomorphisme ---------- 155
3 - Cas d'un opérateur triangulaire ---------- 157

Suite CHAPITRE - VI

II Décomposition de Gauss et Choleski ---------- 159

1 - Décomposition de Gauss ---------- 159

2 - Décomposition de Choleski ---------- 166

III Algorithme LR ---------- 167

1 - Algorithme ---------- 167

2 - Convergence ---------- 168

3 - CS pour que les vecteurs propres de T forment une base -------- 171

4 - Application au calcul de T_∞ ---------- 172

IV Décomposition de Schmidt, algorithmes QR et Choleski ---------- 173

1 - Décomposition de Schmidt ---------- 173

2 - Algorithme QR ---------- 176

3 - Algorithme de Choleski ---------- 179

4 - Application au calcul de T_∞ ---------- 180

BIBLIOGRAPHIE 181

Cette thèse a été préparée à l'Institut de Mathématiques Appliquées de Grenoble.

Monsieur KUNTZMANN a bien voulu accepter de présider le Jury : je lui exprime ma respectueuse gratitude. Qu'il trouve ici un témoignage de ma reconnaissance pour l'existence même de cet Institut et pour la foi avec laquelle il oeuvre pour le développement des Mathématiques Appliquées.

Ma reconnaissance particulière va à Monsieur GASTINEL qui anime l'équipe d'Analyse Numérique avec une ardeur incessante. Il a su, par son enseignement, me donner le goût de l'Analyse Numérique. Ses conseils, ses encouragements et ses critiques ne m'ont pas manqué.

Je remercie Messieurs CEA et MARONI qui m'ont fait l'honneur de venir à Grenoble participer au Jury.

Je remercie aussi tous mes camarades de l'équipe d'Analyse Numérique qui m'ont aidé, en particulier Madame A. DESSENNE qui avec une patience toujours souriante a mis au point de nombreux programmes.

Je tiens beaucoup à remercier mes parents qui ont su former et entretenir mon goût pour les mathématiques.

Enfin, que Mademoiselle G. BICAÏS et le Service Tirage soient remerciés pour le soin et la gentillesse apportés à la réalisation matérielle de cette thèse, dans un temps très bref.

METHODES NUMERIQUES DE CALCUL DES VALEURS PROPRES ET VECTEURS PROPRES D'OPERATEURS LINEAIRES

INTRODUCTION

Un problème fréquemment posé au mathématicien appliqué consiste à résoudre un problème d'éléments propres :

T étant un opérateur linéaire dans un espace de dimension finie ou infinie, trouver, s'ils existent, les vecteurs $x \neq 0$ et les scalaires λ tels que :

$$Tx = \lambda x.$$

λ et x sont éléments propres de T : λ est valeur propre et x vecteur propre.
Sauf dans le cas où la résolution est explicite par les moyens de l'analyse classique, lorsque la dimension de l'espace est infinie, les méthodes numériques employées consistent à remplacer l'opérateur par un opérateur "approché" dans un espace de dimension finie, dont on calcule les éléments propres.

Deux problèmes se posent alors :

1) disposer d'algorithmes de calcul des éléments propres de matrices et connaître la précision obtenue,

2) préciser la notion d'opérateur "approché" de manière à ce que les éléments propres de cet opérateur soient des approximations des éléments propres cherchés.

Le travail qui suit a deux pôles principaux : l'étude dans $\mathbb{C}^n$ d'algorithmes de calcul des éléments propres (chapitres 1 et 2), et l'étude dans un espace de Banach de méthodes d'approximation des éléments propres d'opérateurs (chapitres 4 et 5).

Plus précisément, le premier chapitre est consacré à la présentation d'une décomposition matricielle généralisant la décomposition de Gauss en produit d'une matrice triangulaire inférieure par une matrice triangulaire supérieure, ainsi que la décomposition de Schmidt en produit d'une matrice unitaire par une matrice triangulaire supérieure. On en déduit un algorithme (noté MR) de calcul des éléments propres analogue aux méthodes LR de Rutishauser et QR de Francis. Cet algorithme fait intervenir des paramètres pour lesquels on propose un choix qui rende la convergence plus rapide que celle de LR.

D'après un théorème de Householder, toute matrice est unitairement semblable à une matrice de Hessenberg supérieure (la partie inférieure de la matrice, jusqu'à la première sous-diagonale non inclue est constituée de 0). Les algorithmes LR, QR, MR sont systématiquement appliqués à des matrices préalablement mises sous forme de Hessenberg supérieure, ce qui présente le double avantage suivant : du point de vue théorique, on simplifie les conditions de convergence des algorithmes, du point de vue pratique on réduit considérablement le nombre d'opérations à effectuer à chaque itération ainsi que le nombre de mot-mémoires nécessaires en machine. Une étude au 1er ordre de la convergence de LR et QR permet de proposer des algorithmes modifiés dont la convergence est du 2ème ordre. Cette accélération de la convergence peut se combiner avec les translations d'origine de Wilkinson [71].

Dans le deuxième chapitre, on étudie, pour des matrices hermitiennes ou normales, la précision obtenue sur les éléments propres en les calculant à l'aide d'algorithmes tels que QR ou Jacobi qui forment une suite de matrices unitairement semblables, suite qui converge vers une matrice diagonale. Dans ce but, on considère $A' = A+H$, où A est une matrice d'éléments propres connus, et H une matrice de norme "petite". On approche les éléments propres $\{\lambda',\varphi'\}$ de A' par $\{\lambda,\varphi\}$ éléments propres de A. On obtient pour $|\lambda'-\lambda|$ et $\|\varphi'-\varphi\|$ des majorations en fonction de $\|H\|$ ou de $\|H\varphi\|$, ainsi que de la distance de λ aux autres valeurs propres de A.

On retrouve au passage, en les replaçant dans le même cadre, des résultats bien connus de Davis |8|, Gavurin |14|, Kato |24|, Golub |15|, Ortega |11| et Wilkinson |70|. Les majorants obtenus sont réalistes et aisément calculables au cours de l'algorithme.

En étudiant la sensibilité de la précision obtenue sur les valeurs propres à la distance de séparation entre ces valeurs propres, on a mis en évidence le phénomène suivant : soit à calculer la valeur propre λ lorsque λ est une valeur propre double, ou lorsque la distance de séparation d est petite, il faut alors arrêter l'itération relative à λ lorsque $\|H\varphi\|$ est suffisamment grand ou petit par rapport à d, car la précision sur λ reste à peu près stationnaire lorsque $\|H\varphi\|$ est voisin de d.

Le troisième chapitre traite d'un problème inverse du problème classique de recherche des valeurs propres : étant donné une matrice A, hermitienne si elle est complexe, ou réelle non nécessairement symétrique, existe-t-il une diagonale réelle V telle que A+V admette des valeurs propres réelles données. Ce problème d'ajustement de paramètres dans une matrice dont on connait les valeurs propres se rencontre assez souvent en physique (spectrométrie physique, physique nucléaire, théorie des réseaux, systèmes dynamiques en chimie physique). On donne des conditions suffisantes pour que ce problème admette au moins une solution ; ces conditions recouvrent partiellement celles données par Hadeler [18] et Oliveira [44].

Pour l'étude des opérateurs linéaires dans un espace de dimension infinie, nous avons choisi le cadre général de la théorie des opérateurs fermés dans un espace de Banach séparable, cadre dans lequel figurent les principales classes d'opérateurs différentiels ou intégraux de la physique mathématique. La notion d'opérateur approché considérée (Aubin [4]) recouvre les méthodes d'approximation (différences finies, interpolation,...) et les méthodes de projection (Ritz, Galerkin, Aronzajn...).

Le quatrième chapitre présente une étude de la stabilité des méthodes d'approximation des éléments propres d'opérateurs linéaires dans un espace de Banach : on donne des conditions suffisantes pour que les valeurs propres et les sous-espaces invariants (dans certains cas les vecteurs propres) de l'opérateur approché convergent vers les valeurs propres et sous-espaces invariants de l'opérateur considéré. On trouve dans la littérature la convergence de méthodes particulières ou appliquées à certaines classes particulières d'opérateurs (cf. par exemple [1, 23, 26, 28, 36, 41, 48, 51]). La notion de stabilité, qui est générale, recouvre celle d'opérateur régulier introduite par Polskii [52] et développée par Medvedev [37] pour la convergence de la méthode de Galerkin appliquée à des opérateurs dans un Hilbert.

La stabilité forte entraîne l'approximation des valeurs propres avec conservation de l'ordre de multiplicité, ainsi que l'approximation uniforme des sous-espaces invariants. Cette condition est d'un emploi simple pour les opérateurs autoadjoints.

On donne des conditions suffisantes de stabilité et de stabilité forte. Entrent dans le premier cas les approximations uniformes d'opérateurs compacts, les approximations compactes (Anselone [3], Vainikko [67]) d'opérateurs bornés, et dans le deuxième cas, certaines approximations d'opérateurs autoadjoints dans un Hilbert, compacts ou bornés inférieurement à résolvante compacte.

Dans le cinquième chapitre nous étudions la précision obtenue par une méthode d'approximation stable. Là encore, il n'existe que des résultats particuliers, citons par exemple les travaux de Saulev [56], Keller [26] pour des méthodes aux différences finies appliquées à des problèmes elliptiques, ceux de Krasnoselskii-Vainikko [28] pour la méthode de Galerkin appliquée à des opérateurs compacts. Nous donnons un résultat qui étend la formule valable pour des matrices. Pour des opérateurs autoadjoints bornés dans un espace de Hilbert, nous pouvons étendre les formules du chapitre II et obtenir des majorations satisfaisantes.

Dans le sixième chapitre, nous étudions dans un espace de Hilbert les deux problèmes suivants, importants en analyse numérique linéaire : décomposition de Gauss d'une matrice, en vue du calcul de l'inverse et recherche de la forme triangulaire de Schur à l'aide de matrices unitaires, en vue du calcul des valeurs propres. Ces deux problèmes ont été traités pour des opérateurs dans un espace de Hilbert. Göhberg et Krein [16] pour le premier, Lifchitz [35] pour le second, ont défini des classes d'opérateurs pour lesquels ces problèmes admettent une solution. Nous abordons ces problèmes dans l'optique suivante : pour certaines classes d'opérateurs à spectre discret dans un espace de Hilbert nous définissons les décompositions de Gauss et de Schmidt et les algorithmes LR et QR. L'algorithme QR permet en effet de déterminer l'opérateur unitaire à l'aide duquel l'opérateur considéré est semblable à un opérateur triangulaire.

NOTATIONS

Nous précisons les notations couramment utilisées dans le texte.

Dans l'ensemble $\mathcal{M}_{nm}(\mathbb{C})$ des matrices à n lignes et m colonnes, à éléments complexes :

$A = (a_{ij})$ désigne la matrice d'élément général a_{ij} , $i = 1,\dots,n$, $j = 1,\dots,m$.

$A_{.i} = (a_{ki})$, $k = 1,\dots,n$ est le $i^{\text{ème}}$ vecteur colonne de A.

$A_{i.} = (a_{ik})$, $k = 1,\dots,m$ est le $i^{\text{ème}}$ vecteur ligne de A.

E_{ij} est la matrice de $\mathcal{M}_{nm}$ admettant un 1 en $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne et des 0 ailleurs.

Si $n = m$, $\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$ est la trace de A.

$A^H = (\bar{a}_{ji})$ est la matrice transposée conjuguée de A.

$\rho(A) = \max(|\lambda|$; λ valeur propre de A) est le rayon spectral de A.

On utilise les normes de matrices suivantes :

$$S_{\varphi\varphi}(A) = \max_{\substack{x \neq 0 \\ x \in \mathbb{C}^n}} \frac{\varphi(Ax)}{\varphi(x)}$$, où φ est une norme sur $\mathbb{C}^n$.

$\varphi_p(x)$ désigne la norme de Hölder :

$$\varphi_p(x) = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$, si $x = (x_i)$,

$$S_{\varphi_p\varphi_p} = \|A\|_p.$$

$$|||A||| = \operatorname{tr}^{1/2}(A^H A) = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}$$ est la norme de Frobenius de A.

CHAPITRE - I

CALCUL DES VALEURS PROPRES D'UNE MATRICE : ALGORITHME MR

Plusieurs méthodes numériques de calcul des valeurs propres d'une matrice sont fondées sur le comportement asymptotique des puissances successives de cette matrice: méthode de la puissance [12], méthode LR [71] de Rutishauser, méthode QR [71] de Francis, par exemple. Nous présentons ici un algorithme basé sur la factorisation MR relative à N d'une matrice, factorisation qui généralise celle de Gauss et de Schmidt. Au cours de l'algorithme intervient, en paramètre, une suite convergente de matrices données, que l'on choisira de manière à accélérer si possible la convergence de l'algorithme.

I - FACTORISATION MR RELATIVE A N

1 - Notations

$\mathcal{M}_{n,n}(\mathbb{C})$ est l'ensemble des matrices carrées d'ordre n à éléments complexes.
I est la matrice unité de $\mathcal{M}_{n,n}$.
On note R une matrice triangulaire supérieure, L une matrice triangulaire inférieure à diagonale unité, et Q une matrice unitaire.

2 - Factorisation MR

A est une matrice régulière donnée.

Définition

A est décomposée en MR relativement à N, matrice régulière donnée ou calculée, s'il existe des matrices R et L telles que

$$A = MR \quad , \quad NM = L.$$

Explicitons ceci. Ecrire que A = MR, où R est triangulaire supérieure, équivaut pour les vecteurs colonnes $A_{.j}$ et $M_{.k}$ de A et M, aux relations :

$$A_{.j} = \sum_{k=1}^{j} r_{kj} M_{.k} \qquad (j = 1,\dots,n)$$

$M_{.j}$ appartient au sous-espace vectoriel engendré par $A_{.1},\dots,A_{.j}$.
On achève la détermination des $M_{.j}$ en imposant à M des conditions qui déterminent pour chaque $M_{.j}$, les inconnues r_{ij} , $i \leq j$:

$$N_{i.} \cdot M_{.j} = \delta_{ij} \quad , \quad i \leq j \ ,$$

où δ_{ij} est le symbole de Kronecker.
Ce qui équivaut à la condition :

il existe L telle que NM = L.

La matrice N est donnée, ou construite par l'algorithme de détermination de M.
Comment calculer M et R ? Cela se fait par récurrence, comme on va le voir.
Supposons $M_{.1},\dots,M_{.j-1}$ déterminés, ainsi que $R_{.1},\dots,R_{.j-1}$.

a) N est donnée.
Pour $i \leq j$, $N_{i.} \cdot M_{.j} = \delta_{ij}$ implique :

(1) $$r_{ij} = N_{i.}(A_{.j} - \sum_{k<i} r_{kj} M_{.k})$$

SERVICE POLYCOPIE
MATHÉMATIQUES APPLIQUÉES
Université de GRENOBLE

(2) $$M_{.j} = \frac{1}{r_{jj}} (A_{.j} - \sum_{k<j} r_{kj} M_{.k}) \quad \text{si } r_{jj} \neq 0.$$

La décomposition MR relative à N donnée équivaut à la décomposition de Gauss $NA = LR$, et $M = N^{-1}L$. Elle existe si et seulement si les déterminants principaux de NA sont différents de zéro, et alors, elle est unique.

b) N est construite.

Pour $i < j$, $N_{i.} \cdot M_{.j} = 0$ implique (1).

Pour $i = j$, soit $B_{.j} = A_{.j} - \sum_{k<j} r_{kj} M_{.k}$.

Si $N_{j.} = F_{j.}(M_{.j})$, où $F_{j.}$ est une application de $\mathbb{C}^n$ dans $\mathbb{C}^n$, $N_{j.} \cdot M_{.j} = 1$ entraîne que r_{jj} est racine de

(1') $$F_{j.}\left(\frac{B_{.j}}{r}\right) \frac{B_{.j}}{r} = 1$$

On suppose que cette équation en r a au moins une solution $\neq 0$.

Alors $M_{.j}$ est déterminé par (2).

(1), (2) et (1') éventuellement, sont les formules de calcul de la décomposition MR relative à N.

<u>Remarque</u> - La décomposition MR relative à N est une généralisation des décompositions de Gauss et de Schmidt :

a) si $N = I$ ou $N = L_1$ on retrouve la décomposition de Gauss

b) si $N = M^H$ on retrouve la décomposition de Schmidt :
$M^H M = L$ hermitienne implique $L = I$, M est unitaire.

3 - Forme de Hessenberg

3.1 - Définition

$A = (a_{ij})$ est une matrice de <u>Hessenberg</u> supérieure si :

$$a_{ij} = 0 \quad , \quad i > j+1 \quad , \quad j = 1,\dots,n-2.$$

Toute matrice est unitairement semblable à une matrice de Hessenberg.
La méthode de Householder [21] permet d'effectuer cette similitude à l'aide de matrices unitaires hermitiennes.

3.2 - Propriétés de la forme de Hessenberg supérieure

Soit $H = (h_{ij})$ une matrice de Hessenberg supérieure telle que $A = QHQ^H$.
Les propriétés suivantes sont immédiates :

1) Si $h_{i\ i-1} = 0$, H peut être partitionnée en deux sous-matrices pour le problème des valeurs propres.

2) Si A est diagonalisable, et si $h_{i\ i-1} \neq 0$, $i = 2,\dots,n$, il ne peut y avoir de valeur propre multiple.

Réciproquement, si A est diagonalisable et a des valeurs propres multiples dont l'ordre de multiplicité le plus élevé est m, il y a au moins m-1 éléments $h_{i\ i-1}$ nuls.

En effet, le produit des $h_{i\ i-1}$ est un mineur d'ordre n-1.
S'il est différent de 0, $A-\lambda I$ est de rang n-1 au moins.
Si A admet λ de multiplicité m, $A-\lambda I$ est de rang n-m.

3) Soit X une matrice des vecteurs propres de H : $H = XDX^{-1}$.
Si A est diagonalisable et $h_{i\ i-1} \neq 0$, $i = 2,\dots,n$, X^{-1} est telle que tous les déterminants principaux sont non nuls.

Montrons que, quel que soit $i = 1,2,\dots,n$, aucun mineur d'ordre i formé à l'aide des i premières colonnes de $Y = X^{-1}$ n'est nul.

Soit $Y^{(i)} \in M_{i,n}(\mathbb{C})$, la matrice obtenue à partir de Y en prenant i lignes quelconques de Y. Soit r, $0 \leq r \leq n$, déterminé par la condition : les r-1 premières colonnes de $Y^{(i)}$: $Y^{(i)}_{.1},\dots,Y^{(i)}_{.r-1}$ sont indépendantes, tandis que les r colonnes $Y^{(i)}_{.1},\dots,Y^{(i)}_{.r}$ sont dépendantes.

Soit : $\sum_{k=1}^{r} \alpha_k Y^{(i)}_{.k} = 0$ avec $\alpha_r \neq 0$

ou encore

$$Y^{(i)}\alpha = 0 \text{ avec } \alpha = (\alpha_1,\ldots,\alpha_r,0,\ldots,0) \in \mathbb{C}^n.$$

Soit $D^{(i)} \in M_{i,i}(\mathbb{C})$, la matrice obtenue à partir de D en prenant les i lignes correspondant à $Y^{(i)}$.

Alors $YH = DY$ implique $Y^{(i)}H = D^{(i)}Y^{(i)}$

et $D^{(i)}Y^{(i)}\alpha = Y^{(i)}H\alpha = 0$.

Soit $\beta = H\alpha$, $\beta = (\beta_1,\ldots,\beta_{r+1},0,\ldots,0) \in C^n$ avec $\beta_{r+1} = h_{r+1r}\alpha_r \neq 0$.

Donc $\sum_{k=1}^{r+1} \beta_k Y^{(i)}_{.k} = 0$: $Y^{(i)}_{.r+1}$ est combinaison linéaire de $Y^{(i)}_{.1},\ldots,Y^{(i)}_{.r}$.

Pour tout $j \geq r$, $Y^{(i)}_{.j}$ est combinaison linéaire de $Y^{(i)}_{.1},\ldots,Y^{(i)}_{.r}$.

Or les i lignes de $Y^{(i)}$ sont indépendantes (ce sont les vecteurs propres à droite de H, matrice diagonalisable). Il existe au moins i colonnes de $Y^{(i)}$ indépendantes. Donc $r > i$: les i premières colonnes de $Y^{(i)}$ sont indépendantes. Aucun mineur d'ordre i formé à l'aide des i premières colonnes de Y n'est nul, en particulier le mineur principal d'ordre i.

4) <u>Supposons que A soit une matrice de Hessenberg supérieure et soit A = MR la factorisation de A relative à N. Alors M est une matrice de Hessenberg supérieure, ainsi que la matrice produit RM.</u>

Cela se déduit immédiatement des formules de calcul (1), (1') et (2).

4 - *Lemme*

L'ensemble des matrices régulières de $M_{n,n}(\mathbb{C})$ *admettant une décomposition* MR *relative à* N, *matrice régulière donnée, est un ouvert* Ω *de* $M_{n,n}(\mathbb{C})$. *L'application :* $A \in \Omega \mapsto M \in M_{n,n}(\mathbb{C})$ *est continue dans* Ω.

Démonstration

N étant donnée régulière, la décomposition A = MR équivaut à :

$$NA = LR,\ M = N^{-1}L.$$

$M_{n,n}(\mathbb{C})$ est normé par $\|.\|$.
Soit $A' = A+B$, avec $\|B\| < \varepsilon$
$NA' = NA(I+A^{-1}B) = L(I+RA^{-1}BR^{-1})R$.
Soit $C = RA^{-1}BR^{-1}$, $\|C\| < \|R\|\ \|A^{-1}\|\ \|R^{-1}\|\varepsilon = k.\varepsilon$
Pour tout ε, $\varepsilon < \frac{1}{k}$, I+C est inversible, donc A' régulière.
Il est facile de voir que, pour ε assez petit, I+C vérifie les conditions d'existence de la décomposition de Gauss (déterminants principaux $\neq 0$) :

$$I+C = \ell r.$$

En explicitant ℓ et r en fonction des éléments de I+C, il vient :

$$\|\ell - I\| \leq k_1\ \varepsilon$$

Donc $NA' = L\ell rR = L'.R'$ et $M' = N^{-1}L' = N^{-1}L\ell$

$$\|M'-M\| \leq \|N^{-1}\|\ \|L\|\ \|\ell - I\| = k_2\varepsilon.$$

II - ALGORITHME MR RELATIF A UNE SUITE $\{P_s\}_{s \in \mathbb{N}}$

1 - Définition

Soit une suite de matrices régulières $P_1,\ldots,P_s,\ldots$ données ou construites par l'algorithme, et A une matrice régulière. On forme la suite A_s de matrices semblables à A :

$$A = A_1 = M_1R_1 \text{ relativement à } N_1 = P_1 \quad ; \quad N_1M_1 = L_1$$

$$A_2 = R_1M_1 = M_1^{-1}A_1M_1 \ ;$$

$$= M_2R_2 \text{ relativement à } N_2 = P_2M_1 \ ; \quad N_2M_2 = L_2$$

$$A_s = M_sR_s \text{ relativement à } N_s = P_sM_1\ldots M_{s-1} ; N_sM_s = L_s .$$

L'algorithme LR correspond à $N_s = I$, soit $P_1 = I$, $P_s = L_{s-1}^{-1}$, ou encore $P_s = I$; l'algorithme QR correspond à $N_s = M_s^H$, soit $P_s = (M_1\ldots M_s)^H$.

2 - Propriétés

1) L'algorithme MR conserve la forme de Hessenberg supérieure.
Si $A = A_1$ est une matrice de Hessenberg supérieure, la suite A_s est constituée de matrices de même type. C'est une conséquence de la propriété 4) du § I.3.

2) A_s converge si et seulement si $M_1\ldots M_s$ converge, et $M_1\ldots M_s$ est une matrice de décomposition de A^s relativement à P_s.

En effet : $A_{s+1} = R_sM_s = M_s^{-1}A_sM_s = (M_1\ldots M_s)^{-1}A_1(M_1\ldots M_s)$.

Et : $(M_1 \ldots M_s)R_s \ldots R_1 = (M_1 \ldots M_{s-1})A_s R_{s-1} \ldots R_1$

$$= A_1 (M_1 \ldots M_{s-1}) R_{s-1} \ldots R_1$$

$$= A^s$$

avec : $N_s M_s = P_s (M_1 \ldots M_s) = L_s$.

3) <u>Convergence lorsque la suite $\{P_s\}$ est donnée.</u>

Puisque l'algorithme conserve la forme de Hessenberg supérieure, on supposera toujours la réduction préalable de A à la forme de Hessenberg H effectuée et l'on appliquera l'algorithme à H. Cela réduit considérablement, à chaque itération, le nombre de termes non nuls à déterminer dans la décomposition $MR(n^2 - n\frac{n-1}{2}$ au lieu de n^2 dans le calcul de M), ainsi que le temps d'exécution de la multiplication RM.

Nous étudions donc la convergence de l'algorithme appliqué à H, matrice de Hessenberg supérieure.

$A = QHQ^H$.

On suppose que les valeurs propres λ_i de A sont simples :

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > 0.$$

Alors H, ainsi que A, est diagonalisable :

$$H = XDX^{-1}.$$

D'après la propriété 1 du § I.3.2, si $h_{i\,i-1} = 0$, H peut être partitionnée en deux sous matrices pour le problème des valeurs propres. On suppose donc $h_{i\,i-1} \neq 0$, $i = 2,\ldots,n$. Alors, puisque H est diagonalisable, d'après la propriété 3, $Y = X^{-1}$ a tous ses déterminants principaux non nuls : Y admet la décomposition de Gauss $Y = L_Y . R_Y$.

On suppose que la suite de matrices régulières P_s tend vers P :

$P_s = P + \varepsilon_s$, $\varepsilon_s \xrightarrow[s\to\infty]{} 0$, et que la décomposition $\begin{cases} X = M_X R_X \\ PM_X = L_X \end{cases}$

relative à P existe.

On suppose encore que, pour tout s, la décomposition $\begin{cases} H_s = M_s R_s \\ N_s M_s = L_s \end{cases}$

relative à $N_s = P_s . M_1 \ldots M_{s-1}$ existe.

Montrons alors que $M_1 \ldots M_s$ tend vers M_X.

$H = XDY$

$H^s = XD^s Y = (M_1 \ldots M_s)(R_s \ldots R_1)$ et $P_s(M_1 \ldots M_s) = L_s$.

Alors $P_s H^s = L_s(R_s \ldots R_1)$: c'est une décomposition de Gauss.

D'autre part,

$$P_s H^s = (P+\varepsilon_s)\ X\ D^s\ Y$$
$$= (I+\varepsilon_s P^{-1})PX(D^s\ L_Y\ D^{-s})D^s\ R_Y.$$

$D^s L_Y D^{-s}$ est une matrice triangulaire inférieure à diagonale unité, d'élément général $\ell_{ij}^Y (\frac{\lambda_i}{\lambda_j})^s$ si $L_Y = (\ell_{ij}^Y)$.

Pour $i > j$, $|\lambda_i| < |\lambda_j|$ et $D^s L_Y D^{-s} \xrightarrow[s\to\infty]{} I$. $PX = L_X R_X$. Pour s assez grand, $R_X D^s L_Y D^{-s} R_X^{-1}$ admet la décomposition de Gauss $\ell_s r_s$. D'après la continuité de l'application $A \mapsto L$, $R_X(D^s L_Y D^{-s})R_X^{-1} \xrightarrow[s\to\infty]{} I$ implique $\ell_s \xrightarrow[s\to\infty]{} I$. $P_s H^s$ s'écrit alors :

$$P_s H^s = L_X(I+L_X^{-1}\varepsilon_s P^{-1} L_X)\ell_s r_s R_X D^s R_Y$$
$$I+L_X^{-1}\varepsilon_s P^{-1} L_X = \ell_\varepsilon\ r_\varepsilon,\ \ell_\varepsilon \xrightarrow[s\to\infty]{} I$$
$$r_\varepsilon \ell_s = \ell_s(\ell_s^{-1} r_\varepsilon \ell_s) = \ell_s \ell r\ ,\ \ell \xrightarrow[s\to\infty]{} I.$$

Finalement :

$$P_s H^s = (L_X \ell_\varepsilon \ell_s \ell)(r\ r_s R_X D^s R_Y) = L_s(R_s \ldots R_1).$$

D'après l'unicité de la décomposition de Gauss, $L_s = L_X \ell_\varepsilon \ell_s \ell$ et $L_s \xrightarrow[s\to\infty]{} L_X$.

Alors $M_1 \ldots M_s = P_s^{-1} L_s \xrightarrow[s\to\infty]{} P^{-1} L_X = M_X$.

Et :

$$H_{s+1} = (M_1 \ldots M_s)^{-1} H (M_1 \ldots M_s) = (M_1 \ldots M_s)^{-1} M_X R_X D R_X^{-1} M_X^{-1} (M_1 \ldots M_s)$$
$$\xrightarrow[s\to\infty]{} R_X D R_X^{-1} = H_\infty.$$

Lorsque s tend vers l'infini, la suite H_s converge vers une matrice triangulaire supérieure admettant les valeurs propres de A, ordonnées par valeur absolue décroissante, sur la diagonale.

D'où le :

THEOREME 1

Etant données une matrice A *régulière mise sous forme de Hessenberg supérieure :* $A = QHQ^H$ *, et une suite* $\{P_s\}$ *de matrices régulières qui vérifient les hypothèses suivantes :*

i) les valeurs propres λ_i *de* A *sont telles que :*

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > 0$$

donc A *est diagonalisable et* $H = XDX^{-1}$, $D = (\lambda_i)$,

ii) la suite P_s *tend vers* P, *matrice régulière,*

iii) quel que soit s, *la décomposition* $H_s = M_s R_s$ *relative à* $N_s = P_s M_1 \ldots M_{s-1}$ *existe,*

iv) la décomposition $X = M_X R_X$ *relative à* P *existe ,*
alors la suite de matrices semblables à A :

$$\begin{cases} H_1 = H = M_1 R_1 & \textit{relativement à } N_1 = P_1 \\ H_s = M_s R_s & \textit{relativement à } N_s = P_s M_1 \ldots M_{s-1} \ , \quad s = 2,3,\ldots \end{cases}$$

converge vers une matrice triangulaire supérieure admettant les valeurs propres λ_i *de* A *sur la diagonale.*

Remarque - On voit que le fait d'appliquer l'algorithme à une matrice, préalablement mise sous forme de Hessenberg, permet de faire disparaître des hypothèses de convergence de l'algorithme la 5ème condition :

la décomposition de Gauss $Y = L_Y R_Y$ existe.

condition qui figure classiquement dans les hypothèses de convergence des algorithmes LR et QR, [71].
A part cela, le schéma de la démonstration est classique [71].

III - ETUDE DE LA CONVERGENCE DE L'ALGORITHME MR AU 1er ORDRE

D'après la démonstration du théorème 1, la convergence de H_{s+1} vers H est déterminée par celle de $M_1 \ldots M_s$ vers M_X, convergence qui dépend de ε_s et des quantités $(\frac{\lambda_i}{\lambda_j})^s$, $i > j$.
Les calculs étant difficiles à expliciter, nous allons les faire au 1er ordre en prenant pour infiniment petits principaux, les éléments de ε_s ainsi que $(\frac{\lambda_{i+1}}{\lambda_i})^s$, $i = 1,\ldots,n-1$.

1 - Convergence de $(M_1 \ldots M_s)$ vers M_X au 1er ordre

1.1 - Si A est une matrice voisine de la matrice unité, en prenant pour infiniment petits principaux a_{ij} si $i \neq j$, ou $a_{ii}-1$, on a facilement la décomposition de Gauss de A au 1er ordre.
Soit A_i la matrice d'éléments : a_{ij} pour $i > j$, 0 pour $i \leq j$.
Soit A_s la matrice d'éléments : a_{ij} pour $i < j$, $a_{ii}-1$ pour $i = j$ et 0 pour $i > j$.
Alors $A = I+A_i+A_s = (I+A_i)(I+A_s)$ au 1er ordre.

$$1.2 - M_1 \ldots M_s = P_s^{-1} L_s = P^{-1}(I+\varepsilon_s P^{-1})^{-1} L_X \ell_\varepsilon \ell_s \ell$$
$$= P^{-1} L_X (I+L_X^{-1}\varepsilon_s P^{-1} L_X)^{-1} \ell_\varepsilon \ell_s \ell$$
$$= M_X r_\varepsilon^{-1} \ell_s \ell$$

Soit $M_1 \ldots M_s - M_X = M_X(r_\varepsilon^{-1} \ell_s \ell - I)$.

Calculons ℓ_s, r_ε^{-1} et ℓ au 1er ordre.

1) $\underline{\ell_s r_s = R_X(D^s L_Y D^{-s}) R_X^{-1}}$

$D^s L_Y D^{-s}$ est une matrice d'élément général $\ell_{ij}^Y (\frac{\lambda_i}{\lambda_j})^s = \alpha_{ij}$.

Au 1er ordre, les seuls éléments différents de 0 sont α_{ii} et $\alpha_{i+1\ i}$.

Un calcul simple montre que, si on pose $\ell_s = (\beta_{ij})$, $R_X = (r_{ij}^X)$,

$$\beta_{i+1i} = \frac{r_{i+1\ i+1}^X}{r_{ii}^X} \ell_{i+1i}^Y \left(\frac{\lambda_{i+1}}{\lambda_i}\right)^s$$

On pose $\ell_s = I+\eta_s$.

2) $\underline{\ell_\varepsilon r_\varepsilon = I+L_X^{-1}\varepsilon_s P^{-1} L_X}$

On pose $r_\varepsilon = I+\tilde{s}$, $\tilde{s}$ étant une matrice triangulaire supérieure dont les éléments sont de l'ordre de ceux de ε_s.

Alors $r_\varepsilon^{-1} = I-\tilde{s}$ au 1er ordre.

3) $\underline{\ell_s^{-1} r_s \ell_s = \ell r}$, $\ell = I$ au 1er ordre.

Alors $r_\varepsilon^{-1} \ell_s \ell - I = -\tilde{s}+\eta_s$ au 1er ordre, où $\tilde{s}$ dépend de la convergence de P_s vers P et η_s de $\frac{\lambda_{i+1}}{\lambda_i}$.

2 - Comparaison avec la convergence de LR

Supposons que l'on applique l'algorithme LR à H. Il converge sous des hypothèses analogues à celles du théorème 1. On veut comparer les vitesses de convergence des deux algorithmes.

Tout ce qui se rapporte à l'algorithme LR est distingué par une barre supérieure.

$$\bar{H}_{s+1} = (\bar{L}_1 \ldots \bar{L}_s)^{-1} \, H(\bar{L}_1 \ldots \bar{L}_s)$$

$$\bar{L}_1 \ldots \bar{L}_s - \bar{L}_X = \bar{L}_X(\bar{\ell}_s - I) \text{ avec } \bar{R}_X(D^s L_Y D^{-s})\bar{R}_X^{-1} = \bar{\ell}_s \bar{r}_s$$

Remarquons que dans les deux algorithmes on utilise la décomposition de Gauss $Y = L_Y R_Y$.

$$\bar{\ell}_s - I = \bar{\eta}_s$$

Supposons la <u>suite</u> P_s <u>constante</u> : $\varepsilon_s = 0$ et $\tilde{s} = 0$.

Alors $M_1 \ldots M_s - M_X = M_X \eta_s$ et $\bar{L}_1 \ldots \bar{L}_s - \bar{L}_X = \bar{L}_X \, \bar{\eta}_s$.

Peut-on avoir $\|\eta_s\| \leq \|\bar{\eta}_s\|$?

Si $\|.\| = S_{\varphi_1 \varphi_1}$, $\|\eta_s\| = \max\limits_{i=1,\ldots n-1} \left| \dfrac{r^X_{i+1\ i+1}}{r^X_{ii}} \, \ell^Y_{i+1\ i} \, \left(\dfrac{\lambda_{i+1}}{\lambda_i}\right)^s \right|$

Or les éléments de η_s et de $\bar{\eta}_s$ ne diffèrent que par les éléments diagonaux de R_X et $\bar{R}_X$.

Comparons R_X et $\bar{R}_X$.

$$X = \bar{L}_X \bar{R}_X \text{ et } PX = L_X R_X$$

Si $P = L_P R_P$, $L_P R_P \bar{L}_X \bar{R}_X = L_X R_X$

En décomposant $R_P \bar{L}_X$ en $L'R'$, il vient :

$$(L_P L')(R' \bar{R}_X) = L_X R_X \text{ , soit } R' \bar{R}_X = R_X.$$

Dans le cas où P est triangulaire inférieure :

$$P = L_P . \Delta_P, \text{ où } \Delta_P \text{ est la diagonale de } P.$$

Alors $\Delta_P \bar{L}_X = L'\Delta_P$ et $R_X = \Delta_P \bar{R}_X$

et

$$\frac{r^X_{i+1\ i+1}}{r^X_{ii}} = \frac{p_{i+1\ i+1}}{p_{ii}} \frac{\overline{r^X_{i+1\ i+1}}}{\overline{r^X_{ii}}}$$ où p_{ii} est l'élément diagonal de P.

Si

$$\left|\frac{p_{i+1\ i+1}}{p_{ii}}\right| < 1,\ i = 1,\ldots,n-1, \quad \|\eta_s\| \leq \|\bar{\eta}_s\|.$$

Nous avons montré la :

Proposition

En prenant pour infiniment petit principal $(\frac{\lambda_{i+1}}{\lambda_i})^s$ *, la convergence de l'algorithme MR relatif à une suite constante* $P_s = P$ *, matrice triangulaire inférieure, est, au 1er ordre, au moins aussi rapide que la convergence de l'algorithme LR, si la diagonale de* P *est constituée d'éléments décroissants en valeur absolue.*

Sur les matrices testées, dont l'ordre varie entre 3 et 20, le gain en précision sur les valeurs propres est de l'ordre de un chiffre significatif sur sept.

IV - ACCELERATION DE LA CONVERGENCE DES ALGORITHMES LR ET QR

Connaissant les expressions, au 1er ordre par rapport à $(\frac{\lambda_{i+1}}{\lambda_i})^s$, des quantités $(L_1 \ldots L_s)-L_X$ pour l'algorithme LR, $(Q_1 \ldots Q_s)-Q_X$ pour l'algorithme QR, on déduit, d'une manière classiquement employée pour la sommation d'une série, des algorithmes modifiés dont la convergence est du 2ème ordre .

1 - Algorithme LR modifié

Puisque $L_1 \dots L_s = L_X(I+\eta_s)$ au 1er ordre, $\Lambda'_s = (L_1 \dots L_s)(I-\eta_s)$ approche L_X au 2ème ordre. De même :

$$H'_{s+1} = (I+\eta_s)R_s L_s(I-\eta_s) = \Lambda'^{-1}_s H_1 \Lambda'_s,$$

approche $H_\infty = L_X^{-1} H_1 L_X$ au 2ème ordre.

η_s se calcule en fonction de s, R_X, L_Y et des valeurs propres λ_i, quantités dont on peut obtenir des approximations pour s assez grand.

R_X, matrice des vecteurs propres de $H_\infty = R_X D R_X^{-1}$, est approchée par V_s, matrice des vecteurs propres de R_s. On peut normer les vecteurs propres en choisissant, par exemple, la ième composante du ième vecteur égale à un, pour i = 1,...,n.
$Y = X^{-1} = R_X^{-1} L_X^{-1}$ est approchée par $Y_s = V_s^{-1} \Lambda_{s-1}^{-1}$.
La décomposition de Gauss $Y_s = L_{Y_s} . R_{Y_s}$ détermine L_{Y_s}, approximation de L_Y.
On prend alors pour η_s la matrice d'éléments :

$$\ell^{Y_s}_{i+1\ i} \left(\frac{h^{(s)}_{i+1 i+1}}{h^{(s)}_{i,i}}\right)^s \quad , \quad i = 1,\dots,n-1, \text{ sinon } 0$$

avec $H_s = (h^{(s)}_{ij})$ et $L_{Y_s} = (\ell^{Y_s}_{ij})$.

L'algorithme LR modifié est alors le suivant : on applique l'algorithme LR à $H_1 = H$ jusqu'au pas S, S étant déterminé par le fait qu'une norme donnée de la partie inférieure de H_S est inférieure à un nombre donné ε. On calcule alors :

$$H'_{S+1} = (I+\eta_S)H_{S+1}(I-\eta_S) \qquad \text{au 1er ordre}$$

et la norme de la partie inférieure de H'_{S+1} est en ε^2.
On peut combiner cette modification de l'algorithme avec les translations d'origine de Wilkinson ([71] p.505).

2 - Algorithme QR modifié

La démarche est tout-à-fait analogue à celle qui a conduit à l'algorithme LR modifié. Nous ne faisons que citer les principaux résultats utiles.
Etant donnée une matrice A régulière, mise sous forme de Hessenberg supérieure : $A = QHQ^H$, et dont les valeurs propres λ_i vérifient :

i) $$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > 0,$$

on démontre que, à une matrice diagonale unitaire près, la suite de matrices H_s, unitairement semblables à H, définies par l'algorithme QR, converge vers une matrice triangulaire supérieure, admettant les valeurs propres de A sur la diagonale.

La démonstration est basée sur la décomposition de Schmidt de :

$$H^s = XD^sY$$

$$= Q_X(R_XD^sL_YD^{-s}R_X^{-1})R_XD^sR_Y.$$

On peut identifier deux décompositions de Schmidt d'une matrice, si, par exemple, les matrices triangulaires supérieures sont à éléments diagonaux positifs.

Il existe une matrice δ_s, diagonale unitaire, telle que $\delta_s(Q_1 \ldots Q_s)$ converge vers Q_X, lorsque s tend vers l'infini.

