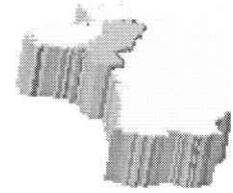


IREM

INSTITUT DE RECHERCHE SUR L'ENSEIGNEMENT DES MATHÉMATIQUES
DES PAYS DE LA LOIRE



NANTA IREMICA N° 11

Méthodes mathématiques modernes utilisées en théorie de l'approximation

Jean DHOMBRES

1976

ISBN 2-86300-010-1



PREFACE

Le développement extraordinaire des moyens de calcul offerts par les ordinateurs contemporains a eu par effet de retour une influence déterminante sur une théorie mathématique assez ancienne, la théorie de l'approximation. Cette théorie connut trois stades jusqu'à présent. Une préhistoire avec l'interpolation polynômiale de Lagrange, la méthode des moindres carrés, les formules de quadrature de Gauss, ainsi que diverses considérations éparses dans la littérature au XIX^{ème} siècle commençant avec le développement en série de Fourier et se terminant avec les théorèmes limites de Weierstrass, une histoire classique avec Čebičev et les premiers résultats sur l'approximation uniforme, les travaux de la Vallée Poussin, Bernstein et Jackson, une histoire moderne enfin avec l'arrivée des calculateurs et qui offre un florilège de publications au niveau de la recherche, de rapports techniques et d'ouvrages de vulgarisation. Voici un exemple supplémentaire de cette dernière catégorie.

Nous avons voulu reprendre la théorie mathématique de l'approximation d'une fonction et à moindre titre celle d'un opérateur, en mettant systématiquement l'accent sur les méthodes de l'analyse fonctionnelle. Par conséquent, on ne pourra bénéficier de ce texte sans une certaine connaissance de cette dernière discipline et un ouvrage particulier, Eléments d'Analyse Fonctionnelle (en abrégé E.A.F.) a été d'ailleurs rédigé à cette intention*. Nous fournissons ici les résultats élémentaires de la théorie désormais classique, du moins ceux qui ne présentent pas un caractère technique trop difficile ou trop long à détailler.

De nombreux exemples jalonnent l'exposé et un chapitre entier porte sur une technique numérique particulière. Plus d'une soixantaine d'exercices sont proposés au cours du texte.

* Publié dans la collection Nanta-Iremica : Volume N°7, suivi d'un volume N°8 consacré à des exercices en analyse fonctionnelle.

Toutefois, rappelons qu'il s'agit d'un texte mathématique dont le but premier est de décrire des idées et leurs enchaînements, de justifier l'utilité des concepts introduits et non seulement d'inventorier des recettes numériques.

Nous souhaitons mettre le lecteur en possession d'un outil puissant et lui permettre, face à une situation concrète de son travail d'ingénieur, de faire un choix critique parmi les diverses méthodes plus ou moins possibles. Ensuite, il pourra se reporter aux bibliothèques d'ordinateurs, ou aux ouvrages spécialisés pour la mise en oeuvre pratique de la procédure adoptée.

Ce texte peut aussi être utilisé par des professeurs de l'enseignement secondaire désirant apprendre la théorie de l'approximation en vue de modifier certains aspects de leur enseignement trop éloigné des techniques numériques. Nous avons utilisé un tel cours dans de nombreux stages de recyclage pour ingénieurs à l'Ecole Nationale Supérieure des Techniques Avancées et à l'Ecole Nationale Supérieure de l'Aéronautique et de l'Espace*. Nous l'avons utilisé avec des stagiaires de l'Institut de Recherche sur l'Enseignement des Mathématiques de l'Université de Nantes. Les nombreuses remarques des auditeurs nous ont fait modifier bien des présentations et nous remercions ces participants.

Monsieur DAUBISSE a dirigé la rédaction de mise en oeuvre numérique des méthodes mathématiques de l'approximation dans le cadre d'un groupe IREM au cours des dernières années. Le texte est fourni à la suite du texte théorique, au chapitre 8. Il est dû au travail en collaboration.

* Depuis 1973

INTRODUCTION

Précisons le problème général de l'approximation. Soit f , une fonction représentant un phénomène physique et que nous voulons exprimer, de la meilleure façon possible, à partir d'autres fonctions $\{\phi_n\}_{n \geq 1}$, tabulées ou considérées comme plus simples

La première question qui se pose est celle de mesurer la précision de l'expression choisie pour remplacer f à partir des fonction ϕ_n . Il nous faut définir, sur un ensemble de fonctions contenant f et les ϕ_n , une notion quantitative mesurant l'écart de deux fonctions. La notion de norme convient à cet effet et nous nous placerons sur un espace vectoriel normé, utilisant constamment le langage de E.A.F. chapitres 1 et 2.

La deuxième question, une fois ce langage posé, est celle de définir mathématiquement l'expression "de la meilleure façon possible". Soit f une fonction, élément d'un espace vectoriel E contenant les fonctions ϕ_n . Appelons F le sous-espace vectoriel de E engendré par les fonctions ϕ_n . Nous chercherons à exprimer f au moyen d'un élément g de l'espace F et cela de la meilleure façon possible. Le mieux alors est de chercher s'il existe un élément g de F tel que la distance de f à g , mesurée par la norme de $f-g$, soit la plus petite possible parmi toutes les distances de f à un élément quelconque h de F . On veut donc résoudre, si cela est possible, l'égalité :

$$\|f-g\| = \inf_{h \in F} \|f-h\| \quad ; \quad g \in F$$

Si un tel point g existe, on l'appelle une meilleure approximation de la fonction f au moyen de combinaisons linéaire des $\{\phi_n\}_{n \geq 1}$

Bien entendu, un tel élément g peut ne pas exister, ou s'il existe ne pas être unique. C'est l'objet du chapitre 1 que de donner les résultats utiles en ce

domaine. Ajoutons que le problème d'existence est toujours résolu par l'affirmative, dès que l'on cherche à approximer au moyen de combinaisons linéaires d'un nombre fini de fonctions ϕ_n . Ceci est de grande importance pratique.

Cependant, lorsqu'il y a existence et unicité, il faut favoriser un moyen pratique de calcul, soit un algorithme, soit une propriété caractéristique de la meilleure approximation. Notre effort portera essentiellement sur deux normes : la plus utile, la norme uniforme, sera traitée au chapitre 6 ; la plus utilisable, la norme quadratique, fera l'objet du chapitre 2. Cette dernière norme présente en effet de nets avantages, d'ailleurs souvent caractéristiques (existence et unicité, même en dehors des cas de dimension finie, caractérisation au moyen de propriétés de nature linéaire grâce au produit scalaire, linéarité de la meilleure approximation, etc.) Sur le plan de la compréhension, la norme quadratique est tellement proche de l'intuition géométrique usuelle que les théorèmes et les démonstrations sont très parlants. Le chapitre 3 détaille la technique numérique des fonctions splines où la norme quadratique joue son plein rôle.

Enfin, il reste à préciser l'erreur commise en remplaçant f par sa meilleure approximation en prenant n fonctions de base $\phi_1, \phi_2, \dots, \phi_n$. Bien entendu, on obtiendra surtout des résultats sur ce qui se passe lorsque n tend vers l'infini. C'est le but du chapitre 4 et du chapitre 5 où l'on rejoint avec la notion de lissage et de régularisation d'une fonction l'importante notion de convolution.

Enfin, on a regroupé en un chapitre 7 ce qui concerne les approximations d'opérateurs linéaires et des exemples de calcul numérique.

TABLE DES MATIERES

INTRODUCTION

CHAPITRE 1 : THEOREMES FONDAMENTAUX SUR LES MEILLEURES APPROXIMATIONS..... p 1

1. Théorèmes d'existence 1
2. Exemples 4
3. Unicité de la meilleure approximation 6
4. Algorithme de Polya 7
5. Autres exemples 10
6. Une propriété caractéristique d'une meilleure approximation 13

EXERCICES SUR LE CHAPITRE 1 : N° 1 à 9

CHAPITRE 2 : APPROXIMATION EN MOYENNE QUADRATIQUE..... 18

1. Un problème de linéarité 18
2. Rappels sur les espaces de Hilbert 22
3. Projection et approximation : théorème de caractérisation 22
4. Application au calcul de la projection 26
5. Utilisation des polynômes orthogonaux et des bases orthonormales 29

EXERCICES SUR LE CHAPITRE 2 : N° 10 à 22

CHAPITRE 3 : FONCTIONS SPLINES..... 40

1. Interpolation linéaire d'une fonction 40
2. Interpolation polynômiale d'une fonction 43
3. Interpolation par les fonctions splines cubiques 45
4. Propriétés des fonctions splines cubiques 49
5. Généralisation 51

EXERCICES SUR LE CHAPITRE 3 : N° 23 à 33

<u>CHAPITRE 4</u> : CONVERGENCE ET ESTIMATION DES METHODES D'APPROXIMATION.....	p	58
1. Introduction		58
2. Théorème de Korovkin		61
3. Application au théorème de Weierstrass		64
4. Estimation de la meilleure approximation		67
5. Estimation de l'approximation avec des fonctions splines cubiques		71
6. Autres théorèmes limites		71
<u>EXERCICES SUR LE CHAPITRE 4</u> : N° 34 à 44		74
 <u>CHAPITRE 5</u> : L'OPERATEUR DE CONVOLUTION.....		79
1. Convolution de deux fonctions périodiques de carré sommable		79
2. Convolution et exemples fonctionnels		86
3. Propriétés de la convolution		93
4. Quelques utilisations de la convolution		96
 <u>EXERCICES SUR LE CHAPITRE 5</u> : N° 45 à 50		113
 <u>CHAPITRE 6</u> : APPROXIMATION UNIFORME.....		117
1. Introduction		117
2. Théorème caractéristique		121
3. Procédé pratique de calcul		129
 <u>EXERCICES SUR LE CHAPITRE 6</u> : N° 51 à 55		131
 <u>CHAPITRE 7</u> : APPROXIMATIONS D'OPERATIONS LINEAIRES ET APPLICATIONS DIVERSES.....		133
1. Une application des polynômes orthogonaux : les méthodes de quadrature		133
2. Une méthode d'approximation : la méthode de Galerkine		136
3. Calcul numérique d'une approximation par discrétisation		143
4. Approximation non linéaire		147
 <u>EXERCICES SUR LE CHAPITRE 7</u> : N° 57 à 66		149

<u>CHAPITRE 8</u> : PROCEDURES NUMERIQUES DE L'APPROXIMATION DES FONCTIONS.....	p 154
8.1	But du travail
8.2	Pourquoi PL/1
8.3	1ère méthode
8.4	2ème méthode : polynômes orthogonaux
8.5	Approximation par les polynômes de Čebičev.
 Appendice	 ET PAN DANS LE MILLE..... p 184
 Bibliographie	 p 198
 Index des expressions	 p 201



LISTE DES PUBLICATIONS DISPONIBLES DE NANTA-IREMICA

- Etude épistémologique et historique des idées de nombre, de mesure et de continu (DHOMBRES - N° 3)
- Eléments d'Analyse Fonctionnelle (DHOMBRES - N° 7)
- Exercices d'Analyse Fonctionnelle (DHOMBRES - N° 8)
- Méthodes mathématiques modernes utilisées en théorie de l'approximation (DHOMBRES N° 11)
- Mathématiques comparées : Angleterre, Québec, Allemagne Fédérale, Chine (BIGARD)
- Sensibilisation aux structures de données (VIVET - N° 15)
- Brochure HP 25 (QUELFETER - N° 17)
- Nantinfo 78 (QUELFETER - N° 19)
- Nombres entiers naturels (Groupe des Professeurs d'Ecole Normale - N° 18)
- Mathématiques en L.E.P. (PAPIN - N° 20)
- Liaison CM2 - 6^e (CHARLOT-LIZE - N° 21)
- Couleurs, électricité et mathématique (CARNEC-SEROUX - N° 22)
- Des applications de la proportionnalité en 6^e et 5^e (POCHE - N° 23)