Le calcul de $\delta_s(Q_1 \ldots Q_s)-Q_X$ au 1er ordre est analogue à celui de η_s :

$$\delta_s(Q_1 \ldots Q_s)-Q_X = Q_X(I-\gamma_s) \quad \text{au 1er ordre}$$

où γ_s est la matrice antihermitienne ($\gamma_s = -\gamma_s^H$) déduite de la partie inférieure de $R_X D^s L_Y D^{-s} R_X^{-1}$ au 1er ordre :

$$\gamma_s = (\gamma_{ij}^{(s)}) \text{ avec } \begin{cases} \gamma_{ij}^{(s)} = 0 \text{ si } j \neq i-1 \text{ et } j \neq i+1 \ ; \\ \gamma_{i+1i}^{(s)} = -\gamma_{i\ i+1}^{(s)} = \dfrac{r_{i+1i+1}^X}{r_{ii}^X} \ \ell_{i+1\ i}^Y \left(\dfrac{\lambda_{i+1}}{\lambda_i}\right)^s \quad i=1,\ldots,n-1 \end{cases}$$

L'algorithme QR modifié découle alors facilement.

3 - Expériences numériques et procédure ALGOL

3.1 - Résultats numériques

Pour les matrices testées, d'ordre 3, 10 et 12 telles que $\max_i \left|\frac{\lambda_{i+1}}{\lambda_i}\right|$ varie entre 0,9 et 0,95, on obtient pour un même nombre d'itérations, un à deux chiffres significatifs de plus sur sept, par l'algorithme modifié.

ALGORITHME LR MODIFIE

	valeurs propres	éléments diagonaux de H_S	éléments diagonaux de H'_S
S=20	18	17,999 093	18,000 002
	12	12,000 898	11,999 997
	6	6,000 014	6,000 007
S=39	10	10,000 817	10,000 050
	9	8,999 530	8,999 983
	8	7,999 774	7,999 980
	7	6,999 886	6,999 994
	6	5,999 945	5,999 998
	5	4,999 975	4,999 999
	4	3,999 993	3,999 999
	3	2,999 998	2,999 998
	2	1,999 999	1,999 999
	1	0,999 999	0,999 999
S=39	12	12,002 154	12,000 685
	11	11,002 144	11,001 482
	10,5	10,498 936	10,499 337
	10	9,997 288	9,998 670
	9	8,999 666	7,99 847
	8	7,999 702	7,999 982
	6	6,000 545	6,000 062
	5,5	5,499 926	5,499 987
	5	4,999 516	4,999 945
	3	2,999 999	2,999 999

ALGORITHME QR MODIFIE

S=17	3	2,999 593	3,000 001
	2	2,000 410	1,999 999
	1	0,999 995	1,000 004

3.2 - Procédure ALGOL

```
procédure AXELR(A,N,EPS,SORTIE,SFIN,D,X) ;
comment cette procédure calcule, par l'algorithme LR modifié, les valeurs
propres D et les vecteurs proprex X d'une matrice A d'ordre N sous forme
de Hessenberg supérieure. Lorsque la décomposition de Gauss est impos-
sible, la sortie de la procédure se fait par SORTIE. L'algorithme s'ar-
rête lorsque la somme des valeurs absolues des termes sous-diagonaux de
la matrice itérée, est inférieure à EPS, ou lorsque le nombre d'itéra-
tions est supérieur à SFIN ;
value N ;
label SORTIE ; integer N,SFIN ; real EPS ; array A,D,X ;
begin integer I,J,K,S ; real SOM, C ;
      array A1,R,LB, LB1,V[1:N,1:N], L[1:N-1] ;
      for I := 2 step 1 until N do
      begin for J := 1 step 1 until I-1 do
            A1[I,J] := R[I,J] := V[I,J] := 0.0 ;
            V[I,I] := 1.0 ;
      end ;
      V[1,1] := 1.0 ;
      for I := 1 step 1 until N-1 do
      begin for J := I+1 step 1 until N do
            LB[I,J] := LB1[I,J] := 0.0 ;
            LB[I,I] := LB1[I,I] := 0.0 ;
      end ;
      LB[N,N] := LB1[N,N] := 1.0 ;
      S:= 1 ;
```

```
        comment decomposition LR ;
ALG :   if A[1,1] = 0.0 then goto SORTIE ;
        L[1] := A[2,1]/A[1,1] ;
        for J := 1 step 1 until N do R[1,J] := A[1,J];
        for J := 2 step 1 until N-1 do
        begin for I := 2 step 1 until J do
              R[I,J] := A[I,J]-R[I-1,J]*L[I-1];
              if R[J,J] = 0.0 then goto SORTIE ;
              L[J] := A[J+1,J]/R[J,J] ;
        end ;
        for I := 2 step 1 until N do
        R[I,N] := A[I,N]-R[I-1,N]*L[N-1];
        comment produit RL ;
        for I := 1 step 1 until N-1 do for J := I step 1 until N-1 do
        A1[I,J] := R[I,J]+R[I,J+1]*L[J];
        for I := 2 step 1 until N do
        begin A1[I,I-1] := R[I,I]*L[I-1];
              A1[I,N] := R[I,N] ;
              for J := 1 step 1 until I-1 do
              LB1[I,J] := LB[I,J]+LB[I,J+1]*L[J] ;
        end ;
        A1[1,N] := R[1,N] ;
TEST:   if S > SFIN then goto FINALG ;
        SOM := 0.0 ;
        for I := 2 step 1 until N do
        SOM := SOM+ABS(A1[I,I-1]) ;
        if SOM > EPS then
        begin S := S+1 ;
              for I := 2 step 1 until N do
              begin for J := 1 step 1 until I-1 do
                    LB[I,J] := LB1[I,J] ;
                    for J := I-1 step 1 until N do
                    A[I,J] := A1[I,J];
              end ;
              for J := 1 step 1 until N do A[1,J] := A1[1,J] ;
        end ; goto ALG ;
```

```
        comment calcul de V matrice des vecteurs propres à droite de R ;
MODIF:  for I := 1 step 1 until N-1 do for J := I+1 step 1 until N do
        begin C := 0.0 ;
              for K := I step 1 until J-1 do
              C := C+R[K,J]*V[I,K] ;
              V[I,J] := C/(R[I,I]-R[J,J]) ;
        end ;
        comment inversion de LB1 ;
        for J := 1 step 1 until N-1 do for I := J+1 step 1 until N do
        begin C := 0.0 ;
              for K := J step 1 until I-1 do
              C := C+LB1[I,K]*LB[K,J] ;
              LB[I,J] := -C ;
        end ;
        comment calcul de Y ;
        for I := 1 step1until N do for J := 1 step 1 until N do
        begin C := 0.0 ;
              for K := 1 step 1 until N do
              C := C+V[I,K]*LB[K,J] ;
              A[I,J] := C ;
        comment decomposition de Gauss  de Y ;
        if A[1,1] = 0.0 then goto SORTIE ;
        for J := 1 step 1 until N do
        begin R[1,J] := A[1,J] ;
              LB[J,1] := A[J,1]/A[1,1] ;
        end ;
        for J := 2 step 1 until N do
        begin for I := 2 step 1 until J do
              begin C := 0.0 ;
                    for K := 1 step 1 until I-1 do
                    C := C+R[K,J]*LB[I,K] ;
                    R[I,J] := A[I,J]-C ;
              end;
```

```
            if R[J,J] = 0.0 then goto SORTIE ;
            for I := J+1 step 1 until N do
            begin C := 0.0 ;
                  for K := 1 step 1 until J-1 do
                  C := C+R[K,J]*LB[I,K] ;
                  LB[I,J] := (A[I,J]-C)/R[J,J] ;
            end ;
      end ;
      comment calcul de ETA ;
      for I := 1 step 1 until N-1 do
      L[I] := (A1[I+1,I+1]/A1[I,I])**S*LB[I+1,I] ;
      for I := 2 step 1 until N do
      begin for J := 1 step 1 until I-1 do
            LB[I,J] := LB1[I,J]-LB1[I,J+1]*L[J] ;
            for J := 1 step 1 until N-1 do
            A[I,J] := A1[I,J]-A1[I,J+1]*L[J]+A1[I-1,I]*L[I-1] ;
            A[I,N] := A1[I,N]+A1[I-1,N]*L[I-1];
      end ;
      for J := 1 step 1 until N-1 do
            A[1,J] := A1[1,J]-A1[1,J+1]*L[J] ;
            A[1,N] := A1[1,N] ;
      comment calcul de X matrice des vecteurs propres de A ;
FINALG :for J := N step -1 until 2 do for I := J-1 step -1 until 1 do
      begin C := 0.0 ;
            for K := I+1 step 1 until J-1 do
            C := C+A[I,K]*V[K,J]
            V[I,J] := C/(A[J,J]-A[I,I]) ;
      end ;
      for I := 1 step 1 until N do
      begin D[I] := A[I,I] ;
            for J := 1 step 1 until N do
```

```
        comment calcul de V matrice des vecteurs propres à droite de R ;
MODIF:  for I := 1 step 1 until N-1 do for J := I+1 step 1 until N do
        begin C := 0.0 ;
              for K := I step 1 until J-1 do
              C := C+R[K,J]*V[I,K] ;
              V[I,J] := C/(R[I,I]-R[J,J]) ;
        end ;
        comment inversion de LB1 ;
        for J := 1 step 1 until N-1 do for I := J+1 step 1 until N do
        begin C := 0.0 ;
              for K := J step 1 until I-1 do
              C := C+LB1[I,K]*LB[K,J] ;
              LB[I,J] := -C ;
        end ;
        comment calcul de Y ;
        for I := 1 step1until N do for J := 1 step 1 until N do
        begin C := 0.0 ;
              for K := 1 step 1 until N do
              C := C+V[I,K]*LB[K,J] ;
              A[I,J] := C ;
        comment decomposition de Gauss  de Y ;
        if A[1,1] = 0.0 then goto SORTIE ;
        for J := 1 step 1 until N do
        begin R[1,J] := A[1,J] ;
              LB[J,1] := A[J,1]/A[1,1] ;
        end ;
        for J := 2 step 1 until N do
        begin for I := 2 step 1 until J do
              begin C := 0.0 ;
                    for K := 1 step 1 until I-1 do
                    C := C+R[K,J]*LB[I,K] ;
                    R[I,J] := A[I,J]-C ;
              end;
```

CHAPITRE - II

PERTURBATION D'UNE MATRICE HERMITIENNE

APPLICATION AU CALCUL DE VALEURS ET VECTEURS PROPRES

Etant donnée une matrice hermitienne A dont on suppose connus les valeurs propres et les vecteurs propres, on étudie l'effet sur ces éléments d'une perturbation hermitienne H. Dans le cas d'une matrice d'ordre 2, la résolution du problème est explicite (équation du deuxième degré), et l'on a cherché, dans le cas d'une matrice d'ordre n, à obtenir des majorations qui soient le plus près possible des résultats exacts pour n = 2.

Les résultats obtenus sont appliqués à l'étude de la précision d'algorithmes de recherche des éléments propres tels que les méthodes QR et Jacobi. Ils généralisent et présentent d'une manière uniforme certains résultats épars dans la littérature (Davis [8], Gavurin [14], Golub [15], Kato-Temple [24], Ortega [11], Wilkinson [70]). On obtient un majorant réaliste de la précision obtenue sur les éléments propres, aisément calculable au cours des algorithmes et qui permet d'arrêter le calcul dès qu'il est inférieur à la précision voulue.

A la fin du chapitre, on étudie la sensibilité de la précision obtenue à la distance de séparation entre les valeurs propres de A. Lorsque cette distance devient très petite, la recherche des éléments propres de A+H est un problème mal conditionné relativement aux valeurs propres, les vecteurs propres restant orthogonaux.

Remarque - La théorie faite pour des matrices hermitiennes s'étend sans difficulté aux matrices normales.

I - PERTURBATION D'UNE MATRICE HERMITIENNE

1 - Préliminaire

1.1 - L'espace $\mathbb{C}^n$ est muni du produit scalaire euclidien (,).
Soit $A' = A+H$. $\{\lambda,\varphi\}$ représente un couple d'éléments propres de A, (λ valeur propre, φ vecteur propre normé), $\{\lambda',\varphi'\}$ un couple d'éléments propres de A'.

Nous notons d la distance de λ aux autres valeurs propres de A. Et nous rappelons un résultat classique en théorie des perturbations ([25] p. 88) :

THEOREME 1

Si A *admet* λ *pour valeur propre de multiplicité* m *et si* $2\|H\| < d$, *il existe exactement* m *valeurs propres de* A', *distinctes ou confondues, dans l'intervalle* $[\lambda-\|H\|.\lambda+\|H\|]$.

Nous aurons besoin aussi du :

THEOREME 2 (Krylov-Weinstein, [72] p. 321)

Pour toute valeur propre λ *de* A, *il existe une valeur propre* λ' *de* A' *telle que :*

$$|\lambda'-(A'\varphi,\varphi)| \leq \sqrt{\|H\varphi\|^2-(H\varphi,\varphi)^2}$$

$(A'\varphi,\varphi) = \lambda+(H\varphi,\varphi)$.
Si $(H\varphi,\varphi) = 0$, l'intervalle défini par le théorème 2 est inclus dans celui défini par le théorème 1.

Si $(H\varphi,\varphi) \neq 0$, ces intervalles se chevauchent en général.

1.2 - Etude pour une matrice symétrique 2 × 2

Supposons que

$A = \begin{pmatrix} a_1 & h \\ h & a_2 \end{pmatrix}$, matrice symétrique, donc à éléments réels, admette les valeurs propres $\lambda_1 > \lambda_2$ et les vecteurs propres normés φ_1, φ_2.

$$\lambda_1 = \frac{a_1+a_2}{2} + \frac{1}{2}\sqrt{(a_1-a_2)^2+4h^2}$$

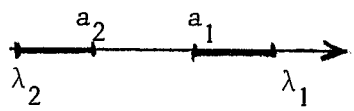
$$\lambda_2 = \frac{a_1+a_2}{2} - \frac{1}{2}\sqrt{(a_1-a_2)^2+4h^2}$$

a) Supposons $a_1 > a_2$ et posons $r = \dfrac{2|h|}{a_1-a_2}$

$$\lambda_1-a_1 = \frac{a_2-a_1}{2} + \frac{1}{2}\sqrt{(a_1-a_2)^2+4h^2} = \frac{1}{r}\left(\sqrt{1+r^2}-1\right)|h|$$

$$\lambda_2-a_2 = \frac{a_1-a_2}{2} - \frac{1}{2}\sqrt{(a_1-a_2)^2+4h^2} < 0$$

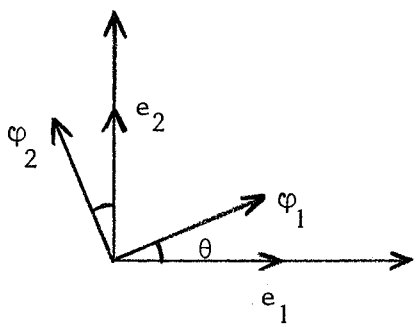
$$|\lambda_2-a_2| = \frac{1}{r}\left(\sqrt{1+r^2}-1\right)|h|$$



Soit θ l'angle aigu entre φ_1, e_1 ou φ_2, e_2

$$\operatorname{tg}\theta = \frac{\lambda_1-a_1}{|h|} = \frac{|\lambda_2-a_2|}{|h|} = \frac{1}{r}\left(\sqrt{1+r^2}-1\right)$$

$\theta < \dfrac{\pi}{4}$.



La distance entre les valeurs propres de A et ses éléments diagonaux s'exprime donc facilement en fonction du rapport entre la norme de la partie hors diagonale de A : $|h|$, et la distance entre les éléments diagonaux : a_1-a_2. Nous allons étendre ce résultat à une matrice hermitienne d'ordre quelconque.

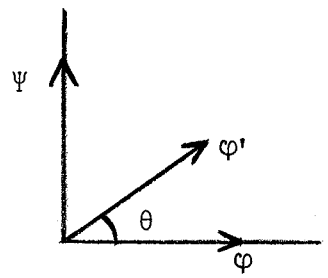
b) Si $a_1 = a_2 = a$, $\lambda_1-a = a-\lambda_2 = |h|$, $\theta = \frac{\pi}{4}$.

2 - Majorant de $|\lambda'-(A'\varphi,\varphi)|$

λ est une valeur propre simple de A. Il existe λ' valeur propre de A' que l'on suppose simple, telle que $|\lambda'-(A'\varphi,\varphi)| \leq \sqrt{\|H\varphi\|^2-(H\varphi,\varphi)^2} \leq \|H\varphi\|$. φ et φ' sont des vecteurs propres normés associés respectivement à λ et λ'. Soit θ l'angle aigu entre φ et φ'.

L'hypothèse $2\|H\| < d$ va permettre de déterminer un coefficient $C < 1$ tel que :

$$|\lambda'-(A'\varphi,\varphi)| \leq C\|H\varphi\| \text{ et } tg\theta \leq C.$$



Soit π la projection orthogonale de $\mathbb{C}^n$ sur le plan engendré par φ et φ'. Dans ce plan, Ψ est normé et orthogonal à φ.

Soit T' l'application linéaire de matrice A', dans la base canonique de $\mathbb{C}^n$. A la somme de matrices A' = A+H correspond la somme d'applications T' = T+U. Dans une base orthonormée dont les deux premiers vecteurs sont φ et Ψ, $\pi T'\pi$, restriction de T' au plan engendré par φ, φ', est représenté par une matrice dont les seuls éléments non nuls sont ceux de :

$$\tilde{A}' = \begin{pmatrix} (A'\varphi,\varphi) & (A'\Psi,\varphi) \\ (A'\varphi,\Psi) & (A'\Psi,\Psi) \end{pmatrix} \quad \text{où :}$$

$$(A'\varphi,\varphi) = \lambda+(H\varphi,\varphi) = \lambda+h_1 = \alpha$$

$$(A'\varphi,\Psi) = (H\varphi,\Psi) = h \qquad \text{puisque } (\varphi,\Psi) = 0$$

$$(A'\Psi,\varphi) = \bar{h}$$

$$(A'\Psi,\Psi) = (A\Psi,\Psi)+(H\Psi,\Psi) = a+h_2 = \beta$$

2.1 - <u>$\tilde{A}'$ vérifie les propriétés suivantes</u> :

1) <u>λ' est valeur propre de $\tilde{A}'$</u>

En effet, $\varphi' = \cos\theta\varphi+\sin\theta\Psi$. Il suffit de transcrire l'égalité $A'\varphi' = \lambda'\varphi'$ dans la base dont les deux premiers vecteurs sont φ et Ψ. Il vient :

$$\tilde{A}' \begin{pmatrix}\cos\theta\\ \\ \sin\theta\end{pmatrix} = \lambda' \begin{pmatrix}\cos\theta\\ \\ \sin\theta\end{pmatrix}$$

On note μ' la deuxième valeur propre de $\tilde{A}'$.

2) <u>$\tilde{A}'$ est symétrique</u>

Explicitons l'égalité ci-dessus :

$$\alpha\cos\theta+h\sin\theta = \lambda'\cos\theta$$

$$\bar{h}\cos\theta+\beta\sin\theta = \lambda'\sin\theta .$$

Soit $\bar{h} = (\lambda'-\beta)\mathrm{tg}\theta$, ce qui implique $h = \bar{h}$: h est réel.

3) Expression de $|\lambda'-\alpha|$ et $tg\theta$

$\tilde{A}'$ étant une matrice symétrique 2×2, on peut calculer exactement $|\lambda'-\alpha|$ et $tg\theta$ en fonction de $r = \frac{2|h|}{|\alpha-\beta|}$.

Supposons $\alpha > \beta$. Comment placer μ' par rapport à λ' ?

On a les relations :

1 - $(\lambda'-\alpha)\cos\theta - h\sin\theta = 0$ donc $(\lambda'-\alpha)h > 0$

2 - $(\alpha-\beta)\cos\theta\sin\theta + h(\sin^2\theta - \cos^2\theta) = 0$

donc . $0 \leq \theta < \frac{\pi}{4} \iff \sin^2\theta - \cos^2\theta < 0 \iff (\alpha-\beta)h > 0$

. $\frac{\pi}{4} < \theta \leq \frac{\pi}{2} \iff$ " > 0 " < 0

D'où la disposition :

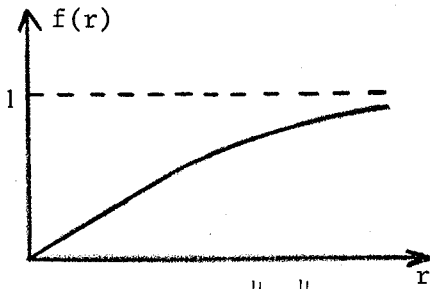
$\theta < \frac{\pi}{4}$ μ' β α λ' et $|\lambda'-\alpha| = \frac{1}{r}(\sqrt{1+r^2}-1)|h| < |h|$

$\theta > \frac{\pi}{4}$ λ' β α μ' et $|\lambda'-\alpha| \geq |\alpha-\beta|$

Si l'on sait que $\theta < \frac{\pi}{4}$ alors :

$$|\lambda'-\alpha| = \frac{1}{r}(\sqrt{1+r^2}-1)|h| \ , \quad r = \frac{2|h|}{|\alpha-\beta|}$$

$$tg\theta = \frac{1}{r}(\sqrt{1+r^2}-1)$$



2.2 - La fonction $r \overset{f}{\mapsto} \frac{1}{r}(\sqrt{1+r^2}-1)$ est croissante. Nous allons chercher une condition suffisante pour que $\theta < \frac{\pi}{4}$, ainsi qu'un majorant de r en fonction de d et $\|H\|$, ou $\|H\varphi\|$. En effet une majoration en fonction de $\|H\varphi\|$ est intéressante car $\|H\varphi\|$ peut être très inférieure à $\|H\|$ pour certaines valeurs propres ; d'autre part $\|H\varphi\|$ est aisément calculable, tandis qu'on ne peut obtenir facilement qu'un majorant de $\|H\|$.

1) Minorant de $|\alpha-\beta| = |\lambda-a+h_1-h_2|$

Nous minorons $|\lambda-a|$ et majorons $|h_1-h_2|$ en gardant $|\lambda-a| > |h_1-h_2|$.

a) $|\lambda-a|$

Appelons $\bar{A}$ la restriction de $A-\lambda I$ au complémentaire orthogonal de φ : $\bar{A}$ admet des valeurs propres de module supérieur ou égal à d.

$a-\lambda = ((A-\lambda I)\Psi,\Psi) = (\bar{A}\Psi,\Psi)$ puisque $(\Psi,\varphi) = 0$

D'où : $|a-\lambda| = |(\bar{A}\Psi,\Psi)| \geq d$.

b) $|h_1-h_2|$

$$\tilde{H} = \begin{pmatrix} h_1 & h \\ h & h_2 \end{pmatrix}$$ admet les valeurs propres : $\frac{h_1+h_2}{2} \pm \frac{1}{2}\sqrt{(h_1-h_2)^2+4h^2}$

$$\tilde{\rho} = \rho(\tilde{H}) = \frac{|h_1+h_2|}{2} + \frac{1}{2}\sqrt{(h_1-h_2)^2+4h^2}$$

et $|h_1-h_2| \leq 2\tilde{\rho}\sqrt{1-(\frac{h}{\tilde{\rho}})^2} \leq 2\tilde{\rho}$,

$|h| \leq \tilde{\rho}$ et $|h| = |(H\varphi,\Psi)| \leq \|H\varphi\|$.

D'autre part $\tilde{\rho} = \rho(\tilde{H}) \leq \rho = \rho(H)$. En effet, $\tilde{H}$ est la matrice représentant $\pi U \pi$ dans une base orthonormée dont les 2 premiers vecteurs sont φ et Ψ.

$$\rho(\tilde{H}) = \|\pi U \pi\| \leq \|U\| \, \|\pi\|^2 \leq \|U\| = \|H\| = \rho(H).$$

$\|\pi\| \leq 1$ puisque π est une projection orthogonale.

Par hypothèse $2\|H\| < d$, donc $|h_1-h_2| < d$ et $|\lambda-a| > d$.

2) <u>condition $\theta < \frac{\pi}{4}$</u>

On a vu que $|\lambda'-\alpha| \leq \|H\varphi\| \leq \|H\| = \rho$

Si $\theta > \frac{\pi}{4}$, $|\lambda'-\alpha| > |\alpha-\beta| > d-2\rho$.

Cela est impossible tant que $d-2\rho > \rho$, soit $\|H\| < \frac{d}{3}$.

3) <u>Majorant de r</u>

$$r = \frac{2|h|}{|\alpha-\beta|} \leq \frac{2|h|}{d-2\rho\sqrt{1-(\frac{h}{\rho})^2}}$$

a) Cherchant un majorant de r en fonction de $s = \frac{2\|H\|}{d} < \frac{2}{3}$, nous calculons le maximum de :

$$\frac{2|h|}{d-2\rho\sqrt{1-(\frac{h}{\rho})^2}} \text{ , lorsque } |h| \text{ décrit } [0,\rho]$$

Il est atteint pour $|h| = \rho\sqrt{1-(\frac{2\rho}{d})^2}$.
Alors :

$$r_{max} = \frac{s}{\sqrt{1-s^2}} .$$

b) Cherchant un majorant de r en fonction de $t = \frac{2\|H\varphi\|}{d} \leq s$, nous écrivons : $\rho = s\,\frac{d}{2}$

$$\frac{2|h|}{d-2\rho\sqrt{1-(\frac{h}{\rho})^2}} = \frac{2|h|}{d\left[1-s\sqrt{1-(\frac{2h}{ds})^2}\right]} \leq \frac{t}{1-\sqrt{s^2-t^2}} = r_{max} = r_s .$$

La pente à l'origine est $\frac{1}{1-s}$. Lorsque $s \to 0$, $r_s \to t$, la borne correspond à la formule exacte obtenue pour une matrice 2×2 : a_1-a_2 est remplacé par d, et $|h|$ par $\|H\varphi\|$.

Nous avons vu que $s < \frac{2}{3}$ est une C.S. pour que $\theta < \frac{\pi}{4}$. Mais d'après la majoration obtenue, lorsque s croît jusqu'à 1, θ croît jusqu'à $\frac{\pi}{4}$ sans atteindre cette valeur. Donc les résultats sont valables pour $\underline{s < 1}$.
D'où le :

THEOREME 3

A admet un élément propre $\{\lambda, \varphi\}$ *,* λ *valeur propre simple séparée des autres par* d. *Alors* $A' = A+H$ *admet un élément propre* $\{\lambda', \varphi'\}$ *,* λ' *supposée simple, tel que si :*

$$s = \frac{2\|H\|}{d} < 1 \quad , \quad t = \frac{2\|H\varphi\|}{d} \leq s,$$

$$\begin{cases} |\lambda' - (A'\varphi, \varphi)| \leq f(r) \ \|H\varphi\| \\ \\ \operatorname{tg}\theta \leq f(r) \end{cases}$$

avec

$$f(r) = \frac{1}{r}(\sqrt{1+r^2} - 1)$$

$$(3.1) \quad r = \frac{s}{\sqrt{1-s^2}}$$

$$(3.2) \quad r_s = \frac{t}{1-\sqrt{s^2-t^2}}$$

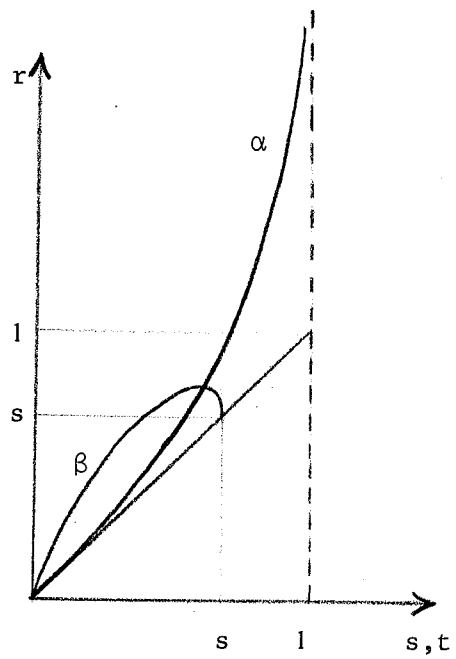
SERVICE POLYCOPIE
MATHÉMATIQUES APPLIQUÉES
Université de GRENOBLE

Le résultat (3.1) a été donné sans démonstration par Gavurin [14]. Il est meilleur que celui de Ortega [11].

Nous donnons ci-contre les courbes représentant :

α) r fonction de s,

β) r fonction de t, s étant un paramètre < 1.



3 - Cas particulier où $(H\varphi,\varphi) = 0$

Nous obtenons un certain nombre de simplifications par rapport au cas général. En particulier on doit s'attendre à trouver des bornes plus précises pour les valeurs propres, puisque, d'après un principe mis en évidence par Davis [8] : lorsqu'une application autoadjointe T est perturbée par une application U, dans la base des vecteurs propres de T, la partie diagonale de U perturbe essentiellement les valeurs propres et la partie hors-diagonale de U, les vecteurs propres.

D'autre part $(A'\varphi,\varphi) = \lambda$ et l'intervalle $[\lambda-\|H\varphi\|,\lambda+\|H\varphi\|]$ est contenu dans $[\lambda-\|H\|,\ \lambda+\|H\|]$.

3.1 - Modifications par rapport au § I.2

$$(A'\varphi,\varphi) = \lambda\ ,\ h_1 = 0, \qquad \tilde{H} = \begin{pmatrix} 0 & h \\ h & h_2 \end{pmatrix}$$

$$|\lambda'-\lambda| = \frac{1}{r}\left(\sqrt{1+r^2}-1\right)|h|, \quad r = \frac{2|h|}{|\lambda-\beta|}$$

$$|\lambda-\beta| = |\lambda-a-h_2|$$

$$\tilde{\rho} = \rho(\tilde{H}) = \frac{|h_2|}{2} + \frac{1}{2}\sqrt{|h_2|^2+4h^2} \quad \text{et} \quad |h_2| = \tilde{\rho}\left(1-\left(\frac{h}{\tilde{\rho}}\right)^2\right) \leq \tilde{\rho}$$

3.2 - Majorant de r

Si $\theta > \frac{\pi}{4}$, $|\lambda'-\lambda| \geq |\lambda-\beta| \geq d-\rho \geq \frac{d}{2}$, ce qui est incompatible avec :

$$|\lambda'-\lambda| \leq \|H\varphi\| \leq \|H\| < \frac{d}{2}.$$

a) Le maximum de $r = \dfrac{2|h|}{d-\rho\left(1-\left(\frac{h}{\rho}\right)^2\right)}$ lorsque $|h|$ décrit $[0,\rho]$ est atteint pour $|h| = \rho$, et $r_{max} = s$.

b) $\rho = s\,\frac{d}{2}$

$$r = \frac{2|h|}{d-\rho(1-(\frac{h}{\rho})^2)} = \frac{2|h|}{d\left[1-\frac{s}{2}(1-(\frac{2h}{ds})^2)\right]} \leq \frac{2t}{2-s+t^2} = r_s$$

La pente à l'origine est $\frac{1}{1-\frac{s}{2}}$, et $r_s \xrightarrow[s\to 0]{} t$.

D'où le :

THEOREME 4

A *admet un élément propre* $\{\lambda,\varphi\}$ *où* λ *est une valeur propre simple, séparée des autres par* d, *et où* φ *vérifie* $(H\varphi,\varphi) = 0$. *Si* $2\|H\| < d$, $A' = A+H$ *admet une valeur propre simple* λ' *dans* $[\lambda-\|H\|,\ \lambda+\|H\|]$, *telle que, si :*

$$s = \frac{2\|H\|}{d} < 1 \quad , \quad t = \frac{2\|H\varphi\|}{d} \leq s,$$

$$\begin{cases} |\lambda'-\lambda| \leq f(r)\ \|H\varphi\| \\ \\ \operatorname{tg}\theta \leq f(r) \end{cases}$$

avec

$$f(r) = \frac{1}{r}\,(\sqrt{1+r^2}-1)$$

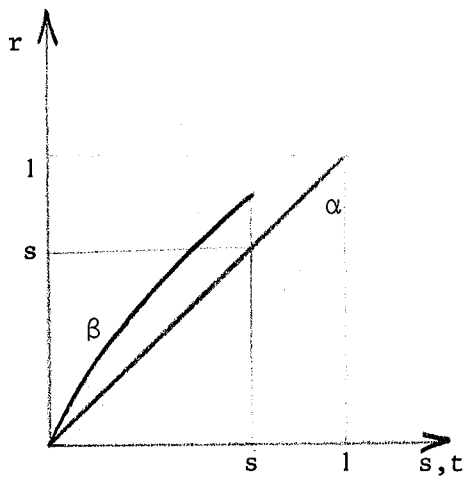
(4.1) $r = s$

(4.2) $r_s = \frac{2t}{2-s+t^2}$

Nous donnons ci-contre les courbes représentant

α) r fonction de s,

β) r fonction de t, s étant un paramètre < 1.



Par application de l'inégalité de Kato-Temple [24], sous l'hypothèse $2\|H\| < d$, on obtient :

$$|\lambda'-\lambda| \leq \frac{1}{2}\,\frac{2t}{2-s}\,\|H\varphi\|$$

inégalité moins fine que celle obtenue par (4.2) puisque

$$f(r) \leq \frac{r}{2} \text{ et } 2-s+t^2 \geq 2-s.$$

Lorsque t varie dans $[0,s]$, $s < 1$, on a toujours :

$$\frac{2t}{2-s+t^2} \leq \frac{2t}{2-s} \leq 2t.$$

D'où le :

Corollaire

Lorsque $2\|H\| < d$, $\quad |\lambda'-\lambda| \leq \dfrac{2\|H\varphi\|^2}{d}$

et $\quad \mathrm{tg}\theta \leq \dfrac{2\|H\varphi\|}{d}$.

On a vu que lorsque $s \to 0$, $r_s \to t$:
les majorants de $|\lambda'-\lambda|$ et $\mathrm{tg}\theta$ tendent respectivement vers $\dfrac{\|H\varphi\|^2}{d}$ et $\dfrac{\|H\varphi\|}{d}$.

4 - Valeurs propres multiples

Nous regardons d'abord le cas de valeurs propres doubles (pour la clarté de l'exposé). Nous ne nous intéressons plus aux inégalités portant sur les vecteurs propres puisqu'il existe une infinité de vecteurs propres correspondant à la valeur propre double, le problème de couplage avec 2 vecteurs propres de la matrice perturbée est délicat.

4.1 - Cas général : $(H\varphi,\varphi) \neq 0$

1) A admet une valeur propre double λ

Soient φ_1, φ_2 deux vecteurs indépendants du plan propre associé à λ. Il existe λ'_1, λ'_2 valeurs propres de A' distinctes ou confondues, telles que :

$$|\lambda'_1-(A'\varphi_1,\varphi_1)| \leq \|H\varphi_1\|$$

et

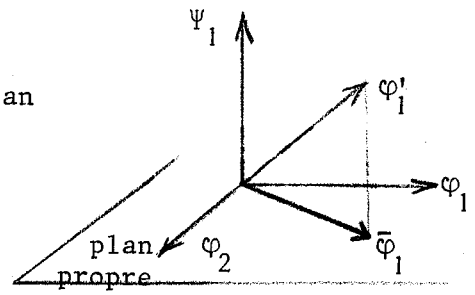
$$|\lambda'_2-(A'\varphi_2,\varphi_2)| \leq \|H\varphi_2\|$$

a) majoration de $|\lambda'_1-(A'\varphi_1,\varphi_1)|$

Si le vecteur Ψ_1, orthogonal à φ_1 dans le plan engendré par φ_1 et φ'_1 est aussi orthogonal à φ_2, le quotient de Rayleigh $|((A-\lambda I)\Psi_1,\Psi_1)|$, sera supérieur ou égal à d, distance de λ aux autres valeurs propres de A. Et rien ne sera changé par rapport au § I.2, dans le calcul du majorant de $r = \frac{2|h|}{|\alpha-\beta|}$.

Si φ_1, φ'_1 et φ_2 ne sont pas coplanaires, on définit $\bar{\varphi}_1$ vecteur normé colinéaire à la projection orthogonale de φ'_1 sur le plan propre. Alors Ψ_1 défini à l'aide de $\bar{\varphi}_1$ et φ'_1 est orthogonal au plan propre.

Si φ_1, φ'_1 et φ_2 sont coplanaires, on prendra $\bar{\varphi}_1 = \varphi'_1$.



b) majoration de $|\lambda'_2-(A'\varphi_2,\varphi_2)|$

On peut de la même façon définir $\bar{\varphi}_2$ et Ψ_2 et l'on aura

$$|((A-\lambda I)\Psi_2,\Psi_2)| \geq d.$$

Dans la pratique, on ne peut pas déterminer de tels vecteurs $\bar{\varphi}_1$ et $\bar{\varphi}_2$ puisque φ'_1 et φ'_2 sont inconnus.

Néanmoins, nous pouvons énoncer le :

THEOREME 5

A *admet la valeur propre double* λ, *séparée des autres par* d. *Soient* φ_1 *et* φ_2 *deux vecteurs indépendants dans le plan propre associé à* λ. $A' = A+H$ *admet alors deux valeurs propres* λ'_1 *et* λ'_2 *telles que :*

$$|\lambda'_1-(A'\varphi_1,\varphi_1)| \leq \|H\varphi_1\|$$

$$|\lambda'_2-(A'\varphi_2,\varphi_2)| \leq \|H\varphi_2\|$$

Il existe dans le plan propre deux vecteurs $\bar{\varphi}_1$ *et* $\bar{\varphi}_2$ *tels que si :*

$$s = \frac{2\|H\|}{d} < 1$$

$i = 1,2, |\lambda'_i-(A'\bar{\varphi}_i,\bar{\varphi}_i)|$ *vérifient les inégalités du théorème 3 où* t *est remplacé par* $t_i = \frac{2\|H\bar{\varphi}_i\|}{d} \leq s.$

2) <u>A' admet une valeur propre double λ'</u>

Supposons que, parmi les éléments propres de A, il en existe deux $\{\lambda_1,\varphi_1\}$ et $\{\lambda_2,\varphi_2\}$ tels que

$$|\lambda'-(A'\varphi_1,\varphi_1)| \leq \|H\varphi_1\|$$

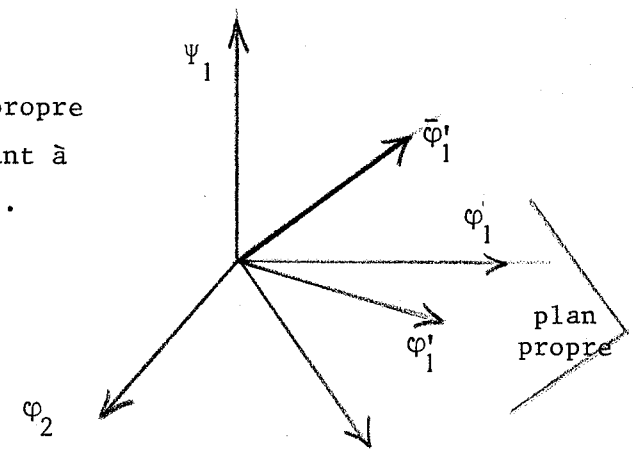
$$|\lambda'-(A'\varphi_2,\varphi_2)| \leq \|H\varphi_2\|$$

Soit $d_i = \min_{j\neq 1,2} |\lambda_i-\lambda_j|, \quad i = 1,2.$

De même qu'au 1) si Ψ_1 est orthogonal à φ_1 et φ_2, le quotient de Rayleigh :

$$|((A-\lambda_1 I)\Psi_1,\Psi_1)| \text{ est } \geq d_1$$

Pour cela, on choisira dans le plan propre associé à λ', un vecteur $\bar{\varphi}'_1$ appartenant à l'hyperplan orthogonal à φ_2. D'où Ψ_1.
On peut aussi déterminer $\bar{\varphi}'_2$ et Ψ_2.



D'où le :

THEOREME 6

Si A' *admet une valeur propre double* λ' *et si* A *admet deux éléments propres* $\{\lambda_i,\varphi_i\}$, $i = 1,2$, *tels que :*

$$|\lambda'-(A'\varphi_i,\varphi_i)| \leq \|H\varphi_i\|,$$

si on pose $d_i = \min\limits_{j\neq 1,2} |\lambda_i-\lambda_j|$,

$|\lambda'-(A'\varphi_i,\varphi_i)|$ *vérifient les inégalités du théorème 3 où* d *est remplacé par* d_i, *dès que*

$$s_i = \frac{2\|H\|}{d_i} < 1.$$

Au contraire du théorème 5, ce théorème donne des majorations de $|\lambda'-(A'\varphi_i,\varphi_i)|$ en fonction des éléments propres de A qui sont supposés connus.

N.B. - On a vu le rôle que joue l'existence du plan propre associé à λ', pour déterminer des majorations en fonction de la distance d de l'ensemble $\{\lambda_1,\lambda_2\}$ aux autres valeurs propres de A. Ceci a une importance pratique : pour deux valeurs propres voisines λ_1 et λ_2, on aura des bornes d'erreur faisant intervenir 1/d et non $\frac{1}{\lambda_1-\lambda_2}$, dans le cas où l'on saura qu'elles approchent une valeur propre double λ'. Ce point sera étudié plus en détail au § II.2.

4.2 - Cas particulier : $(H\varphi,\varphi) = 0$

1) - A admet une valeur propre double

Si $2\|H\| < d$, il existe exactement 2 valeurs propres distinctes ou confondue λ_1' et λ_2' de A' dans $[\lambda-\|H\|,\lambda+\|H\|]$. Si $(H\varphi_1,\varphi_1) = (H\varphi_2,\varphi_2) = 0$ elles sont telles que :

$$|\lambda_1'-\lambda| \leq \|H\varphi_1\|$$

$$|\lambda_2'-\lambda| \leq \|H\varphi_2\|$$

Mais le théorème 5 ne donne rien de plus car, en général, $(H\bar{\varphi}_1,\bar{\varphi}_1)$ et $(H\bar{\varphi}_2,\bar{\varphi}_2)$ sont différents de 0.

2) - A' admet une valeur propre double λ'

On montre facilement le :

THEOREME 7

Si A' *admet une valeur propre double* λ' *séparée des autres par* d' *et si* $2\|H\| < d'$ *il existe exactement 2 valeurs propres* λ_1 *et* λ_2 *de* A *dans* $[\lambda'-\|H\|,\lambda'+\|H\|]$.

Soit $d_i = \min_{j\neq 1,2} |\lambda_i-\lambda_j|$, $i = 1,2$.

On suppose que $(H\varphi_i,\varphi_i) = 0$, $i = 1,2$.

Alors $|\lambda'-\lambda_i|$ *vérifie les inégalités du théorème 4 avec* d_i *dès que*

$$s_i = \frac{2\|H\|}{d_i} < 1.$$

4.3 - Cas de valeurs propres multiples

Il se traite de manière analogue. Nous gardons les mêmes notations, $i = 1,\dots,m$.

1) A admet la valeur propre λ d'ordre m.
   $\bar{\varphi}_i$ est un vecteur normé, colinéaire à la projection orthogonale de φ'_i sur le sous-espace propre associé à λ.

2) A' admet la valeur propre λ' d'ordre m.
   Dans le sous-espace propre associé à λ' on choisit $\bar{\varphi}'_i$ appartenant au complémentaire orthogonal de $\varphi_2,\dots,\varphi_m$.

Nous énonçons l'analogue des théorèmes 6 et 7 :

THEOREME 8

Si A' *admet une valeur propre multiple* λ' *de multiplicité* m *et si* A *admet* m *éléments propres* $\{\lambda_i,\varphi_i\}$, $i = 1,\dots,m$, *tels que :*

$$|\lambda'-(A'\varphi_i,\varphi_i)| \leq \|H\varphi_i\| ,$$

si on pose

$$d_i = \min_{j\neq 1,\dots,m} |\lambda_i-\lambda_j|$$

$|\lambda'-(A'\varphi_i,\varphi_i)|$ *vérifient les inégalités du théorème 3 avec* d_i, *dès que*

$$s_i = \frac{2\|H\|}{d_i} < 1$$

Si de plus, $(H\varphi_i,\varphi_i) = 0$, $|\lambda'-\lambda_i|$ *vérifient les inégalités du théorème 4 avec* d_i *sous les mêmes hypothèses.*

V - MATRICES NORMALES

5.1 - Etude pour une matrice normale 2 × 2 (sur $\mathbb{C}$)

Ecrivant que $A = \begin{pmatrix} a_1 & h_3 \\ h_4 & a_2 \end{pmatrix}$ est normale, on obtient les relations :

$$\begin{cases} |h_3| = |h_4| \\ (a_1-a_2)\bar{h}_3 = (\overline{a_1-a_2})h_4. \end{cases}$$

Donc $\quad (a_1-a_2)^2\overline{h_3h_4} = |a_1-a_2|^2|h_4|^2$ est réel positif.

De même $\quad \dfrac{h_3h_4}{(a_1-a_2)^2} = \dfrac{|h_4|^2}{|a_1-a_2|^2}$

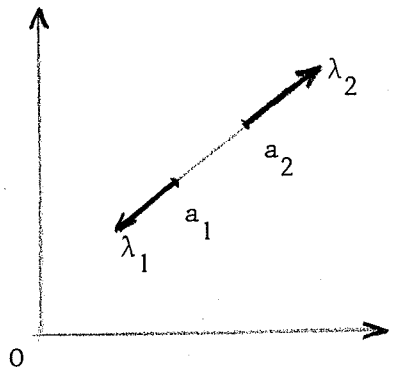
A admet les valeurs propres :

$$\lambda_1 = \frac{a_1+a_2}{2} + \frac{1}{2}\sqrt{(a_1-a_2)^2+4h_3h_4}$$

$$\lambda_2 = \frac{a_1+a_2}{2} - \frac{1}{2}\sqrt{(a_1-a_2)^2+4h_3h_4}$$

et grâce à la remarque ci-dessus on aboutit facilement à la conclusion suivante : <u>les valeurs propres λ_1 et λ_2 d'une matrice normale sont dans le plan complexe, alignées avec les éléments diagonaux a_1 et a_2.</u>

Posans $|h| = |h_3| = |h_4|$ et $r = \dfrac{2|h|}{|a_1-a_2|}$, on obtient pour $|\lambda_1-a_1|$, $|\lambda_2-a_2|$ et $\mathrm{tg}\theta$, les expressions valables dans le cas d'une matrice symétrique 2 × 2. Et la disposition des valeurs propres est celle de la figure ci-contre



5.2 - Majorant de $|\lambda'-(A'\varphi,\varphi)|$

Avec les mêmes notations qu'au § I.2, si $A' = A+H$ est normale, $\tilde{A}'$ matrice 2×2 associée est normale.

Tout ce qui a été établi dans ce paragraphe se transpose aisément

$$\tilde{H} = \begin{pmatrix} h_1 & h_3 \\ h_4 & h_2 \end{pmatrix} \text{ admet pour valeurs propres : } \frac{h_1+h_2}{2} \pm \frac{1}{2}\sqrt{1+\alpha^2}(h_1-h_2)$$

où l'on a posé : $|h_3| = |h_4| = |h|$,

$$\alpha = \frac{2|h|}{|h_1-h_2|} .$$

Donc

$$\tilde{\rho} = \rho(\tilde{H}) \geq \frac{1}{2}\sqrt{1+\alpha^2}\,|h_1-h_2| = \frac{1}{2}\sqrt{|h_1-h_2|^2+4|h|^2}$$

Soit

$$|h_1-h_2| \leq 2\tilde{\rho}\sqrt{1-\left(\frac{|h|}{\tilde{\rho}}\right)^2} \leq 2\tilde{\rho}.$$

On retrouve alors la majoration pour r (cf. § I.2) :

$$r \leq \frac{2|h|}{d-2\rho\sqrt{1-\left(\frac{|h|}{\rho}\right)^2}} \quad , \quad \text{où } \rho = \rho(H).$$

II - APPLICATION AU CALCUL DES ELEMENTS PROPRES - RESULTATS NUMERIQUES

1 - Algorithmes QR et Jacobi

Soit à calculer les éléments propres $\{\lambda_i, \varphi_i\}$, d'une matrice hermitienne A, par les méthodes QR ou Jacobi. Dans les deux cas, on construit une suite de matrices hermitiennes, semblables à A :

$$A_1 = A \quad , \quad A_k = Q_{k-1}^H A_{k-1} Q_{k-1}$$

qui converge (sous certaines hypothèses) vers une matrice diagonale dont les éléments sont les valeurs propres de A.

Posons $A_k = D_k + H_k$, où D_k est la diagonale de A_k et H_k la partie hors diagonale. Généralement on arrête l'algorithme dès qu'une norme de H_k est plus petite qu'un nombre ε fixé à l'avance et on prend les éléments de D_k pour approximation des valeurs propres de A_k, donc de A. Quelle précision est-on alors en droit d'attendre sur les valeurs propres et les vecteurs propres en fonction de ε? Des majorations ont été données pour les valeurs propres à l'aide de l'inégalité de Hadamard et celle de Krylov-Weinstein, ainsi que des bornes d'erreur a postériori (Wilkinson [70]). A l'aide du § I, nous obtenons des bornes très précises que nous testons.

1.1 - $A_k = D_k + H_k$

Considérons la décomposition : $A_k = D_k + H_k$.

D_k est diagonale et admet la base canonique $\{e_i\}$ pour base de vecteurs propres.

H_k, matrice de diagonale nulle est telle que :

$$(H_k e_i, e_i) = 0 \quad , \quad i = 1, \ldots, n.$$

$$A = A_1 = (Q_1 \ldots Q_{k-1}) A_k (Q_1 \ldots Q_{k-1})^H.$$

A admet les éléments propres : λ_i, φ_i $\quad i = 1,\dots,n$

A_k admet les éléments propres : λ_i, Ψ_i^k

D_k admet les éléments propres : a_{ii}^k, e_i

et l'on a les relations : $\varphi_i = (Q_1 \dots Q_{k-1})\Psi_i^k$, vecteurs propres exacts

$\varphi_i^k = (Q_1 \dots Q_{k-1})e_i$, vecteurs propres approchés.

Les Q_i étant unitaires, θ_i^k angle aigu entre φ_i et φ_i^k est égal à l'angle aigu entre Ψ_i^k et e_i.

$$\|\varphi_i - \varphi_i^k\| = \|\Psi_i^k - e_i\| = 2 \sin \frac{\theta_i^k}{2} .$$

Appliquons les résultats du théorème 4 si A admet des valeurs propres simples.

<u>*Proposition 1*</u>

Soit la suite $\begin{cases} A = A_1 \\ A_k = Q_{k-1}^H A_{k-1} Q_{k-1} = D_k + H_k, \end{cases} \quad k = 2,3,\dots$

qui converge vers la diagonale $D = (\lambda_i)$ *des valeurs propres simples de* A : $\|H_k\| \xrightarrow[k\to\infty]{} 0$.

(les quantités ci-après se rapportent à A_k *et l'indice* k *est omis).*

Soit

$$d_i = \min_{j \neq i} |a_{ii} - a_{jj}| > 0$$

$$s_i = \frac{2\|H\|}{d_i} \quad , \quad t_i = \frac{2\|He_i\|}{d_i}$$

pour chaque i *tel que* $s_i < 1$ *on a les inégalités :*

$$\begin{cases} |\lambda_i - a_{ii}| \leq f(r_i) \ \|He_i\| \\ \operatorname{tg}\theta_i \leq f(r_i) \end{cases}$$

avec $f(r) = \frac{1}{r}(\sqrt{1+r^2}-1)$

(4.1) $r_i = s_i$

(4.2) $r_i = \dfrac{2t_i}{2-s_i+t_i^2}$.

Si λ_i est valeur propre multiple d'ordre m de A, isolée des autres par $\delta_i = \min(|\lambda_i-\lambda_j|,\ \lambda_i \neq \lambda_j)$, il existe m nombres a_{ii}^k dans $[\lambda_i-\|H_k\|,\lambda_i+\|H_k\|]$ dès que $2\|H_k\| < \delta_i$.

Soit $\Lambda_i^{(k)}$ cet ensemble de m nombres.
Le théorème 8 montre que la proposition précédente s'applique avec $d_i^k = \min(|a_{ii}^k-a_{jj}^k|\ ,\ a_{jj}^k \notin \Lambda_i^{(k)}) > 0$.

Dans le cas de <u>valeurs propres simples ou multiples</u>, lorsque $\|H_k\| \to 0$, $a_{ii}^k \to \lambda_i$ et $d_i^k \to \delta_i = \min(|\lambda_i-\lambda_j|,\ \lambda_i \neq \lambda_j)$, distance de séparation de λ_i des autres valeurs propres de A.
D'après le corollaire du théorème 4 :

- pour chaque valeur propre λ_i, la convergence $a_{ii}^k \to \lambda_i$ est <u>quadratique</u> par rapport à $\|H_k\|$: elle est réglée par $\dfrac{2\|H_k e_i\|^2}{\delta_i}$

- pour chaque vecteur propre φ_i, la convergence $tg\theta_i^k \to 0$ est <u>linéaire</u> en $\|H_k\|$: elle est réglée par $\dfrac{2\|H_k e_i\|}{\delta_i}$.

Les formules déduites de (4.2), faisant intervenir $\|He_i\|$ et non $\|H\| \geq \|He_i\|$ sont très intéressantes dans la pratique car la convergence des valeurs propres n'est en général par uniforme : $\|He_i\|$ peut être beaucoup plus petit pour une valeur propre que pour une autre. Ces bornes d'erreur rendent donc beaucoup mieux compte de la situation réelle que toute formule avec $\|H\|$, (nous en verrons un exemple plus loin).

Nous allons illustrer ceci à l'aide d'exemples numériques, sur l'algorithme QR (sans translation [71] p. 485) à titre d'exemple.

1.2 - Algorithme QR - Résultats numériques

Nous appliquons l'algorithme QR à une matrice hermitienne préalablement mise sous forme tridiagonale par la méthode de Householder [21].

A_k converge vers D s'il n'existe pas de valeurs propres opposées pour A. Les calculs sont faits en double précision sur IBM 360.

1er exemple :

$$n = 5,\ A = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & 0 & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix} \text{ admet les valeurs propres : } \begin{matrix} 2+\sqrt{3} \\ 3 \\ 2 \\ 1 \\ 2-\sqrt{3} \end{matrix}$$

a) Comparons, pour k = 15, les résultats obtenus par Golub [15] (à partir de l'inégalité de Kato-Temple) et ceux obtenus à partir de (4.2) pour la majoration de $|\lambda_i - a_{ii}^k|$:

Golub	Borne 2
$1,4\ 10^{-1}$	$3,4\ 10^{-3}$
$1,8\ 10^{-1}$	$3,5\ 10^{-3}$
$2\ 10^{-2}$	$4,2\ 10^{-5}$
$3\ 10^{-4}$	$1\ 10^{-8}$
$6\ 10^{-9}$	$1\ 10^{-15}$

b) Pour des valeurs de k croissantes nous comparons :

- pour les valeurs propres :
  la différence exacte $|\lambda_i - a_{ii}^k|$, avec les majorations de cette erreur obtenue par Krylov, (4.1) et (4.2).

- pour les vecteurs propres :
  la différence exacte $\|\varphi_i - \varphi_i^k\|$, avec les majorations obtenues à partir de (4.1) et (4.2).

$\|H\|$ est majoré par la norme de Frobenius : $\||H|\| = \left(\sum_{i,j} |h_{ij}|^2\right)^{1/2}$

$2\||H|\| < d = \min_i d_i$ est vérifié dès que $k > 8$.

Valeurs propres

k	Différence	Krylov	borne 1	borne 2	
k = 13	$7,4 \quad 10^{-3}$	$7,3 \quad 10^{-2}$	$1,1 \quad 10^{-2}$	$8,5 \quad 10^{-3}$	
	$7,2 \quad 10^{-3}$	$7,4 \quad 10^{-2}$	$1,1 \quad 10^{-2}$	$8,8 \quad 10^{-3}$	
	$1,9 \quad 10^{-4}$	$1,4 \quad 10^{-2}$	$1,5 \quad 10^{-3}$	$1,1 \quad 10^{-4}$	
	$1,1 \quad 10^{-7}$	$3,3 \quad 10^{-4}$	$4,7 \quad 10^{-5}$	$1,8 \quad 10^{-7}$	
	$2,2 \quad 10^{-15}$	$4,6 \quad 10^{-8}$	$6,6 \quad 10^{-9}$	$3,4 \quad 10^{-15}$	=> λ_5
k = 25	$4,0 \quad 10^{-5}$	$5,4 \quad 10^{-3}$	$5,6 \quad 10^{-5}$	$4,0 \quad 10^{-5}$	
	$4,0 \quad 10^{-5}$	$5,4 \quad 10^{-3}$	$5,6 \quad 10^{-5}$	$4,0 \quad 10^{-5}$	
	$1,2 \quad 10^{-8}$	$1,1 \quad 10^{-4}$	$8,3 \quad 10^{-7}$	$1,2 \quad 10^{-8}$	
	$3,0 \quad 10^{-15}$	$8,1 \quad 10^{-8}$	$8,5 \quad 10^{-10}$	$9,1 \quad 10^{-15}$	=> λ_4
k = 42	$2,4 \quad 10^{-8}$	$1,3 \quad 10^{-4}$	$3,4 \quad 10^{-8}$	$2,4 \quad 10^{-8}$	
	$2,4 \quad 10^{-8}$	$1,3 \quad 10^{-4}$	$3,4 \quad 10^{-8}$	$2,4 \quad 10^{-8}$	
	$5 \quad 10^{-15}$	$1,1 \quad 10^{-7}$	$2,0 \quad 10^{-11}$	$1,2 \quad 10^{-14}$	=> λ_3
k = 68	$3 \quad 10^{-13}$	$4,5 \quad 10^{-7}$	$3,9 \quad 10^{-13}$	$3 \quad 10^{-13}$	=> λ_1
	$2,5 \quad 10^{-13}$	$4,5 \quad 10^{-7}$	$3,9 \quad 10^{-13}$	$2,8 \quad 10^{-13}$	λ_2

vecteurs propres

k	Différence	borne 1	borne 2	
k = 13	$1,0\ 10^{-1}$	$1,4\ 10^{-1}$	$1,1\ 10^{-1}$	
	$1,0\ 10^{-1}$	$1,4\ 10^{-1}$	$1,1\ 10^{-1}$	
	$1,4\ 10^{-2}$	$1,0\ 10^{-1}$	$1,6\ 10^{-2}$	
	$3,3\ 10^{-4}$	$1,4\ 10^{-1}$	$5,3\ 10^{-4}$	
	$6,3\ 10^{-8}$	$1,4\ 10^{-1}$	$7,3\ 10^{-8}$	
k = 25	$7,4\ 10^{-3}$	$1,1\ 10^{-2}$	$7,4\ 10^{-3}$	
	$7,4\ 10^{-3}$	$1,1\ 10^{-2}$	$7,5\ 10^{-3}$	
	$1,1\ 10^{-4}$	$7,6\ 10^{-3}$	$1,1\ 10^{-4}$	
	$8,1\ 10^{-8}$	$1,1\ 10^{-2}$	$1,1\ 10^{-7}$	
	$1\ 10^{-14}$	$1,1\ 10^{-2}$	$1\ 10^{-14}$	=> φ_5
k = 42	$1,8\ 10^{-4}$	$2,5\ 10^{-4}$	$1,8\ 10^{-4}$	
	$1,8\ 10^{-4}$	$2,5\ 10^{-4}$	$1,8\ 10^{-4}$	
	$1,1\ 10^{-7}$	$1,9\ 10^{-4}$	$1,1\ 10^{-7}$	
	$6,2\ 10^{-13}$	$2,5\ 10^{-4}$	$8,5\ 10^{-13}$	
k = 49	$3,9\ 10^{-5}$	$5,5\ 10^{-5}$	$3,9\ 10^{-5}$	
	$3,9\ 10^{-5}$	$5,5\ 10^{-5}$	$3,9\ 10^{-5}$	
	$6,4\ 10^{-9}$	$4,0\ 10^{-5}$	$6,4\ 10^{-9}$	
	$5\ 10^{-15}$	$5,5\ 10^{-5}$	$6\ 10^{-15}$	=> φ_4
k = 68	$6,2\ 10^{-7}$	$8,8\ 10^{-7}$	$6,2\ 10^{-7}$	
	$6,2\ 10^{-7}$	$8,8\ 10^{-7}$	$6,2\ 10^{-7}$	
	$2,9\ 10^{-12}$	$6,4\ 10^{-7}$	$2,9\ 10^{-12}$	

Pour λ_5 (resp. λ_4) la borne obtenue par Krylov : $\|H_k e_5\|$ (resp. $\|H_k e_4\|$) est de l'ordre de 10^{-15} pour k = 26 (resp. 48). Alors que cette précision est garantie par la borne 1 dès que k = 22 (resp. 37) et par la borne 2 dès que k = 13 (resp. 25).

Il y a donc un gain appréciable du nombre des itérations nécessaires pour obtenir une précision donnée sur les valeurs propres, à se servir de la majoration obtenue à l'aide de (4.2). On remarquera la différence de précision entre la borne 1 calculée à partir de $\|H\|$ et la borne 2, à partir de $\|H\varphi\|$.

<u>2ème exemple</u> :

$$n=5,\ A = \begin{pmatrix} 6,16 & -2,64 & -1,84 & -1,04 & -0,24 \\ -2,64 & 5,56 & -0,64 & 0,16 & 0,96 \\ -1,84 & -0,64 & 5,16 & 0,96 & 1,76 \\ -1,04 & 0,16 & 0,96 & 4,76 & 2,56 \\ -0,24 & 0,96 & 1,76 & 2,56 & 4,36 \end{pmatrix}$$

admet les valeurs propres 10, 7, 5, 3, 1.

On constate le même phénomène :

1°) Il faut environ deux fois plus d'itérations pour obtenir par l'inégalité de Krylov une borne d'erreur de l'ordre de celle donnée par (4.2).

2°) Les bornes données par (4.2) sont très voisines de $|\lambda_i - a_{ii}^k|$ et de $\|\varphi_i - \varphi_i^k\|$, la précision étant d'autant meilleure que $\|H_k\|$ est plus petit.

2 - Valeurs propres simples, séparées ou voisines - Valeurs propres multiples

Quelle est l'influence de la distance δ de λ aux autres valeurs propres de A, sur la précision obtenue en approchant λ par un élément diagonal de A ?
Les formules données contiennent le facteur $\frac{1}{\delta}$, et semblent indiquer que plus δ est petit, moins la précision est bonne.

Or dans la pratique, on peut constater que bien souvent il n'en est rien. Appliquons par exemple la méthode de Jacobi à la matrice d'ordre 21, d'élément général :

$$a_{ii} = \begin{cases} 11-i & , \quad i = 1,\dots,11 \\ i-11 & , \quad i = 12,\dots,21 \end{cases}$$

$$a_{i,i+1} = a_{i+1,i} = 1$$

$$a_{ij} = a_{ji} \qquad = 0 \quad j \neq i,i+1.$$

(Wilkinson [71] p. 485) qui admet des paires de valeurs propres voisines : les 2 plus grandes valeurs propres ont 15 décimales identiques, les 2 suivantes 11, etc... la différence augmente.
En double longueur, les résultats obtenus sont de précision comparable pour toutes les valeurs propres.

Pour comprendre ce qui se passe, il suffit de regarder le développement limité au 2ème ordre de la valeur propre λ_i de A au voisinage de l'élément diagonal a_{ii} :

$$A = D+H$$

Si $2\|H\| < \delta_i$, on peut écrire :

$$\lambda_i = a_{ii}+(He_i,e_i)+\sum_{j\neq i}\frac{(He_i,e_j)^2}{a_{ii}-a_{jj}} + O(\|H\|^3)$$

$$= a_{ii}+\sum_{j\neq i}\frac{h_{ij}^2}{a_{ii}-a_{jj}} + O(\|H\|^3).$$

Lorsque $\|H\| \to 0$, si les h_{ij} sont de même ordre, l'erreur est bien de l'ordre de $\frac{h_{ik}^2}{a_{ii}-a_{kk}}$, avec :

$$|a_{ii}-a_{kk}| = \min_{j\neq i}|a_{ii}-a_{jj}|.$$

Mais si les h_{ij} sont d'ordre très différents, le terme prépondérant dans l'erreur peut ne pas correspondre à $|a_{ii}-a_{kk}|$.

C'est bien ce que l'on constate dans l'exemple cité plus haut.

<u>Remarque</u> - Si H est tridiagonale (cas de la méthode QR),

$$\lambda_i = a_{ii}+\frac{h_{i,i-1}^2}{a_{ii}-a_{i-1,i-1}}+\frac{h_{i,i+1}^2}{a_{ii}-a_{i+1,i+1}} + O(\|H\|^4).$$

2.1 - Précision théorique sur λ en fonction de δ et $\|H\|$

On veut étudier la précision obtenue sur λ en fonction de δ et de $\|H\|$. Pour cela, soit λ_1,λ_2 deux valeurs propres simples de A, telles que $\delta = \lambda_1-\lambda_2 > 0$. Les autres valeurs propres de A sont à une distance $d > \delta$ de λ_1 et λ_2.
λ_1 est approché par a_{11}, et nous établissons des majorations en $O(\|H\|^3)$.

1°) $\underline{\delta > 2\|H\|}$ $\qquad |\lambda_1 - a_{11}| \leq \dfrac{\|He_1\|^2}{\delta} + O(\|H\|^3)$

C'est le cas de valeurs propres <u>séparées.</u>

2°) $\underline{\delta < 2\|H\|}$

On ne peut plus appliquer de résultat faisant appel à la théorie des perturbations.

Néanmoins :

$$|\lambda_1 - a_{11}| \leq \|He_1\|$$

C'est le cas de valeurs propres <u>voisines.</u>

<u>N.B.</u> - Les notions de valeurs propres séparées ou voisines sont <u>relatives à $\|H\|$.</u>

3°) $\underline{\delta = 0}$ $\quad$ Si $2\|H\| < d$, $\quad |\lambda_1 - a_{11}| \leq \dfrac{\|He_1\|^2}{d} + O(\|H\|^3)$

C'est le cas de valeurs propres <u>doubles</u>.

Lorsque δ est suffisamment petit par rapport à $\|H\|$, on peut étendre ce dernier résultat, en considérant que A est voisine, à $\frac{\delta}{2}$ près, d'une matrice admettant la valeur propre double $\frac{\lambda_1+\lambda_2}{2}$.

<u>*LEMME*</u>

Si A *admet deux valeurs propres simples* λ_1 *et* λ_2 *telles que* $\lambda_1 - \lambda_2 = \delta > 0$ *il existe une matrice* $A' = A + H'$ *admettant la valeur propre double* $\lambda' = \frac{\lambda_1+\lambda_2}{2}$ *telle que* $\|H'\| = \frac{\delta}{2}$. *Les* n-2 *autres valeurs propres de* A' *sont celles de* A.

Démonstration

Soit $\Delta = (\lambda_i)$ la diagonale des valeurs propres de A. et considérons :

$$\Delta' = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & & \\ & & & \lambda_n \end{pmatrix} + \begin{pmatrix} -\frac{\delta}{2} & & & \\ & \frac{\delta}{2} & & \\ & & 0 & \\ & & & 0 \end{pmatrix} = \Delta + \Delta_1 .$$

Si $A = Q\Delta\, Q^H$, $A' = Q\Delta' Q^H = A+H'$ admet les valeurs propres :

$$\lambda' = \frac{\lambda_1+\lambda_2}{2} \text{ double} \quad , \quad \lambda_i, \quad i \neq 1,2.$$

Et $\|H'\| = \|Q\Delta_1 Q^H\| = \frac{\delta}{2}$.

Que peut-on dire alors de $|\lambda'-a_{11}|$?

$A' = D+H+H' = D+K$ admet la valeur propre double λ'. Appliquons le théorème 6 à A' et D.

Nous posons :

$$d_1 = \min_{j\neq 1,2} |a_{11}-a_{jj}|$$

Si $\quad 2\|K\| < d_1, \quad |\lambda'-(A'e_1,e_1)| \leq \dfrac{\|Ke_1\|^2}{d_1} + O(\|K\|^3)$

$$\leq \frac{\|Ke_1\|^2}{d} + O(\|K\|^3).$$

D'autre part,

$$(A'e_1,e_1) = a_{11}+(H'e_1,e_1)$$

$$|(H'e_1,e_1)| \leq \|He_1\| \leq \frac{\delta}{2}.$$

Donc

$$|\lambda'-a_{11}| \leq \frac{\delta}{2} + \frac{\|Ke_1\|^2}{d} + O(\|K\|^3)$$

$$\|Ke_1\| = \|(H+H')e_1\| \leq \|He_1\| + \frac{\delta}{2}$$

$$\frac{\delta}{2} + \frac{(\|He_1\|+\frac{\delta}{2})^2}{d} = \frac{\delta}{2} + \frac{\|He_1\|^2}{d} + \frac{\|He_1\|\delta}{d} + \frac{\delta^2}{4d}$$

$$= g(\delta, \|He_1\|).$$

Posons $\varepsilon = \|He_1\|$.

La majoration par g est intéressante tant que $g(\delta,\varepsilon) < \varepsilon$, soit :

$$\delta < 2\varepsilon - 8\,\frac{\varepsilon^2}{d} + O(\|H\|^3).$$

Alors l'hypothèse $2\|K\| < d$ est vérifiée dès que $d > 2(\|H\|+\varepsilon)$.

α) à ε fixé, g fonction de δ est représenté par la parabole :

$$\delta \mapsto \frac{\delta^2}{4d} + \left(\frac{1}{2} + \frac{\varepsilon}{d}\right)\delta + \frac{\varepsilon^2}{d}\ .$$

β) à δ fixé, g fonction de ε est représenté par la parabole :

$$\varepsilon \mapsto \frac{\varepsilon^2}{d} + \varepsilon\,\frac{\delta}{d} + \frac{\delta}{2} + \frac{\delta^2}{4d}\ .$$

THEOREME 9

A = D+H *admet deux valeurs propres simples* λ_1 *et* λ_2 *telles que* $\lambda_1-\lambda_2 = \delta > 0$. λ_1 *est approché par* a_{11} *et l'on pose* $\varepsilon = \|He_1\|$. *Les autres valeurs propres de* A *sont à une distance* $d > 2(\|H\|+\varepsilon)$. *A l'ordre* $O(\|H\|^3)$, *on a les majorations :*

$\delta > 2\|H\|$: $|\lambda_1 - a_{11}| \le \dfrac{\varepsilon^2}{\delta}$
<u>les valeurs propres sont séparées.</u>

$2\varepsilon - 8\,\dfrac{\varepsilon^2}{d} \le \delta \le 2\|H\|$: $|\lambda_1 - a_{11}| \le \varepsilon$

$\varepsilon^3 \le \delta \le 2\varepsilon - 8\,\dfrac{\varepsilon^2}{d}$: $\left|\dfrac{\lambda_1+\lambda_2}{2} - a_{11}\right| \le g(\delta,\varepsilon)$
<u>les valeurs propres sont voisines.</u>

$0 \le \delta \le \varepsilon^3$: $|\lambda_1 - a_{11}| \le \dfrac{\varepsilon^2}{d}$
<u>la valeur propre est double.</u>

avec

$$g(\delta,\varepsilon) = \frac{\delta}{2} + \frac{\varepsilon^2}{d} + \frac{\varepsilon\delta}{d} + \frac{\delta^2}{4d}\ .$$

2.2 - Résultats numériques

Les formules du théorème 8 indiquent que, lorsque ε est fixé et $\delta \to 0$, il existe une zone intermédiaire pour laquelle la précision sur λ_1 est stationnaire. Pour obtenir une bonne précision, δ doit être, soit suffisamment grand, soit suffisamment petit par rapport à ε, les valeurs intermédiaires de δ, correspondent à deux valeurs propres voisines relativement à ε, donnant une précision médiocre. Nous allons voir qu'il en est bien ainsi dans la pratique.

Exemple :

Pour éviter les h_{ij} d'ordre très différents, nous avons pris pour exemple :

$$n = 5,\ A = \begin{pmatrix} 4+\delta & \varepsilon & & & \\ \varepsilon & 4 & \varepsilon & 0 & \\ & \varepsilon & 6 & \varepsilon & \\ & 0 & \varepsilon & 8 & \varepsilon \\ & & & \varepsilon & 10 \end{pmatrix} \quad , \ \|H\| = \varepsilon\sqrt{3} \quad , \quad d = 2.$$

et nous avons calculé les valeurs propres de A par la méthode de Jacobi en faisant varier :

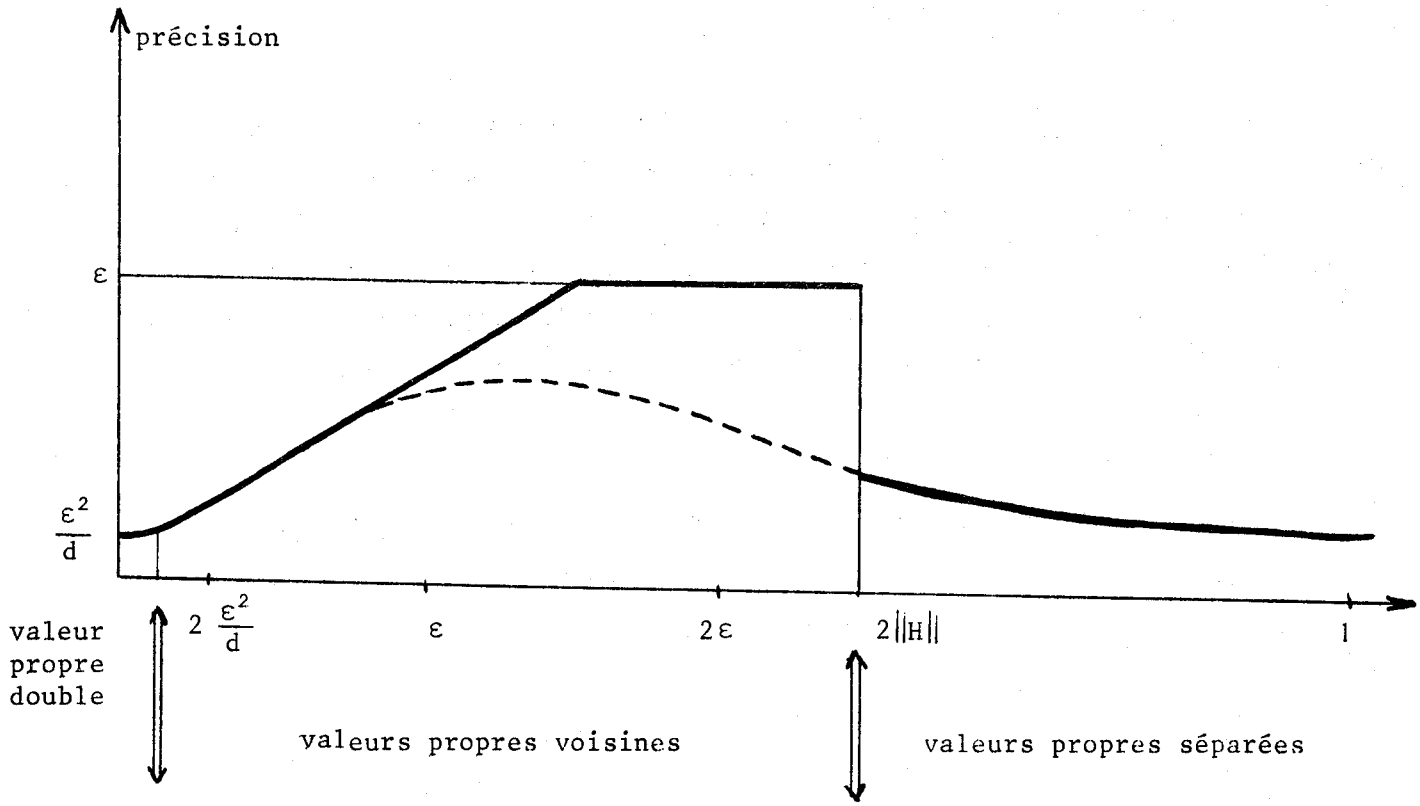
α) pour $\varepsilon = 10^{-5}$, δ dans $[0,1]$

β) pour $\delta = 10^{-3}$, ε dans $[0,0.1]$.

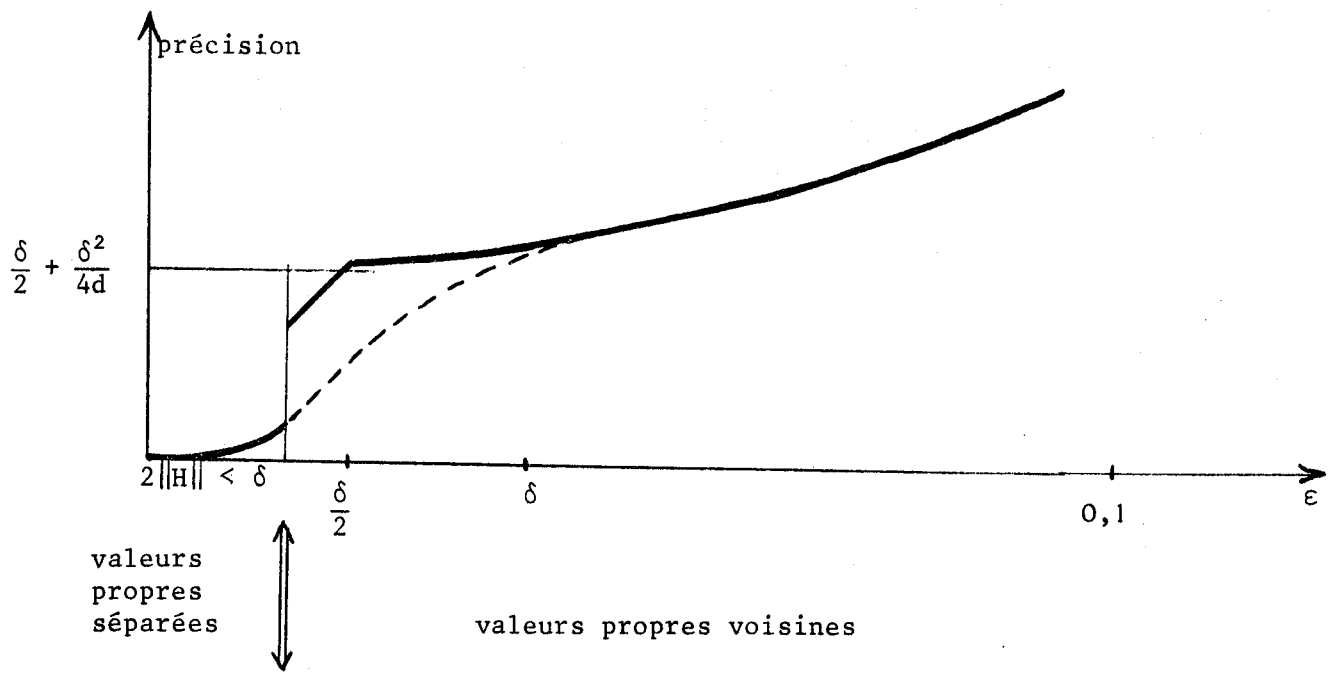
Les courbes ci-dessous indiquent, pour λ_1,

- la précision théorique en trait plein
- la précision effective en pointillé.

α) ε fixé



β) δ fixé



On voit que la précision théorique donnée par le théorème 9 est très bonne, sauf pour la région correspondant à :

$$\varepsilon \leq \delta \leq 2\|H\|.$$

La courbe donnant la variation de la précision obtenue sur λ_1, lorsque ε tend vers 0, la distance δ de λ_1 aux autres valeurs propres étant constante, s'interprète de la manière suivante :

a) si $\varepsilon > \delta$, la précision varie selon la parabole $\varepsilon \mapsto \frac{\varepsilon^2}{d} + \varepsilon\frac{\delta}{d} + \frac{\delta}{2} + \frac{\delta^2}{4d}$, et ne peut être inférieure à $\frac{\delta}{2} + \frac{\delta^2}{4d}$.

b) si $\varepsilon < \frac{\delta}{2}$, la précision varie selon la parabole $\varepsilon \mapsto \frac{\varepsilon^2}{\delta}$, et tend vers 0 avec ε.

Lors de l'utilisation d'algorithmes tels que QR ou Jacobi, on peut mettre en évidence, pour chaque valeur propre un intervalle $[k_1,k_2]$ tel que si le nombre k d'itérations appartient à cet intervalle, le passage d'une itération à la suivante n'augmente pas la précision sur la valeur propre considérée de façon "rentable". On a alors intérêt, soit à arrêter les calculs avant que le nombre d'itérations n'atteigne k_1, soit à les poursuivre jusqu'à ce que le nombre d'itérations dépasse k_2.

3 - Application pratique

Lors de l'utilisation d'algorithmes de calcul d'éléments propres de matrices, l'intérêt de tests d'arrêt de l'algorithme dès que la précision théorique obtenue est inférieure à un nombre fixé, est évident. Et l'on a avantage à disposer de majorations les plus fines possible pour les valeurs propres et les vecteurs propres. Les majorations proposées au paragraphe II.1 sont bien adaptées à ce but : les bornes données sont facilement calculables au cours de l'algorithme et sont réalistes.

Nous présentons ci-dessous une procédure ALGOL de calcul de la précision sur les valeurs propres, ainsi que l'organigramme de l'utilisation de cette procédure à chaque itération de l'algorithme QR.