LISTE DES PUBLICATIONS PARUE A L'IREM DE NANTES, MAIS EPUISEES

Volume 1	Introduction à la logique	MM. DURAND et VAN DEN BOSSCHE
Volume 2	Introduction à la théorie des Ensembles	MM. DURAND et VAN DEN BOSSCHE
Volume 4	Documents relatifs au volume 3	J. DHOMBRES
Volume 5	Le langage BASIC	M. BELHACHE
Volume 6	Echec en mathématiques	A. BIGARD
Volume 9	Algèbre linéaire et géométrie vectorielle	R. SEROUX
Volume 10	Analyse et Topologie	Melle VENARD et J. DHOMBRES
Volume 12	Introduction à la géométrie et propositions pour la classe de 4è	J.P. LETOURNEUX
Volume 13	Le PL/1 Optimiseur	M. BELHACHE
	Quelques difficultés pédagogiques dans l'enseignement de l'analyse : Second cycle des lycées	MM. FOUQUES et SEROUX
	Trigonométrie, algèbre linéaire optique	R. SEROUX
Volume 14	Mathématique et langage informatique (BASIC)	J. BETREMA

- CHAPITRE 1 -

THEOREMES FONDAMENTAUX SUR LES MEILLEURES APPROXIMATIONS

I. THEOREMES D'EXISTENCE.

Nous utiliserons les notions d'analyse fonctionnelle pour poser et résoudre le problème annoncé dans l'introduction. A priori, nous nous plaçons donc sur un espace vectoriel normé E , dont les éléments sont des fonctions. Nous nous donnons un sous-espace vectoriel F , de dimension finie et nous cherchons, f étant un élément fixé de E , un élément g de F tel que

$$\|f - g\| = \inf_{h \in F} \|f - h\|$$

Cet élément g est appelé une meilleure approximation de f dans F .

Pour utiliser l'hypothèse de la dimension finie de F , il est relativement commode de partir d'une base $\phi_1, \phi_2, \dots, \phi_n$ engendrant F . Le problème se situe alors de la façon suivante :

Chercher n constantes réelles a_1, a_2, \dots, a_n (ou complexes si l'espace E vectoriel normé sur lequel on travaille est complexe) pour lesquelles

$$\|f - \sum_{k=1}^{n} a_k \phi_k\| = \inf_{b_1, b_2, \dots, b_n \in \mathbb{R}^n} \|f - \sum_{k=1}^{n} b_k \phi_k\|$$

Exemple : choisissons pour E l'espace vectoriel des fonctions continues et à valeurs réelles, définies sur le segment $[0,1]$: $E = C[0,1]$. Nous munissons E de l'une des normes définies en EAF, chap. 1 pour constituer un espace normé du type $C[0,1]$ ou $C_p[0,1]$

A priori f est une fonction continue quelconque et les $\phi_1, \phi_2, \dots, \phi_n$ sont n fonctions continues supposées indépendantes. Cette dernière expression signifiant que s'il existe n constantes réelles b_1, b_2, \dots, b_n telles que l'expression $\sum_{k=1}^n b_k \phi_k(x)$

soit identiquement nulle sur $[0,1]$, alors on a nécessairement les égalités

$$b_1 = b_2 = \dots = b_n = 0$$

On peut énoncer le théorème très général suivant :

Théorème d'existence 1 : il existe une meilleure approximation de tout élément f d'un espace normé au moyen d'un sous-espace de dimension finie.

L'idée de la démonstration est simple. Tout d'abord, f étant fixé dans E , nous constatons que l'expression $\|f - \sum_{k=1}^n b_k \phi_k\|$ est positive et définit une application de F dans \mathbb{R}^+ ($x \in \mathbb{R}^+$ si $x \geq 0$)

Elle admet donc une borne inférieure notée α . Le problème est de montrer maintenant que cette borne inférieure est effectivement atteinte pour n constantes réelles $\alpha_1, \alpha_2, \dots, \alpha_n$. A priori, il semble possible d'obtenir un tel résultat (quand on songe au problème analogue : montrer qu'une fonction continue sur $[0,1]$ atteint sa borne inférieure) à condition toutefois de montrer que les b_1, b_2, \dots, b_n ne varient pas sur tout l'espace \mathbb{R}^n :

$$\alpha = \inf_{b_1, b_2, \dots, b_n \in \mathbb{R}^n} \left\| f - \sum_{k=1}^n b_k \phi_k \right\|$$

D'une part, cette constante, positive ou nulle, est majorée par $\|f\|$ (faire $b_1 = b_2 = \dots = b_n = 0$). Nous allons approcher le minimum au moyen d'une suite g_m (g_m , élément de F) telle que

$$\alpha < \|f - g_m\| < \alpha + \frac{1}{m}$$

La suite g_m est caractérisée par ses n coordonnées dans la base $\phi_1, \phi_2, \dots, \phi_n$ de l'espace F , c'est-à-dire

$$g_m = \sum_{k=1}^n b_{k,m} \phi_k$$

Nous allons montrer que l'on peut alors extraire de g_m une sous-suite g_{m_e} qui converge dans F vers un élément g de coordonnées $[\alpha_k]$. Cet élément g sera justement un élément réalisant la meilleure approximation.

Reprenons donc la démonstration. L'expression

$$\{b_1, b_2, \dots, b_n\} \longrightarrow \left\| \sum_{k=1}^n b_k \phi_k \right\| = N(b_1, b_2, \dots, b_n)$$

constitue une norme sur l'espace vectoriel \mathbb{R}^n .

En effet

$$N(\lambda b_1, \lambda b_2, \dots, \lambda b_n) = |\lambda| N(b_1, b_2, \dots, b_n)$$

$$N(b_1 + b'_1, \dots, b_n + b'_n) \leq N(b_1, \dots, b_n) + N(b'_1, \dots, b'_n)$$

$$N(b_1, b_2, \dots, b_n) = 0 \text{ est équivalent à } b_1 = b_2 = \dots = b_n = 0$$

(Ce dernier point est acquis grâce à l'indépendance linéaire des fonctions $\Phi_1, \Phi_2, \dots, \Phi_n$ qui forment une base de F).

Grâce au théorème d'équivalence des normes sur l'espace \mathbb{R}^n , nous assurons l'existence de deux constantes positives A et B pour lesquelles

$$(1) \quad A \max_{1 \leq k \leq n} |b_k| \leq N(b_1, b_2, \dots, b_n) \leq B \max_{1 \leq k \leq n} |b_k|$$

Nous avons donc limité ainsi le champ de variation des constantes b_1, b_2, \dots, b_n

Nous approchons maintenant le minimum a au moyen de la suite g_m . La première inégalité de (1) fournit :

$$\max_{1 \leq k \leq n} |b_{k,m}| \leq \frac{a + \|f\| + \frac{1}{m}}{A}$$

car on a

$$N(b_{1,m}, b_{2,m}, \dots, b_{n,m}) \leq \|f - \sum_{k=1}^n b_{k,m} \Phi_k\| + \|f\|$$

Ceci prouve que la suite $B_m = \{b_{1,m}, b_{2,m}, \dots, b_{n,m}\}$ est une suite bornée dans \mathbb{R}^n . Certes nous ne savons pas a priori si la suite B_m converge, mais le théorème de Bolzano-Weierstrass permet d'extraire de la suite B_m une sous-suite B_{m_e} convergente dans l'espace \mathbb{R}^n vers un point A .

$$A = \{a_1, a_2, \dots, a_n\}$$

La convergence de la suite B_{m_e} implique la convergence de la suite de fonctions $g_{m_e} = \sum_{k=1}^n b_{k,m_e} \Phi_k$ au sens de la norme sur l'espace E , grâce à la deuxième inégalité (1).

En effet :

$$\left\| \sum_{k=1}^{k=n} b_{k,m_e} \Phi_k - \sum_{k=1}^{k=n} a_k \Phi_k \right\| \leq B \max_{1 \leq k \leq n} |b_{k,m_e} - a_k|$$

Ceci implique aussitôt grâce à l'inégalité triangulaire

$$\left\| f - \sum_{k=1}^n a_k \Phi_k \right\| = \lim_{m_e \rightarrow +\infty} \left\| f - \sum_{k=1}^n b_{k,m_e} \Phi_k \right\|$$

Or :

$$\left\| f - \sum_{k=1}^n b_{k,m_e} \Phi_k \right\| \leq a + \frac{1}{m_e}$$

qui tend vers a lorsque m tend vers l'infini. La suite a_1, a_2, \dots, a_n réalise donc :

$$a = \left\| f - \sum_{k=1}^n a_k \Phi_k \right\| = \inf_{h \in F} \| f - h \|$$

L'existence d'une meilleure approximation est ainsi assurée conformément au théorème énoncé. (Une autre démonstration bien plus rapide utilise la compacité d'une boule fermée de R^n , notion non définie dans E.A.F.)

2. EXEMPLES.

On applique ce théorème au cas $[(0,1)]$ ou encore $]\!]\!][a,b]$, a et b étant finis ou non.

Exemple 1 considérons l'espace $[(0,1)]$ et différentes normes sur cet espace :

$$\| f \|_{\infty} = \sup_{x \in [0,1]} |f(x)|$$

et

$$\| f \|_p = \left(\int_0^1 |f(x)|^p dx \right)^{\frac{1}{p}}, \quad p \geq 1, \quad (\text{Cf. E.A.F. Chapitre 1})$$

Cherchons à déterminer une meilleure approximation, pour ces normes, de la fonction $f : x \longrightarrow x$ lorsque l'on choisit $n = 1$ et $\Phi_1(x) = 1$.

Il s'agit de trouver une constante a telle que $\| f - a \|$ soit minimum.

Traisons d'abord le cas p quelconque mais fini. Des inégalités

$$a < 0 \quad x-a \geq x \geq x+a \quad \text{sur } [0,1]$$

et

$$x+a-1 < x+a$$

on déduit :

$$\|f-a\|_p \geq \|f\|_p \geq \|f+a\|_p \quad \text{si } a < 0$$

et

$$\|f+a-1\|_p \leq \|f+a\|_p \quad \text{si } a > 1$$

c'est-à-dire que l'on peut supposer a compris entre 0 et 1. Dès lors

$$\begin{aligned} \|f-a\|_p^p &= \int_0^1 |x-a|^p dx = \int_{-a}^{1-a} |x|^p dx = \frac{a^{p+1} + (1-a)^{p+1}}{p+1} \\ \frac{d}{da} (\|f-a\|_p^p) &= a^p - (1-a)^p \end{aligned}$$

L'unique valeur de a réalisant le minimum est donnée par $a^p = (1-a)^p$, c'est-à-dire $a = \frac{1}{2}$.

Maintenant il est facile de vérifier que $a = \frac{1}{2}$ est l'unique solution réalisant le minimum pour la norme de la convergence uniforme.

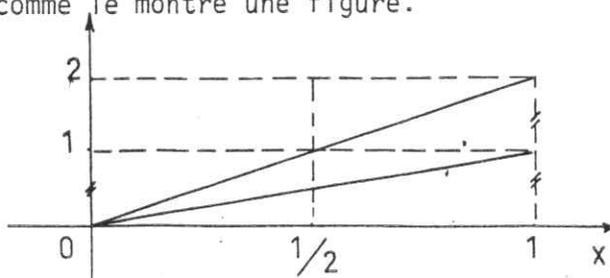
Exemple 2: Toutefois l'unicité de la meilleure approximation n'est pas toujours exacte comme le montre l'exemple suivant. Supposons que nous voulions réaliser la meilleure approximation de la fonction f valant 1 sur $[0,1]$ en prenant $n = 1$ et $\Phi_1(x) = x$.

$$\|f - ax\|_\infty = \|1 - ax\|_\infty$$

D'une part

$$\|1 - ax\|_\infty \geq f(0) = 1$$

Cette borne inférieure est atteinte pour toutes les valeurs de a comprises entre 0 et 2, comme le montre une figure.