3.1 - Procédure ALGOL de calcul de la précision sur les valeurs propres

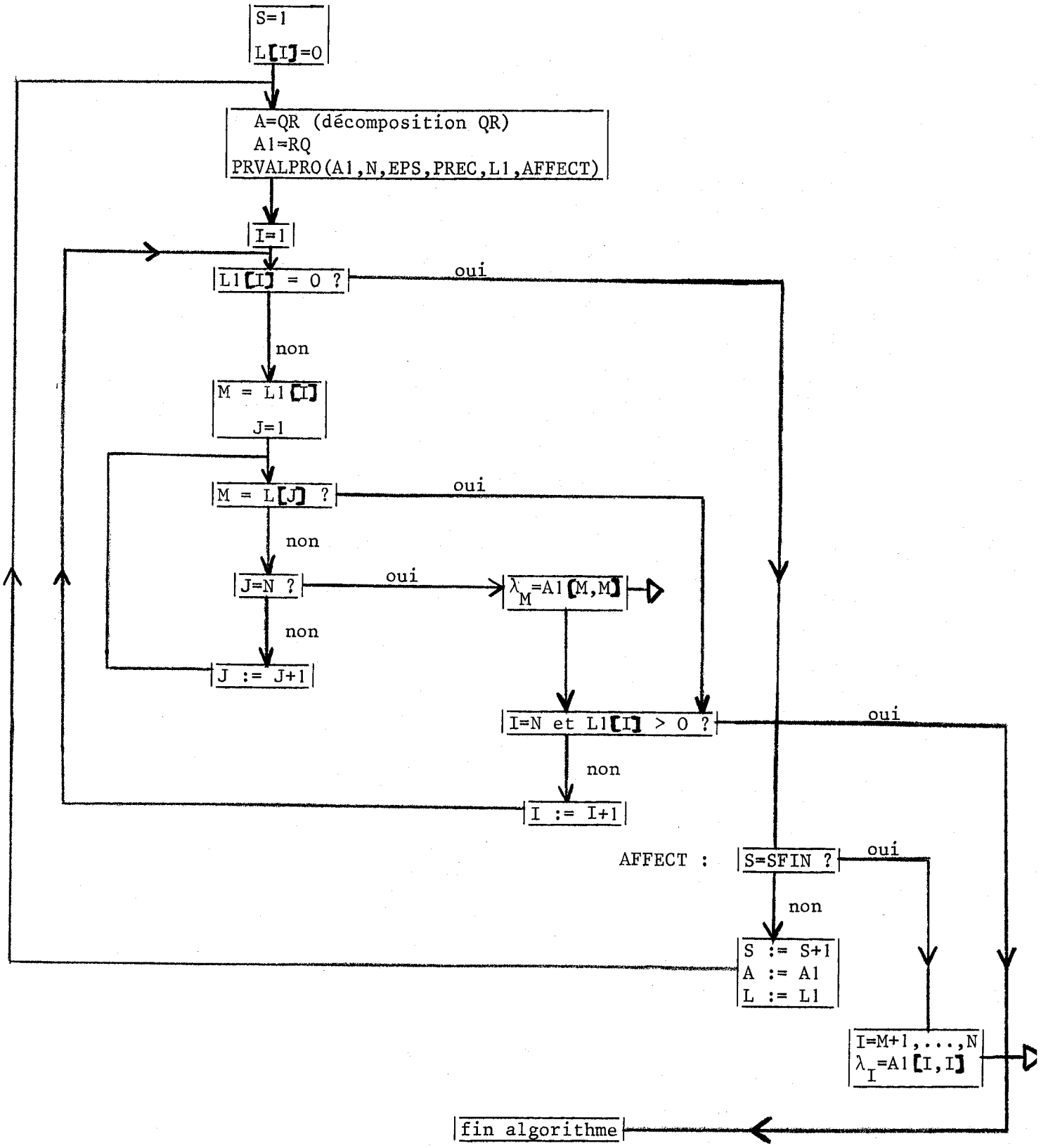
```
procédure PRVALPRO(A,N,EPS,PREC,L,SORTIE) ;
comment cette procédure calcule la précision sur les valeurs propres, d'une
matrice symétrique tridiagonale A = D+H, lorsque A est approchée par D.
Si ||H|| est trop grand, le calcul est impossible, la sortie de la procédure
se fait par SORTIE.
Si ||H|| est assez petit, le tableau PREC est chargé avec la précision calculée.
Le tableau L contient, ordonnés en croissant, les indices de valeurs propres
pour lesquelles la précision calculée est inférieure à EPS, nombre donné,
sinon 0 ;

value N ;
integer array L ; array A,PREC ; integer N ; real EPS ;
begin integer I,J ;
      real B,C,SOM ;
      array K,D,T,DELT,R,S[1,N] ;
      comment calcul de KI ;
      K[1] := A[2,1] ;
      for I := 2 step 1 until N-1 do
      K[I] := SQRT(A[I,I-1]*A[I,I-1]+A[I+1,I]*A[I+1,I]) ;
      K[N] := A[N,N-1] ;
      comment calcul de DI,TI,DELTI,RI,SI ;
      D[1] := ABS(A[1,1]-A[2,2]) ;
      for I := 2 step 1 until N do
      begin B := ABS(A[1,1]-A[I,I]) ;
            if B < D[1] then D[1] := B ;
            for J := 1 step 1 until I-1, I+1 step 1 until N do
            begin C := ABS(A[I,I]-A[J,J])
                  D[I] := if C < B then C else B
            end
```

```
      end ;
      SOM := 0.0 ;
      for I := 1 step 1 until N-1 do
      SOM := SOM+A[I+1,I]*A[I+1,I] ;
      SOM := SQRT(2.0*SOM) ;
      for I := 1 step 1 until N do
      begin T[I] := 2.0*SOM/D[I] ;
            if T[I] < 1.0 then goto ET else goto SORTIE ;
      ET :  DELT[I] := 2.0*K[I]/D[I] ;
            B := DELT[I] ;
            R[I] := 2.0*B/(2.0+B*B-T[I]) ;
            C := R[I] ;
            S(I) := if C > '-4 then(SQRT(1.0+C*C)-1.0)/C
                    else C*(0.5-C*C/8.0) ;
            PREC[I] := S[I]*K[I] ;
      end ;
      for I := 1 step 1 until N do L[I] := 0 ;
      I := 1 ; J := 1 ;
      ETQ : if PREC[I] < EPS then
            begin L[J] := I ;
                  J := J+1
            end ;
            if I < N then begin I := I+1 ;
                                goto ETQ
                          end
end procedure PRVALPRO ;
```

3.2 - Cette procédure peut être utilisée à chaque itération de l'algorithme QR appliqué à une matrice symétrique tridiagonale, selon l'organigramme ci-dessous. On fait écrire les valeurs propres pour lesquelles la précision EPS donnée est atteinte, si elles n'ont pas été écrites à l'itération précédente.

L'algorithme s'arrête :

- soit dès que toutes les valeurs propres ont été calculées à la précision EPS.
- soit lorsque le nombre d'itération atteint une valeur maximum fixée SFIN. On écrit alors les valeurs propres non encore calculées et la précision correspondante.


S=1
L[I]=0
A=QR (décomposition QR)
A1=RQ
PRVALPRO(A1,N,EPS,PREC,L1,AFFECT)
I=1
L1[I] = 0 ?
oui
non
M = L1[I]
J=1
M = L[J] ?
oui
non
J=N ?
oui
λ_M=A1[M,M]
non
J := J+1
I=N et L1[I] > 0 ?
oui
non
I := I+1
AFFECT :
S=SFIN ?
oui
non
S := S+1
A := A1
L := L1
I=M+1,...,N
λ_I=A1[I,I]
fin algorithme

CHAPITRE - III

PROBLEME INVERSE D'UN PROBLEME DE VALEURS PROPRES

On considère le problème suivant, rencontré dans beaucoup d'applications (systèmes dynamiques en chimie physique, réseaux électriques finis, spectrométrie physique, physique nucléaire etc...).

(P) : étant donnée une matrice carrée A d'ordre n (hermitienne si elle est complexe, ou réelle non nécessairement symétrique), à quelle condition est-il possible de trouver une diagonale réelle V telle que A+V admette des valeurs propres réelles données ?

C'est un problème inverse de l'habituel problème de recherche de valeurs propres.
Il peut aussi se formuler ainsi : peut-on trouver une diagonale réelle V et une matrice non singulière X telles que $A+V = XSX^{-1}$, si S est la diagonale des valeurs propres. Or la relation entre les éléments d'une matrice diagonalisable et les vecteurs propres est très complexe et mal connue. On en a une idée en remarquant que le problème posé équivaut, algébriquement, à écrire que le polynôme caractéristique de A+V (de degré n) admet n zéros donnés.

Ce problème, où A est donnée et V inconnue est différent de celui, analogue dans sa formulation, où A est inconnue et V donnée. Pour ce dernier problème, Horn [20] puis Mirsky [40] ont établi une CNS le résolvant entièrement. Dans le cas présent, nous n'obtiendrons qu'une condition suffisante [33].

On peut immédiatement faire les remarques suivantes :

1) Si V est une solution pour A, V-D est une solution pour A+D, quelle que soit la matrice diagonale D. On peut donc supposer A de diagonale nulle.

2) Les n valeurs propres données $s_1,\ldots,s_n$ forment un ensemble non ordonné, tandis que lorsqu'on écrit $A+V = XSX^{-1}$, cela présuppose un ordre sur S.

I - UNE CONDITION NECESSAIRE ET SUFFISANTE

Un premier résultat est constitué par la :

Proposition 1

Une CNS pour que l'ensemble des valeurs propres de A+V *, lorsque* V *parcourt l'ensemble des matrices diagonales d'ordre* n*, soit* $\mathbb{R}^n$ *est que* A *soit nulle.*

La C.S. est évidente.
La C.N. découle du fait que, pour que A+V soit semblable à la diagonale S, il faut que

$$\text{(1)} \qquad \operatorname{tr} S^2 - \frac{1}{n}(\operatorname{tr} S)^2 \geq \operatorname{tr} A^2$$

où tr A désigne la trace de A.
On peut choisir S tel que le 1er terme de (1) soit nul.
La démonstration de (1) peut se trouver, dans le cas où A est hermitienne dans Hadeler [18], dans le cas où A est quelconque dans Oliveira [44].

II - ETUDE DE (P) LORSQUE A EST HERMITIENNE

1 - Conditions suffisantes pour que (P) admette une solution

Reprenant une idée de Hadeler [18], nous allons chercher V par une méthode de point fixe.

A tout vecteur $v = (v_i)$ de $\mathbb{R}^n$, on associe la matrice $V = (v_i \delta_{ij})$, diagonale de A+V. On note $\lambda = (\lambda_i)$ le vecteur de $\mathbb{R}^n$ dont les composantes sont les valeurs propres de A+V. On désigne par F l'application : $v \overset{F}{\mapsto} F(v) = \lambda$, et on applique le théorème du point fixe à : $v \overset{T}{\mapsto} v+s-F(v)$.

1.1 - Soit $\mathcal{C}$ le demi-cône de $\mathbb{R}^n$ défini par :

$$\mathcal{C} = \{v \ ; \ v_1 \leq v_2 \leq \ldots \leq v_n\}.$$

Supposons que les λ_i sont numérotés de sorte que $\lambda \in \mathcal{C}$. Alors la restriction de F à $\mathcal{C}$ est à valeur dans $\mathcal{C}$ et continue.
Soit s un vecteur donné de $\mathcal{C}$. On pose

$$d(s) = \min_i (s_{i+1} - s_i).$$

Dire que A+V admet $s_1, \ldots, s_n$ pour valeurs propres équivaut à dire que $F(v) = s$, ou encore $T(v) = v$, c'est-à-dire v est point fixe de T.
Soit B la boule :

$$B = \{v \ ; \ \varphi_\infty(v-s) \leq \rho(A)\}$$

B est contenue dans $\mathcal{C}$ dès que :

$$(2) \qquad d(s) \geq 2\rho(A).$$

On suppose cette condition sur s remplie, la restriction de T à $\mathcal{C}$ est alors à valeurs dans $\mathcal{C}$. Elle applique continûment B dans B puisque :

$$\varphi_\infty(T(v)-s) = \varphi_\infty(F(v)-v) \leq \rho(A)$$

d'après un résultat bien connu [21] sur la localisation des valeurs propres d'une matrice hermitienne.
T admet donc un point fixe dans B. D'où le :

THEOREME 1

A étant une matrice hermitienne à diagonale nulle, une condition suffisante pour qu'il existe une diagonale réelle V telle que A+V admette les valeurs propres données $s_1,\ldots,s_n$ est que :

$$(2) \qquad d(s) = \min_i (s_{i+1}-s_i) \geq 2\rho(A).$$

D'après la 1ère remarque de l'introduction, on peut affiner ce résultat en cherchant une diagonale D telle que A' = A+D ait un rayon spectral $\rho(A')$ inférieur à celui de A.
Si on prend $D = \alpha I$, avec $\alpha = \frac{\lambda_1+\lambda_n}{2}$
$\rho(A') = \frac{\lambda_n-\lambda_1}{2}$ est toujours inférieur ou égal à $\rho(A)$.
D'une manière générale, la quantité :

$$\ell(A) = \max_{i\neq j} |\lambda_i-\lambda_j|$$

largeur du spectre de A, a été étudiée par Mirsky [39].
Il a montré que :

$$\ell(A) \leq (2\||A|\|^2 - \frac{2}{n}\,|\mathrm{tr}\ A|^2)^{1/2}$$

et si A est hermitienne :

$$\ell(A) \geq \max_{i\neq j} (|a_{ii}-a_{jj}|^2+4|a_{ij}|^2)^{1/2}.$$

Dans notre cas :

$$2 \max_{i\neq j}|a_{ij}| \leq \lambda_n-\lambda_1 \leq \sqrt{2}\,||\,|A|\,||$$

D'où le :

THEOREME 2

A étant une matrice hermitienne de diagonale nulle et de valeurs propres $\lambda_1,\ldots,\lambda_n$ *, il existe une diagonale réelle* V *telle que* A+V *admette* $s_1,\ldots,s_n$ *pour valeurs propres, si :*

$$(3) \qquad d(s) \geq \max_{i\neq j}|\lambda_i-\lambda_j|$$

Remarques -

1. Par rapport à la proposition du § I, les théorèmes 1 et 2 peuvent s'interpréter comme donnant une borne supérieure, fonction de S, de la perturbation hermitienne hors-diagonale de V.

2. Dans le cas où n = 2, (2) <=> (3) est une CNS.

3. K.P. Hadeler [18] est parti du résultat suivant sur la localisation des valeurs propres d'une matrice hermitienne A, en posant :

$$d = \min_{i\neq j}|a_{ii}-a_{jj}| \quad , \quad k(A) = \max_i\Big(\sum_{j\neq i}|a_{ij}|^2\Big)^{1/2}$$

$$|\lambda_i-a_{ii}| \leq \frac{1}{2}\left(d-\sqrt{d^2-4k^2(A)}\right) \quad , \quad i = 1,\ldots,n$$

dès que les disques de Krylov-Weinstein :

$$\{z \;;\; |z-a_{ii}| \leq k_i\} \quad \text{avec} \quad k_i = \Big(\sum_{j\neq i}|a_{ij}|^2\Big)^{1/2},$$

sont disjoints, c'est-à-dire dès que $d \geq 2k(A)$.

Il a alors obtenu la condition suffisante :

(4) $$d(s) \geq 2\sqrt{3}\, k(A).$$

Comparons ce résultat à (2) et (3).
On peut montrer [61] que :

$$n^{-1/2}\, \rho(A) \leq k(A) \leq \rho(A),$$

l'inégalité de gauche étant atteinte pour des matrices de rang 1.

1) pour $n \leq 3$, $\rho(A) \leq \sqrt{3}\, k(A)$ et (2) est moins restrictif que (4).

2) pour $n > 3$, pour que (2) soit moins restrictif que (4), il suffit que :

$$\frac{\rho(A)}{k(A)} \leq n^{1/2}\, \frac{\rho(A)}{\|A\|} = \left(\frac{n}{1+\Sigma\left(\frac{\lambda_i}{\rho(A)}\right)^2}\right)^{1/2} < \sqrt{3}.$$

Soit :

$$1+\Sigma\left(\frac{\lambda_i}{\rho(A)}\right)^2 > \frac{n}{3}.$$

Le membre de gauche est une somme de n nombres inférieurs ou égaux à 1.
Il suffit donc que :

$$\frac{1}{\rho(A)} \min_i |\lambda_i| > \frac{1}{\sqrt{3}} \sqrt{\frac{n-3}{n-1}}$$

c'est-à-dire :

$$1 \geq \gamma(A) > \frac{1}{\sqrt{3}} \sqrt{\frac{n-3}{n-1}}$$

où

$$\gamma(A) = \frac{|\lambda|_{min}}{|\lambda|_{max}} = \frac{1}{S_{\varphi_2\varphi_2}(A)\, S_{\varphi_2\varphi_2}(A^{-1})}$$

est le conditionnement euclidien de A.

<u>Exemple</u> : soit $A = (a_{ij})$ avec

$$a_{ij} = 0 \text{ sauf } a_{i+1,i} = a_{i,i+1} = 1, \quad i = 1,\dots,n-1$$

A admet pour valeurs propres $\lambda_i = 2\cos \frac{i\pi}{n+1}$, $i = 1,\dots,n$.

Comparons les conditions suffisantes données : $k(A) = \sqrt{2}$ et $\rho(A) = 2\cos \frac{\pi}{n+1} < 2$

(2) $\Leftrightarrow$ (3): $d(s) \geq 2 \cos \frac{\pi}{n+1}$

(4) : $d(s) \geq 2\sqrt{6} = 4,9$.

1.2 - On peut déduire une autre condition suffisante à partir du théorème de Krylov-Weinstein [72]. Si les disques de Krylov de centre a_{ii} sont disjoints, il existe une numérotation des valeurs propres de A telle que :

$$|\lambda_i - a_{ii}| \leq \Big(\sum_{j\neq i} |a_{ij}|^2\Big)^{1/2}, \quad i = 1,\dots,n.$$

Posons $k_i = \Big(\sum_{j\neq i} |a_{ij}|^2\Big)^{1/2}$.

On suppose qu'il existe une numérotation des s_i telle que :

(5) $$|s_i - s_j| > 2(k_i + k_j), \quad i \neq j.$$

Considérons alors l'hyperparallélépipède :

$$B = \{v \ ; \ |v_i - s_i| \leq k_i \ , \ i = 1,\dots,n\}.$$

Quel que soit v dans B, les disques de Krylov pour A+V sont disjoints et

$$|\lambda_i - v_i| \leq k_i, \quad i = 1,\dots,n$$

ou encore $T(B) \subset B$.

T admet un point fixe dans B. D'où le :

THEOREME 3

A étant une matrice hermitienne à diagonale nulle, il existe une diagonale réelle V telle que A+V admette des valeurs propres s_i données s'il existe un ordre sur les s_i tel que :

$$(5) \qquad |s_i - s_j| \geq 2(k_i + k_j), \quad i \neq j$$

2 - Convergence de la méthode des approximations successives

En vue de la résolution numérique du problème, nous étudions la convergence de l'itération :

$$(I) \qquad \begin{cases} v_o \\ v_{n+1} = T(v_n) = v_n - s + F(v_n) \ , \quad n = 1,2,\ldots \end{cases}$$

On suppose

$$(1) \qquad d(s) \geq 2\rho(A)$$

ou

$$(3) \qquad d(s) \geq \max_{i \neq j} |\lambda_i - \lambda_j|$$

Il existe alors $\tilde{v}$, solution de $F(\tilde{v}) = s$: soit $A + \tilde{V} = QSQ^H = M$.

Soit $J = (\frac{\partial \lambda_i}{\partial v_j})_{\tilde{v}}$ le jacobien de F en $\tilde{v}$.

Alors le jacobien de T est I-J. S'il existe une solution $\tilde{v}$, il existe un voisinage de $\tilde{v}$ dans lequel (I) converge, dès que

$$\rho(I-J) < 1.$$

Calculons J. D'après l'ordre choisi pour les valeurs propres, elles sont des fonctions des v_j, différentiables par morceaux en général [25]. (2) implique qu'en $\tilde{v}$, les valeurs propres sont bien séparées ; il existe donc un voisinage de $\tilde{v}$ dans lequel les λ_i sont continûment différentiables.

On pose :

$$q = q_{.i}\ ,\ \lambda = \lambda_i, \qquad ' = \frac{\partial}{\partial v_j}$$

Alors :

$$Aq+Vq = \lambda q$$

$$Aq'+V'q+Vq' = \lambda' q+\lambda q'$$

$$(Aq,q')+(Vq,q') = \lambda(q,q')$$

$$(Aq',q)+(V'q,q)+(Vq',q) = \lambda'(q,q)+\lambda(q',q)$$

D'où :

$$\lambda' = (V'q,q) = q_{ji}^2$$

et

$$J = (q_{ji}^2)$$

Or

$$\|I-J\|_\infty = \max_j \sum_{i\neq j} (|q_{ji}|^2+1-|q_{jj}|^2) = 2 \max_j (1-|q_{jj}|^2)$$

$$\|I-J\|_\infty < 1 \text{ implique } \rho(I-J) < 1$$

et

$$\|I-J\|_\infty < 1 \iff |q_{ii}|^2 > \frac{1}{2} \qquad i = 1,\dots,n.$$

Soit $\min_i |q_{ii}| > \frac{1}{\sqrt{2}}$.

Si θ_i désigne l'angle aigu entre e_i et $q_{.i}$,

$$|q_{ii}| = \frac{|(q_{.i},e_i)|}{\varphi_2(q_{.i})\varphi_2(e_i)} = \cos\theta_i \qquad i = 1,\dots,n.$$

En appliquant les résultats du chapitre II :

a) dans le cas de l'inégalité (2) : $M = \tilde{V}+A$

$$\alpha = \frac{2\rho(A)}{d(s)} < 1 \Rightarrow \operatorname{tg}\theta_i \leq \frac{1}{\alpha}\left(\sqrt{1+\alpha^2}-1\right) < 0{,}414$$

puisque A est à diagonale nulle.

b) dans le cas de l'inégalité (3) : $M = \tilde{V}-D+(A+D)$

$$= \tilde{V}'+A'$$

$$\alpha' = \frac{2\rho(A')}{d(s)} < 1 \Rightarrow tg\theta_i \leq \frac{\alpha}{1+\sqrt{1-\alpha^2}} < 1$$

puisque A' est à diagonale non nulle.

Dans les deux cas :

$$tg\theta_i < 1 \ , \quad \theta_i < \frac{\pi}{4} \quad \text{et} \quad \cos\theta_i > \frac{1}{\sqrt{2}}$$

D'où la :

Proposition 2

La condition

(2') $d(s) > 2\rho(A)$

ou

(3') $d(s) > \max\limits_{i\neq j} |\lambda_i-\lambda_j|$

assure la convergence de (I) dans un voisinage de la solution.

Remarque - On retrouve la condition suffisante :

$$|q_{ii}| > \frac{1}{\sqrt{2}}$$

donnée par Downing et Householder [10] pour que l'itération du 1er ordre proposée converge.

3 - Exemple

Nous étudions les différentes conditions proposées, sur l'exemple de Hadeler [18] :

$$A = \begin{pmatrix} 0 & 4 & 3 & 2 & 1 \\ 4 & 0 & 3 & 2 & 1 \\ 3 & 3 & 0 & 2 & 1 \\ 2 & 2 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} ,$$

Le vecteur s prend les différentes valeurs :

$$s_\nu = (0,8)^\nu \times \begin{pmatrix} 0 \\ 32 \\ 60 \\ 90 \\ 115 \end{pmatrix} , \quad \nu = 0,1,2,\ldots$$

$$d(s) = (0,8)^s \times 25$$

$1 - k(A) = \sqrt{30}$, $\rho(A) = 8,616$, $\|\,|A|\,\| = 10$, $\ell(A) = 12,616$

(2) $\quad d(s) \geq 2 \times 8,616 \Rightarrow \nu = 0,1$

(3) $\quad d(s) \geq 12,616 \quad \Rightarrow \nu = 0,1,2,3$

$\quad d(s) \geq 2\sqrt{10} \quad \Rightarrow \nu = 0,1,2$

(4) $\quad d(s) \geq 2\sqrt{3}\sqrt{30} \quad \Rightarrow \nu = 0$

alors que l'itération (I) converge pour $\nu = 0,1,\ldots,7$ et diverge à partir de $\nu = 8$.

2 - $k_1 = k_2 = 5,48$; $k_3 = 4,80$; $k_4 = 3,60$; $k_5 = 2$.
Avec l'ordre :

$$(s_o)_1 = 0,\ (s_o)_2 = 90,\ (s_o)_3 = 32,\ (s_o)_4 = 60,\ (s_o)_5 = 115$$

(5) est vérifié pour $\nu = 0,1,2$.

III - A EST REELLE

1 - Conditions suffisantes

Nous considérons maintenant A comme une perturbation réelle de V, matrice diagonale réelle. A+V admet-elle encore des valeurs propres réelles ?

Chaque valeur propre λ de A+V appartient au moins à l'un des disques :

$$\{z \in \mathbb{C} \ ; \ |z-v_i| \leq \|A\|_2\} \ , \ i = 1,\dots,n$$

d'après l'inégalité de Bauer-Fike [21].
A+V étant réelle, si elle admet une valeur propre complexe, elle admet la complexe conjuguée. Si les n disques précédents sont disjoints, chacun ne contient qu'une valeur propre qui ne peut donc être que réelle ; c'est le cas si :

$$d(v) = \min_{i \neq j} |v_i - v_j| \geq 2\|A\|_2 .$$

Nous gardons les notations du paragraphe précédent.
s étant donné dans C, soit B la boule :

$$B = \{v \ ; \ \varphi_\infty(v-s) \leq \|A\|_2\}$$

L'hypothèse

$$(5) \qquad d(s) \geq 4\|A\|_2$$

assure que tout v de B appartient à C et vérifie

$$d(v) \geq 2\|A\|_2 .$$

T est alors une application continue de B dans B.

D'où le :

THEOREME 4

A étant une matrice réelle de diagonale nulle, une condition suffisante pour qu'il existe une diagonale réelle V telle que A+V admette les valeurs propres données $s_1,\dots,s_n$ *est que :*

$$(5) \qquad d(s) = \min_i |s_{i+1}-s_i| \geq 4\|A\|_2 .$$

Remarques

1 - La condition (5) n'est que suffisante comme on peut le voir en explicitant le cas où n = 2.

2 - On peut faire un raisonnement analogue à partir de la localisation par les disques de Gershgorin. On retrouve la condition suffisante de Oliveira [44] :
il existe un ordre sur les s_i tel que :

$$|s_i-s_j| \geq 2(P_i+P_j) \text{ avec } i \neq j, \qquad P_i = \sum_{j\neq i} |a_{ij}|$$

IV - GENERALISATION DE (P) A UNE VARIETE LINEAIRE

Considérons maintenant le problème :

(P') : On se donne une matrice hermitienne A, ainsi que N matrices hermitiennes $M_1,\dots,M_N$ et on cherche s'il existe N réels $v_1,\dots,v_N$ tels que, si $V = \sum_1^N v_i M_i$, A+V admette des valeurs propres réelles données $s_1,\dots,s_n$.

Nous avons étudié au § II le cas où :

$$N = n,\ M_i = E_{ii},\ V = \sum_1^n v_i E_{ii} \quad \text{est diagonale.}$$

E_{ii} est la matrice admettant un 1 en ième ligne et ième colonne et des 0 ailleurs.

On suppose $s_1 \leq \ldots \leq s_n$ et on note (ν_i) les valeurs propres de V.

On considère la variété linéaire $\mathcal{V}$ engendrée par $M_1,\ldots,M_N$ dans $\mathbb{C}^{n\frac{n+1}{2}}$.

On remarque qu'il est facile de passer du <u>problème affine</u> : trouver V dans la variété telle que A+V ait des valeurs propres données (s_i), au <u>problème linéaire</u> : trouver V dans la variété, telle qu'elle admette des valeurs propres données (ν_i).

Reprenant le schéma de démonstration du § II, si $d(s) \geq 2\rho(A)$ le n-uple $\nu_1 \leq \ldots \leq \nu_n$ des valeurs propres de V est déterminé par la méthode des approximations successives.

Il reste à déterminer les coefficients v_i de la combinaison linéaire : $V = \sum_1^N v_i M_i$, pour que V admette les ν_i pour valeurs propres.

Ce problème est défini par la correspondance $(\nu_i)_{1,\ldots n} \longmapsto (v_i)_{1,\ldots,N}$ qui est, dans le cas général, très difficile à expliciter.

Soit $D = (\nu_i)$ la diagonale de V, il s'agit de trouver une matrice unitaire Q et N nombres $v_1,\ldots,v_N$, tels que :

$$(6) \qquad V = \sum_1^N v_i M_i = QDQ^H \iff D = \sum_1^N v_i\ Q^H\ M_i\ Q.$$

Explicitons deux exemples simples :

1 - $N \geq n$ et $\mathcal{V}$ contient les diagonales

$$M_i = E_{ii},\ i = 1,\ldots,n \quad \text{implique} \quad Q = I,\ v_i = \nu_i.$$

2 - $N = n$ et $\mathcal{V}$ est l'ensemble des matrices de vecteurs propres donnés

$$M_i = M^{i-1},\ i = 1,\dots,n,\ V = \sum_1^n v_i\, M^{i-1}$$

On note $m_1,\dots,m_n$ les valeurs propres de M.

L'application $v_i \mapsto \nu_i$ est définie par la matrice de Vandermonde construite sur les points $m_1,\dots,m_n$. Elle est inversible si et seulement si les m_i sont distincts.

CHAPITRE IV

ETUDE DE LA STABILITE DES METHODES D'APPROXIMATION D'ELEMENTS PROPRES D'OPERATEURS LINEAIRES

Pour calculer numériquement les valeurs propres et les vecteurs propres d'un opérateur linéaire T dont le domaine de définition et l'image sont dans un espace de Banach séparable X, on les approche par les éléments propres d'un opérateur $\mathcal{T}$ dans un espace de dimension finie. Il existe essentiellement deux types de méthodes : les méthodes d'approximation (méthode aux différences finies, interpolation...) où $\mathcal{T}$ est un opérateur, défini dans un espace de dimension finie, qui approche T ; les méthodes de projection (Ritz, Galerkin, Aronzajn...) qu'on peut définir lorsqu'il existe dans X, une suite croissante de sous-espaces vectoriels dont la réunion est dense dans X. Le deuxième type de méthodes n'est qu'un cas particulier du premier et nous ne l'étudions de manière spécifique que lorsque nous pouvons obtenir des résultats plus précis.

Nous étudions une condition suffisante générale portant sur l'approximation $\mathcal{T}$, pour que les éléments propres de $\mathcal{T}$ approchent ceux de T. Une approximation $\mathcal{T}$ vérifiant cette condition suffisante est dite <u>stable</u>.

I - RAPPELS (Kato [25])

1 - Soit T un opérateur linéaire fermé de domaine D(T) et d'image Im(T) dans X, espace de Banach séparable sur $\mathbb{C}$. 1 représente l'application identité sur X. $z \in \mathbb{C}$ est dit point régulier de T si l'opérateur résolvant $R(z) = (T-z1)^{-1}$ est borné, de domaine X.

L'ensemble des points réguliers de T forme $\rho(T)$, ensemble résolvant de T.

L'ensemble complémentaire $\sigma(T)$ de $\rho(T)$ dans $\mathbb{C}$ est appelé le spectre de T. Il est formé du spectre ponctuel P_σ, du spectre continu C_σ et du spectre résiduel R_σ.

Le spectre ponctuel est constitué des valeurs propres.

Une condition nécessaire et suffisante pour que λ appartienne à P_σ est que l'équation $Tx = \lambda x$ admette une solution $x \neq 0$, c'est-à-dire $\mathrm{Ker}(T-\lambda 1) \neq \{0\}$.

$N_\lambda = \mathrm{Ker}(T-\lambda 1)$ est le sous-espace propre associé à λ, c'est l'ensemble des vecteurs propres. $\gamma = \dim \mathrm{Ker}(T-\lambda 1)$ est la multiplicité géométrique de λ.

$\lambda \in P_\sigma$ est une valeur propre isolée si λ est un point isolé de $\sigma(T)$.
Soit alors Γ une courbe fermée simple de Jordan entourant λ et aucun autre point de $\sigma(T)$. L'opérateur $P = \frac{-1}{2i\pi} \int_\Gamma R(z)dz$ est la projection spectrale associée à λ.

$M_\lambda = PX$ est le sous-espace invariant associé à λ. $m = \dim PX$ est la multiplicité algébrique de λ.

λ est dite valeur propre simple si $\gamma = m = 1$,
semi simple si $\gamma = m$.

Dans ce qui suit, nous nous intéresserons uniquement aux valeurs propres isolées, de multiplicité algébrique finie.

2 - Nous utiliserons les notions de convergence suivantes :

a) pour des opérateurs de X :

1) convergence simple :

T, T_n étant des opérateurs fermés de X : $n \in \mathbb{N}$

$T_n \to T$ <u>simplement</u> si tout x de D(T) appartient à $D(T_n)$ pour n assez grand et si $T_n x \to Tx$.

2) convergence en norme :

$T, T_n \in \mathcal{B}(X)$

$T_n \to T$ <u>en norme</u> si $\|T-T_n\| \to 0$.

b) Pour des ensembles de points de $\mathbb{C}$ (Kuratowski [29]), $\{E_n\}$ est une suite de parties de $\mathbb{C}$.

1) limite supérieure :

$\overline{\lim}\, E_n = \{x \in E \; ; \; \exists x_k, \; x_k \in E_{n_k}, \; x_k \to x, \quad n_k \in \mathbb{N}\}$

2) limite inférieure :

$\underline{\lim}\, E_n = \{x \in E \; ; \; \exists x_n, \; x_n \in E_n, \; x_n \to x\}$

On a toujours $\underline{\lim}\, E_n \subset \overline{\lim}\, E_n$ et lorsque $\underline{\lim}\, E_n = \overline{\lim}\, E_n$, on appelle cet ensemble $\lim E_n$.

II - DEFINITIONS

Soit T un opérateur fermé de X de domaine de définition et d'image dans X, espace de Banach séparable.

Soit h un paramètre positif. A chaque h on associe [4] :

- X_h un espace vectoriel de dimension finie normé par $\|.\|_h$
- r_h un opérateur dit de restriction : $X \xrightarrow{r_h} X_h$
- T_h un opérateur linéaire de X_h dans lui-même.

On fait les deux hypothèses suivantes :

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "r_h"'] & X \arrow[d, "r_h"] \\ X_h \arrow[r, "\mathcal{T}_h"] & X_h \end{tikzcd}$$

1) la norme $\|.\|_h$ de X_h est <u>compatible</u> avec la norme $\|.\|$ de X :

$\|r_h x\|_h \xrightarrow[h\to 0]{} \|x\|$, pour $x \in X$

La norme d'opérateur dans X_h est aussi notée $\|.\|_h$

2) $\mathcal{T}_h$ est une <u>approximation de T</u> :

$\|(r_h T - \mathcal{T}_h r_h)x\|_h \xrightarrow[h\to 0]{} 0$ pour $x \in D(T)$.

<u>N.B.</u> On dit aussi que $\mathcal{T}_h$ est un schéma discrètement consistant à T.

Soit λ une valeur propre isolée de T : il existe R tel que le disque $\Gamma(\lambda,R) = \{z \; ; \; |z-\lambda| \leq R\}$ isole λ, c'est-à-dire ne contient aucun autre point du spectre de T.
On pose la définition suivante : [7]

<u>Définition</u> :

$\mathcal{T}_h$ est stable dans le disque qui isole λ si :

$\forall r : 0 < r < R, \; \exists h_o, M > 0 : \|(\mathcal{T}_h - z1)^{-1}\|_h < M,$

pour $h < h_o$ et $z \in \{z \; ; \; r \leq |z-\lambda| \leq R\}$.

<u>N.B.</u> On dira aussi que $\mathcal{T}_h$ est stable dans $\Gamma(\lambda,R)$, ou encore que $\mathcal{T}_h$ est stable au voisinage de λ.

Puisque le rayon spectral d'un opérateur borné est inférieur ou égal à la norme de cet opérateur, il vient :

$$q_h(z) \leq \min_{\lambda^h \in \sigma(T_h)} |\lambda^h - z|.$$

Lemme 1

$q_h(z)$ *est une fonction continue de* z.

1) Si $q_h(z) = 0$, il existe $\lambda^h \in \sigma(T_h)$ tel que $\lambda^h = z$ et

$$q_h(z') \leq \min_{\lambda^h \in \sigma(T_h)} |\lambda^h - z'| \leq |z-z'|$$

2) Si $q_h(z) > 0$, nous allons montrer que

$$(1) \qquad \forall \varepsilon > 0,\ \exists \eta : |z'-z| < \eta \Rightarrow |q_h(z') - q_h(z)| < \varepsilon$$

Pour cela, nous considérons, d'une manière générale, T un opérateur fermé de X dans X, et z un point de $\rho(T)$.

$$T - z'1 = (T - z1)(1 - R(z)(z'-z)1) \quad \text{puisque } R(z) \text{ existe.}$$

Donc

$$R(z') = (1 - R(z)(z'-z)1)^{-1} R(z)$$

est borné si l'on peut définir le facteur $(\ )^{-1}$ par une série absolument convergent Cela est possible dès que :

$$\|R(z)\| |z'-z| < 1.$$

Alors :

$$R(z') = \left(\sum_{k=0}^{\infty} R^k(z)(z'-z)^k\right) R(z)$$

$R(z')$ appartient à $\mathcal{L}(X)$ et :

$$\|R(z')-R(z)\| \leq |z'-z|\,\|R(z)\| \sum_{k=0}^{\infty} |z'-z|^k \|R(z)\|^k < 2|z'-z|\,\|R(z)\|$$

$$\text{si } |z'-z|\,\|R(z)\| < \frac{1}{2}$$

Pour tout $\varepsilon > 0$ tel que $\varepsilon < \frac{1}{2\|R(z)\|}$, $|z'-z| < \varepsilon$ implique :

$$\big|\,\|R(z')\|-\|R(z)\|\,\big| \leq \|R(z')-R(z)\| < 2\varepsilon\|R(z)\|.$$

Remarquons que :

$$\left|\frac{1}{\|R(z')\|} - \frac{1}{\|R(z)\|}\right| = \frac{\big|\,\|R(z)\|-\|R(z')\|\,\big|}{\|R(z')\|\,\|R(z)\|} \leq \frac{2\varepsilon\|R(z)\|}{\|R(z)\|^2(1+2\varepsilon)} = \frac{2}{\|R(z)\|}\,\frac{\varepsilon}{1+2\varepsilon}$$

Appliquant ce résultat à T_h, on en déduit (1).

Soit λ une valeur propre de T isolée par $\Gamma(\lambda,R)$.

Lemme 2

Si T_h est une approximation de T, stable dans $\Gamma(\lambda,R)$, il existe dans tout voisinage de λ, une valeur propre de T_h pour h assez petit.

Pour tout r, $0 < r < R$, on définit le disque $D = \{z \;;\; |z-\lambda| \leq r\}$ de frontière $\gamma = \{z \;;\; |z-\lambda| = r\}$.
Par hypothèse :

$$\|(r_h T - T_h r_h)x\|_h \xrightarrow[h\to 0]{} 0, \quad \forall x \in D(T).$$

Au point φ, vecteur propre normé associé à λ : $(T-\lambda I)\varphi = 0$ et

$$\|(r_h(T-\lambda 1)-(T_h-\lambda 1)r_h)\varphi\|_h \to 0$$

$$\|(T_h-\lambda 1)r_h\varphi\|_h \to 0 \text{ et } \|r_h\varphi\|_h \to \|\varphi\| = 1.$$

Alors :

$$q_h(\lambda) = \min_{x\in X_h} \frac{\|(T_h-\lambda 1)x\|_h}{\|x\|_h} \leq \frac{\|(T_h-\lambda 1)r_h\varphi\|_h}{\|r_h\varphi\|_h}$$

et

$$q_h(\lambda) \xrightarrow[h\to 0]{} 0.$$

Puisque T_h est stable dans $\Gamma(\lambda,R)$:

$$\forall r,\ 0 < r < R,\ \exists h_o, M : \|(T_h-z1)^{-1}\|_h < M \text{ pour } h < h_o \text{ et } z \in \gamma$$

r étant fixé, soit h_1 l'indice ($h_1 \leq h_o$) déterminé par :

(2) $\quad h < h_1$ implique $q_h(z) > \frac{1}{M}$ pour $z \in \gamma$ et $q_h(\lambda) < \frac{1}{M}$.

Soit maintenant h fixé, inférieur à h_1. Lorsque $z \in D$, $q_h(z)$, fonction positive ou nulle, continue d'après le lemme 1, admet un minimum en un point λ^h de D, intérieur à γ d'après (2). Si ce minimum est positif, T_h-z1 n'admet pas de valeur propre dans D, $(T_h-z1)^{-1}$ est holomorphe en z dans D et $\|(T_h-z1)^{-1}\|_h$ admet un maximum en λ^h, point intérieur de D. Ceci n'est possible que si la fonction $(T_h-z1)^{-1}$ est constante en z sur D, ce qui n'est pas le cas, à cause de (2).

Donc $q_h(z)$ ne peut avoir de minimum différent de 0 dans D :

$$\forall r,\ 0 < r < R,\ \exists h_1 : h < h_1 \Rightarrow \exists \lambda^h : |\lambda-\lambda^h| < r.$$

Lemme 3

Si T_h est une approximation de T stable dans $\Gamma(\lambda,R)$, λ est le seul point limite possible de toute suite de valeurs propres de T_h appartenant à $\Gamma(\lambda,R)$.

La stabilité de T_h dans $\Gamma(\lambda,R)$ s'écrit :

$$\forall r,\ 0 < r < R,\ \exists h_o, M : q_h(z) > \frac{1}{M}$$

$$\text{pour } h < h_o \text{ et } z \in \{z\ ;\ r \leq |z-\lambda| \leq R\}.$$

Soit : $\min\limits_{\lambda^h \in \sigma(T_h)} |z-\lambda^h| > \frac{1}{M}$ pour $z \in \{z\ ;\ r \leq |z-\lambda| \leq R\}$.

Les valeurs propres de T_h ne peuvent être qu'à l'extérieur de cette couronne. Elles vérifient donc $|\lambda^h-\lambda| < r$ ou $|\lambda^h-\lambda| > R$: λ est le seul point limite possible de toute suite de valeurs propres de T_h dans le disque $\Gamma(\lambda,R)$.

3 - <u>*THEOREME 1*</u>

Soit σ_1 *un ensemble fini de valeurs propres isolées, de multiplicité finie, de* T : $\lambda_1,\ldots,\lambda_s$. *Soient* $\Gamma(\lambda_i,R_i)$ *des disques isolant* λ_i $i = 1,\ldots,s$. *Pour tout* $h > 0$ σ_{s1}^h *est l'ensemble des valeurs propres de* T_h *appartenant à* $D = \bigcup\limits_{i=1}^{s} \Gamma(\lambda_i,R_i)$ *où* T_h *est une approximation de* T *stable dans* D. *Alors :*

$$\sigma_1 = \underline{\lim}\ \sigma_1^h.$$

Ce théorème est une conséquence directe des lemmes 2 et 3 appliqués en chaque point de σ_1.

Nous verrons au § VI des conditions suffisantes de stabilité. On peut comparer la notion de stabilité classique pour la résolution de l'équation

$$Tu = f, \quad u,f \in X$$

en l'approchant par :

$$T_h u_h = r_h f, \quad u_h \in X_h$$

avec celle de stabilité au voisinage d'une valeur propre isolée, que nous venons d'introduire.
La première s'écrit : $\|T_h^{-1}\|_h < M$, et l'on sait que si T^{-1} existe, est borné :

$$\|u_h - r_h u\|_h \to 0 \iff \|T_h^{-1}\|_h < M.$$

Cela revient à dire que si $0 \in \rho(T)$, on fait, sur les valeurs propres λ^h de T_h, l'hypothèse :

$$\min_{\lambda^h \in \sigma(T_h)} |\lambda^h| > \frac{1}{M}.$$

La notion de stabilité classique est donc une notion de stabilité au voisinage de 0, elle entraîne que 0, point régulier de T, ne peut être limite d'une suite de valeurs propres de T_h.

4 - Exemples d'instabilité

4.1 - des valeurs propres de T ne sont pas obtenues comme limite de valeurs propres de T_h, lorsque $h \to 0$.

Exemple 1 : valeurs propres non isolées.

$X = \ell^p$, $1 \leq p \leq \infty$. Soit $\{e_i\}$ la base canonique.
T, opérateur de translation à gauche, est défini par :

$$\begin{cases} Te_1 = 0 \\ Te_i = e_{i-1}, \ i > 1. \end{cases}$$

Le spectre : $\sigma(T) = \{z \ ; \ |z| \leq 1\}$ se divise en spectre ponctuel : $P_\sigma = \{z \ ; \ |z| < 1\}$ et spectre continu : $C_\sigma = \{z \ ; \ |z| = 1\}$.

Soit X_n l'espace engendré par les n premiers vecteurs de la base, muni de la norme de Hölder φ_p :

L'opérateur $\mathcal{T}_n$, représenté par la matrice d'ordre n :

$$A_n = \begin{pmatrix} 0 & 1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$

est une approximation de T.
Or $\sigma(\mathcal{T}_n) = \{0\}$.

Exemple 2

$X = C(0,1)$ avec la norme uniforme

$D = \{ x \in X \ ; \ x' \in X \text{ et } x(0) = x(1)\}$

$T : x \mapsto x', \ x \in D$

$\sigma(T) = P_\sigma = \{2ik\pi, \ k \in \mathbb{Z}\}$, les valeurs propres sont simples.

$[0,1]$ est divisé en n parties égales aux points $t_i = \frac{i}{n} = hi$, $i = 0,1,\ldots,n$, et $x(t_i) = x_i$.

Le problème approché dans $\mathbb{R}^{n+1}$, avec la norme du maximum, est pris de la forme :

$$\begin{aligned} x_o &= x_n \\ x_1 - x_o &= \lambda h \ x_o \\ x_1 - x_o &= \lambda h \ x_1 \\ x_3 - x_2 &= \lambda h \ x_2 \\ &\text{--------} \\ x_n - x_{n-1} &= \lambda h \ x_{n-1} \end{aligned}$$

L'opérateur $\mathcal{T}_h$, de matrice $A_h = \frac{1}{h} \begin{pmatrix} -1 & 1 & & & 0 \\ -1 & 1 & & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots \\ 0 & & & & 1 \\ & & & & -1 \end{pmatrix}$ d'ordre n, admet pour valeurs propres 0 d'ordre 2 et $\frac{-1}{h}$ d'ordre n-2. Et $\underline{\lim}\ \sigma(\mathcal{T}_h) = \{0,\infty\}$.

Le point à l'infini peut être considéré comme valeur propre de T qui n'est pas borné ([25] p. 176).

4.2 - des points réguliers de T sont obtenus comme limites de valeurs propres de T_h, lorsque $h \to 0$.

Exemple 3 [25] Valeur propre de multiplicité infinie

Soit une suite de projections P_n de X convergeant vers 0 :

$$\|P_n x\| \to 0, \quad \forall x \in X.$$

En général $\sigma(P_n) = \{0,1\}$, tandis que $\sigma(0) = \{0\}$.

Exemple 4

$X = C(0,1)$ avec la norme uniforme

$D = \{x \in X \; ; \; x' \in X \text{ et } x'(0) = \alpha \, x(0)\}$.

$T : x \mapsto x'$, $x \in D$.

$\sigma(T) = \{\alpha\}$, α est valeur propre simple.

$[0,1]$ est divisé en n parties égales aux points $t_i = \frac{i}{n} = hi$, $i = 0,1,\ldots,n$ et $x(t_i) = x_i$.

Le problème approché dans $\mathbb{R}^{n+1}$, avec la norme du maximum, est choisi de la forme

$$\begin{aligned}
x_1 &= (1+\alpha h)x_o \\
x_1 - x_o &= \lambda h \, x_1 \\
x_{n-2} - x_{n-3} &= \lambda h \, x_{n-2} \\
x_n - x_{n-1} &= \lambda h \, x_{n-1} \\
x_n - x_{n-1} &= \lambda h \, x_n .
\end{aligned}$$

L'opérateur $\mathcal{T}_h$, de matrice $A_h = \frac{1}{h}\begin{pmatrix} \frac{\alpha h}{1+\alpha h} & & & & \\ -1 & 1 & & 0 & \\ & \ddots & \ddots & & \\ & & -1 & 1 & \\ 0 & & & -1 & 1 \\ & & & -1 & 1 \end{pmatrix}$ d'ordre n, admet pour valeurs propres 0 d'ordre 2, $\frac{1}{h}$ d'ordre n-3 et $\frac{\alpha}{1+\alpha h}$ d'ordre 1.

$\underline{\lim}\ \sigma(\mathcal{T}_h) = \{0,\alpha,\infty\}$.

5 - Continuité de la notion de stabilité

Soit toujours σ_1 une partie du spectre de T, constitué d'un nombre fini de valeurs propres isolées de multiplicité finie : $\lambda_1,\ldots,\lambda_s$. σ_1 est isolé par $D = \bigcup_{i=1}^{s} \Gamma(\lambda_i,R_i)$.

On considère les deux familles d'opérateurs :

1- pour $\varepsilon > 0$, T^ε est un opérateur de X dans X,

2- pour ε et $h > 0$, $\mathcal{T}_h^\varepsilon$, est un opérateur de X_h dans X_h,

vérifiant les 4 propriétés suivantes :

i) T^ε converge vers T :

$$\forall\eta\ ,\ \exists\varepsilon_o\ :\ \varepsilon < \varepsilon_o \Rightarrow \|(T-T^\varepsilon)x\| < \eta \text{ pour } x \in D(T) \subset D(T^\varepsilon).$$

ii) les valeurs propres λ_i^ε de T^ε dans D sont telles que :

$$\forall\eta,\ \exists\varepsilon_o\ :\ \varepsilon < \varepsilon_o \Rightarrow |\lambda_i-\lambda_i^\varepsilon| < \eta, \qquad i = 1,\ldots,s$$

iii) $\mathcal{T}_h^\varepsilon$ est une approximation de T^ε, stable dans D :

a) $\|(r_h T^\varepsilon - \mathcal{T}_h^\varepsilon r_h)x\|_h \xrightarrow[h\to 0]{} 0$ pour $x \in D(T^\varepsilon)$

b) pour toute courbe fermée γ inclue dans $\rho(T^{\varepsilon}) \cap D$:

$$\exists M_{\gamma} : \|(T_h^{\varepsilon}-z1)^{-1}\|_h < M_{\gamma}, \quad \text{pour } z \in \gamma.$$

iv) T_h^{ε} converge en norme vers une limite T_h, uniformément en h :

$$\forall \eta, \exists \varepsilon_o : \varepsilon < \varepsilon_o \Rightarrow \|(T_h^{\varepsilon}-T_h)\xi\|_h < \eta\|\xi\|_h, \text{ pour } \xi \in X_h.$$

On montre alors la :

Proposition

Si les familles T^{ε} *et* T_h^{ε} *vérifient les 4 propriétés ci-dessus,* T_h *est une approximation de* T *, stable dans* D .

Démonstration

Nous montrons d'abord que T_h est une approximation de T. Pour tout x dans D(T), on peut écrire :

$$\|(r_hT-T_hr_h)x\|_h \leq \|r_h(T-T^{\varepsilon})x\|_h+\|(r_hT^{\varepsilon}-T_h^{\varepsilon}r_h)x\|_h+\|(T_h^{\varepsilon}-T_h)r_hx\|_h$$

$$\forall \eta, \exists \varepsilon_o : \varepsilon < \varepsilon_o \Rightarrow \|(T-T^{\varepsilon})x\| < \eta \text{ et } \|(T_h^{\varepsilon}-T_h)r_hx\|_h < \eta\,\|r_hx\|_h.$$

Soit ε_o ainsi fixé. Si l'on fait tendre h vers 0 :

$$\|r_h(T-T^{\varepsilon})x\|_h \to \|(T-T^{\varepsilon})x\|$$

et

$$\|(r_hT^{\varepsilon}-T_h^{\varepsilon}r_h)x\|_h \to 0$$

Donc

$$\|(r_hT-T_hr_h)x\|_h \to 0 \text{ quand } h \to 0.$$

Etudions maintenant la stabilité de T_h.
Soit γ une courbe fermée inclue dans $\rho(T) \cap D$. Puisque les valeurs propres de T dans D tendent vers celles de T lorsque $\varepsilon \to 0$, il existe ε_1 tel que pour tout $\varepsilon < \varepsilon_1$, γ est inclue dans $\rho(T^\varepsilon) \cap D$.

$$\|(T_h - z1)\xi\|_h \geq \|(T_h^\varepsilon - z1)\xi\|_h - \|(T_h^\varepsilon - T_h)\xi\|_h \quad \text{pour } \xi \text{ dans } X_h \text{ et } z \text{ appartenant à } \gamma.$$

$$\geq \left(\frac{1}{M_\gamma} - \eta\right)\|\xi\|_h$$

$\|(T_h - z1)^{-1}\|_h$ est uniformément borné en h pour z appartenant à γ.

<u>Remarque</u> : si $\|T - T^\varepsilon\| \xrightarrow[\varepsilon \to 0]{} 0$, i) et ii) sont vérifiées.

IV - APPROXIMATION DES SOUS-ESPACES PROPRES ET INVARIANTS

Soit λ une valeur propre de T de multiplicité finie, isolée par $\Gamma(\lambda,R)$, et T_h une approximation stable dans $\Gamma(\lambda,R)$. On appelle λ^h les valeurs propres de T_h appartenant à $\Gamma(\lambda,R)$. On appelle Γ la frontière de $\Gamma(\lambda,R)$.

1 - Approximation des sous-espaces invariants

Soit P (resp. P_h) la projection spectrale dans X (resp. X_h) associée à λ (resp. λ^h) :

$$P = \int_\Gamma R(z)dz \quad , \quad P_h = \int_\Gamma R_h(z)dz.$$

1.1 - *THEOREME 2*

Soit λ une valeur propre de T isolée par $\Gamma(\lambda,R)$ et soit T_h une approximation de T stable dans $\Gamma(\lambda,R)$, on appelle λ^h les valeurs propres de T_h appartenant à $\Gamma(\lambda,R)$.
Si P est la projection spectrale associée à λ, P_h celle associée aux λ^h:

$$\|(r_h P - P_h r_h)x\|_h \xrightarrow[h \to 0]{} 0, \text{ pour } x \in X.$$

Démonstration

Par hypothèse :

$$\|(r_h T - \mathcal{T}_h r_h)x\|_h \to 0, \text{ pour } x \in D(T)$$

et

$$\forall r > 0,\ \exists h_o, M : \|(\mathcal{T}_h - z1)^{-1}\|_h < M \quad \text{pour } h < h_o,$$
$$z \in K = \{z\ ;\ r \leq |z-\lambda| \leq R\}.$$

Nous notons $R(z)$ et $R_h(z)$ les inverses, lorsqu'ils existent, de $T-z1$ et $\mathcal{T}_h - z1$.

Alors :

$$\|(r_h R(z) - R_h(z) r_h)x\|_h \to 0 \quad \text{pour } x \in X \text{ et } z \in K.$$

En effet,

$$(R_h(z) r_h - r_h R(z))(T-z1)x = R_h(z)(r_h T - \mathcal{T}_h r_h)x$$

Si $z \in \rho(T)$, $\mathrm{Im}(T-z1) = D(R(z)) = X$ et la convergence vers 0 découle de la borne uniforme de $\|R_h(z)\|_h$ lorsque $z \in K$.

On en déduit que :

$$\|(r_h P - P_h r_h)x\|_h \leq R \sup_{z \in \Gamma} \|(r_h R(z) - R_h(z) r_h)x\|_h$$

d'où la convergence vers 0.

Le théorème 2 signifie que, pour tout vecteur du sous-espace invariant associé à λ, isolée par Γ, il existe un vecteur du sous-espace invariant associé aux valeurs propres de $\mathcal{T}_h$ contenues dans Γ, qui converge, dans X_h, vers la restriction du vecteur initial à X_h.

1.2 - On suppose défini un opérateur $p_h : X_h \to X$, dit de prolongement [4], tel que :

$$p_h r_h x \to x \quad \text{pour } x \in X.$$

Dans la suite du chapitre, cette hypothèse sur p_h sera supposée vérifiée.

On dira que p_h est stable lorsque :

$$\exists C_1 : \|p_h \xi\| \leq C_1 \|\xi\|_h \text{ pour } \xi \in X_h$$

THEOREME 3

Avec les hypothèses du théorème 2, si p_h est un opérateur de prolongement stable,

$$\|(P-p_h P_h r_h)x\| \xrightarrow[h\to 0]{} 0, \text{ pour } x \in X.$$

Démonstration

$$(P-p_h P_h r_h)x = (1-p_h r_h)Px + p_h(r_h P - P_h r_h)x.$$

Et :

$$\|(1-p_h r_h)Px\| \to 0,$$

$$\|p_h(r_h P - P_h r_h)x\| \leq C_1 \|(r_h P - P_h r_h)x\|_h \to 0.$$

Nous avons montré que pour tout vecteur du sous-espace invariant associé à λ, isolée par Γ, il existe un vecteur du sous-espace invariant associé aux valeurs propres de T_h contenues dans Γ, dont le prolongement converge dans X vers le vecteur initial.

2 - Approximation des sous-espace propres

1) Si λ et λ^h sont semi-simples, les sous-espaces propres et invariants sont identiques. En particulier, si λ et λ^h sont des valeurs propres simples, les vecteurs propres approchés convergent vers la restriction du vecteur propre associé à λ :

$$\|(1-P_h)r_h \varphi\|_h \xrightarrow[h\to 0]{} 0, \text{ pour } \varphi \in \mathrm{Ker}(T-\lambda 1).$$

2) Sinon, la situation est beaucoup plus compliquée . On peut voir sur un exempl de Polskii [5 1] qu'il existe des vecteurs propres de λ dont les restrictions à X_h ne peuvent être obtenues comme limites de combinaisons linéaires de vecteurs propres de λ^h.

3 - Cas où dim $P_hX_h \leq$ dim $PX = m$: stabilité forte

Nous allons montrer avec cette hypothèse que les convergences énoncées aux théorèmes 2 et 3 sont uniformes en x. De plus, pour h assez petit, dim P_hX_h=dim PX

3.1 - On suppose l'opérateur de restriction $r_h : X \to X_h$, linéaire surjectif et stable :

$$\exists C_2 : \|r_h x\|_h \leq C_2 \|x\| , \quad \text{pour } x \in X.$$

Soit λ une valeur propre isolée de multiplicité m.
On définit :

$M = PX$, le sous-espace invariant associé à λ.

$m = \dim M$.

$M_h = P_hX_h$, le sous-espace invariant associé aux λ^h

$m_h = \dim M_h$

$\bar{M}_h = r_h M$, le sous-espace de X_h, restriction de M

$\bar{m}_h = \dim \bar{M}_h$

Lemme 4

Pour h *assez petit,* $\bar{m}_h = m$.

Il est clair que $\bar{m}_h \leq m$.
Soit $x_1,\ldots,x_m$ une base de M. Si $\bar{m}_h < m$, il existe au moins une combinaison linéaire des x_i (et au plus m-1) telle que :

$$\|r_h(\sum_{i=1}^{m} \alpha_i x_i)\|_h = 0,$$

tandis que

$$\|\sum_{i=1}^{m} \alpha_i x_i\| > 0.$$

Or la compatibilité des normes implique que :

$$\exists h_o : h < h_o \Rightarrow \|r_h(\sum_{1}^{m} \alpha_i x_i)\|_h > 0.$$

SERVICE POLYCOPIE
MATHÉMATIQUES APPLIQUÉES
Université de GRENOBLE

Donc $\bar{m}_h = m$ pour h assez petit.

THEOREME 4

Soit λ une valeur propre de T isolée par $\Gamma(\lambda,R)$.

Si i) T_h est une approximation de T , stable dans $\Gamma(\lambda,R)$; on appelle λ^h les valeurs propres de T_h appartenant à $\Gamma(\lambda,R)$.

ii) r_h est linéaire surjectif et stable

iii) les projections spectrales P, associée à λ, et P_h , associée aux λ_h sont telles que :

$$\dim P_h X_h \leq \dim PX$$

la convergence de $\|(r_h P - P_h r_h)x\|_h$ vers 0 est uniforme en x et $\dim P_h X_h = \dim PX$, pour h assez petit.

Ce théorème a été énoncé par Kato ([25] p. 438) dans le cas de projections définies dans un même espace de Banach.

Démonstration

Soit $\{x_i\}$, $i = 1,\dots,m$ une base normée de M.

$y_i^h = r_h x_i$, $i = 1,\dots,m$ sont m vecteurs de $\bar{M}_h$, indépendants pour h assez petit puisque $\bar{m}_h = m$, d'après le lemme 4. $\|y_i^h\|_h \to 1$.

Soit, dans X_h^*, $\{e_j^h\}$, $j = 1,\dots,m$, m vecteurs tels que :

$$(y_i^h, e_j^h)_h = \delta_{ij} \quad , \quad i,j = 1,\dots,m.$$

$\|e_j^h\|_h \to 1$. Les hypothèses i) et ii) impliquent que :

$$\|(r_h P - P_h r_h)x\|_h \to 0 \quad , \quad \forall x \in X$$

Donc :

$$\|r_h x_i - P_h r_h x_i\|_h \to 0$$

$$|(y_i^h - P_h y_i^h, e_j^h)_h| \leq \|y_i^h - P_h y_i^h\|_h \|e_j^h\|_h \to 0$$

ou encore

$$(P_h y_i^h, e_j^h)_h \to \delta_{ij} .$$

Le déterminant de Gram des m vecteurs $P_h y_i^h$ tend vers 1 : ce sont des vecteurs indépendants de M_h, donc il y en a au plus m_h :

$$m \leq m_h .$$

L'hypothèse $m_h \leq m$ implique donc $\underline{m = m_h}$.

Soit $\{f_j^h\}$ la base de $M_h^* = P_h^* X_h^*$ adjointe de $\{P_h y_i^h\}$:

$$(P_h y_i^h, f_j^h)_h = \delta_{ij} \quad , \quad i,j = 1,\dots,m.$$

$$\|f_j^h\|_h \to 1 \text{ et } e_j^h = \sum_{i=1}^{m} (e_j^h, P_h y_i^h)_h f_i^h ,$$

donc

$$\|e_j^h - f_j^h\|_h \leq \max_i |(e_j^h, P_h y_i^h)_h - \delta_{ij}| \to 0 \quad , \quad j = 1,\dots,m.$$

Puisque r_h est surjectif, il est facile de voir, comme au lemme 4, que l'image par r_h de la décomposition $X = PX+(1-P)X$ est une décomposition en somme directe de X_h. Alors :

$$r_h Px = \sum_{i=1}^{m} (r_h x, e_i^h)_h y_i^h .$$

D'autre part :

$$P_h r_h x = \sum_{i=1}^{m} (r_h x, f_i^h)_h P_h y_i^h$$

$$(r_h P - P_h r_h)x = \sum_{i=1}^{m} [(r_h x, e_i^h - f_i^h)_h y_i^h + (r_h x, f_i^h)_h (y_i^h - P_h y_i^h)]$$

$$\|(r_h P - P_h r_h)x\|_h \leq \|r_h x\|_h \left(\sum_{i=1}^{m} [\|e_i^h - f_i^h\|_h \|y_i^h\|_h + \|f_i^h\|_h \|y_i^h - P_h y_i^h\|_h] \right)$$

$$\leq C_2 \|x\| \left(\qquad\qquad\qquad\qquad \right)$$

et la parenthèse tend vers 0 avec h. La convergence de $\|(r_h P - P_h r_h)x\|_h$ vers 0 est uniforme en x.

$m = m_h$ implique qu'il y a exactement m valeurs propres λ^h de T_h, distinctes ou confondues, à l'intérieur de Γ. Et le résultat est valable lorsque le rayon R de Γ décroît, pourvu que h soit assez petit. Ces valeurs propres convergent donc vers λ lorsque h tend vers 0.

3.2 - On suppose défini un opérateur de prolongement $p_h : X_h \to X$ linéaire injectif et stable.

Nous montrons l'analogue du théorème 4.

THEOREME 5

Si i) T_h *est une approximation de* T, *stable dans* $\Gamma(\lambda,R)$ *; on appelle* λ^h *les valeurs propres de* T_h *appartenant à* $\Gamma(\lambda,R)$.

ii) p_h *est linéaire injectif et stable,*

iii) les projections spectrales P, *associée à* λ, *et* P_h *associée aux* λ^h, *sont telles que :*

$$\dim P_h X_h \leq \dim PX,$$

alors

$$\|P-p_h P_h r_h\| \xrightarrow[h\to 0]{} 0$$

et $\dim P_h X_h = \dim PX$ pour h assez petit.

Démonstration

Les notations sont les mêmes qu'au paragraphe précédent.
Soit $\{e_j\}$ $i = 1,\dots,m$ la base adjointe de $\{x_i\}$ dans $M^* = P^* X^*$

$$(x_i,e_j) = \delta_{ij} \ , \quad i,j = 1,\dots,m$$

Comme précédemment, les m vecteurs $P_h y_i^h$ forment une base de M_h.
On pose $z_i^h = p_h P_h y_i^h$.
Puisque p_h est injectif, les m vecteurs z_i^h forment une base de $N_h = p_h M_h$.
Par hypothèse :

$$\|(P-p_h P_h r_h)x\| \to 0 \quad , \ \forall x \in X$$

Donc :

$$\|x_i - p_h P_h r_h x_i\| \to 0$$

$$|(x_i - z_i^h, e_j)| \leq \|x_i - z_i^h\| \ \|e_j\| \to 0$$

ou encore

$$(z_i^h, e_j) \to \delta_{ij}.$$

Soit, dans X^*, $\{g_j^h\}$, $j = 1,\dots,m$, m vecteurs tels que :

$$(z_i^h, g_j^h) = \delta_{ij} \quad , \quad i,j = 1,\dots,m$$

$$\|g_j^h\| \to 1 \quad \text{et} \quad \|e_i - g_i^h\| \to 0 \quad , \quad i = 1,\dots,m.$$

Alors

$$Px = \sum_{i=1}^{m} (x,e_i)x_i$$

$$p_h P_h r_h x = \sum_{i=1}^{m} (x,g_i^h) z_i^h$$

$$(P - p_h P_h r_h)x = \sum_{i=1}^{m} \left[(x,e_i - g_i^h)x_i + (x,g_i^h)(x_i - z_i^h)\right]$$

$$\|P - p_h P_h r_h\| \leq \sum_{i=1}^{m} \left[\|e_i - g_i^h\| \; \|x_i\| + \|g_i^h\| \; \|x_i - z_i^h\|\right]$$

$$\to 0$$

<u>Remarque</u> : D'après ce que nous venons de voir, si λ est une valeur propre de multiplicité m, isolée par $\Gamma(\lambda,R)$, il existe au moins m valeurs propres de T_h dans $\Gamma(\lambda,R)$. Si l'on sait par ailleurs, qu'il ne peut y en avoir plus de m, on déduit l'existence de m valeurs propres de T_h, distinctes ou confondues, dans un voisinage arbitraire de λ.

On pose alors la :

<u>Définition</u>

Soit λ une valeur propre de T isolée par $\Gamma(\lambda,R)$
T_h est une approximation de T, fortement stable dans le disque qui isole λ, si :

i) *T_h est stable dans ce disque*

ii) *r_h est linéaire surjectif et stable, ou p_h est linéaire injectif et stable.*

iii) $\dim P_h X_h \leq \dim PX$

où $P = \int_\Gamma R(z)dz$ et $P_h = \int_\Gamma R_h(z)dz$ et Γ *est la frontière de* $\Gamma(\lambda,R)$.

D'après la remarque précédente, si T_h est une approximation de T fortement stable dans $\Gamma(\lambda,R)$ et si λ est une valeur propre de multiplicité m, il existe, pour h assez petit, exactement m valeurs propres de T_h distinctes ou confondues, dans tout voisinage de λ: la stabilité forte de l'approximation entraîne la convergence des valeurs propres avec conservation de l'ordre de multiplicité algébrique.

Exemples d'approximations non fortement stables

Les exemples 2 et 4 donnés au § III.4 sont des approximations non fortement stables, car la propriété ii) n'est pas vérifiée.

V - CAS PARTICULIERS DANS UN ESPACE DE HILBERT

Deux classes d'opérateurs fermés jouissent de propriétés particulières :

1) Si T est normal :

$$\forall z \in \rho(T) \qquad \|R(z)\| = \min_{\lambda \in \sigma(T)} \frac{1}{|\lambda - z|}$$

2) Si T est autoadjoint, il vérifie la propriété précédente. De plus $\sigma(T)$ est réel et T admet la décomposition spectrale $T = \int_{-\infty}^{+\infty} \lambda dE(\lambda)$.

Dans ces deux cas, la théorie précédente peut être simplifiée, comme on va le voir. On considère l'espace de Hilbert H, et on choisit dans H_h une norme hilbertienne compatible avec celle de H.

1 - T_h est normal

$$q_h(z) = \min_{\lambda^h \in \sigma(T_h)} |\lambda^h - z|$$

Si λ est une valeur propre isolée de T, d'après la démonstration du lemme 2, $q_h(\lambda) \xrightarrow[h\to 0]{} 0$, c'est-à-dire dans tout voisinage de λ, il existe une valeur propre de T_h pour h assez petit.
D'où le

THEOREME 6

Soit σ_1 un ensemble fini de valeurs propres isolées de multiplicité finie de T . Si T_h est une approximation normale de T opérateur fermé dans un Hilbert H , alors

$$\sigma_1 \subset \underline{\lim}\, \sigma_1^h$$

THEOREME 7

Soit σ_1 un ensemble fini de s valeurs propres λ_i de T, isolées par les disques $\Gamma(\lambda_i, R_i)$, $i = 1,..,s$. Soit σ_1^h l'ensemble des valeurs propres de T_h appartenant à $D = \bigcup_{i=1}^{s} \Gamma(\lambda_i, R_i)$.
Si T_h est une approximation normale de T, opérateur fermé dans un Hilbert la stabilité de T_h dans D est une condition nécessaire et suffisante pour que

$$\sigma_1 = \underline{\lim}\, \sigma_1^h$$

Démonstration

Nous avons vu que la condition est suffisante au théorème 1. Montrons qu'elle est nécessaire.
L'approximation de σ_1 par σ_1^h implique que si λ est un point de σ_1 isolé par $\Gamma(\lambda,R)$, pour tout $\varepsilon > 0$, $\varepsilon < R$, il existe h_o tel que pour $h < h_o$, toutes les valeurs propres de T_h contenues dans $\Gamma(\lambda,R)$ sont telles que :

$$|\lambda^h-\lambda| < \varepsilon$$

Pour tout r, $0 < r < R$, soit z dans la couronne $\{z \; ; \; r \leq |z-\lambda| \leq R\}$.
On considère h_o déterminé par $\varepsilon < r$.
Alors :

$$q_h(z) \geq r-\varepsilon \quad \text{pour } h < h_o$$

ou encore

$$\|(T_h-z1)^{-1}\|_h \leq \frac{1}{r-\varepsilon} \quad \text{pour } h < h_o \text{ et } z \in \{z \; ; \; r \leq |z-\lambda| \leq R\}.$$

C'est la définition de la stabilité dans $\Gamma(\lambda,R)$.
Le même raisonnement s'applique pour chaque point de σ_1.

2 - T et T_h sont autoadjoints

Définition

T *est autoadjoint si* $\overline{D(T)} = H$ et $T^* = T$.

2.1 - approximation du spectre

Le théorème 7 s'applique toujours et la propriété du théorème 6 est valable pour tout point du spectre de T :

THEOREME 8

Soit T *un opérateur autoadjoint fermé dans* H, *et* T_h *une approximation autoadjointe de* T *dans* H_h , *alors :*

$$\sigma(T) \subset \underline{\lim}\,\sigma(T^h).$$

Démonstration

Soit $\lambda \in \sigma(T)$. Posons $z = \lambda+i\varepsilon$, où ε est un nombre > 0 donné.

Puisque T est autoadjoint :

$$\|R(z)\| = \|(T-z1)^{-1}\| = \frac{1}{\mathrm{dist}(z,\sigma(T))} = \frac{1}{\varepsilon} .$$

De même :

$$\|R_h(z)\|_h \leq \frac{1}{\varepsilon} .$$

La convergence simple :

$$\|(R_h(z)r_h - r_h R\ (z))x\|_h \to 0, \text{ pour } x \in H$$

se déduit de la borne uniforme de $\|R_h(z)\|_h$, comme lors de la démonstration du théorème 2. On a donc :

$$\|R(z)x\| = \lim_h \|r_h R(z)x\|_h = \lim_h \|R_h(z)r_h x\|_h \ , \ \forall x \in H$$

$$\|R(z)\| = \sup_{x\neq 0} \frac{\|R(z)x\|}{\|x\|} = \frac{1}{\varepsilon} .$$

Soit x fixé, $\|x\| = 1$ tel que $\|R(z)x\| \geq \frac{1}{2\varepsilon}$.
Alors pour h assez petit

$$\|R_h(z)r_h x\|_h \geq \frac{1}{3\varepsilon} \quad \text{et} \quad \|r_h x\|_h \leq 2,$$

donc

$$\|R_h(z)\|_h = \frac{1}{\mathrm{dist}(z,\sigma(T_h))} \geq \frac{\|R_h(z)r_h x\|_h}{\|r_h x\|_h} \geq \frac{1}{6\varepsilon} .$$

Il existe $\lambda^h \in \sigma(T_h)$ tel que $|\lambda^h - z| \leq 6\varepsilon$ et $|\lambda^h - \lambda| \leq 7\varepsilon$.

Donc, quel que soit $\lambda \in \sigma(T)$, il existe un point de $\sigma(T_h)$ aussi voisin qu'on veut de λ, pour h assez petit.

Exemple : Opérateur de multiplication

$$X = L^2(0,1),\ x(t) \overset{T}{\mapsto} \cos \pi t . x(t).$$

T est un opérateur autoadjoint et $\sigma(T) = [-1,1]$.
Soit, dans $L^2(0,1)$, la base orthonormale $e_i = \sqrt{2}\sin i\pi t$, $i = 1,2,\ldots$
Soit X_n l'espace engendré par les n premiers vecteurs de la base, muni de la norme euclidienne.
Soit $T_n : X_n \to X_n$ l'opérateur représenté par la matrice d'ordre n :

$$A_n = (a_{ij})_{i\leq i,j\leq n} \quad \text{où } a_{ij} = (\cos\pi t\, e_j, e_i) = 2\int_0^1 \cos\pi t \sin i\pi t \sin j\pi t\, dt$$

$$a_{i\,i+1} = a_{i+1\,i} = \frac{1}{2} \quad i = 1,\ldots,n-1$$

$$\text{sinon } a_{ij} = 0.$$

$\sigma(T_n) = \{\cos\pi \frac{i}{n+1},\ i = 1,\ldots n\}$. Tout point de $[-1,+1]$ est limite d'un point de $\sigma(T_n)$, lorsque $n \to \infty$.

2.2 - Approximation des sous-espaces invariants

1) Rappels sur la décomposition spectrale d'un opérateur autoadjoint ([25] p.356)
Tout opérateur autoadjoint T peut s'écrire

$$T = \int_{-\infty}^{+\infty} \lambda dE(\lambda)$$

où la famille de projections orthogonales $E(\lambda)$ vérifie :

i) $E(\lambda') \leq E(\lambda'')$ pour $\lambda' < \lambda''$

ii) $E(\lambda) \to 0$ si $\lambda \to -\infty$
$E(\lambda) \to 1$ si $\lambda \to +\infty$

En général on suppose aussi que $E(\lambda)$ est continue à droite :

$$E(\lambda+0) = E(\lambda) \quad , \quad \lambda \in \mathbb{R}.$$

λ est valeur propre de T si et seulement si $E(\lambda)-E(\lambda-0) \neq 0$, et $P(\lambda) = E(\lambda)-E(\lambda-0)$ est la projection spectrale associée à λ.
La famille spectrale $E(\lambda)$ est bien déterminée par T et l'on a la relation ([25] p. 359) :

$$\lim_{\substack{\varepsilon\to 0\\ A\to\infty}} U_{\varepsilon,A}(\lambda) = U(\lambda) = 1-E(\lambda)-E(\lambda-0)$$

avec

$$U_{\varepsilon,A}(\lambda) = \frac{1}{\pi}\left(\int_{-A}^{-\varepsilon} + \int_{\varepsilon}^{A}\right)R(\lambda+i\nu)d\nu = \frac{1}{\pi}\int_{\varepsilon}^{A}\left[R(\lambda+i\nu)-R(\lambda-i\nu)\right]d\nu$$

$$= \frac{2}{\pi}\int_{\varepsilon}^{A} (T-\lambda 1)((T-\lambda 1)^2+\nu^2 1)^{-1}d\nu$$

Si T est un opérateur autoadjoint dans un espace de dimension finie, de valeurs propres $\lambda_1,\ldots,\lambda_s$ et de projections spectrales $P_1,\ldots,P_s$, on a :

$$E(\lambda) = \sum_{\lambda_k\leq\lambda} P_k$$

$$E(\lambda_k)-E(\lambda_k-0) = P_k$$

2) *THEOREME 9*

Soit T *un opérateur autoadjoint dans* H *et* T_h *une approximation autoadjointe de* T *dans* H_h, *de familles spectrales* $\{E(\lambda)\}$ *et* $\{E_h(\lambda)\}$ *respectivement. Si l'opérateur de restriction* r_h *est stable :*

$$\|(r_h E(\mu)-E_h(\mu)r_h)x\|_h \xrightarrow[h\to 0]{} 0 \text{ pour } x \in X \text{ et } \mu \text{ non valeur propre de T.}$$

Ce théorème est énoncé par Kato ([25] p. 432) dans le cas d'opérateurs dans un même espace de Hilbert.

Démonstration

Nous faisons la démonstration en supposant $\mu = 0$ non valeur propre de T. On peut écrire :

$$(1-E_h(0)-E_h(-0))\mathcal{T}_h(\mathcal{T}_h^2+1)^{-1} = \frac{1}{\pi}\int_0^\infty [R_h(i\nu)-R_h(-i\nu)]\,[R_h(i)-R_h(-i)]\,d\nu$$

car l'intégrale du membre de droite est absolument convergente :

$$\|(R_h(i\nu)-R_h(-i\nu))(R_h(i)-R_h(-i))\|_h \leq 2\min(1, \frac{1}{\nu}).$$

D'autre part, $E_h(0)-E_h(-0)$ est la projection sur $\mathrm{Ker}\mathcal{T}_h$, et cet opérateur commute avec $\mathcal{T}_h$. Donc $(E_h(0)-E_h(-0))\mathcal{T}_h(\mathcal{T}_h^2+1)^{-1} = 0$. Et l'on peut remplacer, dans le membre de gauche, $E_h(-0)$ par $E_h(0)$.
On a aussi

$$(1-2E(0))T(T^2+1)^{-1} = \frac{1}{\pi}\int_0^\infty (R(i\nu)-R\ (-i\nu))(R\ (i)-R\ (-i))d\nu.$$

En tout point z tel que $\mathfrak{Im}\ z \neq 0$, on a la borne uniforme :

$$\|R_h(z)\|_h \leq \frac{1}{|\mathfrak{Im} z|}$$

D'où l'on déduit (cf. démontration du théorème 2) que :

$$\|(r_h R(z)-R_h(z)r_h)x\|_h \to 0 \text{ pour } x \in X \text{ et } z \text{ tel que } \mathfrak{Im}\ z \neq 0.$$

Il vient alors :

$$r_h(R(i)-R(-i))x-(R_h(i)-R_h(-i))r_h x \to 0$$

$$[r_h(R(i\nu)-R(-i\nu))-(R_h(i\nu)-R_h(-i\nu))r_h](R(i)-R(-i))x \to 0$$

et

$$(1-2E_h(0))(R_h(i)-R_h(-i))r_h x-r_h(1-2E(0))(R(i)-R(-i))x \to 0$$

ou encore

$$\left[(1-2E_h(0))r_h-r_h(1-2E(0))\right](R(i)-R(-i))x \to 0$$

$$(r_hE(0)-E_h(0)r_h)(R(i)-R(-i))x \to 0.$$

$\mathrm{Im}(R(i)-R(-i)) = \mathrm{Im}\ T(T^2+1)^{-1} = D((T^2+1)T^{-1})$ est dense dans H puisque T est autoadjoint et n'admet pas la valeur propre 0.
Puisque r_h est stable, la famille $r_hE(0)-E_h(0)r_h$ est uniformément bornée en h.
On en déduit la convergence simple cherchée.

3) Stabilité forte

THEOREME 10

Soit T *un opérateur autoadjoint fermé et* T_h *une approximation autoadjointe de* T. λ *est une valeur propre de* T, *de multiplicité* m, *isolée dans* $[\mu_1,\mu_2]$, *on appelle* λ^h *les valeurs propres de* T_h *dans* $[\mu_1,\mu_2]$
Si i) r_h *est linéaire surjectif stable ou* p_h *linéaire injectif stable,*

ii) les projections spectrales P, *associé à* λ, *et* P_h *associée aux* λ^h *sont telles que*

$$\dim P_hH_h \leq \dim PH,$$

il existe dans tout voisinage de λ *inclus dans* $[\mu_1,\mu_2]$, m *valeurs propres de* T_h, *distinctes ou confondues, pour* h *assez petit.*
D'autre part ,

$$\|(r_hP-P_hr_h)x\|_h \xrightarrow[h\to 0]{} 0 \text{ uniformément en } x$$

ou

$$\|P-p_hP_hr_h\| \xrightarrow[h\to 0]{} 0$$

Démonstration :

Soit λ une valeur propre isolée de multiplicité m. L'intervalle $[\mu_1,\mu_2]$ contient λ et aucun autre point de $\sigma(T)$.

$P = E(\mu_2)-E(\mu_1)$ est la projection spectrale associé à λ

$P_h = E_h(\mu_2)-E_h(\mu_1)$ est la projection spectrale associée aux valeurs propres de T_h appartenant à $[\mu_1,\mu_2]$.

r_h est stable, donc si μ n'est pas valeur propre de T,
$\|(r_hE(\mu)-E_h(\mu)r_h)x\|_h \to 0$, ce qui implique que :

$$\|(r_hP-P_hr_h)x\|_h \to 0.$$

Supposons que $\dim P_hH_h \leq \dim PH$.

Si r_h est linéaire surjectif stable, ou si p_h est linéaire injectif stable, on peut appliquer le théorème 4 ou 5 :

$$\|(r_hP-P_hr_h)x\|_h \to 0 \text{ uniformément en } x$$

ou

$$\|P-p_hP_hr_h\| \to 0$$

et

$$\dim P_hH_h = \dim PH.$$

Il y a alors, dans tout voisinage de λ inclus dans $[\mu_1,\mu_2]$, m valeurs propres λ^h de T_h, distinctes ou confondues, pour h assez petit.

Dans le cas d'une approximation autoadjointe d'un opérateur autoadjoint, la condition

$$\underline{\dim P_hH_h \leq \dim PH}$$

est une condition suffisante pour que les valeurs propres et les sous-espaces invariants approchés convergent (au sens défini plus haut) respectivement vers λ et le sous-espace invariant associé.

Les opérateurs autoadjoints ont des propriétés particulières relativement à l'hypothèse de stabilité. En effet, on a vu au théorème 7 que la stabilité de T_h est une CNS pour que les valeurs propres approchées convergent ; du théorème 9 on peut déduire la convergence des sous-espaces invariants pour tout opérateur autoadjoint.

Néanmoins la stabilité forte de T_h reste une condition suffisante de convergence des valeurs propres avec conservation de l'ordre de multiplicité, lorsqu'on a posé la définition suivante :

Définition

Soit λ une valeur propre de l'opérateur autoadjoint T, *isolée par* $\Gamma(\lambda,R)$.
Soit T_h une approximation autoadjointe de T
T_h *est fortement stable dans le disque qui isole λ si les deux conditions :*

ii) r_h *est linéaire surjectif et stable ou* p_h *est linéaire injectif et stable*

iii) $\dim P_h H_h \leq \dim PH$

sont vérifiées, où $P = \int_\Gamma R(z)dz$ et $P_h = \int_\Gamma R_h(z)dz$, Γ *étant la frontière de* $\Gamma(\lambda,R)$.

Exemple : approximation fortement stable

Soit $H = L^2(0,1)$, $D = \{x \in H \; ; \; x'' \in H \text{ et } x(0) = x(1) = 0\}$

$T : x \mapsto -x''$, $x \in D$, est un opérateur autoadjoint.

a) approximation

$[0,1]$ est divisé en n parties égales aux points $t_i = \frac{i}{n} = hi$, $i = 0,\dots,n$
$x(t_i) = x_i$

$H_h = \mathbb{R}^{n-1}$ est muni de la norme euclidienne, compatible avec la norme de H.
$r_h x = (x_i)$, $i = 1,\dots,n-1$
r_h est linéaire et stable : $\|r_h x\|_h \leq \|x\|$.

L'opérateur $\mathcal{T}_h$, représenté par la matrice hermitienne d'ordre n-1 :

$$A_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ 0 & & & -1 & 2 \end{pmatrix}$$

est une approximation autoadjointe de T.

b) éléments propres.
T admet les valeurs propres simples $\lambda_k = k^2\pi^2$, $k = 1,2,\dots$ et les vecteurs propres $\varphi_k(t) = \sin k\pi t$.

$\mathcal{T}_h$ admet les valeurs propres simples $\lambda_k^h = \frac{2}{h^2}(1-\cos k\pi h)$, $k = 1,\dots,n-1$ et les vecteurs propres $\varphi_k^h = (\sin ik\pi h)$, $i = 1,\dots,n-1$.
La stabilité forte est vérifiée puisque les valeurs propres sont simples et l'on peut constater la convergence des valeurs propres et des vecteurs propres.

<u>Application pratique</u> :

Soit à calculer les valeurs propres d'un opérateur autoadjoint T, dont on sait qu'il en existe s dans l'intervalle I.
Il y a au moins s valeurs propres λ^h d'une approximation autoadjointe $\mathcal{T}_h$, appartenant à I.

a) si, par le calcul, on obtient s valeurs propres séparées dans I pour $\mathcal{T}_h$, c'est que les valeurs propres de T dans I sont simples et approchées par celles de $\mathcal{T}_h$, approximation fortement stable.

b) s'il y en a $k > s$, deux cas sont à envisager : ou bien les valeurs propres calculées sont séparées et le schéma n'est pas stable pour k-s d'entre elles, ou bien il y a k-s valeurs propres voisines et l'on pourra conclure à la stabilité forte si l'on sait que T admet une valeur propre d'ordre k-s près des k-s valeurs propres voisines de T_h. On concluera aussi que les autres valeurs propres de T dans I sont simples et approchées par celles de T_h.

VI - CONDITIONS SUFFISANTES DE STABILITE

Deux types d'approximations vérifient une condition suffisante de stabilité : les approximations uniformes d'opérateurs bornés et les approximations supérieures d'opérateurs autoadjoints.