On note que $a = 2$ correspond à l'unique meilleure approximation pour les normes $\|\cdot\|_p$

3. UNICITE DE LA MEILLEURE APPROXIMATION.

Nous allons voir cependant que pour de nombreuses normes fonctionnelles, il y a unicité de la meilleure approximation. Ce résultat est d'importance pour le traitement par machines.

Définition : on dit qu'une norme fonctionnelle est strictement convexe lorsqu'elle possède la propriété suivante :

si $\|f_1\| = \|f_2\|$ et si α est un nombre strictement positif et strictement inférieur à 1, alors

$$\|\alpha f_1 + (1 - \alpha)f_2\| < \|f_1\| = \|f_2\| \text{ lorsque } f_1 \neq f_2$$

Théorème d'unicité 2 : il y a unicité de la meilleure approximation lorsque la norme fonctionnelle est strictement convexe.

Supposons que deux suites de nombres $\{a_1, a_2, \dots, a_n\}$ et $\{a'_1, a'_2, \dots, a'_n\}$ réalisent une meilleure approximation de la fonction f . Envisageons la suite $\frac{a_1+a'_1}{2}, \frac{a_2+a'_2}{2}, \dots, \frac{a_n+a'_n}{2}$ et

l'approximation que réalise cette suite. Il vient :

$$\|f - \sum_{k=1}^n (\frac{a_k+a'_k}{2}) \phi_k\| = \|\frac{1}{2}(f - \sum_{k=1}^n a_k \phi_k) + \frac{1}{2}(f - \sum_{k=1}^n a'_k \phi_k)\|$$

Mais nous avons par définition d'une meilleure approximation

$$\|f - \sum_{k=1}^n a_k \phi_k\| = \|f - \sum_{k=1}^n a'_k \phi_k\| = \inf_{(b_1, \dots, b_n) \in \mathbb{R}^n} \|f - \sum_{k=1}^n b_k \phi_k\|$$

et en appliquant la définition d'une norme strictement convexe avec $\alpha = \frac{1}{2}$

$$\|f - \sum_{k=1}^n (\frac{a_k+a'_k}{2}) \phi_k\| < \inf_{(b_1, \dots, b_n) \in \mathbb{R}^n} \|f - \sum_{k=1}^n b_k \phi_k\|$$

Cette dernière inégalité est impossible si l'on a $\{a_1, \dots, a_n\} \neq \{a'_1, \dots, a'_n\}$ d'où le théorème 2.

L'exemple donné après le théorème 1 montre que sur l'espace $[[0,1]]$ la norme de la convergence uniforme n'est pas strictement convexe puisque nous avons trouvé plusieurs meilleures approximations. On montrerait de même que la norme $\|f\|_1$ n'est pas strictement convexe. Toutefois, les autres normes $\|f\|_p$ sont strictement convexes.

Théorème 3 : Pour $1 < p < +\infty$ les normes $\| \cdot \|_p$ sont strictement convexes.

Pour ce théorème, il importe d'abord de préciser l'inégalité de Minkowsky en cherchant les cas d'égalité, dans le cas des normes $\| \cdot \|_p$. Lors de la démonstration de

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \quad \text{avec } 1 < p < \infty$$

nous avons vu qu'il y a égalité seulement si f et g sont des fonctions linéairement dépendantes (Cf. E.A.F.).

Dès lors si $\|f_1\|_p = \|f_2\|_p$, alors pour un nombre α ($0 < \alpha < 1$), on a

$$\|\alpha f_1 + (1-\alpha)f_2\|_p \leq \alpha \|f_1\|_p + (1-\alpha)\|f_2\|_p = \|f_1\|_p = \|f_2\|_p$$

L'égalité a lieu seulement si αf_1 et $(1-\alpha)f_2$ sont linéairement dépendantes et de même signe. Mais comme f_1 et f_2 ont même norme, cela implique $f_1 = f_2$. Nous avons bien vérifié la définition d'une norme strictement convexe (ici les espaces fonctionnels sont supposés réels).

La norme $\| \cdot \|_\infty$ a une relation limite avec les normes $\| \cdot \|_p$. Nous allons voir laquelle et comment on peut utiliser une telle relation pour la recherche d'une meilleure approximation au sens de la norme de la convergence uniforme.

4. ALGORITHME DE POLYA.

Énonçons le théorème en vue sous la forme la plus simple.

Théorème 4 : soit f une fonction continue sur le segment $[0,1]$. On a :

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)| = \lim_{p \rightarrow \infty} \|f\|_p$$

Ce théorème est facile. D'une part $\|f\|_p \leq \|f\|_\infty$ d'après l'inégalité de Hölder (Cf. E.A.F.). D'autre part, plaçons-nous autour d'un point x_0 où f atteint son maximum. Nous supposons, sans perdre de généralité, (changer f en $-f$) que :

$$\|f\|_\infty = f(x_0)$$

Il existe un nombre η positif tel que pour $|x - x_0|$ inférieur à η alors $|f(x) - f(x_0)|$ est inférieur à ε . Il vient, pour tout $\varepsilon > 0$ assez petit

$$\begin{aligned} \int_0^1 |f(x)|^p dx &> \int_{x_0-\eta}^{x_0+\eta} |f(x)|^p dx \geq (\|f\|_\infty - \varepsilon)^p \int_{x_0-\eta}^{x_0+\eta} dx \\ &\geq 2\eta (\|f\|_\infty - \varepsilon)^p \end{aligned}$$

D'où

$$\|f\|_\infty > \|f\|_p > (\|f\|_\infty - \varepsilon) (2\eta)^{\frac{1}{p}}$$

Lorsque p tend vers l'infini, $(2\eta)^{\frac{1}{p}}$ tend vers 1, ce qui établit le théorème 4, le résultat étant vrai pour tout ε positif assez petit.

Appliquons ceci au problème de l'approximation. Nous savons que pour la norme L_p ($p > 1$), il existe une unique meilleure approximation de f sur F , que je note $P_p(f)$. Que peut-on dire de $\lim_{p \rightarrow +\infty} P_p(f)$?

Heuristiquement, on peut imaginer que $\lim_{p \rightarrow +\infty} P_p(f)$ est l'une des meilleures approximations de f au sens de la norme de la convergence uniforme.

Le résultat n'est pas exact sous cette forme. Nous avons bien

$$(1) \quad \|f - P_p(f)\|_p < \|f - g\|_p \quad \text{pour tout } g \text{ dans } F$$

et en particulier pour $g = 0$

$$\|f - P_p(f)\|_p \leq \|f\|_p$$

ce qui montre que, en utilisant les relations entre les normes L_p (cf EAF chap 2§5

$$\|P_p f\|_1 \leq \|P_p(f)\|_p \leq 2\|f\|_p \leq 2\|f\|_\infty$$

Mais $P_p(f)$ est une suite de l'espace vectoriel de dimension finie F . Comme cette suite est bornée, en norme L_1 , en reprenant l'utilisation du théorème de Bolzano-Weierstrass comme dans la démonstration du théorème d'unicité 1, on montre que l'on peut extraire une sous-suite $P_{p_j}(f)$ telle que $P_{p_j}(f)$ converge au sens de L_1 et au sens de la norme de la convergence uniforme (car sur F toutes les normes sont équivalentes), vers une

fonction de F que je note f_0 . Montrons que f_0 réalise une meilleure approximation de f au sens de la norme de la convergence uniforme. En effet on a

$$\|f - P_{p_\ell}(f)\|_{p_\ell} \leq \|f - g\|_{p_\ell} \quad \text{pour tout } g \in F$$

Donc :

$$\|f - f_0\|_{p_\ell} \leq \|f - P_{p_\ell}(f)\|_{p_\ell} + \|P_{p_\ell}(f) - f_0\|_{p_\ell}$$

Mais la norme de la convergence uniforme est plus fine que la norme L_p , d'où

$$\|f - f_0\|_{p_\ell} \leq \|f - g\|_{p_\ell} + \|f_0 - P_{p_\ell}(f)\|_\infty$$

Faisons tendre ℓ vers l'infini, il vient grâce au théorème 4 :

$$\begin{aligned} \lim_{\ell \rightarrow \infty} \|f - f_0\|_{p_\ell} &= \|f - f_0\|_\infty \\ \text{et} \quad \lim_{\ell \rightarrow +\infty} \|f - g\|_{p_\ell} &= \|f - g\|_\infty \end{aligned}$$

et grâce à la construction de $P_{p_\ell}(f)$

$$\lim_{\ell \rightarrow +\infty} \|f_0 - P_{p_\ell}(f)\|_\infty = 0$$

d'où

$$\|f - f_0\|_\infty \leq \|f - g\|_\infty \quad \text{pour tout } g \text{ dans } C[0,1]$$

ce qui démontre le théorème suivant.

Théorème 5 : soit $P_p(f)$ la meilleure approximation de f au sens de $L_p[0,1]$ sur un espace vectoriel de dimension finie F . Toute sous-suite $P_{p_\ell}(f)$ convergente en norme uniforme, converge vers une meilleure approximation de f au sens de la norme uniforme.

Obtient-on de cette manière toutes les meilleures approximations pour la norme uniforme ? La réponse en général est négative et l'on peut avoir une limite pour la suite $P_p(f)$ elle-même bien que f n'ait pas une unique meilleure approximation au sens de la norme uniforme. Un tel cas est fourni par l'exemple 2, page 5.

Cependant ce théorème 5 permet le calcul de la meilleure approximation lorsque cette dernière est unique, compte tenu du fait que les espaces L_p sont uniformément convexes (voir plus loin).

5. AUTRES EXEMPLES.

Comme nous nous sommes donnés beaucoup de liberté dans le choix de la norme, il peut être intéressant de chercher les meilleures approximations lorsque la norme fait intervenir f et ses dérivées.

Ainsi prenons $C^1[0,1]$ l'espace des fonctions continûment dérivables sur le segment $[0,1]$.

On peut se poser le problème de la meilleure approximation d'une fonction f en choisissant une norme particulière.

Exemple : sur $C^1[0,1]$, nous prenons trois normes différentes :

$$\|f\|_2 = \left(\int_0^1 f(x)^2 dx \right)^{\frac{1}{2}} \quad \text{norme de la convergence en moyenne quadratique}$$

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)| \quad \text{norme uniforme}$$

et $\|f\| = \left(\int_0^1 |f(x)|^2 dx + \int_0^1 |f'(x)|^2 dx \right)^{\frac{1}{2}}$ laquelle fait jouer un rôle simultané à f et à sa dérivée f' .

Recherchons la meilleure approximation de la fonction $y = x^4$ par exemple en prenant pour F l'espace des fonctions linéaires.

Pour la première norme, il y a unicité et on trouve facilement (Cf Chapitre suivant)

$$\frac{4}{5}x - \frac{1}{5} = 0,8x - 0,2$$

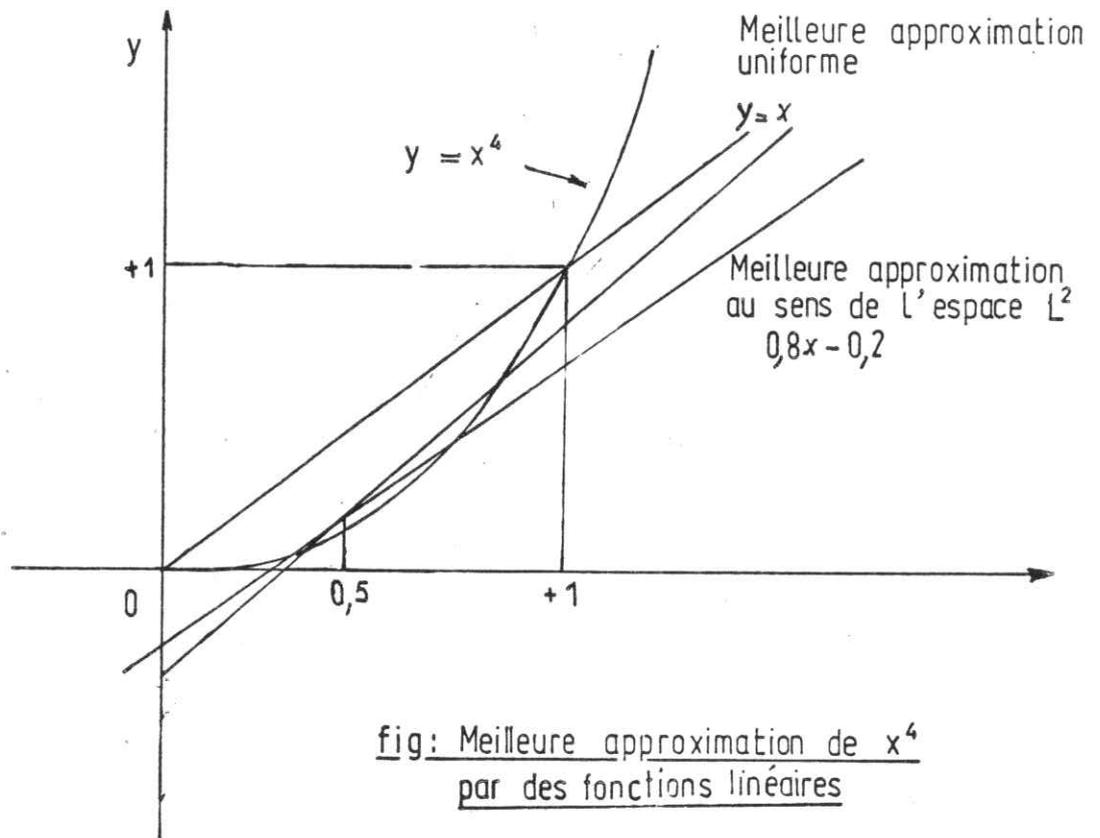
Pour la troisième norme, il y a également unicité (cette norme dérive d'un produit scalaire, et, grâce à l'identité du parallélogramme, les espaces préhilbertiens sont strictement convexes). On trouve :

$$\frac{64}{65}x - \frac{19}{65} = 0,98x - 0,29$$

Pour la norme uniforme, on trouve

$$x - \frac{3}{16} \sqrt[3]{2} = x - 0,236$$

Naturellement ces trois meilleures approximations sont différentes.



La troisième norme de cet exemple est assez intéressante, elle dérive d'un produit scalaire et le complété de l'espace $C^1[0,1]$ pour cette norme est un espace de Hilbert appelé communément espace H^1 et sur lequel on pose souvent les problèmes d'équations différentielles avec problèmes aux limites.

Nous verrons au Chapitre 2 que l'on peut caractériser les meilleures approximations sur un espace de Hilbert, et que l'on dispose ainsi d'un moyen de calcul.

Si l'on démontre que la norme $\|\cdot\|_{H^1}$ est plus fine (Cf E.A.F.) que la norme uniforme, la connaissance de la meilleure approximation pour la norme de H^1 donne une bonne approximation en norme uniforme, non optimale en général. Ceci est appréciable quand on connaît les difficultés de l'approximation en norme uniforme (Cf. Chapitre 6).

Ces quelques remarques montrent que le choix d'une norme est une chose importante pour l'approximation d'une fonction continue. Donnons une figure suggestive qui fera comprendre en outre le problème d'unicité. Nous nous plaçons dans l'espace \mathbb{R}^2 et choisissons comme normes les normes L_p à savoir pour $x = (x_1, x_2)$:

$$\|x\|_p = (|x_1|^p + |x_2|^p)^{\frac{1}{p}} \text{ pour } \infty > p > 1$$

et $\|x\|_\infty = \text{Sup} (|x_1|, |x_2|)$

Les boules unit   B_p pour ces normes peuvent ˆtre repr  sent  es de la fa on suivante :

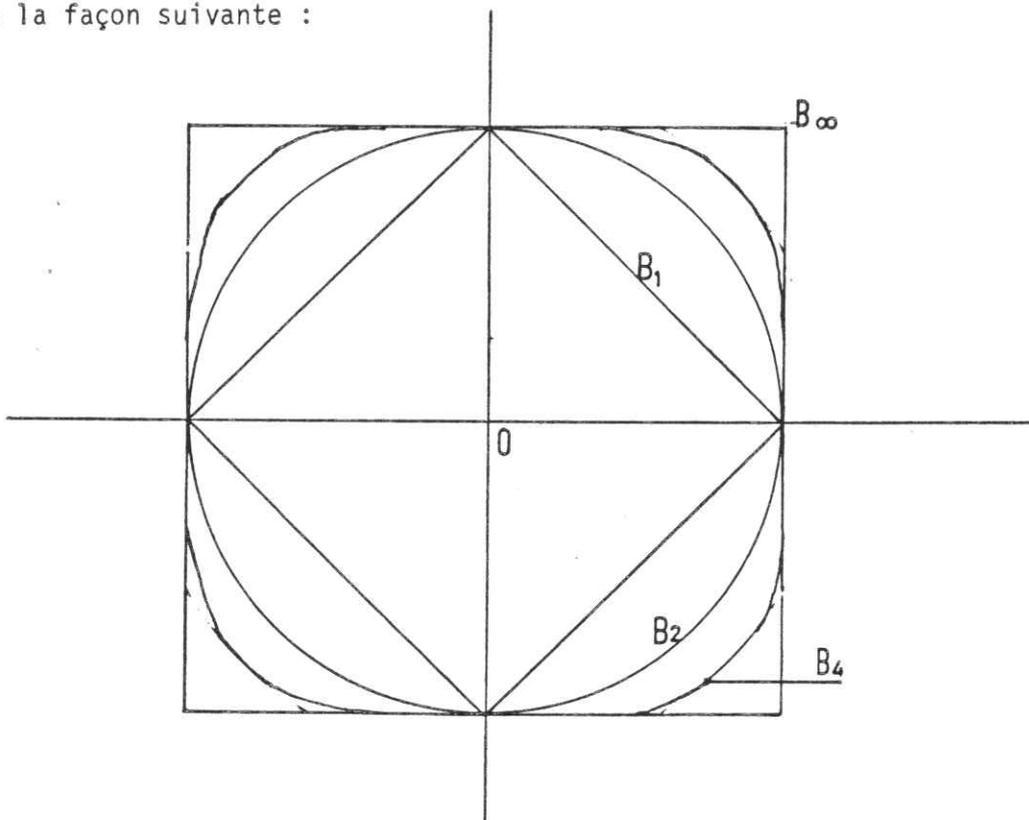
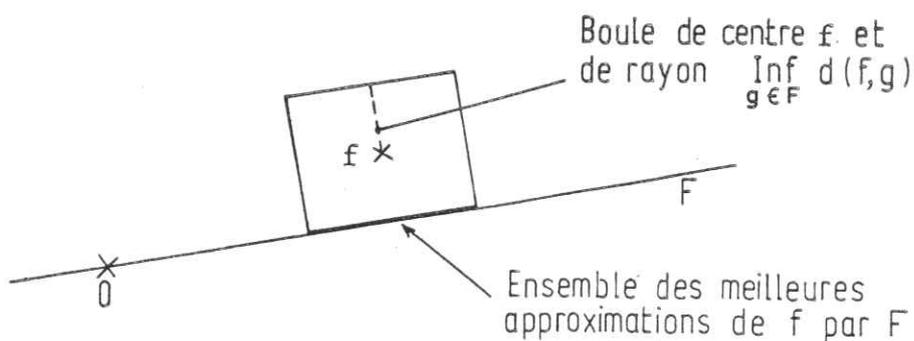


Fig: Boules unit  es en norme L_p

Puisque les fronti eres de B_1 et de B_∞ pr  sentent des portions rectilignes, il ne peut y avoir unicit   de la meilleure approximation au sens de ces normes. En effet, les  l ements de meilleure approximation d'un point x de E par un sous-espace vectoriel F sont   l'intersection du sous-espace F et de la sph re de centre x et dont le rayon est la plus courte distance de x   F . Si la sph re en question poss de une portion de segment, ou plus g n ralement une plage convexe, on n'a pas unicit  , comme la figure ci-dessus l'indique bien.



6. UNE PROPRIÉTÉ CARACTÉRISTIQUE D'UNE MEILLEURE APPROXIMATION.

L'existence d'une meilleure approximation est donc acquise dans le cas du moins où F est de dimension finie. La question qui se pose maintenant est le calcul effectif d'une meilleure approximation. La démonstration du théorème d'existence n'est pas constructive puisqu'elle ne fournit pas un algorithme convergeant vers une meilleure approximation. En effet on a seulement dit qu'il était possible d'extraire une sous-suite convergente en se basant sur le théorème de Bolzano-Weierstrass, mais on n'a pas donné de moyen pratique pour une telle construction : on n'a pas encore programmé sur machine l'obtention d'une meilleure approximation.

Une démarche naturelle consiste alors à chercher des propriétés caractéristiques des meilleures approximations avec l'espoir que ces propriétés s'avèreront constructives. De fait, de telles caractérisations seront envisagées avec succès pour des espaces normés particuliers : les espaces hilbertiens. Ainsi au Chapitre 2, pour la norme de la convergence en moyenne quadratique, une excellente caractérisation sera obtenue grâce à la propriété d'orthogonalité.

Au Chapitre 6, on passera à la norme de la convergence uniforme pour laquelle une caractérisation moins maniable sera obtenue au moyen des propriétés d'alternance. Pour la norme L^1 , nous étudierons également des propriétés d'alternances et de convexité.

Toutefois, nous allons terminer ce chapitre par une caractérisation fort générale.

Théorème 6: soit E un espace normé sur \mathbb{R} (ou sur \mathbb{C}) et F un sous-espace vectoriel fermé de E . Soit x un point de E n'appartenant pas à F . Un point y_0 de F est une meilleure approximation de x par les éléments de F si et seulement s'il existe une forme linéaire continue L sur E possédant les propriétés suivantes :

$$(1) \quad \|L\| = 1$$

$$(2) \quad L(z) = 0 \quad \text{pour tout } z \text{ de } F$$

$$(3) \quad L(x - y_0) = \|x - y_0\|$$

Démonstration : La condition est suffisante. En effet s'il existe un y_0 de F et une forme linéaire continue sur E ayant les propriétés requises, on a

$$\|x - y_0\| = |L(x - y_0)| = |L(x - z)|$$

puisque $L(y_0) = L(z) = 0$. Par suite, pour tout z de F , on a bien

$$\|x - y_0\| \leq \|L\| \|x - z\| = \|x - z\|$$

ce qui caractérise bien une meilleure approximation.

La condition est nécessaire. Puisque x n'appartient pas à F et que F est un sous-espace fermé de E , on dispose de

$$\inf_{z \in F} \|x - z\| = a > 0$$

Considérons le sous-espace vectoriel G de E , engendré par le sous-espace F et l'élément x . Tout élément y de G s'écrit d'une façon unique sous la forme

$$y = z + \lambda x$$

où λ est un scalaire réel si E est réel, complexe si E est complexe.

On définit une forme linéaire L' sur G en posant pour tout y de G

$$L'(y) = \lambda a$$

En particulier $L'(z) = 0$ pour tout z de F .

Cette forme linéaire est continue sur G puisque

$$|L'(y)| \leq \|y\|$$

En effet, cette inégalité est évidente si $\lambda = 0$. Pour $\lambda \neq 0$, on a une minoration puisque z/λ appartient à l'espace F .

$$\|y\| = \|z + \lambda x\| = |\lambda| \|z/\lambda + x\| \geq |\lambda| a = |L'(y)|$$

Il est facile de voir que la norme de L' , considérée comme forme linéaire continue sur G , est exactement 1 grâce au choix de a . Dès lors, un théorème fondamental d'analyse fonctionnelle, le théorème de Hahn-Banach, assure qu'il existe un prolongement L de L' en une forme linéaire et continue sur tout E . En outre L' a même norme que L . Nous admettrons

ce théorème assez intuitif et il devient alors clair que L est la forme, linéaire et continue cherchée.