1 - Approximation uniforme d'un opérateur borné

1.1 - On suppose qu'il existe un opérateur de prolongement stable tel que :

$$r_h p_h \xi = \xi \text{ , pour } \xi \in X_h .$$

Exemple : approximation de Galerkin dans un espace de Hilbert séparable

Soit $H = L^2(0,1)$

$[0,1]$ est divisé en $n = \frac{1}{h}$ parties égales aux points $t_i = \frac{i}{n}$, $i = 0,1,\ldots,n$.

Soit y_h^i la fonction caractéristique de l'intervalle $]t_{i-1},t_i]$.

H_h est le sous-espace vectoriel de H des fonctions étagées :

$$H_h = \{x, x(t) = \sum_{i=1}^{n} x(t_i) y_h^i(t)\}.$$

p_h est l'injection canonique, et r_h est défini par :

$$(r_h x)(t_i) = \frac{1}{h} \int_{t_{i-1}}^{t_i} x(t)dt \quad , \quad i = 1,\ldots,n$$

Il est clair que $r_h p_h x = x$ pour $x \in H_h$.

On considère $T_h = p_h \mathcal{T}_h r_h$, opérateur de X dans X.
T_h est une approximation de T :

$$T-T_h = (1-p_h r_h)T+p_h(r_h T-\mathcal{T}_h r_h)$$

$$\|(T-T_h)x\| \leq \|(1-p_h r_h)Tx\|+C_1\|(r_h T-\mathcal{T}_h r_h)x\|_h$$

$$\to 0 \text{ pour } x \in D(T).$$

Numériquement, les éléments propres que l'on calcule sont ceux de $\mathcal{T}_h$, mais nous allons obtenir des conditions suffisantes de stabilité portant sur T_h. Nous étudions tout d'abord le lien entre les éléments propres de $\mathcal{T}_h$ et ceux de T_h.

On suppose de plus p_h et r_h linéaires, p_h injectif et r_h surjectif. Il est facile de voir que $\mathcal{T}_h$ et la restriction de T_h à X/Ker r_h ont même structure. D'où la :

Proposition 1

Si r_h *et* p_h *sont des opérateurs linéaires tels que* p_h *est injectif,* r_h *surjectif et* $r_h p_h = 1$ *,* $\mathcal{T}_h$ *et la restriction de* T_h *à* X/Ker r_h *ont les mêmes valeurs propres, avec le même ordre de multiplicité.*

Si la norme de X_h est choisie telle que

$$\|\xi\|_h = \|p_h \xi\| \text{ pour } \xi \in X_h,$$

$$\|(T_h-z1)p_h\xi\| = \|(p_h \mathcal{T}_h r_h - z1)p_h\xi\| = \|(\mathcal{T}_h-z1)\xi\|_h$$

Si $\mathcal{T}_h$ est stable, il en est de même de T_h et réciproquement.
Nous supposons, dans la suite du paragraphe, les hypothèses ci-dessus réalisées.

1.2 - Approximation uniforme d'un opérateur compact

Les seuls opérateurs approchés en norme par des opérateurs de rang fini sont les opérateurs compacts.

Proposition 2

Si T_h *est une approximation de l'opérateur compact* T *uniforme en* x *et si* p_h *injectif,* r_h *surjectif vérifient :*

$\|p_h\xi\| = \|\xi\|_h$ pour $\xi \in X_h$ et $r_h p_h = 1$

T_h *est stable au voisinage de toute valeur propre de* T.

Démonstration : par hypothèse :

$\forall\varepsilon > 0,\ \exists h_o : h < h_o \Rightarrow \|(r_h T - T_h r_h)x\|_h \leq \varepsilon\|x\|$, pour $x \in D(T)$.

Alors :

$$\|(T-T_h)x\| \leq \|(I-p_h r_h)Tx\| + \|p_h(r_h T - T_h r_h)\|$$

$1-p_h r_h$ tend vers 0 uniformément sur le précompact $T(S_X)$, image par T de la boule unité S_X de X.

D'autre part :

$$\|p_h(r_h T - T_h r_h)x\| \leq C_1\ \varepsilon\|x\|$$

Pour tout z de $\rho(T)$, il existe $\delta > 0$ tel que

$$\|(T-zI)x\| \geq \delta\|x\| \quad , \quad \text{pour } x \in X.$$

Soit ξ dans X_h, $p_h\xi = x_h$ appartient à X

$$(T-zI)x_h = (T-T_h)x_h + (T_h - zI)x_h$$

$\forall \varepsilon > 0,\ \varepsilon < \delta\ ,\ \exists h_o$:

$$h < h_o \Rightarrow \|(T_h - z1)\xi\|_h = \|(T_h - z1)p_h\xi\| \geq \|(T-z1)x_h\| - \|(T-T_h)x_h\|$$

$$\geq (\delta-\varepsilon)\|\xi\|_h$$

Soit

$$\|(T_h - z1)^{-1}\|_h \leq \frac{1}{\delta-\varepsilon}$$

ce qui démontre la stabilité de T_h.

Ceci est un résultat bien connu (Kantorovitch [23] p. 553, Medvedev [37]) que nous avons présenté à l'aide de la notion de stabilité.
Les conditions données par Kantorovitch sont un peut différentes :

i) $\|(T_h r_h - r_h T p_h r_h)x\|_h \leq \varepsilon \|x\|$

ii) $\|T(1-p_h r_h)x\| \leq \varepsilon_1 \|x\|$

Remarquons que ii) est vérifié si $(p_h r_h)^* \to 1$ dans X^*, espace adjoint de X.

Nous verrons, en exemple de ceci, la méthode de Galerkin au § VIII.3.1. Un exemple numérique est traité au chapitre 5, § II.2.3.

Proposition 3

Soit $T = U+V$ *un opérateur borné où* U *est compact.*
L'approximation $T_h = U_h + V_h$ *est telle que :*

i) $\|U-U_h\| \xrightarrow[h\to 0]{} 0$

ii) V_h *est stable au voisinage de la valeur propre* $\lambda \neq 0$ *de* T

Alors T_h *est stable au voisinage de cette valeur propre.*

Démonstration :

Supposons qu'il existe une suite $\{x_h\}$ telle que :

$$\|x_h\| = 1, \quad \|(T_h - z1)x_h\| \xrightarrow[h\to 0]{} 0 \quad \forall z \in \rho(T)$$

alors que :

$$\|(T-z1)x_h\| > \delta > 0.$$

Puisque U est compact, à un changement d'indice près ,

$$\|Ux_h\| \to 0.$$

D'autre part $\|U-U_h\| \to 0$, donc $\|U_h x_h\| \to 0$.
Alors, pour tout z de $\rho(T)$:

$$\|(V_h - z1)x_h\| \leq \underbrace{\|(T_h - z1)x_h\| + \|U_h x_h\|}_{\to 0}$$

ce qui contredit la stabilité de V_h au voisinage de λ.

1.3 - Approximation compacte d'un opérateur borné

Au paragraphe précédent, T est compact et T_h est de rang fini donc $T-T_h$ est compact : l'image $\{(T-T_h)x_h\}$ de toute suite bornée $\{x_h\}$ contient une sous-suite convergente.
On peut obtenir une condition suffisante plus large que l'approximation uniforme en ne retenant que cette hypothèse.
C'est ce qu'ont fait, indépendamment, Anselone [1,2,3] et Vainikko [67] :

1) Ensembles collectivement compacts d'opérateurs [2]

T et T_h sont des opérateurs bornés de X.
L'approximation T_h de T est définie ainsi :

i) $T_h x \to Tx$, pour $x \in X$

ii) l'ensemble $\{T-T_h\}$ est <u>collectivement compact</u>, c'est-à-dire : l'image $\{(T-T_h)x_h\}$ de toute suite bornée $\{x_h\}$ contient une sous-suite convergente.

La démonstration de la stabilité de T_h est dans [3].
[1] contient l'exemple suivant :

$X = C(0,1)$ avec la norme uniforme.
On considère l'équation intégrale de Fredholm de 2ème espèce :

$$f(x)-\int_0^1 K(x,y)f(y)dy = g(x)$$

où K est une fonction continue sur $[0,1] \times [0,1]$, f, g sont continues sur $[0,1]$.
On considère une formule de quadrature telle que :

$$\begin{cases} \sum_{k=1}^{n} w_{nk} f(t_{nk}) \xrightarrow[n\to\infty]{} \int_0^1 f(t)dt, \ f \in X \\ \sum_{k=1}^{n} |w_{nk}| \leq B, \ n \geq 1, \text{ et } B < \infty . \end{cases}$$

$S : f(x) \mapsto \int_0^1 K(x,y)f(y)dy$, est un opérateur compact ,

$S_n : f(x) \mapsto \sum_{k=1}^{n} w_{nk} K(x,t_{nk})f(t_{nk})$.

S_n est une approximation de S et l'ensemble $\bigcup_1^\infty \{g = S_n f, \|f\| \leq 1\}$ est relativement compact.

Alors $T_n = I-S_n$ est une approximation de $T = I-S$ vérifiant les hypothèses i) et ii).

2) Approximation compacte [67]

$\{X_n\}$ est une suite de sous-espaces fermés de X.
On définit $X_\infty = \{x \in X \ ; \ \mathrm{dist}(x,X_n) \xrightarrow[n\to\infty]{} 0\}$, sous-espace fermé de X.
La suite d'opérateurs bornés $T_n : X_n \to X_n$ est une approximation compacte de l'opérateur borné $T : X \to X$, si :

i) $TX \subset X_\infty$

ii) pour toute sous-suite n_k, $x_{n_k} \in X_{n_k} \to x$ implique $Tx_{n_k} \to Tx$

iii) l'image $\{(T-T_n)x_n\}$ de toute suite bornée $\{x_n\}$ de X_n contient une sous-suite convergente dans X.

La démonstration de la stabilité est dans [67] p. 746

2 - Approximation supérieure d'un opérateur autoadjoint

Nous avons vu au théorème 10 que dans le cas d'opérateurs autoadjoints dans un espace de Hilbert H lorsque r_h est linéaire surjectif stable ou p_h linéaire injectif stable, que la stabilité forte de $\mathcal{T}_h$, approximation autoadjointe de T dans H_h, au voisinage de λ, entraîne la convergence des valeurs propres de $\mathcal{T}_h$ voisines de λ vers λ.

Nous donnons ci-dessous des conditions suffisantes de stabilité forte.

2.1 - Principe du minimax

Soit T un opérateur autoadjoint dans $\mathbb{C}^n$ de valeurs propres :

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

comptées avec leur ordre de multiplicité.

Le principe du minimax ([25] p. 60) affirme que :

$$\lambda_i = \max_{S,\operatorname{codim}S=i-1} \min_{\substack{x\in S\\ x\neq 0}} \frac{(Tx,x)}{\|x\|^2}$$

La démonstration étant basée sur la décomposition de T dans la base de ses vecteurs propres, il est facile de voir que ce principe s'étend à un opérateur T autoadjoint dans un espace de Hilbert H qui vérifie la propriété :

P : T est borné inférieurement et T est à résolvante compacte.

<u>Remarque</u> - Les valeurs propres non négatives d'un opérateur autoadjoint compact, ordonnées en décroissant, vérifient :

$$\lambda_i^+ = \max_{S,\operatorname{codim}S=i-1} \max_{x\in S} \frac{(Tx,x)}{\|x\|^2} \quad , \; i = 1,2,\ldots$$

Les valeurs propres non positives, ordonnées en croissant vérifient :

$$\lambda_i^- = \max_{S,\operatorname{codim}S=i-1} \min_{x\in S} \frac{(Tx,x)}{\|x\|^2} \quad , \; i = 1,2,\ldots$$

2.2 - <u>*Définition*</u>

T_h est une <u>approximation supérieure</u> de l'opérateur autoadjoint T *si :*

i) T_h est une approximation autoadjointe de T

ii) l'une des deux conditions ci-dessous est remplie :

$$S_1 \begin{cases} \text{a) } r_h : X \to X_h \text{ est linéaire surjectif et vérifie } \|r_h x\|_h \leq \|x\| \text{ pour } x\in X \\ \text{b) } (T_h r_h x, r_h x)_h \geq (Tx,x) \text{ pour } x \in D(T) \end{cases}$$

$$S_2 \begin{cases} \text{a) } p_h : X_h \to X \text{ est linéaire, injectif et vérifie} \\ \quad \|\xi\|_h \leq \|p_h\xi\| \leq C_1 \|\xi\|_h \text{ pour } \xi \in X_h, \\ \text{b) } (T_h\xi,\xi)_h \geq (Tp_h\xi, p_h\xi) \text{ pour } \xi \in X_h \text{ tel que } p_h\xi \in D(T). \end{cases}$$

Les valeurs propres λ_i (resp. λ_i^h) de T (resp. T_h) sont ordonnées en croissant.

Lemme 5

Soit T *un opérateur autoadjoint fermé vérifiant* P, *et* T_h *une approximation autoadjointe de* T *vérifiant* S_1 , *il existe* h_o *tel que :*

$$h < h_o \Rightarrow \lambda_i^h \geq \lambda_i \quad , \quad i = 1,2,\ldots$$

Soit S le sous-espace de codimension i-1, pour lequel le maximum figurant dans l'expression de λ_i est atteint

$$\lambda_i = \min_{\substack{0\neq x \in S \\ \text{codim } S=i-1}} \frac{(Tx,x)}{\|x\|^2} \leq \min_{\substack{0\neq x \in S \\ \text{codim} S=i-1}} \frac{(T_h r_h x, r_h x)_h}{\|r_h x\|_h^2}$$

d'après S_1.

On note :

$$S_h = r_h S,$$

$$S_1 = S\perp, \quad S_{1h} = r_h S_1 .$$

D'après le lemme 4, puisque S_1 est de dimension finie, dim S_{1h} = dim S_1 pour h assez petit, ou encore codim S_h = codim S.

Il vient :

$$\min_{\substack{0\neq x \in S \\ \text{codim} S=i-1}} \frac{(T_h r_h x, r_h x)_h}{\|r_h x\|_h^2} = \min_{\substack{0 \neq \xi \in S_h \\ \text{codim } S_h=i-1}} \frac{(T_h \xi,\xi)_h}{\|\xi\|_h^2} \leq \lambda_i^h$$

car S_h n'est pas nécessairement un sous-espace de X_h de codimension i-1 pour lequel le maximum est atteint.

Lemme 6

Soit T *un opérateur autoadjoint fermé vérifiant* P_1 *et* T_h *une approximation autoadjointe de* T *vérifiant* S_2, *il existe* h_o *tel que :*

$$h < h_o \Rightarrow \lambda_i^h \geq \lambda_i, \quad i = 1,2,\ldots$$

$$\lambda_i = \min_{\substack{0 \neq x \in S \\ \text{codim } S = i-1}} \frac{(Tx,x)}{\|x\|^2} \leq \min_{\substack{0 \neq x \in S \cap p_h H_h \\ \text{codim } S = i-1}} \frac{(Tp_h\xi, p_h\xi)}{\|p_h\xi\|^2}$$

On note :

$$S_1 = S^\perp \ , \ S_{1h} = r_h S_1$$

Comme précédemment, dim S_{1h} = dim S_1 pour h assez petit.
Et puisque p_h est injectif, dim S_{1h} = dim $p_h S_{1h}$ = dim S_1.
Si on pose $S_h = (p_h S_{1h})^\perp$, codim S_h = codim S = i-1.
Alors :

$$\min_{\substack{0 \neq x \in S \cap p_h H_h \\ \text{codim } S = i-1}} \frac{(Tp_h\xi, p_h\xi)}{\|p_h\xi\|^2} \leq \min_{\substack{0 \neq x \in S_h \\ \text{codim } S_h = i-1}} \frac{(Tp_h\xi, p_h\xi)}{\|p_h\xi\|^2}$$

$$\leq \min_{\substack{0 \neq \xi \in S_{1h}^\perp \\ \dim S_{1h} = i-1}} \frac{(T_h\xi,\xi)_h}{\|\xi\|_h^2} \leq \lambda_i^h$$

car $S_{1h}^\perp$ n'est pas nécessairement un sous-espace de X_h de codimension i-1 pour lequel le maximum est atteint.

SERVICE POLYCOPIE
MATHÉMATIQUES APPLIQUÉES
Université de GRENOBLE

Proposition 4

Soit T *un opérateur autoadjoint fermé borné inférieurement à résolvante compacte, et* T_h *une approximation autoadjointe de* T *vérifiant* S_1 *ou* S_2 *, alors* T *est fortement stable au voisinage de toute valeur propre de* T.

Démonstration

Soit λ une valeur propre de T, isolée par Γ qui coupe l'axe réel en μ_1 et μ_2

Si $P_h = \int_\Gamma R_h(z)dz$ et $P = \int_\Gamma R(z)dz$, dim P_hH_h est le nombre de valeurs propres de T_h appartenant à $[\mu_1,\mu_2]$. dim PH est le nombre analogue pour T. Puisque $\lambda_i^h \geq \lambda_i$, $i = 1,2,\ldots$, il est facile de voir que :

$$\dim P_hH_h \leq \dim PH.$$

T_h est une approximation de T, fortement stable.

Remarque

On peut démontrer de la même façon la stabilité forte d'une approximation supérieure d'un opérateur compact, au voisinage des valeurs propres négatives.

3 - Exemples d'approximations fortement stables d'opérateurs autoadjoints

Planchard a étudié ([48],[49],[50]) pour certaines classes d'opérateurs autoadjoints des méthodes d'approximation des éléments propres pour lesquelles il a montré la convergence. Nous reprenons ces exemples en montrant que les approximations sont fortement stables.
La convergence des éléments propres découlera donc immédiatement.

3.1 - Opérateurs différentiels elliptiques du 2ème ordre

Pour simplifier l'exposé, nous considérons des problèmes à une variable (cela ne change rien quant à la démonstration de la stabilité).

On considère $H = L^2(0,1)$, avec le produit scalaire $(x,y)_1 = \int_0^1 x(t)\bar{y}(t)dt$.
Soit $H^1(0,1)$ le sous-espace de $L^2(0,1)$ des fonctions x telles que la dérivée $\frac{dx}{dt}$ au sens des distributions, appartienne à $L^2(0,1)$.
Avec le produit scalaire :

$$(x,y)_2 = (x,y)_1 + \int_0^1 \frac{dx}{dt}\frac{d\bar{y}}{dt}\,dt$$

c'est un espace de Hilbert, dense dans $L^2(0,1)$.

$V = H_0^1(0,1)$ est l'ensemble des fonctions de $H^1(0,1)$ qui s'annullent en 0 et 1.
Remarquons que $\|x\|_1 \leq \|x\|_2$.

On considère dans H le problème de valeurs propres avec conditions aux limites de Dirichlet suivant :

$$\begin{cases} Tx = \lambda x \\ x(0) = x(1) = 0 \end{cases}$$

où T est l'opérateur différentiel : $-\frac{d}{dt}(a\frac{d}{dt})+a_0$. a et a_0 sont des fonctions mesurables bornées sur $[0,1]$ prenant des valeur $>\alpha$, constante > 0.

Faisant le produit scalaire des 2 membres par une fonction y quelconque de V, le problème est équivalent, par intégration par parties à :

$$\int_0^1 a(t)\frac{dx}{dt}\frac{d\bar{y}}{dt}\,dt + \int_0^1 a_0(t)x\bar{y}\,dt = \lambda\int_0^1 x\bar{y}\,dt\ , \quad \text{pour } y \in V.$$

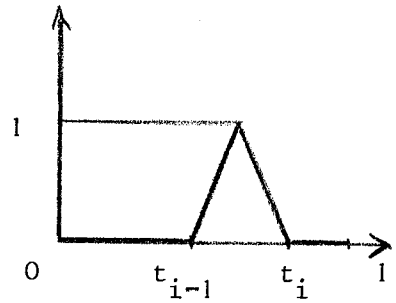
Appelons $t(x,y)$ la forme sesquilinéaire continue sur $V \times V$ qui est au 1er membre.
L'opérateur différentiel T, à résolvante compacte, admet des valeurs propres ordonnées en croissant définies par :

$$\lambda_i = \max_{S,\operatorname{codim} S\leq i-1}\ \min_{x\in S\subset H} \frac{t(x,x)}{\|x\|_1^2}\ , \quad i = 1,2,\ldots$$

1) <u>Approximation par les splines fonctions</u> [48]

On divise $[0,1]$ en $n = \frac{1}{h}$ parties égales par les points $t_i = \frac{i}{h}$, $i = 0,\dots,n$
H_h est l'espace $\mathbb{R}^{n-1}$.

On considère les n fonctions continues :



$$t \mapsto \theta_h^i(t) = \begin{cases} 1 \text{ si } t = \dfrac{t_{i-1}+t_i}{2} \\ 0 \text{ si } \left|t - \dfrac{t_{i-1}+t_i}{2}\right| \geq \dfrac{h}{2} \\ \text{linéaire par morceaux ailleurs} \end{cases}$$

L'opérateur de prolongement p_h est défini par :

$$\xi = (\xi_i) \mapsto p_h\xi = \sum_{i=1}^{n-1} \xi_i \theta_h^i.$$

On démontre que p_h est injectif. La norme sur H_h est définie par :

$$\|\xi\|_{1h} = \|p_h\xi\| \quad , \quad \text{pour} \quad \xi \in \mathbb{R}^{n-1}.$$

L'opérateur de restriction r_h est défini par :

$$x \mapsto r_h x = ((r_h x)_i)$$

avec

$$(r_h x)_i = \frac{1}{h}\int_0^1 \theta_h^i(t)x(t)dt.$$

On montre que $p_h r_h x \xrightarrow[h\to 0]{} x$ dans V, donc dans H.

La forme sesquilinéaire hermitienne continue sur $H_h \times H_h$:

$$t_h(\xi,\eta) = t(p_h\xi, p_h\eta)$$

définit $\mathcal{T}_h$ par :

$$(\mathcal{T}_h\xi,\eta)_{1h} = t_h(\xi,\eta) \qquad \text{pour } \xi,\eta \in H_h.$$

Les valeurs propres de $\mathcal{T}_h$ ordonnées en croissant vérifient :

$$\lambda_i^h = \max_{S,\operatorname{codim} S\leq i-1} \ \min_{x\in S} \frac{t_h(\xi,\xi)}{\|\xi\|_{1h}^2} \quad , \qquad i = 1,2,\ldots$$

La démonstration du lemme 6 s'applique, où l'on remplace les formes (Tx,x) et $(\mathcal{T}_h\xi,\xi)$ respectivement par $t(x,x)$ et $t_h(\xi,\xi)$. Alors $\lambda_i^h \geq \lambda_i$, $i = 1,2,\ldots$ pour h assez petit. D'où la stabilité forte de l'approximation $\mathcal{T}_h$.

2) <u>Approximation par la méthode des différences finies</u> [49]

Soit toujours $[0,1]$ divisé en $n = \frac{1}{h}$ parties égales, et $H_h = \mathbb{R}^{n-1}$.
Soit χ_h^i la fonction caractéristique de $[t_{i-1},t_i]$, $i = 1,\ldots,n$.
On introduit

$$r_h \text{ défini par } (r_h x)_i = \frac{1}{h}\int_0^1 \chi_h^i(t)\, x(t)\, dt$$

$$p_h \text{ défini par } p_h\xi = \sum_{i=1}^{n-1} \xi_i \chi_h^i$$

$$\|\xi\|_{1h} = \|p_h\xi\|.$$

On montre que p_h est injectif et $p_h r_h x \to x$ dans H.
La dérivée discrète est définie par :

$$\delta p_h\xi(t) = \frac{1}{h}\left[p_h\xi(t+\frac{h}{2}) - p_h\xi(t-\frac{h}{2})\right]$$

La forme sesquilinéaire hermitienne continue sur $H_h \times H_h$

$$t_h(\xi,\eta) = \int_0^1 a(t)\delta p_h\xi\overline{\delta p_h\eta}\, dt + \int_0^1 a_0(t)\, p_h\xi\, \overline{p_h\eta}\, dx$$

définit T_h par $(T_h\xi,\eta)_{1h} = t_h(\xi,\eta)$ pour $\xi,\eta \in H_h$.

Montrons que l'on a :

$$(T_h\xi,\xi)_{1h} \geq t(p_h\xi,p_h\xi) \quad \text{pour } \xi \in H_h$$

$$t(p_h\xi,p_h\xi) = \int_o^1 a(t)\left|\frac{d}{dt}\, p_h\xi\right|^2 dt + \int_o^1 a_o(t)\left|p_h\xi\right|^2 dt$$

Il suffit donc de montrer que :

$$\int_o^1 a(t)\left|\delta p_h\xi\right|^2\, dt \geq \int_o^1 a(t)\left|\frac{d}{dt}\, p_h\xi\right|^2\, dt$$

Or

$$\delta p_h\xi = \frac{1}{h}\sum_{i=0}^{n-1} (\xi_{i+1}-\xi_i)\tilde{\chi}_h^i$$

où $\tilde{\chi}_h^i$ est la fonction caractéristique de $\begin{cases} \left[t_i-\frac{h}{2},\ t_i+\frac{h}{2}\right] & , i = 1,\dots,n-1 \\ \left[0,\ \frac{h}{2}\right] & \text{pour } i = 0 \end{cases}$

avec $\xi_o = \xi_n = 0$.

Et

$$\frac{d}{dt}\, p_h\xi = \sum_{i=0}^{n-1} (\xi_{i+1}-\xi_i)\delta(t_i)$$

où $\delta(t_i)$ est la distribution de Dirac au point t_i.

Sur chaque intervalle $\left[t_i-\frac{h}{2},\ t_i+\frac{h}{2}\right]$, $i = 1,\dots,n-1$, l'inégalité cherchée se réduit à :

$$|\xi_{i+1}-\xi_i|^2\int_{t_i-\frac{h}{2}}^{t_i+\frac{h}{2}} \frac{a(t)}{h^2}\, dt \geq |\xi_{i+1}-\xi_i|^2\, a(t_i).$$

Il est facile de voir que ceci est vérifié pour $h < h_o$ assez petit.

On en déduit comme précédemment que le lemme 6 s'applique et que $\lambda_i^h \geq \lambda_i$, $i = 1,2,\dots$ pour h assez petit.

3.2 - Opérateurs compacts positifs [50]

Nous reprenons le problème de la diffusion des neutrons en régime stationnaire, étudié par Planchard. Nous nous plaçons dans le cas où les fonctions s_1 et s_2 sont égales à la constante $\sigma > 0$.
Nous avons $H = L^2(0,1)$ et $V = H_0^1(0,1)$.
Le problème revient à trouver un scalaire $\lambda > 0$ et une fonction φ de V positive, tels que : $Q\varphi = \lambda\varphi$, où $Q = \sigma^2 G_1 G_2$ et G_1 et G_2 sont les opérateurs inverses des opérateurs différentiels T_1 et T_2 définis sur V par les formes sesquilinéaires continues coercives de $V \times V$:

$$t_k(x,y) = \int_0^1 d_k(t) \frac{d}{dt} x \frac{d}{dt} y \, dt + \int_0^1 s_{A,k}(t) \, xy \, dt, \quad k = 1,2.$$

Q est un opérateur compact positif, dont on démontre qu'il admet une valeur propre simple positive λ, plus grande en module que toutes les autres valeurs propres de Q. L'unique vecteur propre positif de Q est φ, vecteur propre associé L'équation $Q\varphi = \lambda\varphi$ s'écrit :

$$\sigma^2 G_1 G_2 \varphi = \lambda\varphi \text{ , ou encore :}$$

$$\sigma^2 G_2 \varphi = \lambda T_1 \varphi.$$

Posons $U = \sigma^2 G_2$.
Si l'on ordonne les valeurs propres de Q en décroissant, le théorème du minimax généralisé donne :

$$\lambda_i = \min_{\substack{S \\ \text{codim } S = i-1}} \max_{x \in S \subset H} \frac{(Ux,x)_1}{t_1(x,x)} \quad , \quad i = 1,2,\ldots$$

1) Approximation par les splines fonctions

En reprenant p_h et r_h au paragraphe 3.1.1.) on définit dans H_h :
$Q_h = \sigma^2 G_{1h} G_{2h}$ où G_{1h} et G_{2h} sont les inverses des opérateurs définis par les formes sesquilinéaires continues sur $H_h \times H_h$:

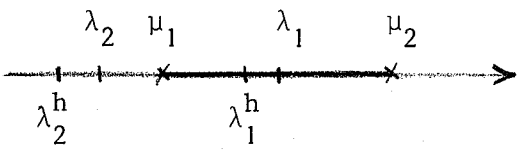
$$t_{kh}(\xi,\eta) = t_k(p_h\xi,p_h\eta) \text{ pour } \xi,\eta \in H_h,\ k = 1,2.$$

$U_h = \sigma^2 G_{2h}$.

La ième valeur propre de Q_h vérifie :

$$\lambda_i^h = \min_{\substack{S \\ \text{codim } S=i-1}} \ \max_{x\in S} \frac{(U_h\xi,\xi)_{1h}}{t_{1h}(\xi,\xi)}$$

Il est alors facile de voir que $\lambda_i \geq \lambda_i^h$, $i = 1,2,\ldots,$ pour h assez petit.
On démontre que λ_1^h est simple (à l'aide du théorème de Perron-Frobenius).
Soit un intervalle $[\mu_1,\mu_2]$ isolant λ du reste du spectre de T.
$[\mu_1,\mu_2]$ ne contient que la valeur propre λ_1^h de T_h, pour h assez petit.



On en déduit que λ_1 est valeur propre simple et que Q_h est une approximation de Q fortement stable au voisinage de toute valeur propre de Q.

2) <u>Méthode des différences finies</u>

p_h, r_h, H_h sont définis au paragraphe 3.1.2).
On obtient de la même façon $\lambda_i \geq \lambda_i^h$ lorsque :

$$(U_h\xi,\xi)_{1h} = (U\ p_h\xi,p_h\xi)_1$$

et

$$t_{1h}(\xi,\xi) \geq t_1(p_h\xi,p_h\xi).$$

VII - EXEMPLE SIMPLE DE CALCUL DE LA CONSTANTE DE STABILITE

Nous allons expliciter la condition de stabilité au voisinage d'une valeur propre, dans un cas bien connu, où les calculs peuvent être menés à leur terme.

$X = C(0,1)$ muni de la norme uniforme.

$D = \{x \in C(0,1) : x'' \in C(0,1) \text{ et } x(0) = x(1) = 0\}$.

$T : x \mapsto -x''$, $x \in D$.

Nous allons retrouver dans ce cas particulier le résultat de Keller [26] sur la convergence des valeurs propres et des vecteurs propres avec la norme du maximum en vérifiant la condition de stabilité.

1 - Approximation

$[0,1]$ est divisé en n parties égales aux points $t_i = \frac{i}{h} = hi$, $i = 0,\dots,n$, et $x(t_i) = x_i$.

x'' est approché par $\dfrac{x_{i+1}-2x_i+x_{i-1}}{h^2}$.

$X_h = \mathbb{R}^{n-1}$ est muni de la norme du maximum $\|.\|_h$

$$x_h = r_h x, \quad \|x_h\|_h = \max_{i=1,\dots,n-1} |x(t_i)| \to \|x\| = \max_{t\in[0,1]} |x(t)|$$

T_h est représenté par la matrice d'ordre n-1 :

$$A_h = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix}$$

On vérifie aisément que $\|(r_h T - \mathcal{T}_h r_h)x\|_h \to 0$, $\forall x \in X$.

Les éléments propres de T et $\mathcal{T}_h$ sont les mêmes que dans l'exemple situé au § V.2.2.3).

2 - Stabilité

Nous la vérifions pour z décrivant $[\lambda_1+r, \lambda_2-r]$:

$$\forall r > 0,\ \exists h_o,\ M : \|(A_h - z1)^{-1}\|_h < M(r) \qquad z \in [\lambda_1+r, \lambda_2 r].$$

A_h est une matrice hermitienne. Les vecteurs $\{\sqrt{2h}\ \varphi_k^h\}_{k=1,\dots,n-1}$ forment une base orthonormée. $(A_h - z1)^{-1}$ admet la même base de vecteurs propres et pour valeurs propres $\dfrac{1}{\frac{2}{h^2}(1-\cos k\pi h)-z}$.

Soit $\alpha_{ij}(z)$ le terme général de $(A_h - z1)^{-1}$

$$\alpha_{ij}(z) = 2h \sum_{k=1}^{n-1} \frac{1}{\lambda_j^h - z} \sin ik\pi h \sin jk\pi h$$

Alors $\|(A_h - z1)^{-1}\|_h = \max_i \sum_j |\alpha_{ij}(z)|$.

D'une manière générale si A hermitienne admet les valeurs propres $\lambda_1,\dots,\lambda_n$

$S_{\varphi\varphi}(A) \le \||A|\| = \sqrt{\Sigma|\lambda_i|^2} \le \Sigma\ |\lambda_i|$

D'où :

$$\|(A_h - z1)^{-1}\|_h \le \sum_{j=1}^{n-1} \frac{1}{|\lambda_j^h - z|} = \frac{1}{|\lambda_1^h - z|} + \frac{1}{|\lambda_2^h - z|} + \sum_{j>2} \frac{1}{|\lambda_j^h - z|} .$$

Il existe h_o tel que $\forall h < h_o$:

$$|\lambda_1 - \lambda_1^h| < \frac{r}{2} \Rightarrow |z - \lambda_1^h| \ge |z - \lambda_1| - |\lambda_1 - \lambda_1^h| \ge \frac{r}{2}$$

$$|\lambda_2 - \lambda_2^h| < \frac{r}{2} \Rightarrow |z - \lambda_2^h| \ge \frac{r}{2} .$$

D'autre part

$$|\lambda_j^h - z| > |\lambda_j^h - \lambda_2^h| \ , \quad j > 2$$

d'où

$$\max_z \sum_j \frac{1}{|\lambda_j^h - z|} \leq \frac{2}{r} + \frac{2}{r} + \sum_{j>2} \frac{1}{|\lambda_j^h - \lambda_2^h|}$$

Donc :

$$\frac{h^2}{2} \sum_3^{n-1} \frac{1}{\cos 2\pi h - \cos j\pi h} \leq \frac{h^2}{2} \int_2^{n-1} \frac{dx}{\cos 2\pi h - \cos \pi h x} \leq \frac{1}{2}$$

$$\forall r, \ \exists h_o \ : \ \|(A_h - z1)^{-1}\|_h \leq \frac{4}{r} + \frac{1}{2\pi} \ , \text{ pour } h < h_o, \quad z \in [\lambda_1 + r \ , \lambda_2 - r].$$

VIII - METHODES DE PROJECTION

Nous nous intéressons maintenant au cas particulier des méthodes de projection.

1 - Définitions

Soit une suite de projections $\pi_n : X \to X$, $n = 1,2,\ldots$ telle que $\pi_n x \xrightarrow[n\to\infty]{} x$, $\forall x \in X$.

D'après le théorème de Banach-Steinhaus, il existe M tel que

$$\|\pi_n\| \leq M$$

$X_n = \pi_n X$ est un sous-espace vectoriel de X.
Posons $N = \dim X_n$ et soit $\{w_i^n\}_{1 \leq i \leq N}$ une base de X_n.
π_n est défini par : $x \in X \mapsto \sum_1^N x_i^n \, w_i^n \in X_n$
Si l'on introduit :

- $r_n : x \in X \mapsto (x_1^n, \ldots, x_N^n) \in \mathbb{R}^N$

- $p_n : (x_1, \ldots, x_n) \mapsto \sum_1^N x_i \, w_i^n$

$\pi_n = p_n r_n$.

On définit

$$T_n = \pi_n T \pi_n : X \to X$$

ou

$$\mathcal{T}_n = r_n T p_n : \mathbb{R}^N \to \mathbb{R}^N$$

La norme $\|\cdot\|_n$ sur $\mathbb{R}^N$ est définie par :

$$\|\xi\|_n = \|p_n \xi\| \quad , \quad \forall \xi \in \mathbb{R}^N$$

Il y a compatibilité des normes :

$$\|r_n x\|_n = \|\pi_n x\| \to \|x\| \quad , \quad \forall x \in X$$

Si T est borné, $\mathcal{T}_n$ est une approximation de T :

$$\|(r_n T - \mathcal{T}_n r_n)x\|_n = \|\pi_n T(1-\pi_n)x\| \leq \|\pi_n\| \, \|T\| \, \|(1-\pi_n)x\| \to 0.$$

La restriction de T_n à X_n est représentée, dans la base $\{w_i^n\}$, par la matrice associée à $\mathcal{T}_n$ dans $\mathbb{R}^N$.

Exemple : Dans un espace de Hilbert H muni d'une base orthonormale, on choisit pour π_n la projection orthogonale sur le sous-espace vectoriel engendré par les n premiers vecteurs de la base. Alors :

$$\pi_n = \pi_n^* \quad \text{et} \quad \pi_n \xrightarrow[n\to\infty]{} 1$$

N.B. T sera supposé borné dans toute la suite.

2 - Méthodes classiques

Elles se différencient par la manière de définir π_n, la méthode générale étant la méthode de Galerkin définie au § 1. Petrov [52] a généralisé cette méthode en considérant $T_n = \pi'_n T \pi_n$, où π'_n appartient à une autre suite convergente de projections.

Dans la méthode d'Aronzajn-Bazley [5] appliquée à T autoadjoint dans un Hilbert, à l'aide de la décomposition $T = T_o + U$ telle qu'on connaisse les éléments propres de T_o, la suite des valeurs propres approchées de T est donnée par celles de $T_n = T_o + (\pi_n U^{-1} \pi_n)^{-1}$ où π_n est la projection orthogonale sur le sous-espace vectoriel engendré par les n premiers vecteurs propres de T_o.

3 - Stabilité

Nous reprenons, pour la méthode de Galerkin, les conditions suffisantes données au § VI.

3.1 - Approximation uniforme

Lemme 7

Si $\{\pi_n\}$ *est une suite de projections telle que* $\pi_n x \to x$ *, pour* $x \in X$ *et* $\pi_n^* y \to y$ *pour* $y \in X^*$ *et* T *un opérateur compact, alors :*
$\|\pi_n T \pi_n - T\| \xrightarrow[n\to\infty]{} 0$.

$$\pi_n T \pi_n - T = (\pi_n - 1)T + \pi_n T(\pi_n - 1)$$

T est compact donc $\|(\pi_n - 1)T\| \to 0$.
D'autre part, $\|\pi_n T(\pi_n - 1)\| \leq M\|(\pi_n^* - 1)T^*\| \to 0$.
La proposition 1 s'applique et l'on déduit la

Proposition 5

La méthode de Galerkin, appliquée à un opérateur compact, avec une suite de projections π_n *telles que :*

$$\pi_n x \to x \text{ pour } x \in X \text{ et } \pi_n^* y \to y \text{ pour } y \in X^*$$

est stable au voisinage de toute valeur propre de T.

3.2 - Approximation fortement stable d'un opérateur compact autoadjoint

Soit un opérateur autoadjoint compact dans un espace de Hilbert H. Soit π_n la projection orthogonale sur les n premiers vecteurs d'une base orthonormée de H. Alors :

$$(T_n\xi,\xi)_n = (r_n T p_n \xi,\xi)_n = (T p_n \xi, p_n \xi)$$

On en déduit facilement que :

a) pour les valeurs propres non négatives ordonnées en décroissant :

$$\lambda_i^+ \geq \lambda_i^{h+} \quad , \quad i = 1,2,\ldots$$

b) pour les valeurs propres non positives ordonnées en croissant :

$$\lambda_i^- \leq \lambda_i^{h-} \quad , \quad i = 1,2,\ldots$$

D'où la :

Proposition 6

L'approximation de Galerkin appliquée à un opérateur compact autoadjoint est fortement stable au voisinage de toute valeur propre de l'opérateur.

3.3 - T est un opérateur borné dans H.

Définition : $\Theta(T) = \{z \in \mathbb{C} \; ; \; z = (Tx,x), \quad x \in H, \; \|x\| = 1\}$ est <u>l'image numérique</u> de T.

$\Theta(T)$ est un convexe qui contient $\sigma(T)$ lorsque T est borné ([25] p. 267).

Lemme 8

Une condition suffisante pour que

$$|((T-z1)x,x)| \geq \delta\|x\|^2, \quad \delta > 0, \quad x \in H$$

est que z soit extérieur à $\overline{\Theta(T)}$.

En effet, il existe un disque : $\{t \in \mathbb{C} \; ; \; |t-z| < \delta\}$ autour de z qui ne contient aucun point de $\Theta(T)$.

Proposition 7

La méthode de Galerkin appliquée à un opérateur borné dans un espace de Hilbert est stable en tout point z extérieur à $\overline{\Theta(T)}$.

Démonstration

Soit $x \in H_n = \pi_n H \quad , \quad \|x\| = 1$.

$$\|(\pi_n T \pi_n - z1)x\| = \|\pi_n(T-z1)x\| = \max_{\substack{y \in H_n \\ \|y\|=1}} |((T-z1)x,y)|$$

$$\geq |((T-z1)x,x)| \geq \delta \quad \text{pour } z \notin \overline{\Theta(T)}$$

C'est-à-dire :

$$\|(T_n - z1)^{-1}\| \leq \frac{1}{\delta} .$$

CHAPITRE - V

BORNES D'ERREUR DANS L'APPROXIMATION STABLE DES ELEMENTS PROPRES D'UN OPERATEUR LINEAIRE

Nous étudions maintenant la précision obtenue en calculant les valeurs propres d'un opérateur linéaire par une méthode d'approximation stable. Il existe de nombreux résultats particuliers : Saulev [56, 57] , Keller [26], par exemple, pour les méthodes aux différences finies appliquées à des problèmes différentiels elliptiques autoadjoints ; Krasnoselskii-Vainikko [28] pour les méthodes de projection appliquées à des opérateurs compacts.

Dans l'hypothèse où l'opérateur de restriction de X dans X_h est stable, nous donnons un résultat général qui est analogue au résultat valable pour des matrices finies.

Dans le cas d'opérateurs autoadjoints bornés, nous pouvons étendre les résultats du chapitre 2. Et pour des opérateurs autoadjoints compacts, nous généralisons l'inégalité que Wielandt et Hoffmann [71] ont établi pour des matrices normales.