Le théorème de caractérisation précédent est évidemment utilisable dès que le dual topologique de E est bien connu : c'est le cas de $C[0,1]$ mais également le cas de $\mathcal{C}_p[0,1]$ pour $1 < p < \infty$. Dans ce dernier cas en effet, le dual s'identifie avec $L_q[0,1]$ où $1/p + 1/q = 1$ (Cf E.A.F.) Dans le cas $\mathcal{C}_1[0,1]$ le dual s'identifie à l'espace des fonctions mesurables et bornées sur $[0,1]$ (Cf E.A.F.). Enfin, pour $\mathcal{C}[0,1]$, le dual est l'espace des mesures de Radon sur $[0,1]$, largement étudié en théorie de l'intégration.

Examinons ce qui se passe dans le cas $\mathcal{C}_2[0,1]$, ou ce qui revient au même pour la dualité topologique, dans le cas $L^2[0,1]$. On sait que le dual topologique s'identifie à $L^2[0,1]$ (Cf. E.A.F.). Soit alors f un élément de $L^2[0,1]$ et g_0 un élément d'un sous-espace vectoriel fermé F de $L^2[0,1]$. Cet élément g_0 est la meilleure approximation de f par F si et seulement si il existe une fonction h de $L^2[0,1]$ telle que

$$(1) \quad \|h\|_2 = \sqrt{\int_0^1 |h(x)|^2 dx} = 1$$

$$(2) \quad \int_0^1 g(x) \overline{h(x)} dx = 0 \quad \text{pour tout } g \text{ de } F$$

$$(3) \quad \int_0^1 (f(x) - g_0(x)) \overline{h(x)} dx = \sqrt{\int_0^1 |f(x) - g_0(x)|^2 dx}$$

(1) et (3) fournissent $\int_0^1 (h(x) - \frac{f(x) - g_0(x)}{\|f - g_0\|}) \overline{h(x)} dx = 0$ et grâce au théorème de Pythagore $\|h - \frac{f - g_0}{\|f - g_0\|}\|^2 + \|h\|^2 = \left(\frac{\|f - g_0\|}{\|f - g_0\|}\right)^2 = 1$

donc $h(x) = \frac{f(x) - g_0(x)}{\|f - g_0\|}$ presque partout.

Par suite, l'élément g_0 réalise la meilleure approximation de f par les éléments de F si et seulement si pour tout g de F on a

$$\int_0^1 g(x) \overline{(f(x) - g_0(x))} dx = 0$$

Cette propriété caractéristique est de maniement facile comme nous allons le constater au Chapitre 2. Nous allons d'ailleurs reprendre complètement la démonstration de cette importante propriété dans un cadre plus général.

- EXERCICES DU CHAPITRE 1 -

EXERCICE N° 1.

Soit f une fonction continue sur $[a, b]$, m et n deux entiers fixés positifs. Montrer que le problème de trouver $n+m+2$ nombres réels $a_0, \dots, a_n, b_0, \dots, b_n$ tels que

$$\sup_{x \in [a, b]} \left| f(x) - \frac{a_0 x^n + a_1 x^{n-1} + \dots + a_n}{b_0 x^m + b_1 x^{m-1} + \dots + b_m} \right|$$

soit minimal, admet au moins une solution (ce résultat constitue le premier pas en théorie non linéaire de l'approximation des fonctions).

EXERCICE N° 2.

Calculer la meilleure approximation de $\sqrt{1+x^2}$ sur $[0, 1]$ au moyen d'une fonction linéaire et au sens de la norme en moyenne quadratique (on trouvera $0,955 + 0,414 x$).

Discuter l'erreur commise.

EXERCICE N° 3.

Comparer la meilleure approximation en norme quadratique de e^x par une fonction linéaire à la meilleure approximation de e^x en norme L^1 ou en norme uniforme.

EXERCICE N° 4.

Montrer qu'un espace préhilbertien (Cf E.A.F.) est strictement convexe.

EXERCICE N° 5.

Reprendre la démonstration du théorème d'unicité dans le cas d'un espace de norme strictement convexe E et d'un sous-espace F dont la dimension n'est pas nécessairement finie. On montrera que l'ensemble des meilleures approximations est un sous-ensemble convexe d'une certaine sphère.

Par ailleurs la stricte convexité réduit à un point les sous-ensembles convexes qui sont sous-ensembles d'une sphère.

EXERCICE N° 6.

Etablir le théorème suivant.

Pour qu'il existe une unique meilleure approximation de tout l'élément d'un espace vectoriel normé E sur \mathbb{R} (ou \mathbb{C}) par un sous-espace vectoriel de dimension finie F quelconque, il faut et il suffit que E soit strictement convexe.

EXERCICE N° 7.

Montrer que le théorème de caractérisation d'une meilleure approximation est encore vrai lorsque les conditions (1), (2) et (3) concernant la forme L sont remplacées par les conditions suivantes, où E est complexe :

$$(1') \quad \|L\| = 1$$

$$(2') \quad \operatorname{Re} (L(z)) = 0 \quad \text{pour tout } z \text{ de } F$$

$$(3') \quad |\operatorname{Re} L(x - y_0)| = \|x - y_0\|$$

EXERCICE N° 8.

Justifier l'assertion suivante, implicitement utilisée dans le texte du Chapitre 1:

x appartient à la fermeture de F si et seulement si $\inf_{z \in F} \|x - z\| = 0$

EXERCICE N° 9

Soit ABC un triangle. Existe-t-il un point P du plan de ce triangle tel que la somme des distances de P aux trois sommets A , B et C soit minimale ? Le point est-il unique ? Construire géométriquement le point P lorsque le triangle a tous ses angles aigus. Le point P s'appelle le point de Fermat.

Peut-on généraliser ces résultats ? (cf Appendice : Et pan dans le mille).

- CHAPITRE 2 -

APPROXIMATION EN MOYENNE QUADRATIQUE

I. UN PROBLEME DE LINEARITE.

Jusqu'à présent, nous avons fixé l'élément f dont on recherchait une meilleure approximation au moyen des éléments du sous-espace vectoriel F . Se plaçant dans un espace normé E justiciable d'un théorème d'unicité de la meilleure approximation (c'est-à-dire pour une norme strictement convexe, Cf exercice N° 6), il est intéressant d'examiner la correspondance établie entre l'élément f et sa meilleure approximation, notée Pf , par les éléments de F . On définit donc l'application P .

$$P: \begin{array}{ccc} f & \longrightarrow & Pf \\ \in E & & \in F \end{array}$$

avec la propriété caractéristique

$$\|f - Pf\| \leq \|f - g\| \quad \text{pour tout } g \text{ dans } F$$

Remarque : en toute rigueur, on devrait noter $P_F(f)$ pour expliciter le rôle de F . Nous ne le ferons pas chaque fois que F aura été clairement choisi et fixé.

La première propriété évidente de cette application P est son idempotence

$$(1) \quad P(Pf) = Pf \quad \text{pour tout } f \text{ dans } E$$

Une deuxième propriété facile est l'homogénéité

$$(2) \quad P(\lambda f) = \lambda P(f) \quad \text{pour tout scalaire } \lambda .$$

En effet, si g est un élément de F et pour un scalaire λ non nul, on a puisque g/λ appartient à F

$$\|\lambda f - g\| = |\lambda| \|f - \frac{g}{\lambda}\| \geq |\lambda| \|f - Pf\| = \|\lambda f - \lambda Pf\|$$

Par suite de l'unicité, on déduit bien

$$P(\lambda f) = \lambda P(f)$$

Une troisième propriété concerne la norme. On a

$$(3) \quad \|Pf\| \leq 2\|f\| \quad \text{pour tout } f \text{ de } E .$$

En effet $Pf = Pf - f + f$ et donc $\|Pf\| \leq \|Pf - f\| + \|f\|$, tandis que

$\|Pf - f\| \leq \|f - 0\| = \|f\|$, puisque $0 \in F$, grâce à la propriété de la meilleure approximation. On peut naturellement chercher à améliorer la constante 2 qui figure dans (3), pour des espaces normés particuliers. Nous reviendrons plus tard sur ce point.

On peut également ajouter l'inégalité suivante, facile à vérifier

$$\| \|f - Pf\| - \|g - Pg\| \| \leq \|f - g\| \quad \text{pour tout } f, g \text{ de } E$$

Il est alors tentant de se poser le problème de la linéarité. En général, on n'a pas l'égalité suivante

$$P(f+g) = Pf + Pg$$

Exemple : Prenons pour E l'espace \mathbb{R}^3 muni de deux normes strictement convexes classiques. Pour tout $x = (x_1, x_2, x_3)$ de \mathbb{R}^3 , on pose

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

et
$$\|x\|_4 = (x_1^4 + x_2^4 + x_3^4)^{\frac{1}{4}}$$

Au Chap. 1 §5, nous avons déjà remarqué les propriétés géométriques différentes des boules relatives à ces deux normes.

Nous cherchons une meilleure approximation de x par les éléments du sous-espace F de \mathbb{R}^3 constitué par l'ensemble des $y = (y_1, y_2, y_3)$ où $y_1 = y_2 = y_3$ (axe ternaire).

Pour la norme $\|\cdot\|_2$, il est facile de voir que $y_1 = y_2 = y_3 = \frac{x_1 + x_2 + x_3}{3}$ réalise la meilleure approximation de $x = (x_1, x_2, x_3)$ par F . Ce sera une conséquence facile des résultats ultérieurs, mais on note directement l'égalité

$$(x_1 - y)^2 + (x_2 - y)^2 + (x_3 - y)^2 =$$

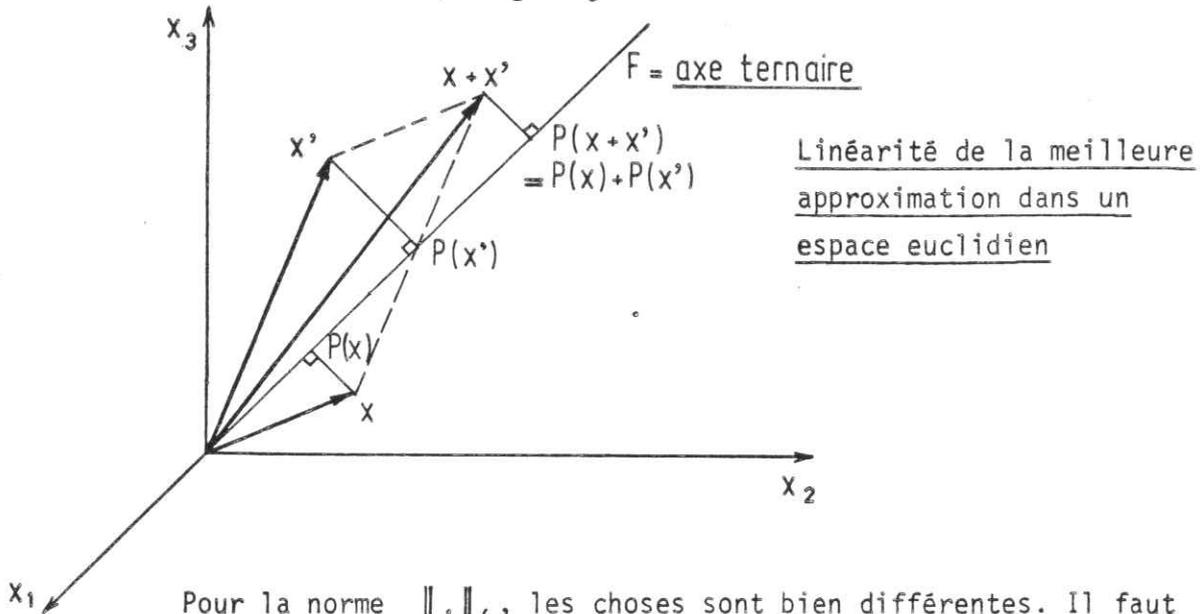
$$(x_1 - \frac{x_1 + x_2 + x_3}{3})^2 + (x_2 - \frac{x_1 + x_2 + x_3}{3})^2 + (x_3 - \frac{x_1 + x_2 + x_3}{3})^2 + 3(y - \frac{x_1 + x_2 + x_3}{3})^2$$

ce qui assure aussitôt le résultat. En outre, l'application $x \rightarrow P(x)$ qui à tout x fait correspondre le point y dont les trois composantes sont égales et valent

$\frac{x_1 + x_2 + x_3}{3}$ est une application linéaire de $E = \mathbb{R}^3$ sur l'espace F

$$P(x+x') = P(x) + P(x')$$

Il ne s'agit d'ailleurs que de la traduction algébrique du théorème de Thalès puisque le point y s'interprète comme la projection orthogonale sur l'axe ternaire du point (x_1, x_2, x_3) .