I - OPERATEURS FERMES DANS UN ESPACE DE BANACH

1 - Supposons qu'on connaisse les fonctions :

a) $(h,x) \mapsto \varepsilon(h,x)$ définie par l'approximation :

$$\|(r_h T - \mathcal{T}_h r_h)x\|_h \leq \varepsilon(h,x) \quad , \quad x \in D(T).$$

b) $r \mapsto M(r)$ définie par la stabilité de T_h au voisinage de λ :

$$\forall r > 0,\ \exists h_o(r),\ M(r) : q_h(z) \geq \frac{1}{M(r)}, \quad h \leq h_o(r)$$
$$z \in \{z\ ;\ r \leq |z-\lambda| \leq R\}.$$

Soit φ un vecteur propre normé associé à λ.

En reprenant la démonstration du lemme 2 du chapitre IV on voit que, pour tout r, il existe $h_1 \leq h_o$ défini par :

$$\varepsilon(h,\varphi) = \frac{1}{M(r)}$$

et qui vérifie la propriété :

$$\forall r,\ \exists h_1(r) : \quad |\lambda-\lambda^h| < r \text{ pour } h < h_1(r).$$

Ce type de borne d'erreur a été développé dans certains cas particuliers par Misovkich [41] et Anselone-Moore [1], à partir d'une idée de Kantorovitch-Akilov ([23], chapitre XIV).

2 - On suppose r_h stable : $\|r_h x\|_h \leq c_2\|x\|$, pour $x \in X$.

Nous établissons des bornes d'erreur sur les valeurs propres et l'angle entre les directions propres dans X_h, lorsque les valeurs propres sont simples.

THEOREME 1

Soit λ *une valeur propre simple de* T, *isolée par* $\Gamma(\lambda,R)$ *de frontière* Γ *et* φ *un vecteur propre normé associé.*
On suppose r_h *stable.*
Soit T_h *une approximation de* T, *stable dans* $\Gamma(\lambda,R)$, *admettant une valeur propre simple dans* $\Gamma(\lambda,R)$, λ^h *qui converge vers* λ. φ_h *est un vecteur propre normé associé.*
Soit θ_h *l'angle aigu entre les directions définies par* $r_h\,\varphi$ *et* φ_h.
Alors :

$$\begin{cases} |\lambda-\lambda^h| \leq C\ \|(r_hT-T_hr_h)\varphi\|_h \\ \sin\theta_h \leq C'\ \|(r_hT-T_hr_h)\varphi\|_h \end{cases}$$

où C *est une constante telle que :*

$$|(\varphi,\Psi)| < \frac{1}{C} \quad , \text{ avec } \Psi \in \mathrm{Ker}(T^*-\bar{\lambda}1),\ \|\Psi\| = 1$$

et C' *est une constante telle que :*

$$(1+C)M < C'$$

M *étant la constante de stabilité de* T_h *définie par :*

$$M = \max_{z\in\Gamma} \|(T_h-z1)^{-1}\|_h .$$

Démonstration

a) inégalité sur les valeurs propres.
Considérons l'égalité :

$$(T_h-\lambda^h 1)r_h\varphi = (T_hr_h-r_hT)\varphi+(\lambda-\lambda^h)r_h\varphi$$

Soit X^* (resp. X_h^*) l'espace adjoint de X (resp. X_h), et T^* (resp. $\mathcal{T}_h^*$) l'opérateur adjoint de T (resp. $\mathcal{T}_h$).

Soit $\Psi_h \in \mathrm{Ker}(\mathcal{T}_h^* - \bar{\lambda}^h 1)$. (,) note la mise en dualité de X, X^* et X_h, X_h^*

$$0 = (r_h\varphi, (\mathcal{T}_h^* - \bar{\lambda}^h 1)\Psi_h) = ((\mathcal{T}_h - \lambda^h 1) r_h\varphi, \Psi_h)$$

$$= ((\mathcal{T}_h r_h - r_h T)\varphi, \Psi_h) + (\lambda - \lambda^h)(r_h\varphi, \Psi_h).$$

Montrons qu'il existe une constante C telle que

$$|(r_h\varphi, \Psi_h)| \geq \frac{1}{C}$$

$(r_h\varphi, \Psi_h) = (\varphi, r_h^*\Psi_h)$ par définition de r_h^*.

Or r_h^*, opérateur de prolongement de X_h^* dans X^* est stable puisque r_h l'est.
D'autre part, $\mathcal{T}_h^*$ est stable au voisinage de $\bar{\lambda}$.
Donc, d'après le théorème 3 du chapitre IV, il existe un vecteur propre normé Ψ de T^*, associé à $\bar{\lambda}$, tel que :

$$\|\Psi - r_h^*\Psi_h\| \to 0$$

lorsqu'on impose, par exemple, $\|r_h^*\Psi_h\| = 1$.

Si l'hypothèse : $|(\varphi, r_h^*\Psi_h)| \geq \frac{1}{C}$ n'était pas vérifiée, il existerait une sous-suite $r_{h_i}^*\Psi_{h_i}$ de $r_h^*\Psi_h$ telle que

$$(\varphi, r_{h_i}^*\Psi_{h_i}) \xrightarrow[h_i \to 0]{} 0$$

Et puisque $\|r_{h_i}^*\Psi_{h_i} - \Psi\| \xrightarrow[h_i \to 0]{} 0$, on aurait $(\varphi, \Psi) = 0$

Or ceci est une condition nécessaire et suffisante pour que l'équation :
$(T - \lambda 1)x = \varphi$ admette une solution $x \neq 0$.
Cette solution vérifie

$$(T - \lambda 1)^2 x = 0.$$

Cela est impossible puisque λ est valeur propre simple de T.
Donc, il existe une constante C telle que

$$|(\varphi, r_h^* \Psi_h)| \geq \frac{1}{C} .$$

On note que $\frac{1}{C} \leq \|\varphi\| \; \|r_h^* \Psi_h\| = 1$.

D'autre part, puique

$$|(\varphi,\Psi)-(\varphi, r_h^* \Psi_h)| \xrightarrow[h \to 0]{} 0$$

la constante C est telle que $C > \frac{1}{|(\varphi,\Psi)|}$
On déduit que :

$$|\lambda-\lambda^h| \leq C \sup_{\substack{\Psi_h \in \mathrm{Ker}(T_h^*-\bar{\lambda}^h 1) \\ \|r_h^* \Psi_h\| = 1}} |((r_h T - T_h r_h)\varphi, \Psi_h)|$$

Soit :

$$|\lambda-\lambda^h| \leq C \; \|(r_h T - T_h r_h)\varphi\|_h .$$

b) Inégalité sur l'angle des directions propres.
Soit P_h la projection spectrale associée à λ^h, et $Q_h = 1-P_h.(T_h-\lambda^h 1)$ commute avec Q_h.
Nous considérons toujours l'égalité :

$$(T_h-\lambda^h 1) r_h \varphi = (T_h r_h - r_h T)\varphi + (\lambda-\lambda^h) r_h \varphi.$$

En multipliant à gauche par Q_h il vient :

$$(T_h-\lambda^h 1) Q_h r_h \varphi = Q_h (T_h r_h - r_h T)\varphi + (\lambda-\lambda^h) Q_h r_h \varphi.$$

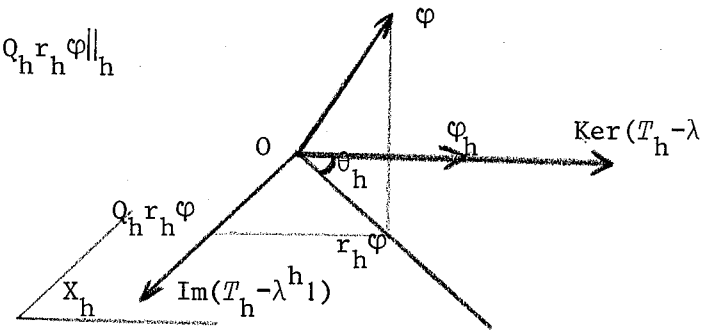
Puisque λ^h est la seule valeur propre de $\mathcal{T}_h$ dans $\Gamma(\lambda,R)$,
$R'_h(z) = R_h(z)Q_h$ est holomorphe dans $\Gamma(\lambda,R)$.
On pose $R'_h(\lambda^h) = \lim\limits_{z\to\lambda^h} R'_h(z)Q_h$

$$\|R'_h(\lambda^h)\|_h \leq \max_{z\in\Gamma} \|R'_h(z)\|_h = \max_{z\in\Gamma} \|R_h(z)Q_h\|_h \leq \max_{z\in\Gamma} \|R_h(z)\|_h = M$$

$\sin\theta_h = \dfrac{\|Q_h r_h \varphi\|_h}{\|r_h\varphi\|_h}$ est alors majoré par

la quantité :

$$\frac{1}{\|r_h\varphi\|_h} \|R'_h(\lambda^h)\|_h \; \|Q_h(\mathcal{T}_h r_h - r_h T)\varphi + (\lambda-\lambda^h)Q_h r_h\varphi\|_h$$



et

$\|r_h\varphi\|_h \xrightarrow[h\to 0]{} \|\varphi\| = 1$. Il vient alors :

$$\sin\theta_h \leq C' \; \|(\mathcal{T}_h r_h - r_h T)\varphi\|_h$$

où C' est une constante telle que

$$C' > (1+C)M$$

<u>Remarque</u> - Si T est autoadjoint , $\varphi = \Psi$ et $C > 1$.

D'autre part $\min\limits_{\substack{\lambda\neq\lambda^h \\ \lambda\in\sigma(\mathcal{T}_h)}} |\lambda-\lambda^h| = \dfrac{1}{\|R'_h(\lambda^h)\|_h}$.

Si λ est une valeur propre de multiplicité algébrique m, isolée par $\Gamma(\lambda,R)$, les inégalités sont un peu modifiées de la manière suivante :

a) Il existe au moins une valeur propre λ^h de $\mathcal{T}_h$ telle que :

$$|\lambda-\lambda^h| \leq C \; \|(r_h T-\mathcal{T}_h r_h)\varphi\|_h^{1/m}$$

avec

$$|(\varphi,\Psi)| > \frac{1}{C} \text{ où } \Psi \in \mathrm{Ker}(T^*-\bar{\lambda}1)^m \text{ et } \|\Psi\| = 1.$$

b) Soit P_h la projection spectrale associée aux valeur propres de $\mathcal{T}_h$ dans $\Gamma(\lambda,R)$

$$\|r_h\varphi-P_h r_h\varphi\|_h \leq C' \; \|(r_h T-\mathcal{T}_h r_h)\varphi\|_h$$

avec

$$C' > (1+C)M^m.$$

Notons que la borne d'erreur obtenue pour les valeurs propres est du même type que celle valable dans le cas de matrices [22].

II - OPERATEURS AUTOADJOINTS DANS UN HILBERT

1 - Perturbation d'un opérateur autoadjoint

Soit T un opérateur autoadjoint dans un espace de Hilbert, λ une valeur propre isolée et φ un vecteur propre normé associé. On s'intéresse au spectre de $T' = T+U$, U étant un opérateur autoadjoint.

D'une manière générale on dispose des deux inégalités suivantes :

a) Krylov-Weinstein ([71], p. 321) :

$$\forall\lambda \in \sigma(T),\ \varphi \in D(U),\ \exists\lambda' \in \sigma(T') :$$

$$|\lambda'-(T'\varphi,\varphi)| \leq \sqrt{\|U\varphi\|^2-(U\varphi,\varphi)^2}$$

b) Kato [24] :

$$\forall\lambda \in \sigma(T),\ \varphi\in D(U),\ \exists\lambda' \in \sigma(T') :$$

$$|\lambda'-(T'\varphi,\varphi)| \leq \frac{\|U\varphi\|^2-(U\varphi,\varphi)^2}{d}$$

où d est la distance de $(T'\varphi,\varphi)$ à $\sigma(T')-\{\lambda'\}$.

c) Par un raisonnement analogue à celui utilisé pour établir l'inégalité de Krylov-Weinstein, on montre que :

$$\forall\lambda \in \sigma(T),\ \varphi \in D(U)\ ,\ \exists\lambda' \in \sigma(T') :$$

$$|\lambda'-\lambda| \leq \|U\varphi\|$$

Nous étudions maintenant le cas où la perturbation U est bornée.
Soit d la distance de λ au reste du spectre de T.
Si $2\|U\| < d$, il existe λ' et φ', éléments propres de T' pour lesquels des majorations de $|\lambda'-(T'\varphi,\varphi)|$ et $\mathrm{tg}\theta$, du type de celles démontrées au chapitre II sont valables, la démonstration pouvant se transcrire mot pour mot dans le cas d'opérateurs autoadjoints (ou normaux).

THEOREME 2

T *admet la valeur propre simple* λ, *séparée du reste du spectre par* d. *Si* $2\|U\| < d$, $T' = T+U$ *admet une valeur propre* λ' *et un vecteur propre* φ' *tels que si :*

$$s = \frac{2\|U\|}{d} \quad , \quad t = \frac{2\|U\varphi\|}{d} \leq s$$

$$\begin{cases} |\lambda'-(T'\varphi,\varphi)| \leq f(r) \ \|U\varphi\| \\ \\ tg\theta \leq f(r) \end{cases}$$

avec $f(r) = \frac{1}{r}\left(\sqrt{1+r^2}-1\right)$

(2.1) $r = \dfrac{s}{\sqrt{1-s^2}}$

(2.2) $r_s = \dfrac{t}{1-\sqrt{s^2-t^2}}$

Si, de plus, $(U\varphi,\varphi) = 0$,

(2.3) $r = s$

(2.4) $r_s = \dfrac{2t}{2-s+t^2}$

2 - Application aux méthodes d'approximation

Nous faisons à nouveau les hypothèses du § VI.1 du chapitre IV : r_h est linéaire surjectif, p_h est linéaire injectif et :

i) $p_h r_h x \to x$ pour $x \in X$

ii) $\|p_h \xi\| \leq C_1 \|\xi\|_h$ pour $\xi \in X_h$

iii) $r_h p_h = 1$.

Si $\mathcal{T}_h$ est une approximation de T, $T_h = p_h \mathcal{T}_h r_h$ en est une aussi :

$$\|(T-T_h)x\| \to 0 \quad \text{pour } x \in D(T)$$

et T_h admet les mêmes valeurs propres que $\mathcal{T}_h$ (sauf 0).

On peut appliquer les résultats du § 1. Le résultat c) par exemple, implique que :

$$\forall \lambda^h, \text{ valeur propre de } T_h, \ \exists \lambda \in \sigma(T) : \ |\lambda-\lambda^h| \leq \|(T-T_h)\varphi_h\|$$

si φ_h est un vecteur propre normé de T_h associé à λ^h.

Si l'on applique <u>la méthode de Galerkin</u> à un opérateur borné T, avec une suite de projections π_n autoadjointes, les propriétés ci-dessus sont vérifiées. $T_n = \pi_n T \pi_n$ est une approximation autoadjointe de T vérifiant :

a) $\|T-T_n\| \xrightarrow[n\to\infty]{} 0$

b) $((T-T_n)x_n, x_n) = 0$ pour $x_n \in \pi_n X$.

Pour chaque valeur propre $\neq 0$ de T_n, il existe N tel que, pour $n > N$, l'hypothèse $2\|T-T_n\| < d$ du théorème 2 est vérifiée. On obtient alors les majorations d'erreur définies à partir de (2.3) et (2.4) puisque la propriété b) ci-dessus est vérifiée pour tout vecteur propre de T_n.

Exemple numérique

Soit à calculer les éléments propres de l'opérateur défini dans $H = L^2(0,1)$ par

$$T : x \mapsto \int_0^1 K(\xi,t)x(t)dt$$

avec

$$K(\xi,t) = \begin{cases} (1-t)\xi & \text{pour } \xi \leq t \\ (1-\xi)t & \text{pour } \xi > t \end{cases}$$

Les valeurs propres sont $\lambda_i = \frac{1}{i^2\pi^2}$, $i = 1,2,\ldots$ et les vecteurs propres normés sont : $\varphi_i(t) = \sqrt{2}\sin i\pi t$, $i = 1,2,\ldots$

On applique la méthode de Galerkin en projetant sur les n premiers vecteurs de la base orthonormée $(1,\sqrt{2}\cos k\pi t)_{k=1,2,\ldots}$) $T_n = \pi_n T \pi_n$ est représenté dans la base canonique de $\mathbb{C}^n$ par la matrice $A_n = ((Te_j,e_i)_{i,j=1,\ldots,n})$. On note $\{\lambda^n,\varphi^n\}$ les éléments propres de A_n.

Les termes non nuls de la partie triangulaire inférieure de A_n sont pour n pair par exemple :

$$a_{11} = \frac{1}{12}$$

$$a_{2p+1,1} = \frac{-1}{2\sqrt{2}p^2\pi^2} \quad , \quad p = 1,2,\ldots,\frac{n}{2}-1$$

$$a_{2p,2p} = \frac{1}{(2p-1)^2\pi^2} - \frac{8}{(2p-1)^4\pi^4} \quad , \quad p = 1,2,\ldots,\frac{n}{2}$$

$$a_{2p+1,2p+1} = \frac{1}{(2p)^2\pi^2} \quad , \quad p = 1,\ldots,\frac{n}{2}-1$$

$$a_{2p,2p} = -\frac{8}{(2p-1)^2(2q-1)^2\pi^4} \quad , \quad p > q \quad , \quad p = 2,\ldots,\frac{n}{2} ,$$

$$q = 1,\ldots,\frac{n}{2}-1.$$

Les calculs ont été faits pour n = 10 et 30.
$\|T-T_n\|$ est majoré par la double norme.
La condition $2\|T-T_n\| < d$ est vérifiée.

a) avec n = 10, pour les 2 plus grandes valeurs propres.

b) avec n = 30, pour les 6 plus grandes.

Pour ces valeurs propres, nous comparons la différence $\lambda-\lambda^n$ aux majorations obtenue à partir de l'inégalité de Krylov et des inégalités (2.3) et (2.4). Et pour les vecteurs propres associés nous comparons $\|\varphi-\varphi^n\|$ aux majorations analogues.

Le calcul des éléments propres de A_n est fait par la méthode de Jacobi.

Valeurs propres

		différence	Krylov	(2.3)	(2.4)
n = 10	λ_1	$3,7\ 10^{-5}$	$1,9\ 10^{-3}$	$9,8\ 10^{-5}$	$5,1\ 10^{-5}$
	λ_2	$2,8\ 10^{-5}$	$9,2\ 10^{-4}$	$2,1\ 10^{-4}$	$6,5\ 10^{-5}$
n = 30	λ_1	$1,1\ 10^{-6}$	$3,2\ 10^{-4}$	$1,6\ 10^{-6}$	$1,3\ 10^{-6}$
	λ_2	$1,0\ 10^{-6}$	$1,4\ 10^{-4}$	$3,7\ 10^{-6}$	$1,4\ 10^{-6}$
	λ_3	$1,1\ 10^{-6}$	$8,6\ 10^{-5}$	$6,5\ 10^{-6}$	$1,6\ 10^{-6}$
	λ_4	$1,0\ 10^{-6}$	$4,5\ 10^{-5}$	$7,1\ 10^{-6}$	$1,0\ 10^{-6}$
	λ_5	$1,1\ 10^{-6}$	$3,8\ 10^{-5}$	$1,0\ 10^{-5}$	$1,7\ 10^{-6}$
	λ_6	$1,0\ 10^{-6}$	$2,2\ 10^{-5}$	$9,2\ 10^{-6}$	$1,3\ 10^{-6}$

Vecteurs propres

		différence	(2.3)	(2.4)
n = 10	φ_1	$1,9\ 10^{-2}$	$5,1\ 10^{-2}$	$2,7\ 10^{-2}$
	φ_2	$3,3\ 10^{-2}$	$2,5\ 10^{-1}$	$7,9\ 10^{-2}$
n = 30	φ_1	$1,4\ 10^{-3}$	$4,9\ 10^{-3}$	$4,2\ 10^{-3}$
	φ_2	$2,7\ 10^{-3}$	$2,6\ 10^{-2}$	$1,0\ 10^{-2}$
	φ_3	$7,2\ 10^{-3}$	$7,5\ 10^{-2}$	$1,9\ 10^{-2}$
	φ_4	$1,0\ 10^{-2}$	$1,6\ 10^{-1}$	$2,3\ 10^{-2}$
	φ_5	$1,5\ 10^{-2}$	$2,7\ 10^{-1}$	$4,4\ 10^{-2}$
	φ_6	$1,7\ 10^{-2}$	$3,9\ 10^{-1}$	$5,9\ 10^{-2}$

On constate toujours que les bornes issues de (2.4) sont meilleures que celles issues de (2.3).

3 - Inégalité de Wielandt-Hoffmann pour des opérateurs compacts autoadjoints

3.1 - Rappels

Von Neumann [59] a introduit les anneaux normés d'opérateurs compacts définis ainsi :

$$\mathfrak{S}_p = \{T \in \mathcal{L}(H)\ ;\ N_p(T) = (\mathrm{tr}(T^*T)^{p/2})^{1/p} < \infty\} \quad , \text{ pour } p > 0.$$

Göhberg et Markus [17] ont démontré les propriétés suivantes, dans le but d'étendre à de tels opérateurs les inégalités de Hardy-Littlewood et Polya entre les éléments d'une matrice.

1) $0 < p < 2$

$T \in \mathfrak{S}_p$ <=> il existe une base orthonormée $\{e_i\}$ de H telle que

$$\sum_i \|Te_i\|^p < \infty.$$

Alors $N_p(T) = \min_{\{e_i\}} (\sum_i \|Te_i\|^p)^{1/p} \leq (\sum_{i,j} |(Te_j,e_i)|^p)^{1/p}$

$\mathfrak{S}_1$ est l'ensemble des opérateurs nucléaires.

2) $p = 2$

$T \in \mathfrak{S}_2 \iff$ il existe une orthonormée $\{e_i\}$ telle que $\sum_i \|Te_i\|^2 < \infty$

$\iff$ pour toute base orthonormée $\{e_i\}$, $\sum_i \|Te_i\|^2 < \infty$

Et

$$N_2(T) = (\sum_i \|Te_i\|^2)^{1/2} = (\sum_{i,j} |(Te_j,e_i)|^2)^{1/2}$$

$\mathfrak{S}_2$ est l'ensemble des opérateurs de Hilbert-Schmidt

3) $p > 2$

$T \in \mathfrak{S}_p \iff$ pour toute base orthonormée $\{e_i\}$, $\sum_i \|Te_i\|^p < \infty$

Et

$$N_p(T) = \max_{\{e_i\}} (\sum_i \|Te_i\|^p)^{1/p} \geq (\sum_{i,j} |(Te_j,e_i)|^p)^{1/p}$$

3.2 - Soient T' et T deux opérateurs compacts autoadjoints dont la différence est un opérateur U de $\mathfrak{S}_p$. Nous étudions les relations entre les valeurs propres λ'_i de T', λ_i de T et γ_i de U, ordonnées par ordre décroissant des valeurs absolues.

THEOREME 3

Soient T', T *et* U *des opérateurs compacts autoadjoints tels que* T' = T+U , U *appartient à* $\mathfrak{S}_p$. λ'_i , λ_i *et* γ_i *sont les valeurs propres respectives, ordonnées par ordre décroissant des valeurs absolues. Elles vérifient :*

$$\sum_{i=1}^{\infty} |\lambda'_i - \lambda_i|^p \leq \sum_{i=1}^{\infty} |\gamma_i|^p.$$

Démonstration

On pose $T(k) = T+kU$, $0 \leq k \leq 1$. On note $\lambda_i(k)$ les valeurs propres de $T(k)$, et $\varphi_i(k)$ les vecteurs propres orthonormés associés. Pour k variant dans $[0,1]$, chaque $\lambda_i(k)$ (ainsi que $\varphi_i(k)$) peut être continuée analytiquement, le long de l'axe réel, à partir de λ_i, tant que cette valeur propre reste isolée, de multiplicité finie ([25] p. 387).

Alors les $\lambda_i(k)$ sont continûment dérivables en k et les vecteurs propres $\varphi_i(k)$ forment un système complet de fonctions dérivables en k.

On ordonne les $\lambda_i(k)$ par ordre de valeur absolue décroissante.
Les graphes des fonctions $k \mapsto \lambda_i(k)$, $i \in \mathbb{N}$, se coupent au plus en une infinité dénombrable de points, ce qui introduit des discontinuités pour $\lambda_i'(k)$ et $\varphi_i(k)$ en infinité dénombrable au plus.

La démonstration faite dans le cas de matrices hermitiennes dans $\mathbb{C}^n$ ([25] p. 125) peut alors se transcrire, et l'on obtient l'inégalité cherchée.

3.3 - Application à la méthode de Galerkin

Soit π_n la projection orthogonale sur les n premiers vecteurs d'une base orthonormée $\{e_i\}$ fixée de H. T est approché par $T_n = \pi_n T \pi_n$.
On note λ_i les valeurs propres de T, $\lambda_i^{(n)}$ celles de T_n (avec $\lambda_i^{(n)} = 0$ pour $i > n$).
Appliquons le théorème 3 à $T = T_n+(T-T_n)$ lorsque $T-T_n$ appartient à $\mathfrak{S}_p$. Puisque $T-T_n$ est autoadjoint :

$$N_p(T-T_n) = \left(\sum_{i=1}^{n} |\gamma_i|^p\right)^{1/p}$$

Il vient alors :

$$\sum_{i=1}^{\infty} |\lambda_i-\lambda_i^{(n)}|^p \leq (N_p(T-T_n))^p.$$

Si $N_p(T-T_n) \xrightarrow[n\to\infty]{} 0$, le vecteur $(\lambda_i^{(n)})$ tend vers (λ_i) dans ℓ^p : on a alors une estimation globale de la convergence des valeurs propres de T_n.

Posons $t_{ij} = (Te_j, e_i)_{i,j \in \mathbb{N}}$. Nous allons donner à l'aide des résultats du § 3.2, des conditions sur les éléments t_{ij} pour que $N_p(T-T_n) \xrightarrow[n\to\infty]{} 0$.

1) $0 < p < 2$

$\sum_{i,j} |t_{ij}|^p < \infty$ est une condition suffisante pour que $T \in \mathfrak{S}_p$.

Et $(N_p(T-T_n))^p \leq \left(\sum_{i=1}^{n} \sum_{j>n} |t_{ij}|^p + \sum_{i>n} \sum_{j} |t_{ij}|^p \right) \xrightarrow[n\to\infty]{} 0$

En particulier si $T \in \mathfrak{S}_1$ est défini positif :

$$N_1(T) = \sum_1^\infty \lambda_i = \sum_1^\infty t_{ii} < \infty$$

donc

$$N_1(T-T_n) = \sum_{i>n} t_{ii} \xrightarrow[n\to\infty]{} 0.$$

2) $p = 2$

$N_2(T) = \left(\sum_{i,j} |t_{ij}|^2 \right)^{1/2}$ est une CNS pour que $T \in \mathfrak{S}_2$.

Alors $N_2(T-T_n) \xrightarrow[n\to\infty]{} 0$.

3) $p > 2$

$\sum_{i,j} |t_{ij}|^p < \infty$ n'est pas suffisant pour que $T \in \mathfrak{S}_p$.

Si la base $\{e_i\}$ choisie dans H pour définir la méthode de Galerkin réalise le maximum de $\sum_i \|Te_i\|^p$, alors $N_p(T-T_n) \xrightarrow[n\to\infty]{} 0$.

Exemple : Reprenons l'opérateur intégral T défini au § II.2 de ce chapitre, dans $H = L^2(0,1)$. T est défini positif et appartient à $\mathfrak{S}_1$:

$$\sum_{i=1}^{\infty} \lambda_i = \sum_{i=1}^{\infty} \frac{1}{i^2\pi^2} < \infty$$

Les valeurs propres $\lambda_i^{(n)}$ calculées par la méthode de Galerkin vérifient

$$\sum_{i=1}^{\infty} |\lambda_i - \lambda_i^{(n)}| \leq \sum_{i>n} t_{ii} < \frac{1}{\pi^2} \sum_{i>n} \frac{1}{(i-1)^2} < \frac{1}{\pi^2} \frac{1}{n-1} .$$

CHAPITRE - VI

ALGORITHMES LR ET QR APPLIQUES A CERTAINES CLASSES D'OPERATEURS LINEAIRES DANS UN ESPACE DE HILBERT

En algèbre linéaire, les problèmes généraux qui se posent au sujet des matrices tournent autour de deux pôles. Si A est une matrice $n \times n$ il s'agit de :

α) l'inversion de A

β) la recherche des éléments propres de A.

Pour résoudre α) une méthode classique consiste à se ramener à l'inversion d'une matrice triangulaire, à l'aide de la décomposition de Gauss $A = LR$, ou de Schmidt $A = QR$. (cf. Chapitre I).

Pour résoudre β) les algorithmes LR et QR sont très souvent employés (cf. Chap. I). L'algorithme QR permet, en outre, de déterminer la matrice triangulaire supérieure R unitairement semblable à A (théorème de Schur [12]).

Dans la théorie des opérateurs, ces deux problèmes ont été étudiés sous l'angle suivant : dans un espace de Hilbert séparable,

α'') Göhberg-Krein [16]et d'autres ont étudié l'analogue de la décomposition de Gauss pour des opérateurs différant de l'identité par un opérateur compact.

β'') Lifchitz [35] et Sahnovitch [58] ont étudié l'analogue du théorème de Schur pour des opérateurs dont la partie imaginaire est un opérateur compact de $\mathfrak{S}_1$ ou $\mathfrak{S}_2$.

Nous présentons ici, pour des opérateurs à spectre discret admettant une base de vecteurs propres, une généralisation des décompositions de Gauss et Schmidt, ainsi que des algorithmes LR et QR. Nous obtenons, avec un matériel technique plus simple, des résultats recoupant ceux cités ci-dessus.

I - ISOMORPHISMES DANS UN HILBERT

1 - Définitions et notations

Soit H un espace de Hilbert séparable admettant la base orthonormée $\{e_i\}$. Soit π_n la projection orthogonale de H sur le sous-espace engendré par les n premiers vecteurs $e_1,\dots,e_n$.

Soit T un opérateur borné dans H. On définit (cf. chapitre IV, § VIII) :

$$r_n : x = \sum_i x_i e_i \mapsto \xi = (x_1,\dots,x_n)$$

$$p_n : \xi = (x_1,\dots,x_n) \mapsto p_n\xi = \sum_1^n x_i e_i$$

$\mathcal{T}_n = r_n T p_n$, opérateur de $\mathbb{C}^n$, représenté dans la base canonique de $\mathbb{C}^n$ par la matrice (notée de la même façon) :

$$\mathcal{T}_n = (t_{ij}), \quad 1 \le i,j \le n \quad \text{avec} \quad t_{ij} = (Te_j, e_i)$$

$T_n = p_n \mathcal{T}_n r_n = \pi_n T \pi_n$, puisque $\pi_n = p_n r_n$.

$\mathbb{C}^n$ est muni du produit scalaire euclidien : $(\xi,\eta)_n = (p_n\xi, p_n\eta)$.

2 - Condition suffisante pour que T soit un isomorphisme de H

Nous démontrons le :

THEOREME 1

Si une suite de matrices emboitées $\{\mathcal{T}_n\}_{n=1,2,\dots}$ *vérifie* $\|\mathcal{T}_n\|_n \le K$ *et* $\|(\mathcal{T}_n)^{-1}\|_n \le K_1$ *, il existe un isomorphisme* T *tel que* $\|T\| \le K$ *et* $\|T^{-1}\| \le K_1$ *, vérifiant :*

$$t_{ij} = (Te_j, e_i) \quad , \quad i,j \in \mathbb{N}.$$

Démonstration

1) La suite de matrices emboitées $\mathcal{T}_n$ définit un opérateur T si et seulement si $\|\mathcal{T}_n\|_n \leq K$. En effet :

a) Supposons qu'il existe T borné tel que $t_{ij} = (Te_j, e_i)$, alors $\|\mathcal{T}_n\|_n = \|T_n\| \leq \|\pi_n\|^2 \|T\| \leq \|T\|$.

b) Supposons que $\|\mathcal{T}_n\|_n \leq K$.

Soit x quelconque dans H et $x_m = \pi_m x$

Pour tout $n > m$, $y_n = T_n x_m = \sum_{i=1}^{n} (\sum_{j=1}^{m} t_{ij} x_j) e_i$ et $\|T_n x_m\| \leq K\|x_m\|$.

Donc $y_n \xrightarrow[n\to\infty]{} y = \sum_{1}^{\infty} (\sum_{j=1}^{m} t_{ij} x_j) e_i = T x_m$ par définition.

L'ensemble des éléments de la forme x_m est dense dans H.

La suite $\{T_n x_m\}$ converge sur un ensemble dense et $\|T_n\| \leq K$ donc $T_n x$ converge pour tout x de H vers Tx.

A la limite, $\|Tx\| \leq K\|x\|$ et $\|T\| \leq K$.

2) Supposons que $\|\mathcal{T}_n\|_n \leq K$ et $\|(\mathcal{T}_n)^{-1}\|_n \leq K_1$.

D'après le 1) il existe T borné défini par $t_{ij} = (Te_j, e_i)$.

Pour tout x dans H, on pose $y = Tx$ dans H

et $\xi_n = \mathcal{T}_n^{-1} r_n y$ dans $\mathbb{C}^n$.

Montrons que $\|p_n \xi_n - x\| \to 0$.

$$\|p_n \xi_n - x\| \leq \|p_n \xi_n - \pi_n x\| + \underset{\downarrow \atop 0}{\|(\pi_n - 1)x\|}$$

$$\|p_n \xi_n - p_n r_n x\| = \|\xi_n - r_n x\|_n$$

$$\xi_n - r_n x = \mathcal{T}_n^{-1} . \mathcal{T}_n(\xi_n - r_n x)$$

$$\mathcal{T}_n(\xi_n - r_n x) = r_n y - \mathcal{T}_n r_n x$$

$$p_n(r_n y - \mathcal{T}_n r_n x) = \pi_n y - (y - Tx) - T_n x$$

Donc :

$$\|p_n \xi_n - \pi_n x\| \leq K_1 (\|(\pi_n - 1)y\| + \|(T - T_n)x\|)$$

$$\to 0.$$

Par hypothèse : $\|\mathcal{T}_n \xi\|_n \geq \frac{1}{K_1} \|\xi\|_n$ pour $\xi \in \mathbb{C}^n$.

En ξ_n : $\|p_n \mathcal{T}_n \xi_n\| \geq \frac{1}{K_1} \|p_n \xi_n\|$ quel que soit n.

A la limite : $\|Tx\| \geq \frac{1}{K_1} \|x\|$.

Alors Tx = 0 implique x = 0 : Ker T = {0} et T^{-1} existe :

$$\|T^{-1}x\| \leq K_1 \|x\| \quad , \quad \|T^{-1}\| \leq K_1 \quad .$$

3 - Cas d'un opérateur triangulaire

3.1 - Définitions

Soit ρ_n la projection orthogonale sur le complémentaire orthogonal de $\{e_1, \ldots, e_n\}$. $\pi_n + \rho_n = 1$.

T sera dit [19] :

- triangulaire supérieur (relativement à la base $\{e_i\}$) si :

$$T\pi_n = \pi_n T \pi_n \quad , \quad n \in \mathbb{N}$$

- <u>triangulaire inférieur</u> (relativement à $\{e_i\}$) si :

$$T\rho_n = \rho_n T\rho_n \quad , \quad n \in \mathbb{N}$$

- <u>diagonal</u> (relativement à $\{e_i\}$) si :

$$T\pi_n = \pi_n T\pi_n \text{ et } T\rho_n = \rho_n T\rho_n \quad , \quad n \in \mathbb{N}.$$

En général, nous noterons, par analogie avec les matrices finies,
- R un opérateur triangulaire supérieur,
- L un opérateur triangulaire inférieur tel que $(Le_j, e_i) = 1$.

3.2 - CNS pour que R soit un isomorphisme

Soit R un isomorphisme triangulaire. On pose $S = R^{-1}$, on note :

$\mathcal{R}_n$ la matrice d'élément général (Re_j, e_i), $1 \leq i,j \leq n$

$\mathcal{S}_n$ la matrice d'élément général (Se_j, e_i), $1 \leq i,j \leq n$.

D'après la démonstration du théorème 1 :

$$\|R\| \leq K \text{ et } \|S\| \leq K_1 \iff \|\mathcal{R}_n\|_n \leq K \text{ et } \|\mathcal{S}_n\|_n \leq K_1.$$

Or $\mathcal{S}_n = (\mathcal{R}_n)^{-1}$.

En effet, $\mathcal{R}_n$ et $\mathcal{S}_n$ sont les matrices représentant $r_n R p_n$ et $r_n S p_n$.

$1 = RS$ implique $\pi_n = (\pi_n R)(S\pi_n) = (\pi_n R)(\pi_n S\pi_n)$

soit : $p_n r_n = p_n (r_n R p_n)(r_n S p_n) r_n$

ou encore : $I_n = \mathcal{R}_n . \mathcal{S}_n$, où I_n est la matrice identité de $\mathbb{C}^n$.

Donc $\|(\mathcal{R}_n)^{-1}\|_n \leq K_1$.

D'où le :

THEOREME 2

Pour qu'un opérateur R *triangulaire soit un isomorphisme, il faut et il suffit que la suite de matrices emboitées* $\{R_n\}_{n=1,2,\dots}$ *, définies par* $r_{ij} = (Re_j, e_i)$, $1 \le i,j \le n$, *vérifie*

$$\|R_n\|_n \le K \quad \text{et} \quad \|(R_n)^{-1}\|_n \le K_1 .$$

Ce résultat a été énoncé dans ℓ^2 par J.C. Miellou [38].

II - DECOMPOSITIONS DE GAUSS ET DE CHOLESKI

Nous cherchons des conditions suffisantes sur l'isomorphisme T, pour que T = LR, produit de deux isomorphismes triangulaires.

1 - Décomposition de Gauss

1.1 - Ecriture formelle

Les suites de matrices emboitées $\{L_n\}$ et $\{R_n\}$ peuvent se définir par récurrence, à partir de la suite $\{T_n\}$. Les T_n sont supposées inversibles.
On pose :

$$T_n = \left(\begin{array}{c|c} T_{n-1} & \tau_n \\ \hline t_n^T & t_{nn} \end{array}\right), L_n = \left(\begin{array}{c|c} L_{n-1} & 0 \\ \hline \ell_n^T & 1 \end{array}\right), R_n = \left(\begin{array}{c|c} R_{n-1} & r_n \\ \hline 0 & r_{nn} \end{array}\right)$$

$T_n = L_n R_n$ équivaut à :

$$(1) \quad \begin{cases} T_{n-1} = L_{n-1} \cdot R_{n-1} \\ \tau_n = L_{n-1}\, r_n \\ t_n^T = \ell_n^T R_{n-1} \\ t_{nn} = \ell_n^T\, r_n + r_{nn} \end{cases}$$

ou encore :

$$(2) \quad \begin{cases} T_{n-1} = L_{n-1} \cdot R_{n-1} \\ r_n = (L_{n-1})^{-1} \tau_n \\ \ell_n^T = t_n^T (R_{n-1})^{-1} \\ r_{nn} = -\ell_n^T r_n + t_{nn}. \end{cases}$$

Puisque T_n est inversible, det T_n = det $R_n \neq 0$ donc $r_{nn} \neq 0$.
R_n est inversible.

Les formules (2) permettent donc de calculer L_n et R_n de façon unique, à partir de $(L_{n-1})^{-1}$, $(R_{n-1})^{-1}$ et T_n.

1.2 - C.S. pour que L et R soient des isomorphismes

Nous démontrons un résultat énoncé dans ℓ^2 par N. GASTINEL [13] :

THEOREME 3

T *vérifie les conditions :*

$$(C)\quad \begin{cases} C_i) & T \text{ est un isomorphisme tel que } \|T_n\|_n \leq K \text{ et } \|(T_n)^{-1}\| \leq K_1 \\ C_{ii}) & T = D+U, \text{ où } D \text{ est un opérateur diagonal borné et } U \text{ est un opérateur compact de } \mathfrak{S}_2, \end{cases}$$

alors T *admet la décomposition de Gauss* T = LR , *où* L *et* R *sont des isomorphismes et* L-1 *appartient à* $\mathfrak{S}_2$.

Démonstration :

Nous avons vu au chapitre V, § II.3, que :

$$U \in \mathfrak{S}_2 \Leftrightarrow N_2(U) = \left(\sum_{i,j} |(Ue_j,e_i)|^2\right)^{1/2} < \infty$$

Par analogie avec les matrices finies (norme de Frobenius), nous noterons la norme de Schmidt (ou double norme) de U : $N_2(U) = \||U|\|$.

L est un isomorphisme si et seulement si L_n et $(L_n)^{-1}$ sont uniformément bornés en n.

On pose :

$$L_n^{-1} = \left(\begin{array}{c|c} L_{n-1}^{-1} & 0 \\ \hline \nu_n^T & 1 \end{array}\right), \qquad R_n^{-1} = \left(\begin{array}{c|c} R_{n-1}^{-1} & \rho_n \\ \hline 0 & 1/r_{nn} \end{array}\right)$$

$L_n^{-1}L_n = I_n$ implique :

$$\nu_n^T L_{n-1} + \ell_n^T = 0$$

$$\nu_n^T = -\ell_n^T L_{n-1}^{-1} = -t_n^T R_{n-1}^{-1} L_{n-1}^{-1} = -t_n^T T_{n-1}^{-1} .$$

$R_n^{-1}R_n = I_n$ implique :

$$R_{n-1}^{-1} r_n + r_{nn}\rho_n = 0$$

$$\rho_n = \frac{1}{r_{nn}} R_{n-1}^{-1} r_n = \frac{1}{r_{nn}} T_{n-1}^{-1} \tau_n .$$

Nous étudions $|||L_n - I_n|||_n$ et $|||L_n^{-1} - I_n|||_n$.

$$|||L_n - I_n|||_n = \left(\sum_{i=2}^{n} \|\ell_i\|_i^2\right)^{1/2} \leq \left(\max_{i=1,\dots,n-1} \|R_i^{-1}\|_i\right)\left(\sum_{i=2}^{n} \|t_i\|_i^2\right)^{1/2}$$

$$\leq \left(\max_{i=2,\dots,n} \|R_{i-1}^{-1}\|_i\right) |||U|||$$

$$|||L_n^{-1} - I_n|||_n = \left(\sum_{i=2}^{n} \|\nu_i\|_i^2\right)^{1/2} \leq \left(\max_{i=1,\dots,n-1} \|T_i^{-1}\|_i\right)\left(\sum_{i=2}^{n} \|t_i\|_i^2\right)^{1/2}$$

$$\leq K_1 \; |||U|||.$$

Montrons que les R_n^{-1} sont uniformément bornés en n.
On appelle D_n la diagonale de R_n : $D_n = (r_{ii})$. Celle de R_n^{-1} est D_n^{-1}.

D'autre part $T_n^{-1} e_n = \left(\frac{\rho_n}{1/r_{nn}}\right)$
donc

$$\|T_n^{-1} e_n\|_n = \left(\|\rho_n\|_n^2 + \frac{1}{|r_{nn}|^2}\right)^{1/2} \leq K_1$$

et

$$\frac{1}{|r_{ii}|} \leq K_1 \quad , \quad i = 1,\dots,n$$

$$\|D_n^{-1}\|_n = \max_{i=1,\dots,n} \frac{1}{|r_{ii}|} \leq K_1$$

$$|||R_n^{-1} - D_n^{-1}|||_n = \left(\sum_{i=2}^{n} \|\rho_i\|_i^2\right)^{1/2} \leq \max_{i=2,\dots,n} \frac{1}{|r_{ii}|} \cdot K_1 \left(\sum_{i=2}^{n} \|\tau_i\|_i^2\right)^{1/2}$$

$$\leq K_1^2 \; |||U|||$$

Donc $\|R_n^{-1}\|_n \leq K_1 \; (1 + K_1 |||U|||)$.

On a ainsi déterminé des bornes uniformes pour $\|\mathcal{L}_n\|_n$ et $\|\mathcal{L}_n^{-1}\|_n$.
L est un isomorphisme ainsi que $R = L^{-1}T$.

Notons que nous avons démontré au passage que :

1) $$|||L-1||| = (\sum_{i=2}^{\infty} \|\ell_i\|_i^2)^{1/2} \leq K_1(1+K_1|||U|||)|||U||| \quad : L-1 \in \mathfrak{S}_2$$

2) $$|||L^{-1}-1||| = (\sum_{i=2}^{\infty} \|\nu_i\|_i^2)^{1/2} \leq K_1 \; |||U||| \quad : L^{-1}-1 \in \mathfrak{S}_2$$

Soit D_1 l'opérateur diagonal tel que $\begin{cases} (D_1 e_j, e_i) = 0 \text{ si } i \neq j \\ (D_1 e_i, e_i) = r_{ii} \end{cases}$

3) $$|||R-D_1||| = (\sum_{i=2}^{\infty} \|r_i\|_i^2)^{1/2} \leq (1+K_1|||U|||)(|||U||| \quad : R-D_1 \in \mathfrak{S}_2$$

4) $$|||R^{-1}-D_1^{-1}||| = (\sum_{i=2}^{\infty} \|\rho_i\|_i^2)^{1/2} \leq K_1^2 \; |||U||| \qquad : R^{-1}-D_1^{-1} \in \mathfrak{S}_2$$

<u>Remarque</u> - Göhberg et Krein [16] ont donné des conditions suffisantes pour que $T = 1+U$, U compact, se factorise en LR, la notion d'opérateur triangulaire étant relative à une suite <u>continue</u> de projections.

1.3 - Nous considérons maintenant $T' = RL$. C'est un isomorphisme. Vérifie-t-il les conditions (C) ?

C_i) : Soit $\{\mathcal{T}'_n\}$ la suite de matrices emboitées définie par T'

$$\|\mathcal{T}'_n\|_n \leq \|T'\|$$

$$\pi_n T' \pi_n = (\pi_n R \pi_n)(\pi_n L \pi_n) + \pi_n R(1-\pi_n)(1-\pi_n) L \pi_n$$

Soit $S = R-D_1$

$$|||\pi_n R(1-\pi_n)||| \leq |||S-\pi_n S\pi_n||| \xrightarrow[n\to\infty]{} 0 \quad \text{puisque } S \in \mathfrak{S}_2$$

De même, $M = L-1$,

$$|||(1-\pi_n)L\pi_n||| \leq |||M-\pi_n M\pi_n||| \xrightarrow[n\to\infty]{} 0 \quad \text{puisque } M \in \mathfrak{S}_2$$

Si nous posons $\tilde{R}_n = \pi_n R(1-\pi_n)$ et $\tilde{L}_n = (1-\pi_n)L\pi_n$,

$$T'_n = R_n L_n + \tilde{R}_n \tilde{L}_n \quad \text{avec} \quad |||\tilde{R}_n||| \xrightarrow[n\to\infty]{} 0,$$

$$\text{et} \quad |||\tilde{L}_n||| \xrightarrow[n\to\infty]{} 0.$$

Les matrices associées dans $\mathbb{C}^n$, vérifient :

$$T'_n = R_n L_n + \tilde{R}_n \tilde{L}_n$$

$$= R_n L_n (I_n + L_n^{-1} R_n^{-1} \tilde{R}_n \tilde{L}_n)$$

$$\forall \alpha < 1, \ \exists N : n > N \Rightarrow \|L_n^{-1} R_n^{-1} \tilde{R}_n \tilde{L}_n\|_n < \alpha$$

Alors T'_n est inversible et

$$\|T'^{-1}_n\|_n < \frac{1}{1-\alpha} \|L_n^{-1}\|_n \|R_n^{-1}\|_n.$$

Donc T' est tel que la suite $\{T'_n\}$ vérifie, à partir d'un certain rang au moins, $\|T'_n\|_n \leq K'$ et $\|(T'_n)^{-1}\|_n \leq K'_1$.

C_{ii}) : $T' = RL = (D_1+S)(1+M) = D_1+S+D_1M+SM = D_1+U'$

et $\|\,|U'|\,\| < \infty$, donc $U' \in \mathfrak{S}_2$.

D'où la :

Proposition 1

T vérifie les conditions (C). Alors T = LR *et* T' = RL *vérifie aussi (C), pour* n *assez grand au moins.*

1.4 - Montrons que si T vérifie (C), il en est de même de T^{-1}.
Si T = LR, $T^{-1} = R^{-1}L^{-1}$, nous avons vu que $L^{-1}-1 = M_1 \in \mathfrak{S}_2$ ainsi que $R^{-1}-D_1^{-1} = S_1$.
Pour C_i), le raisonnement fait pour T' = RL s'applique donc pour n > N.
Quant à C_{ii}) :

$$T^{-1} = (D_1^{-1}+S_1)(1+M_1) = D_1^{-1}+S_1+D_1^{-1}M_1+S_1M_1$$

avec

$$\|\,|S_1+D_1^{-1}M_1+S_1M_1|\,\| < \infty.$$

D'où la :

Proposition 2

Si T *vérifie les conditions (C),* T^{-1} *les vérifie aussi, pour* n *assez grand au moins.*

Remarque - Pour n < N, la situation est analogue au cas des matrices finies. Si une matrice A admet la décomposition de Gauss A = LR, il n'est pas sûr que A' = RL en admette une.
Néanmoins il est facile de montrer que :

l'ensemble des isomorphismes admettant une décomposition de Gauss est un ouvert de l'ensemble des isomorphismes, pour la double norme.

Si $T' = T+U$, $U \in \mathfrak{S}_2$

$$= T(1+T^{-1}U) = L(1+RT^{-1}UR^{-1})R = L(1+V)R$$

et

$$|||V||| \leq C\ |||U|||.$$

Donc $1+V$ vérifie (C) pour $|||U|||$ assez petit, $V = \ell r$ et $T' = (L\ell)(rR)$.

2 - Décomposition de Choleski

T est un isomorphisme autoadjoint défini positif si et seulement s'il existe $M > m > 0$ tels que :

$$0 < m\|x\|^2 \leq (Tx,x) \leq M\|x\|^2, \quad \forall x \in H.$$

Peut-on définir la décomposition de Choleski $T = R^*R$?

Comme précédemment la suite $\{R_n\}$ est définie par récurrence à partir de $T_n = R_n^H R_n$. Elle est unique si on choisit $r_{ii} > 0$

$$m\|\pi_n x\|^2 \leq (T\pi_n x,\pi_n x) \leq M\|\pi_n x\|^2, \quad \forall x \in H$$

Pour tout ξ de $\mathbb{C}^n$, $(T_n\xi,\xi)_n = \|R_n\xi\|^2_n$

et

$$\sqrt{m}\|\xi\|_n \leq \|R_n\xi\|_n \leq \sqrt{M}\ \|\xi\|_n.$$

R est un isomorphisme.

THEOREME 4

Tout isomorphisme autoadjoint défini positif, admet la décomposition de Choleski $T = R^*R$.

III - ALGORITHME LR

T est un opérateur vérifiant (C).

1 - Algorithme

On forme la suite d'opérateurs semblables :

$$T_1 = T = L_1 R_1,$$

$$T_2 = R_1 L_1 = L_2 R_2,$$

$$T_s = R_{s-1} L_{s-1} = L_s R_s,$$

D'après la proposition 1, les conditions (C) sont vérifiées par T_1, donc par T_s pour s assez grand, $s = 2,3,\ldots$

Comment se comporte T_s lorsque $s \to \infty$?

$$T_{s+1} = L_s^{-1} T_s L_s = (L_1 \ldots L_s)^{-1} T (L_1 \ldots L_s)$$

D'autre part

$$(L_1 \ldots L_s)(R_s \ldots R_1) = T(L_1 \ldots L_{s-1})(R_{s-1} \ldots R_1) = T^s.$$

D'où :

1) T_s converge si et seulement si $L_1 \ldots L_s$ converge.

2) $L_1 \ldots L_s$ est l'opérateur triangulaire inférieur dans la décomposition de Gauss de T^s.

2 - Convergence de l'algorithme

Le schéma de la démonstration est celui adopté au chapitre I.
On suppose que :

i) T admet un spectre purement ponctuel, les valeurs propres sont simples, on peut les ordonner par valeur absolue décroissante :

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_i| > \ldots$$

ii) Les vecteurs propres φ_i de T forment une base de H. (Nous en verrons des conditions suffisantes au § III.3).

Soit $X : e_i \mapsto \varphi_i$, c'est un isomorphisme de H.
On suppose que X vérifie (C). Alors $X^{-1} = Y$ vérifie (C) aussi.

$X = L_X.R_X$ et $Y = L_Y.R_Y$ sont les décompositions de Gauss de X et Y.

Soit Δ l'opérateur diagonal défini par $\begin{cases} (\Delta e_j, e_i) = 0 & \text{pour } i \neq j \\ (\Delta e_i, e_i) = \lambda_i \end{cases}$

$$T = X\Delta X^{-1}$$

$$T^s = X\Delta^s X^{-1} = L_X\left[R_X(\Delta^s L_Y \Delta^{-s})R_X^{-1}\right]R_X\Delta^s R_Y.$$

On pose :

$$U_s = D^s L_Y D^{-s}$$

$$(L_Y e_j, e_i) = \ell_{ij}^Y \ , \quad i,j \in \mathbb{N} \quad \text{et} \quad \ell_{ij}^Y = 0 \text{ pour } i < j$$

$$(U_s e_j, e_i) = \ell_{ij}^Y \left(\frac{\lambda_i}{\lambda_j}\right)^s \ , \quad i,j \in \mathbb{N} \text{ et } \left|\frac{\lambda_i}{\lambda_j}\right| < 1 \text{ pour } i > j$$

Alors

$$\|\,|U_s - 1|\,\| = \left(\sum_{i=2}^{\infty} \sum_{j=1}^{i-1} |\ell_{ij}^Y|^2 \left|\frac{\lambda_i}{\lambda_j}\right|^{2s}\right)^{1/2}.$$

Montrons que $\quad \|\,|U_s-1|\,\| \xrightarrow[s\to\infty]{} 0$

a) $\forall\varepsilon > 0,\ \exists N$:

$$\sum_{i>N}\sum_{j=1}^{i-1} |\ell^Y_{ij}|^2 \left|\frac{\lambda_i}{\lambda_j}\right|^{2s} < \sum_{i>N}\sum_{j=1}^{i-1} |\ell^Y_{ij}|^2 < \varepsilon^2$$

puisque $\|\,|L_Y-1|\,\| < \infty$.

b) Soit N ainsi fixé :

$$\sum_{i=2}^{N}\sum_{j=1}^{i-1} |\ell^Y_{ij}|^2 \left|\frac{\lambda_i}{\lambda_j}\right|^{2s} < \sum_{i=2}^{N} \max_{j=1,\dots,i-1} \left|\frac{\lambda_i}{\lambda_j}\right|^{2s} \sum_{j=1}^{i-1} |\ell^Y_{ij}|^2$$

Or

$$\max_{j=1,\dots,i-1} \left|\frac{\lambda_i}{\lambda_j}\right| = \left|\frac{\lambda_i}{\lambda_{i-1}}\right|$$

$$\forall\varepsilon > 0,\ \exists S : s > S \Rightarrow \max_{i=2,\dots,N} \left|\frac{\lambda_i}{\lambda_{i-1}}\right|^{2s} < \varepsilon^2$$

Alors :

$$\sum_{i=2}^{N}\sum_{j=1}^{i-1} |\ell^Y_{ij}|^2 \left|\frac{\lambda_i}{\lambda_j}\right|^{2s} < \varepsilon^2 \sum_{i=2}^{N}\sum_{j=1}^{i-1} |\ell^Y_{ij}|^2 < \varepsilon^2 \|\,|L_Y-1|\,\|^2$$

En conclusion :

$$\forall\varepsilon > 0,\ \exists S : s > S \Rightarrow \|\,|U_s-1|\,\| < \varepsilon(1+\|\,|L_Y-1|\,\|^2)^{1/2}$$

Soit $V_s = R_X U_s R_X^{-1}$, $V_s-1 = R_X(V_s-1)R_X^{-1}$ et $\|\,|V_s-1|\,\| \xrightarrow[s\to\infty]{} 0$.

Il est alors facile de voir que V_s vérifie (C) :

$V_s = \ell_s r_s$, et il existe deux constantes C, C_1 telles que

$$\|\,|\ell_s-1|\,\| < C\ \|\,|V_s-1|\,\| \quad \text{et} \quad \|\,|\ell_s^{-1}-1|\,\| < C_1\ \|\,|V_s-1|\,\|.$$

Donc

$$T^s = (L_X \ell_s)(r_s R_X D^s R_Y) \text{ est la décomposition de Gauss de } T^s$$

$$= (L_1 \dots L_s)(R_s \dots R_1)$$

et

$$\| |(L_1 \dots L_s) - L_X| \| \leq \|L_X\| \; \| |\ell_s - 1| \| \xrightarrow[s\to\infty]{} 0$$

Revenons à $T_{s+1} = \ell_s^{-1} L_X^{-1} T L_X \ell_s = \ell_s^{-1} R_X \Delta R_X^{-1} \ell_s$

$$\ell_s^{-1} R_X \Delta R_X^{-1} \ell_s - R_X \Delta R_X^{-1} = \ell_s^{-1} R_X \Delta R_X^{-1} (\ell_s - 1) + (\ell_s^{-1} - 1) R_X \Delta R_X^{-1}$$

Pour s assez grand, $\|\ell_s^{-1}\| < 2$, donc :

$$\| |T_{s+1} - R_X \Delta R_X^{-1}| \| \leq (2C + C') \|R_X \Delta R_X^{-1}\| \; \| |V_s - 1| \| \xrightarrow[s\to\infty]{} 0$$

L'opérateur limite $T_\infty = R_X \Delta R_X^{-1}$ est un opérateur triangulaire supérieur tel que : $(T_\infty e_i, e_i) = \lambda_i$, $i \in \mathbb{N}$.

Nous énonçons le :

THEOREME 5

T est un opérateur vérifiant (C). On suppose en outre que :

i) le spectre de T est formé de valeurs propres distinctes, ordonnées :

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_i| > \ldots$$

ii) les vecteurs propres φ_i de T forment une base de H

iii) l'isomorphisme $X : e_i \mapsto \varphi_i$ vérifie (C).

On peut alors définir la suite d'opérateurs semblables :

$$\begin{cases} T_1 = T = L_1 R_1 , \\ T_s = R_{s-1} L_{s-1} = L_s R_s \quad , \quad s = 2,3,\ldots \end{cases}$$

Lorsque s tend vers l'infini, T_s converge en double norme vers T_∞, opérateur triangulaire supérieur tel que $(T_\infty e_i, e_i) = \lambda_i$.

3 - C.S pour que les vecteurs propres de T forment une base

Supposons T compact non autoadjoint :

$$T_r = \frac{T+T^*}{2} = \Re e T \quad , \quad T_i = \frac{T-T^*}{2\,i} = \Im m T,$$

$T = T_r + iT_i$, T_r et T_i sont des opérateurs autoadjoints.

On peut citer les 2 conditions suffisantes suivantes :

(I) Lidskii [34] : T_r et T_i définis positifs appartiennent à $\mathfrak{S}_2$.
T est alors un opérateur de $\mathfrak{S}_2$.

(II) : Sahnovitch [58] : T_r et T_i sont définis positifs, T_i appartient à $\mathfrak{S}_2$.

Remarque - Un opérateur T tel que $\mathcal{R}eT$ défini positif est dit accretif [25].

Soit U un opérateur compact de $\mathfrak{S}_2$ vérifient (I), alors T = 1+U admet un spectre purement ponctuel, les valeurs propres s'accumulent en 1 et les vecteurs propres forment une base de H.

4 - Application au calcul de T_∞

Etant donné un opérateur T = D+U vérifiant (C), on veut calculer une approximation de l'opérateur triangulaire inférieur $T_\infty = \lim_s T_s$. Pour cela on considère $T_N = \pi_N T \pi_N$, où N est choisi tel que $|||U-\pi_N U \pi_N||| = \varepsilon$; ε nombre positif fixé.

On applique l'algorithme LR à la matrice A_N d'ordre N représentant T_N dans la base canonique de $\mathbb{C}^N$:

$$A_N = ((Te_j, e_i)_{1 \le i,j \le N})$$

Soit T_s (resp. $(T_N)_s$) le $s^{\text{è}}$ itéré de T (resp. T_N) par LR.
Quelle est l'erreur commise en approchant $\pi_N T_s \pi_N$ par $(T_N)_s$?

Il est facile de voir, à l'aide des décompositions matricielles, que :

$$|||\pi_N T_2 \pi_N - (T_N)_2||| \le 2\ \varepsilon^2$$

Posons $T_2 = D_2 + U_2$.
On peut majorer $|||U_2 - \pi_N U_2 \pi_N|||$ en fonction de ε (cf. § 1), et l'on déduit une majoration de $|||\pi_N T_3 \pi_N - (T_N)_3|||$ en fonction de ε comme ci-dessus.

Une majoration de $|||\pi_N T_s \pi_N - (T_N)_s|||$ en découle.

IV - DECOMPOSITION DE SCHMIDT, ALGORITHMES QR ET CHOLESKI

1 - Décomposition de Schmidt

Soit T un isomorphisme de H, $\{e_i\}$ une base orthonormée de H. On note $t_i = Te_i$, les $\{t_i\}$ sont indépendants puisque T est injectif, et forment une base de X puisque ImT = H.

D'après le théorème d'orthogonalisation de Schmidt ([72] p. 88) on peut définir par récurrence les suites :

$$u_1 = t_1 \neq 0 \quad , \qquad q_1 = \frac{u_1}{\|u_1\|} \quad ,$$

$$u_2 = t_2 - (t_2, q_1) q_1 \quad , \qquad q_2 = \frac{u_2}{\|u_2\|} \quad ,$$

$$\text{------} \qquad\qquad \text{-------}$$

$$u_{n+1} = t_{n+1} - \sum_{i=1}^{n} (t_{n+1}, q_i) q_i , \qquad q_{n+1} = \frac{u_{n+1}}{\|u_{n+1}\|} ,$$

telles que les $\{q_i\}$ forment une base orthonormée de H.
Pour tout n, t_n est combinaison linéaire de $q_1, \ldots, q_n$:

$$t_n = \sum_{i=1}^{n} r_{in} \, q_i$$

Soit Q l'opérateur unitaire défini par : $e_i \mapsto q_i$.
La relation ci-dessus équivaut à la décomposition de Schmidt :

$$T = Q \, R$$

où $R = TQ^*$ est un isomorphisme triangulaire supérieur.
Par le choix fait dans la définition des q_i : $r_{ii} = \|u_i\| > 0$, la décomposition est unique.

THEOREME 6

Tout isomorphsime T *admet la décomposition de Schmidt :*

$$T = QR.$$

Remarque - T étant un isomorphisme, T^*T est un isomorphisme autoadjoint défini positif. $T = QR$ implique $T^*T = R^*Q^*QR = R^*R$. On retrouve la décomposition de Choleski du théorème 4.

Nous aurons besoin dans la suite de la :

Proposition 3

Soit T *un isomorphisme tel que* $t_{ij} = (Te_j, e_i)$ *, admettant la décomposition de Schmidt :* $T = QR$ *et vérifiant* $\|T\| \leq K_1$ *ainsi que* $\|T^{-1}\| \leq K_2$.

Si $n(T) = (\sum_{i>j} |t_{ij}|^2)^{1/2} < \frac{1}{3K_2}$ *, il existe une consante C et un opérateur diagonal unitaire* d, *tels que :*

$$\||Qd-1|\| \leq C\ n(T).$$

Démonstration

a) $Q = TR^{-1}$

Posons $q_{ij} = (Qe_j, e_i)$ et $\rho_{ij} = (R^{-1}e_j, e_i)$

Pour $i > j$ $\quad q_{ij} = \sum_{k=1}^{j} t_{ik}\rho_{kj}$ ne fait intervenir que les t_{ik}, $i > k$.

Donc $(\sum_{i>j} |q_{ij}|^2)^{1/2} \leq \|R^{-1}\|\ n(T) \leq C_1\ n(T)$ avec $C_1 = K_2$.

b) Soit d l'opérateur diagonal unitaire défini par :

$$(de_i,e_i) = \frac{|q_{ii}|}{q_{ii}}$$

Alors si $Q' = Qd$, $q'_{ii} = |q_{ii}| > 0$.

Remarquons que la multiplication par d ne change pas la double norme.
Soit la décomposition $Q' = Q_i+Q_s+\Delta$, avec :

- Q_i, opérateur triangulaire inférieur, défini par $(Q_ie_j,e_i) = (Q'e_j,e_i)$ pour $i > j$, sinon 0.

- Q_s, opérateur triangulaire supérieur, défini par $(Q_se_j,e_i) = (Q'e_j,e_i)$ pour $i > j$, sinon 0.

- Δ, opérateur diagonal, défini par $(\Delta e_i,e_i) = |q_{ii}|$, sinon 0.

Posons $K = \Delta+Q_s$.

$$1 = Q'^*.Q' = (Q_i^*+K^*)(Q_i+K) = Q_i^*Q_i+K^*Q_i+Q_i^*K+K^*K$$

$$|||1-K^*K||| \leq |||Q_i|||(|||Q_i|||+2|||K|||) \leq C_2\ n(T) \quad \text{avec } C_2 \leq 3C_1$$

Si $1-C_2n(T) > 0$, soit $n(T) < \frac{1}{3K_2}$, il est clair que K^*K est un isomorphisme dont K est l'opérateur triangulaire supérieur de Choleski.
D'après le théorème 3, $|||1-K^*K||| \leq C_2\ n(T)$ implique $|||K-\Delta||| \leq C_3\ n(T)$.

c) Soit $\tilde{Q} = Q_i+Q_s$: $|||\tilde{Q}||| \leq (C_1+C_3)n(T)$.

$$1 = Q'^*Q' = (\Delta+\tilde{Q}^*)(\Delta+\tilde{Q}) = \Delta^2+\tilde{Q}^*\Delta+\Delta\tilde{Q}+\tilde{Q}^*\tilde{Q}$$

$$|||1-\Delta^2||| \leq |||\tilde{Q}|||(2|||\Delta|||+|||\tilde{Q}|||) \leq C_4\ n(T).$$

Or

$$0 \leq |q_{ii}|^2 \leq |q_{ii}| \leq 1$$

et

$$1-|q_{ii}|^2 \geq 1-|q_{ii}|$$

Donc

$$|||1-\Delta||| \leq |||1-\Delta^2||| \leq C_4\ n(T)$$

En conclusion, $Q' = Qd$ vérifie :

$$|||Q'-1||| \leq |||\Delta-1||| + |||Q_i||| + |||Q_s||| \leq (C_1+C_3+C_4)n(T).$$

2 - Algorithme QR

On forme la suite d'opérateurs unitairement semblables :

$$T_1 = T = Q_1R_1 \ ,$$

$$T_2 = R_1Q_1 = Q_2R_2 \ ,$$

$$T_s = R_{s-1}Q_{s-1} = Q_sR_s \ ,$$

De la même façon qu'au § III.1,

1) T_s converge si et seulement si $Q_1 \ldots Q_s$ converge.

2) $Q_1 \ldots Q_s$ est l'opérateur unitaire de la décomposition de Schmidt de T^s.

Pour étudier la convergence, nous faisons les hypothèses i) ii) et iii) du théorème 5. Alors, avec les mêmes notations :

$$T^s = X\Delta^sY = Q_XR_X(\Delta^sL_Y\Delta^{-s})\Delta^sR_Y$$

et

$$|||U_s-1||| \xrightarrow[s\to\infty]{} 0.$$

U_s est un opérateur triangulaire inférieur qui vérifie donc l'hypothèse de la proposition 3 :

$$\begin{cases} n(U_s) = |||U_s-1||| \xrightarrow[s\to\infty]{} 0 \\ \|U_s\| \xrightarrow[s\to\infty]{} 1 \ , \ \|U_s^{-1}\| \xrightarrow[s\to\infty]{} 1 \end{cases} .$$

Si $U_s = q_s.r_s$ est la décomposition de Schmidt de U_s, il existe un opérateur diagonal unitaire d_s tel que :

$$|||q_s d_s - 1||| \leq C_1 \; n(U_s)$$

$W_s = R_X q_s d_s$ admet la décomposition de Schmidt : $\bar{q}_s.\bar{R}_s$.
et

$$W_s - R_X = R_X(q_s d_s - 1) \; .$$

Donc

$$n(W_s) \leq |||W_s - R_X||| \leq \|R_X\| \; |||q_s - 1||| \leq C_2 \; n(U_s)$$

On peut appliquer la proposition 3 à W_s : il existe un opérateur diagonal unitaire $\bar{d}_s$ tel que :

$$|||\bar{q}_s \bar{d}_s - 1||| \leq C_3 \; n(U_s).$$

Alors :

$$T^s = (Q_X \bar{q}_s)(\bar{R}_s d_s^* r_s \Delta^s R_Y) = Q^{(s)} R^{(s)}.$$

Pour pouvoir identifier cette décomposition à celle de Schmidt, il faut introduire l'opérateur diagonal unitaire $d^{(s)}$ défini par :

$$(d^{(s)} e_i, e_i) = \mathrm{sgn}(R^{(s)} e_i, e_i), \text{ sinon } 0.$$

Alors :

$$Q_1 \ldots Q_s = Q^{(s)} d^{(s)} = Q_X \bar{q}_s \bar{d}_s \bar{d}_s^* d^{(s)} \quad .$$

Et

$$|||Q_1 \ldots Q_s - Q_X \bar{d}_s^* d^{(s)}||| \leq \|Q_X\| \; |||q_s d_s - 1|||.$$

Revenant à $T_{s+1} = (Q_1 \ldots Q_s)^* T (Q_1 \ldots Q_s)$:

$$|||T_{s+1} - d^{(s)*} \bar{d}_s R_X \Delta R_X^{-1} \bar{d}_s^* d^{(s)}||| \leq C_4 \; |||U_s - 1||| \xrightarrow[s \to \infty]{} 0.$$

A un opérateur diagonal unitaire près, l'opérateur limite $T_\infty = R_X \Delta R_X^{-1}$ est un opérateur triangulaire supérieur tel que $(T_\infty e_i, e_i) = \lambda_i$, $i \in \mathbb{N}$.

SERVICE POLYCOPIE
MATHÉMATIQUES APPLIQUÉES
Université de GRENOBLE

D'où le :

THEOREME 7

T est un isomorphsime. On suppose en outre que

i) le spectre de T est formé de valeurs propres distinctes, ordonnées : λ_i,

ii) les vecteurs propres φ_i *de T forment une base de H.*

iii) l'isomorphisme $X : e_i \mapsto \varphi_i$ *vérifie (C).*

On peut alors définir la suite d'opérateurs unitairement semblables :

$$\begin{cases} T_1 = T = Q_1 R_1 , \\ T_s = R_{s-1} Q_{s-1} = Q_s R_s \quad , \quad s = 2,3... \end{cases}$$

Lorsque s tend vers l'infini, T_s *converge en double norme, à un opérateur diagonal unitaire près, vers* T_∞ *opérateur triangulaire supérieur tel que* $(T_\infty e_i, e_i) = \lambda_i$.

Lifchitz [35] et Sahnovich [58] ont mis en évidence une représentation triangulaire pour des opérateurs T tels que :

a) $\mathfrak{I}mT \in \mathfrak{S}_1$ (Lifchitz)

b) $\mathfrak{I}mT \in \mathfrak{S}_2$ (Sahnovich).

Dans le cas où T admet un spectre purement ponctuel et une base de vecteurs propres, nous venons de voir qu'il suffit que T soit un isomorphisme pour admettre une représentation triangulaire.

Dans le cas où T admet un spectre continu, il semble qu'on puisse utiliser le même schéma de démonstration en l'appliquant à des opérateurs triangulaires relativement à une suite continue de projections [16].

3 - Algorithme de Choleski

Soit T un isomorphisme autoadjoint défini positif. On définit la suite d'opérateurs semblables :

$$\begin{cases} T_1 = T = R_1^* R_1 , \\ T_s = R_{s-1} R_{s-1}^* = R_s^* R_s \quad , \quad s = 2,3,\ldots \end{cases}$$

Sout les hypothèses i) ii) du théorème 3, on peut définir $U = T^{1/2}$, isomorphisme autoadjoint défini positif, pour lequel on définit la suite des itérés dans l'algorithme QR :

$$\begin{cases} U_1 = \bar{Q}_1 . \bar{R}_1 \\ U_s = \bar{R}_{s-1} \bar{Q}_{s-1} = \bar{Q}_s . \bar{R}_s \end{cases}$$

Alors : $T_s = U_s^* U_s = \bar{R}_s^* . \bar{R}_s$, et en identifiant les décompositions de Choleski : $R_s = \bar{R}_s$.

Il y a donc convergence de l'algorithme de Choleski sous les mêmes hypothèses.

THEOREME 8

Si T *est un isomorphisme autoadjoint positif vérifiant les hypothèses i) ii) iii), la suite d'opérateurs semblables :*

$$\begin{cases} T_1 = T = R_1^* R_1 , \\ T_s = R_{s-1} R_{s-1}^* = R_s^* R_s , \qquad s = 2,3,\ldots \end{cases}$$

converge en double norme vers l'opérateur diagonal T_∞ *tel que* $(T_\infty e_i, e_i) = \lambda_i, i \in \mathbb{N}$.

4 - Application au calcul de T_∞

Le procédé employé au § III.4 pour calculer un majorant de l'erreur $\|\pi_N T_s \pi_N - (T_s)_N\|$ s'applique de la même façon lors des algorithmes QR ou Choleski.

BIBLIOGRAPHIE

1 ANSELONE P.M., MOORE R.H.

Approximate solution of integral and operator equations.
J. Math. Anal. Appl. 9, (1964), 268-277.

2 ANSELONE P.M., PALMER T.W.

Collectively compact sets of linear operators.
MRC Rep. 740, U. Wisconsin, Madison (1967).

3 ANSELONE P.M., PALMER T.W.

Spectral analysis of collectively compact, strongly convergent operator sequences.
MRC Rep. 741, U. Wisconsin, Madison (1967).

4 AUBIN J.P.

Approximation des espaces de distributions et des opérateurs différentiels (Thèse).
Soc. Math. France 12, (1967).

5 BAZLEY N.W.

Lower bounds for eigenvalues.
J. Math and Mech., 10, n° 2 (1961), 289-307.

6 CHATELIN-LABORDE F.

Perturbation d'une matrice hermitienne ou normale.
Application au calcul des valeurs propres.
Colloque d'Analyse Numérique Super-Besse (1970).

7 CHATELIN-LABORDE F.

Méthodes d'approximation des valeurs propres d'opérateurs linéaires dans un espace de Banach
I - Critère de stabilité
II- Bornes d'erreur
C.R.A.S. 271, (1970), 949-952
CRAS 271, (1970), 1006-1009.

8 DAVIS C.

The rotation of eigenvectors by a perturbation I, II
J. Math. Anal. Appl. 6, (1963), 159-173
J. Math. Anal. Appl. 11, (1965), 20-27

9 DAVIS C., KAHAN W.M.

The rotation of eigenvectors by a perturbation III.
SIAM J. Num. Anal. 7, n° 1 (1970), 1-46.

10 DOWNING A.C., HOUSEHOLDER A.S.

Some inverse characteristic value problem.
J.A.C.M. 3, (1956), 203-207.

11 GALLIGANI I.

A comparison of methods for computing the eigenvalues and eigenvectors of a matrix.
Euratom (1968).

12 GASTINEL N.

Analyse Numérique Linéaire
Hermann (1966).

13 GASTINEL N.

Décomposition LR pour une matrice infinie.
Séminaire Anal. Num. IMAG (1966).

14 GAVURIN M.K.

Approximate determination of proper values.
Amer. Math. Soc. Transl. 16, (1960), 385-388.

15 GOLUB G.H.

Bounds for eigenvalues of tridiagonal symmetric matrices computed with the LR method.
Math. Comp. J. 16, (1962), 438-445.

16 GOHBERG I.C., KREIN M.G.

On the factorization of operators in Hilbert spaces.
Usp. Mat. Nauk 13, (1958), 3
Sov. Math. Dokl. 3, (1962), 1578-1582.

17 GOHBERG I.C., MARKUS A.S.

Some relations between eigenvalues and matrix elements of linear operators.
Amer. Math. Soc. Transl. 52, (1966), 201-216.

18 HADELER K.P.

Ein inverses eigenwertproblem.
Lin. alg. Appl 1, (1968), 83-101.

19 HALMOS P.R.

Quasi triangular operators
Acta. Scient. Math. 29, (1968), 283-293.

20 HORN A.

Doubly stochastic matrices and the diagonal of a rotation matrix.
Amer. J. Math. 76, (1954), 620-630.

21 HOUSEHOLDER A.S.

The theory of matrices in numerical analysis.
Blaisdell (1964).

22 ISAACSON E., KELLER H.B.

Analysis of numerical methods .
John Wiley (1966).

23 KANTOROVITCH L.V., AKILOV G.P.

Functional analysis in normed spaces.
Pergamon (1964).

24 KATO T.

On the upper and lower bounds of eigenvalues.
J. Phys. Soc. Japan 4 (1949), 334-339.

25 KATO T.

Perturbation theory for linear operators.
Springer Verlag (1966).

26 KELLER H.B.

On the accuracy of finite difference approximation to the eigenvalues of differential and integral operators.
Num. Mat. 7, (1965), 412-419.

27 KRASNOSELSKII M.A.

Topological methods in the theory of non linear integral equations.
Pergamon (1964).

28 KRASNOSELSKII M.A., VAINIKKO G.M., etc,

Solutions approximatives d'équations opérationnelles.
Moscou (1969).

29 KURATOVSKI C.

Topologie.
Varsovie - Wroclaw (1948).

30 KY FAN

Note on circular disks containing the eigenvalues of a matrix.
Duke Math. J. 25 (1958), 441-445.

31 LABORDE F.
Recherche des valeurs propres d'une matrice : algorithme MR
CRAS 261, (1965), 871-874.

32 LABORDE F.
Opérateurs de ℓ^2 dans ℓ^2. Extension des algorithmes LR et QR
Congrès de l'AFIRO. Lille (1966).

33 LABORDE F.
Sur un problème inverse d'un problème de valeurs propres.
CRAS 268, (1969), 153-156.

34 LIDSKII V.B.
Conditions for completeness of a system of root subspaces for non selfadjoint operators with discrete spectrum.
Amer. Math. Soc. Transl. 34, (1963), 241-282.

35 LIFCHITZ S.M.
On the spectral resolution of linear non selfadjoint operators
Amer. Math. Soc. Transl. 5, (1957), 67-114.

36 MEDVEDEV V.A.
On the convergence of the Bubnov-Galerkin method.
J. App. Math. Mech. 27, (1964), 1769-1774.

37 MEDVEDEV V.A.
Convergence of the projection method in eigenvalue problem.
Sov. Math. Dokl. 5, (1964), 631-633.

38 MIELLOU J.C.
Construction d'opérateurs dans ℓ^2.
Séminaire d'Analyse Numérique IMAG (1966).

39 MIRSKY L.
The spread of a matrix.
Matematika 3, (1956), 127-130.

40 MIRSKY L.
Matrices with prescribed characteristic roots and diagonal elements.
J. London Math. Soc. 33, (1958), 14-21.

41 MISOVSKICH I.P.
On error bounds for approximate methods of estimation of eigenvalues of hermitian kernels.
Amer. Math. Soc. Transl. 35, (1964), 237-249.

42 MISOVSKICH I.P.
On error bounds for eigenvalues calculated by replacing the kernel by an approximated kernel.
Amer. Math. Soc. Transl. 35, (1964), 251-262.

43 MOLER C.B., PAYNE L.E.

Bounds for eigenvalues and eigenvectors of symmetric operators.
SIAM J. Num. Anal. 5, (1968), 64-70.

44 OLIVEIRA G.N. de

Note on an inverse characteristic value problem.
Num. Mat. 15, (1970), 345-347.

45 OSBORN J.E.

Approximation of the eigenvalues of non selfadjoint operators.
J. Math. Phys. 45, (1966), 391-401.

46 OSBORN J.E.

A note on a perturbation theorem for matrix eigenvalue problem.
Num. Mat. 13, (1969), 152-153.

47 PERFECT H.

A theorem of Hardy, Littlewood and Polya and some related results for infinite vectors.
Proc. Camb. Phil. Soc. 63, (1967), 1125-1134.

48 PLANCHARD J.

Sur l'approximation des valeurs propres et vecteurs propres des opérateurs hermitienscompacts.
Bull. EDF, ser. C, 1, (1968), 5-30.

49 PLANCHARD J.

Approximation du spectre des opérateurs différentiels elliptiques autoadjoints du 2ème ordre par la méthode des différences finies (schéma classique).
Bull. EDF, ser. C, 2, (1968), 93-100.

50 PLANCHARD J.

Résolution des équations de la diffusion des neutrons en régime stationnaire.
Bull. EDF, ser. C, 2, (1969), 117-140.

51 POLSKII N.I.

Sur la convergence de méthodes approximatives de l'analyse.
Ukr. Mat. J. 7, (1955), 56-70.

52 POLSKII N.I.

Projective methods in applied mathematics.
Sov. Math. Dokl. 3, (1962), 488-492

53 POZNIAK L.T.

Estimation de la vitesse de convergence d'une variante de la méthode des problèmes intermédiaires.
J. Vichl. Mat., mat. fis. 8, (1968), 1117-1126.

54 RELLICH F.
Perturbation theory of eigenvalue problems.
Gordon Breach (1969).

55 RIESZ F., NAGY B.
Leçons d'analyse fonctionnelle.
Akademiai Kiado (1955).

56 SAULEV V.K.
On the solution of problem of eigenvalues by the method of finite differences.
Amer. Math. Soc. Transl. 8, (1958), 257-287.

57 SAULEV V.K.
On estimation of the error in finding eigenfunctions by the method of finite difference.
Amer. Math. Soc. Transl. 35, (1964), 137-165.

58 SAHNOVITCH L.A.
A study of the "triangular form" of nonselfadjoint operators.
Amer. Math. Soc. Transl. 54, (1966), 85-90.

59 SCHATTEN R.
Norm ideals of completely continuous operators.
Springer Verlag (1960).

60 SCHWARZ J.T.
Perturbation of spectral operators and application.
Pac. J. 4, (1954), 415-458.

61 STONE B.J.
Best possible ratios of certain matrix norms.
Tech. Rep. 19, Appl. Math and Stat. Lab. Stanford U. (1962).

62 TRENOGIN V.A.
Perturbation of eigenvalues and eigenelements of linear operators.
Sov. Math. Dokl 7, (1966), 424-427.

63 TROTSKAIA E.A.
Application de la théorie générale des méthodes d'approximation de l'analyse aux bornes de valeurs propres et vecteurs propres.
Dokl. Akad. Nauk.113, (1957), 998-1001.

64 VARAH J.M.
The computation of bounds for the invariant subspaces of a general matrix operator.
Tech. Rep. 66, Comp. Sc. Dept, Stanf.U., Stanford (1967).

65 VAINIKKO G.M.
On the speed of convergence of approximate methods in the eigenvalue problem.
J. Math. Appl. Phys. Math. 7, (1967), 18-32.

66 VAINIKKO G.M., DEMENTIEVA A.M.

Sur la vitesse de convergence de la méthode des moindres carrés dans le problème des valeurs propres.
J. Vichl. Mat, mat.fis. 8, (1968), 1105-1110.

67 VAINIKKO G.M.

Principe de l'approximation compacte dans la théorie des méthodes d'approximation.
J. Vichl. Mat, mat.fis. 9, (1969), 739-761.

68 WEINBERGER H.F.

Error bounds in the Rayleigh-Ritz approximation of eigenvectors.
J. Res. Nat. Bur. Stand. Ser B, 64, (1960), 217-225.

69 WENDROFF B.

Bounds for eigenvalues of some differential operators by the Rayleigh Ritz method.
Math. Comp. 19, (1965), 218-224.

70 WILKINSON J.H.

Rigorous error bounds for computed eigensystems.
Comp. J. 4, (1961) 230-241.

71 WILKINSON J.H.

The algebraic eigenvalue problem.
Clarendon (1965).

72 YOSHIDA K.

Functional analysis.
Springer-Verlag (1965).