Pour la norme $\|\cdot\|_4$, les choses sont bien différentes. Il faut minimiser $(x_1 - t)^4 + (x_2 - t)^4 + (x_3 - t)^4$ parmi tous les t réels possibles. Une solution annule la dérivée en t , c'est-à-dire satisfait

$$(x_1 - t)^3 + (x_2 - t)^3 + (x_3 - t)^3 = 0$$

ou encore

$$3t^3 - 3(x_1 + x_2 + x_3)t^2 + 3(x_1^2 + x_2^2 + x_3^2)t - (x_1^3 + x_2^3 + x_3^3) = 0$$

Pour $x = (0, 0, 1)$, ou $x = (0, 1, 0)$ ou encore $x = (1, 0, 0)$ le nombre t convenable est le même par symétrie et satisfait l'équation polynomiale :

$$(1) \quad 3t^3 - 3t^2 + 3t - 1 = 0$$

Pour $x = (1, 1, 1)$, élément de F , le nombre t' convenable est évidemment 1. On voit alors qu'il n'est pas possible que $t' = 3t$ puisque $1/3$ n'est pas solution de l'équation (1). On en déduit qu'il n'y a pas linéarité de la meilleure approximation pour la norme $\|\cdot\|_4$.

Il devient dès lors naturel de se poser la question suivante :

Pour quels types d'espaces normés $(E, \|\cdot\|)$ (strictement convexes) a-t-on la propriété de linéarité de la meilleure approximation par les éléments d'un sous-espace de dimension finie quelconque ? Un théorème répond complètement à cette question.

Théorème 1 : Soit E un espace vectoriel normé (réel ou complexe), dont la dimension si elle est finie est au moins égale à trois. On suppose que pour tout sous-espace F de dimension 1, il existe une unique meilleure approximation $P_F(x)$ de tout $x \in E$ par les éléments de F telle que

$$P_F(x+x') = P_F(x) + P_F(x') \quad \forall x, x' \in E$$

Alors la norme de E provient d'un produit scalaire.

Nous ne ferons pas la démonstration complète. On note d'abord qu'il suffit de faire la démonstration dans le cas où la dimension de E est finie et égale à trois. En effet, si le théorème est vrai pour tout espace normé de dimension trois, on en déduit que tous les sous-espaces G de dimension trois de E ont une norme induite par E qui provient d'un produit scalaire. Cette norme, sur G , satisfait alors l'égalité du parallélogramme (Cf E.A.F.)

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2) \quad \forall x, y \in G$$

Mais dans la relation précédente, il est bien clair que l'on peut remplacer la mention : $\forall x, y \in G$ par la mention $\forall x, y \in E$ puisque deux éléments quelconques sont en jeu (et appartiennent donc toujours à un G de dimension 3). Par suite :

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2) \quad \forall x, y \in E$$

Nous avons alors montré en exercice (Cf. E.A.F ; Exercice N° 14) que cette égalité caractérise les normes provenant d'un produit scalaire.

Il reste donc seulement à envisager le cas d'un espace vectoriel normé de dimension trois, ce que nous laisserons sans démonstration car c'est assez pénible.

Il est également intéressant de noter dans la même veine le théorème suivant :

Théorème 2 Soit E un espace vectoriel normé (réel ou complexe) dont la dimension si elle est finie est au moins égale à trois. On suppose que pour tout sous-espace F de dimension 1, il existe une unique meilleure approximation $P_F(x)$ de tout $x \in E$ par les éléments de F telle que

$$\|P_F(x)\| \leq \|x\| \quad \forall x \in E$$

Alors la norme de E provient d'un produit scalaire.

2. RAPPELS SUR LES ESPACES DE HILBERT.

La linéarité de la meilleure approximation conduit donc aux espaces dont la norme provient d'un produit scalaire, c'est-à-dire aux espaces préhilbertiens. Dans E.A.F., nous avons étudié la meilleure approximation dans un espace préhilbertien par les éléments d'un sous-espace complet quelconque F . Regroupons les résultats obtenus en un théorème

Théorème 3 Soit E un espace préhilbertien et F un sous-espace vectoriel complet de E . Il existe une unique meilleure approximation $P_F(x)$ de tout élément x de E par les éléments de F . En outre P_F est un opérateur linéaire, idempotent, auto-adjoint, et de norme 1 sur E .

L'élément $P_F(x)$ est caractérisé par la propriété de projection orthogonale, à savoir :

$$\langle x - P_F(x), y \rangle = 0 \quad \forall y \in F$$

Dans un espace de Hilbert, non seulement il y a existence et unicité de la meilleure approximation pour les sous-espaces F de dimension finie mais également pour les sous-espaces fermés quelconques. En outre, on dispose d'une propriété caractéristique de la meilleure approximation, propriété qui est de manière facile comme on va le montrer car elle ne fait intervenir que des outils linéaires.

3. PROJECTION ET APPROXIMATION: THEOREME DE CHARACTERISATION.

A titre d'exercice, nous allons reprendre les démonstrations d'E.A.F. sur un espace préhilbertien réel et pour un $F = E_n$ de dimension finie.

Soit \mathcal{H} un espace préhilbertien (réel pour fixer les idées) de produit scalaire $\langle x, y \rangle$ et de norme $\|x\| = \langle x, x \rangle^{1/2}$

Cherchons la meilleure approximation Px d'un élément x de \mathcal{H} par les éléments d'un sous-espace vectoriel $E_n: Px \in E_n$ et pour tout $y \in E_n$, on a

$$(1) \quad \|x - Px\|^2 \leq \|x - (Px + y)\|^2$$

Explicitons les produits scalaires

$$\|x - Px\|^2 \leq \|x - Px\|^2 - 2\langle x - Px, y \rangle + \|y\|^2$$

soit la relation

$$(2) \quad 2 \langle x - Px, y \rangle \leq \|y\|^2$$

Remplaçons dans (2) y par λy ($0 < \lambda < 1$). On a encore :

$$2 \langle x - Px, y \rangle \leq \lambda \|y\|^2$$

λ étant aussi petit que l'on veut, ceci implique que :

$$\langle x - Px, y \rangle \leq 0$$

Si nous faisons le même raisonnement en remplaçant l'élément y de E_n par $(-y)$, encore élément de E_n nous trouvons de même :

$$\langle x - Px, -y \rangle \leq 0$$

Donc :

$$\langle x - Px, y \rangle \geq 0$$

Finalement

$$(3) \quad \langle x - Px, y \rangle = 0 \quad \text{pour tout } y \in E_n .$$

Nous avons donc obtenu la condition nécessaire suivante pour que Px soit une meilleure approximation sur un espace préhilbertien :
la différence entre x et sa projection Px sur E_n est orthogonale à E_n .

Montrons que la condition (3) est aussi suffisante. Il suffit de décomposer la somme $x - (Px) + y$ selon $(x - Px) - y$ en utilisant le théorème de Pythagore

$$\|x - (Px) + y\|^2 = \|x - Px\|^2 + \|y\|^2 > \|x - Px\|^2$$

ce qui termine la démonstration. On appelle Px la projection orthogonale de x sur E_n .

Remarque : Le fait que E_n soit de dimension finie n'a pas été utilisé pour établir la condition nécessaire et suffisante qui précède. Cependant, en dimension infinie, l'existence de la meilleure approximation est un théorème profond (Cf § 2 et E.A.F).

Dégageons un certain nombre de conséquences de cette propriété caractéristique de la meilleure approximation.

1°) Unicité : l'unicité a été démontrée. A titre d'exercice, donnons ici une démonstration particulière. Soient alors deux projections P_1 , P_2 de l'élément x de \mathcal{H} sur l'espace E_n .

$$\text{On a : } \quad \langle x - P_1, y \rangle = 0$$

En particulier :

$$\langle x - P_1, P_2 - P_1 \rangle = 0$$

De même :

$$\langle x - P_2, P_2 - P_1 \rangle = 0$$

Soustrayons membre à membre les deux expressions précédentes :

$$\|P_2 - P_1\|^2 = 0$$

Donc :

$$P_2 = P_1$$

2°) Linéarité : montrons que l'opérateur P de projection orthogonale qui à x dans \mathcal{H} fait correspondre Px dans E_n est linéaire.

En effet, pour tout élément z de E_n , la bilinéarité du produit scalaire fournit :

$$\begin{aligned} \langle \lambda x - \lambda Px, z \rangle &= \lambda \langle x - Px, z \rangle = 0 \\ \langle x + y - (Px + Py), z \rangle &= \langle x - Px, z \rangle + \langle y - Py, z \rangle = 0 \end{aligned}$$

ce qui assure la linéarité d'après la propriété caractéristique de la meilleure approximation.

3°) Idempotence : il est facile de constater que $Px = x$ pour tout élément x de l'espace E_n . Par suite, on dispose de la relation :

$$P(Px) = Px$$

puisque Px appartient à E_n pour tout x dans \mathcal{H} , ce que l'on note encore

$$P^2 = P$$

P est alors dit idempotent.

4°) Symétrie : l'opérateur P est auto-adjoint relativement au produit scalaire. Ceci veut dire que l'on a la relation

$$(4) \quad \langle Px, y \rangle = \langle x, Py \rangle$$

pour tout x et y de \mathcal{H} .

En effet, écrivons x suivant la décomposition $x = x^\perp + Px$

où l'on a $\langle x^\perp, z \rangle = 0$ pour tout z de E_n .

De même pour l'élément y

$$y = y^\perp + Py$$

Reportons dans l'expression $\langle Px, y \rangle$

$$\langle Px, y \rangle = \langle Px, y^\perp \rangle + \langle Px, Py \rangle = \langle Px, Py \rangle$$

5°) Enfin, l'on peut montrer que P est un opérateur continu dont on peut même préciser la norme $\|P\|$. Rappelons que

$$\|P\| = \sup_{x \neq 0} \frac{\|Px\|}{\|x\|}$$

Pour x dans E , on a $Px = x$ donc on a un premier renseignement

$$\|P\| \geq 1$$

D'autre part, d'après l'égalité de Pythagore :

$$\|x\|^2 = \|x^\perp\|^2 + \|Px\|^2$$

ce qui fournit

$$\|Px\|^2 \leq \|x\|^2$$

Donc on en déduit aussitôt

$$\|P\| \leq 1$$

D'où le résultat final :

$$\|P\| = 1$$

Récapitulons : l'opérateur P de projection orthogonale est un opérateur idempotent, linéaire, auto-adjoint, borné et de norme égale à 1.

Remarque : On peut montrer réciproquement que ces propriétés caractérisent les opérateurs de projection orthogonale dans le cas d'un espace de Hilbert réel.

Nous allons maintenant donner l'application au calcul explicite de la projection.

4. APPLICATION AU CALCUL DE LA PROJECTION.

Soit \mathcal{H} un espace préhilbertien réel et E_n un sous-espace vectoriel de dimension finie. Soit e_1, e_2, \dots, e_n une base de E_n . Tout élément y de E_n s'écrit d'une façon unique :

$$y = \sum_{j=1}^n y_j e_j$$

Px est un élément de E_n dont les coordonnées dans la base $\{e_j\}$ sont appelées $\{\tilde{x}_i\}$. On a

$$Px = \sum_{i=1}^n \tilde{x}_i e_i$$

Nous cherchons à expliciter les coordonnées \tilde{x}_i de la projection Px de x sur E_n .

Appliquons la propriété caractéristique d'orthogonalité. On a :

$$\langle x - \sum_{i=1}^n \tilde{x}_i e_i, \sum_{j=1}^n y_j e_j \rangle = 0$$

pour tout n -uplet (y_1, y_2, \dots, y_n) de nombres réels.

Ceci s'écrit en utilisant la bilinéarité du produit scalaire :

$$\sum_{j=1}^n y_j \left[\sum_{i=1}^n \tilde{x}_i \langle e_i, e_j \rangle - \langle x, e_j \rangle \right] = 0$$

En prenant successivement $y_j = 1, y_k = 0 (k \neq j)$ pour $j = 1, \dots, n$, on obtient le système (5) de n équations linéaires pour les inconnues \tilde{x}_i :

$$(5) \quad \sum_{i=1}^n \tilde{x}_i \langle e_i, e_j \rangle = \langle x, e_j \rangle \quad (\text{pour } j = 1, \dots, n)$$

Le système (5) permet donc de calculer les coordonnées \tilde{x}_i de la projection Px de x sur E_n à condition que le déterminant de ce système linéaire ne soit pas nul. Mais nous savons, d'après les théorèmes généraux du Chapitre 1, que la projection existe et est unique : le système (5) a donc une solution unique. On peut d'ailleurs le vérifier directement (Cf. Exercice N° 14).

Rappelons que dans le système (5), la base (e_1, e_2, \dots, e_n) est connue, ainsi que le vecteur x . En écriture développée

$$(5) \quad \left\{ \begin{array}{l} \tilde{x}_1 \langle e_1, e_1 \rangle + \tilde{x}_2 \langle e_2, e_1 \rangle + \dots + \tilde{x}_n \langle e_n, e_1 \rangle = \langle x, e_1 \rangle \\ \tilde{x}_1 \langle e_1, e_2 \rangle + \tilde{x}_2 \langle e_2, e_2 \rangle + \dots + \tilde{x}_n \langle e_n, e_2 \rangle = \langle x, e_2 \rangle \\ \vdots \\ \tilde{x}_1 \langle e_1, e_n \rangle + \tilde{x}_2 \langle e_2, e_n \rangle + \dots + \tilde{x}_n \langle e_n, e_n \rangle = \langle x, e_n \rangle \end{array} \right.$$

On a donc un système de Cramer dont la résolution est maintenant une affaire de calcul.

Nous reviendrons d'ailleurs plus loin sur le déterminant de ce système linéaire (déterminant de Gram)

$$\begin{vmatrix} \langle e_1, e_1 \rangle & \dots & \langle e_n, e_1 \rangle \\ \vdots & & \vdots \\ \langle e_n, e_1 \rangle & \dots & \langle e_n, e_n \rangle \end{vmatrix}$$

Lorsque ce déterminant est proche de zéro (mais bien sûr différent de zéro), le calcul numérique même sur machine est délicat.

Exemple : soit à chercher la meilleure approximation en moyenne quadratique de la fonction e^x sur $[0, 1]$ par un polynôme de degré $n-1$ au plus. On doit d'abord calculer les coefficients du système (5) :

$$\langle e_i, e_j \rangle = \int_0^1 x^{i-1} x^{j-1} dx = \frac{1}{i+j-1}$$

Nous avons choisi pour base de l'espace des polynômes de degré $(n-1)$ au plus (dimension n) la famille

$$e_i : x \longrightarrow x^{i-1} \quad \text{pour } i = 1, 2, \dots, n$$

Il faut calculer le second membre du système (5) avec la fonction e^x

$$\int_0^1 e^x x^{j-1} dx = I_j$$

Une intégration par parties fournit facilement la relation de récurrence qui lie I_j et I_{j+1}

$$I_{j+1} = e^{-j} I_j \quad \text{pour } j > 1$$

En écrivant I_j sous la forme $I_j = \alpha_j e + \beta_j$ où α_j et β_j sont des nombres entiers, on dispose donc de

$$\beta_{j+1} + j\beta_j = 0 \quad \text{et} \quad \alpha_{j+1} + j\alpha_j = 1$$

d'où $\beta_j = (-1)^j \frac{1}{j-1!}$ tandis que α_j se détermine par une méthode de variation des constantes. On calcule :

$$I_j = (-1)^{j-1} j^{-1}! \left(1 + \sum_{k=2}^j \frac{(-1)^{k-1}}{k-1!} \right) e + (-1)^j j^{-1}!$$

ou encore en désignant par S_j la somme partielle jusqu'à l'ordre $(j-1)$ du développement de la fonction e^x en série entière, développement calculé au point $x = -1$

$$I_j = (-1)^{j-1} j^{-1}! (e S_j - 1)$$

En particulier $I_1 = e - 1$, $I_2 = 1$, $I_3 = e - 2$ et $I_4 = 6 - 2e$

Calculons maintenant le déterminant de Gram du système

$$\Delta_n = \begin{vmatrix} 1 & \frac{1}{2} & \text{---} & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \text{---} & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \text{---} & \frac{1}{2n-1} \end{vmatrix}$$

Ce déterminant est un déterminant appelé persymétrique car sur toute droite orthogonale à la diagonale principale tous les éléments sont égaux. Un calcul astucieux, que nous laissons au lecteur courageux à titre d'exercice, conduit à :

$$\begin{aligned} \Delta_n &= \frac{(\overline{n-1!})^2 (\overline{n-2!})^2 \text{---} (1!)^2}{[1, 2 \text{---} n][2 \text{---} (n+1)] \text{---} [n(n+1) \text{---} (2n-1)]} \\ &= \frac{(\overline{n-1!})^3 (\overline{n-2!})^3 \text{---} (2!)^3 (1!)^3}{n!(\overline{n+1!}) \text{---} (2n-1!)} \end{aligned}$$

En particulier, on obtient

$$\Delta_1 = 1 \quad ; \quad \Delta_2 = \frac{1}{12} \quad ; \quad \Delta_3 = \frac{1}{2160}$$

Par suite, la meilleure approximation de e^x en moyenne quadratique par un polynôme de degré 1 au plus sur le segment $[0,1]$, c'est-à-dire pour la norme

$\left(\int_0^1 |f(x)|^2 dx \right)^{\frac{1}{2}}$ est donnée par $a+bx$, où a et b s'obtiennent en faisant $n = 2$

dans les formules précédentes. Soit

$$a = \frac{1}{\Delta_2} \begin{vmatrix} e-1 & \frac{1}{2} \\ 1 & \frac{1}{3} \end{vmatrix} = 4e - 10 \quad \text{et} \quad b = -\frac{1}{\Delta_2} \begin{vmatrix} 1 & e-1 \\ \frac{1}{2} & 1 \end{vmatrix} = 18 - 6e$$

soit avec trois décimales significatives .

$$(0,865) + (1,691)x$$

est la meilleure approximation cherchée.

Malheureusement, résoudre un système tel que le précédent est techniquement difficile, même sur machine, car le déterminant de Gram est très proche de zéro, d'autant plus proche que n devient grand. Un tel système est dit mal conditionné et une erreur d'approximation sur un terme du déterminant est grandement amplifiée pour le calcul final. De fait, la quasi-nullité de ce déterminant provient de ce que deux vecteurs colonnes sont quasiment colinéaires et ce d'autant plus que n augmente. En termes vectoriels, cela revient à dire que $e_i = x^{i-1}$ et $e_{i+1} = x^i$ sont "très peu indépendants" sur $L^2[-1,+1]$, comme d'ailleurs l'explique le théorème de Müntz (Cf. Chapitre 4). La base utilisée est donc très mal choisie et il faut en changer. Il importe de remarquer que dans le système (5), l'élément à approximer se trouve au second membre seulement, tandis que le déterminant de Gram ne dépend que de la base choisie. Cela incite à changer de base, et ceci est possible d'une façon intéressante.

5. UTILISATION DES POLYNOMES ORTHOGONAUX ET DES BASES ORTHONORMALES.

5.1. Bases orthonormales.

Le système (5) est aisément résoluble dès que le déterminant ne contient de termes non nuls que sur la diagonale principale, c'est-à-dire si l'on a

$$\langle e_i, e_j \rangle = 0 \quad \text{pour tout} \quad i \neq j .$$

Une famille $\{e_i\}$ possédant une telle propriété est dite orthogonale (Cf E.A.F.). Dans le cas d'une famille orthogonale le système (5) devient

$$\tilde{x}_j \langle e_j, e_j \rangle = \langle x, e_j \rangle \quad j = 1, \dots, n$$

d'où la solution immédiate

$$x_j = \frac{\langle x, e_j \rangle}{\langle e_j, e_j \rangle} \quad j = 1, \dots, n$$

lorsque l'on a en plus $\langle e_j, e_j \rangle = 1$ pour tout j , on dit que la famille est orthonormale (Cf A.E.F.)

Remarque : Nous avons fait les calculs dans le cas d'un espace préhilbertien réel. On peut aussi travailler sur un espace préhilbertien complexe en faisant attention à l'ordre des produits scalaires.

Soit \mathcal{E} un espace préhilbertien (réel ou complexe) et F un sous-espace vectoriel de dimension finie n . Soit $\{e_i\}_{i=1, \dots, n}$ une base orthonormale de F .

La meilleure approximation $P_F(x)$ d'un élément quelconque de \mathcal{E} par les éléments de F est donnée par

$$P_F(x) = \sum_{i=1}^n \langle x, e_i \rangle e_i$$

5.2. Exemple de l'analyse harmonique.

Considérons l'espace de Hilbert $L^2[-\pi, +\pi]$ des fonctions 2π périodiques à valeurs complexes et de carré intégrable (Cf E.A.F), muni du produit scalaire

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x) \overline{g(x)} dx$$

La famille $\{e_n\}_{n \in \mathbb{Z}}$ où $e_n : x \longrightarrow e^{inx}$

est une famille orthonormale sur $L^2[-\pi, +\pi]$. C'est d'ailleurs une base hilbertienne de l'espace $L^2[-\pi, +\pi]$, comme nous l'avons démontré dans les éléments d'analyse fonctionnelle. Appelons F le sous-espace vectoriel de dimension finie des combinaisons linéaires des fonctions exponentielles imaginaires e_k pour k variant de $-n$ à $+n$. Cet espace est de dimension $(2n + 1)$. Soit f un élément quelconque de $L^2[-\pi, +\pi]$. D'après ce qui précède, la meilleure approximation de f au sens de $L^2[-\pi, +\pi]$ par les éléments de F est donnée par

$$P_F(f) = \sum_{k=-n}^{k=+n} \langle f, e_k \rangle e_k$$

Soit

$$P_F(f)(x) = \sum_{k=-n}^{k=+n} c_k(f) e^{ikx}$$

où l'on a posé

$$c_k(f) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x) e^{-ikx} dx$$

Coefficient de Fourier d'ordre k de la fonction f .

Par suite $P_F(f)$ n'est autre que la série de Fourier de f tronquée à l'ordre n .

L'analyse harmonique, c'est-à-dire l'étude des séries de Fourier, est donc en particulier l'étude des meilleures approximations en moyenne quadratique et nous renvoyons le lecteur aux ouvrages spécialisés sur ce sujet.

Il importe de remarquer que l'on a une excellente idée de la différence entre f et sa meilleure approximation, grâce à l'égalité de Parseval (Cf. E.A.F.)

$$\|f - P_F(f)\|_2^2 = \sum_{|k| > n} |c_k(f)|^2$$

L'étude du reste, c'est-à-dire de $f - P_F(f)$, sera d'ailleurs entreprise dans un chapitre spécial (Cf Chapitre 4).

Donnons, pour être pratique, les formules relatives au cas d'une fonction f , de période T et supposée à valeurs réelles. On suppose que f appartient à $L^2[-\frac{T}{2}, +\frac{T}{2}]$, lequel est muni du produit scalaire $\langle f, g \rangle = \int_{-\frac{T}{2}}^{+\frac{T}{2}} f(x) \overline{g(x)} dx$

La meilleure approximation de f par l'espace F des combinaisons linéaires des fonctions $(1, \cos \frac{2\pi}{T}x, \sin \frac{2\pi}{T}x, \dots, \cos \frac{2\pi}{T}nx, \sin \frac{2\pi}{T}nx)$ est donnée par :

$$P_F(f)(x) = \frac{a_0}{2} + \sum_{k=1}^{k=n} (a_k \cos k \frac{2\pi}{T}x + b_k \sin k \frac{2\pi}{T}x)$$

où

$$a_k = \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} f(x) \cos k \frac{2\pi}{T}x dx \quad k > 0$$

et

$$b_k = \frac{2}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} f(x) \sin k \frac{2\pi}{T}x dx \quad k \geq 1$$

et on a la valeur du reste :

$$\|P_F(f) - f\|_2^2 = \frac{1}{2} \sum_{k > n} (a_k^2 + b_k^2)$$

5. 5.3. Orthogonalisation et polynômes orthogonaux.

Lorsque l'espace préhilbertien \mathcal{H} est donné, ainsi que le sous-espace F , le paragraphe précédent montre que l'on a avantage à doter F d'une base orthonormale lorsque l'on veut faire de l'approximation dans F .

Nous avons vu dans E.A.F qu'à partir d'une base $\{e_i\}$ de F , on peut toujours construire une base orthonormale $\{e'_i\}$ par le procédé d'orthonormalisation de Schmidt. Naturellement le calcul explicite de $\{e'_i\}$ à partir de $\{e_i\}$ n'évite pas les éventuels problèmes de mauvais conditionnement. Cependant, il existe des tables de bases orthonormales pour les espaces fonctionnels de Hilbert usuels et le calcul est donc tout fait à l'avance, Il ne dépend bien entendu pas de la fonction. On peut donner des formules théoriques pour l'orthonormalisation de Gram-Schmidt (se reporter au corrigé de l'Exercice N° 34 des Eléments d'Analyse Fonctionnelle).

Pour $n > 1$

$$e'_n = \frac{1}{\sqrt{G(e_1, \dots, e_n) G(e_1, \dots, e_{n-1})}}$$

$$\begin{vmatrix} \langle e_1, e_1 \rangle & \dots & \langle e_n, e_1 \rangle \\ \vdots & & \vdots \\ \langle e_1, e_{n-1} \rangle & \dots & \langle e_n, e_{n-1} \rangle \\ e_1 & \dots & e_n \end{vmatrix}$$

où le développement de ce déterminant est formel et s'entend suivant la dernière ligne (qui est composée de vecteurs et non de nombres). En outre pour

$n = 1$, on a : $e'_1 = \frac{e_1}{\|e_1\|}$

et on a posé $G(e_1, \dots, e_k) = \begin{vmatrix} \langle e_1, e_1 \rangle & \dots & \langle e_k, e_1 \rangle \\ \vdots & & \vdots \\ \langle e_k, e_1 \rangle & \dots & \langle e_k, e_k \rangle \end{vmatrix}$

(déterminant de Gram du système e_1, \dots, e_k).

Ces formules théoriques ne sont pas d'un usage pratique. Un exemple fréquent est l'orthonormalisation de la base usuelle $(1, x, \dots, x^n, \dots)$ pour un produit scalaire donné. On a déjà montré (Cf. E.A.F) qu'il était possible de construire une base orthonormale de polynômes $(P_0(x), P_1(x), \dots,$

$P_n(x), \dots)$. Les propriétés de ces polynômes ne se constatent pas sur les formules précédentes mais par une étude directe. Nous renvoyons le lecteur à E.A.F. pour l'étude des principales propriétés des polynômes orthogonaux (relations de récurrence, équations différentielles et problèmes de valeurs propres à la Sturm-Liouville, propriétés d'alternances des racines, etc.).

Exemple : Soit f un élément de $L^2 [-1,+1]$. Cherchons la meilleure approximation de f au sens de la norme

$$\|f\|^2 = \int_{-1}^{+1} |f(x)|^2 dx$$

au moyen de polynômes de degré inférieur ou égal à n . L'espace F de tels polynômes est de dimension $(n+1)$ et une base orthonormale est fournie par les polynômes de Legendre (du moins à un multiple près : Cf. E.A.F.)

$$P_0(x) = \frac{1}{\sqrt{2}}$$

$$P_1(x) = \sqrt{\frac{3}{2}} x$$

$$P_2(x) = \sqrt{\frac{5}{2}} \cdot \frac{3x^2-1}{2}$$

$$P_3(x) = \sqrt{\frac{7}{2}} \cdot \frac{5x^3-3x}{2} \quad \text{etc ---}$$

Par suite, la meilleure approximation $P_F(f)$ de f est fournie par

$$(1) \quad P_F(f)(x) = \sum_{k=0}^n \langle f, P_k \rangle P_k(x)$$

$$\text{où} \quad \langle f, P_k \rangle = \int_{-1}^{+1} f(x) P_k(x) dx$$

On peut écrire (1) sous une autre forme

$$\begin{aligned} \sum_{k=0}^n \langle f, P_k \rangle P_k(x) &= \sum_{k=0}^n P_k(x) \int_{-1}^{+1} f(y) P_k(y) dy \\ &= \int_{-1}^{+1} f(y) \left(\sum_{k=0}^n P_k(x) P_k(y) \right) dy \end{aligned}$$

Mais nous avons calculé déjà l'expression $\sum_{k=0}^n P_k(x) P_k(y) = K_n(x,y)$

(formule de Darboux-Christoffel).

$$K_n(x,y) = \frac{a_n}{a_{n-1}} \frac{P_{n+1}(x) P_n(y) - P_{n+1}(y) P_n(x)}{x-y}$$

où a_n désigne le coefficient du terme de degré n dans P_n . Finalement

$$P_F(f)(x) = \frac{a_n}{a_{n-1}} \int_{-1}^{+1} \left(\frac{P_{n+1}(x) P_n(y) - P_{n+1}(y) P_n(x)}{x-y} \right) f(y) dy$$

Rappelons que $\frac{a_n}{a_{n-1}} = \frac{1}{A_n}$, où A_n est le coefficient intervenant dans

la relation de récurrence

$$P_n(x) = (A_n x + B_n) P_{n-1}(x) - C_n P_{n-2}(x)$$

On sait que $A_n = \frac{2n-1}{n}$ dans le cas des polynômes de Legendre, ce qui

fournit en définitive

$$P_F(f)(x) = \frac{n}{2n-1} \int_{-1}^{+1} \left(\frac{P_{n+1}(x) P_n(y) - P_{n+1}(y) P_n(x)}{x-y} \right) f(y) dy$$

Exemple : cherchons la meilleure approximation de la fonction $f(x) = e^x$ par un polynôme de degré 2 au plus et au sens de $L^2[-1, +1]$, pour la norme

$$\|f\|_2^2 = \int_{-1}^{+1} |f(x)|^2 dx$$

Il vient

$$\begin{aligned} P_F(f)(x) &= \frac{2}{3} \sqrt{\frac{35}{4}} \int_{-1}^{+1} \frac{(5x^3 - 3x)(3y^2 - 1) - (5y^3 - 3y)(3x^2 - 1)}{4(x-y)} e^y dy \\ &= \frac{\sqrt{35}}{12} \int_{-1}^{+1} (15x^2 y^2 - 5(x^2 + y^2) - 14xy + 3) e^y dy \end{aligned}$$

Tous calculs faits, on trouve

$$= \frac{15}{4} \left(e - \frac{7}{e} \right) x^2 - \frac{3}{e} x + \frac{33}{4} \frac{1}{e} - \frac{3}{4} e$$

Soit avec trois décimales

$$P_F(f)(x) \sim 0,539x^2 + 1,104x + 0,996$$

En général, on vérifiera l'égalité permettant de mesurer la qualité de l'approximation faite

$$f(x) - P_F(f)(x) = \frac{a_n}{a_{n-1}} \int_{-1}^{+1} (P_{n+1}(x) P_n(y) - P_{n+1}(y) P_n(x)) \frac{f(x) - f(y)}{x-y} dy$$

- EXERCICES DU CHAPITRE 2 -

On aura avantage à se reporter également aux exercices du Chapitre 3 de E.A.F.

EXERCICE N° 10.

On dit qu'un élément x d'un espace normé E est orthogonal à un élément y de E si l'on a

$$\|x + \lambda y\| \geq \|x\| \quad \text{pour tout scalaire } \lambda$$

Montrer que g_0 est une meilleure approximation de $f \in E$ par les éléments de F (avec $f \in E \setminus \overline{F}$) où F est un sous-espace vectoriel de E , si et seulement si $f - g_0$ est orthogonal à tout élément g de F .

EXERCICE N° 11.

Démontrer la généralisation suivante du théorème 1 de ce chapitre :

Soit E un espace vectoriel normé, dont la dimension, si elle est finie, est au moins égale à trois. Soit n un entier fixe tel que $1 \leq n \leq (\dim E) - 2$. Supposons que pour tout sous-espace vectoriel F de dimension n , il existe une unique meilleure approximation $P_F(x)$ de tout $x \in E$ par les éléments de F telle que

$$P_F(x + x') = P_F(x) + P_F(x') \quad \forall x, x' \in E$$

Alors la norme de E provient d'un produit scalaire.

EXERCICE N° 12.

Démontrer le théorème suivant :

Soit E un espace vectoriel normé, dont la dimension, si elle est finie, est au moins égale à trois. Soit n un entier fixe tel que $1 \leq n \leq (\dim E) - 2$. Supposons que pour tout sous-espace vectoriel F de dimension n , il existe une unique meilleure approximation $P_F(x)$ de tout $x \in E$ par les éléments de F , de sorte que

$$\|P_F(x)\| \leq \|x\| \quad \forall x \in E$$

Alors la norme de E provient d'un produit scalaire.

(remarque : en général on a seulement $\|P_F(x)\| \leq 2\|x\|$ (Cf. cours)).

EXERCICE N° 13.

Soit E un espace vectoriel normé et F un sous-espace de dimension finie tel que tout $x \in E$ ait une unique meilleure approximation $P_F(x)$ par les éléments de F . Montrer que $I - P_F : E \rightarrow E$ est une application continue

(remarque : P_F n'est pas une application linéaire en général).

EXERCICE N° 14.

1°) Montrer directement que le système (5) (Cf ce chapitre § 4) possède une unique solution

$$(5) \quad \sum_{i=1}^n \tilde{x}_i \langle e_i, e_j \rangle = \langle x, e_j \rangle \quad j = 1, \dots, n$$

(par l'absurde, on supposera que le système homogène possède une solution non nulle $(\xi_1, \xi_2, \dots, \xi_n)$ et on additionnera les n équations obtenues préalablement multipliées par ξ_i).

2°) En déduire une condition nécessaire et suffisante pour que n fonctions, réelles pour fixer les idées, f_1, f_2, \dots, f_n de $L^2[a, b]$ soient linéairement indépendantes

(il faut et il suffit que le déterminant de Gram d'ordre n , dont le terme général est $\int_b^a f_i(x) f_j(x) dx$, soit différent de zéro).

EXERCICE N° 15.

Soient f et g deux éléments de $L^2[-\pi, +\pi]$. Donner une expression intégrale pour la fonction

$$F(x) = 2\pi \sum_{k=-\infty}^{+\infty} C_k(f) C_k(g) e^{ikx}$$

où

$$C_k(f) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x) e^{-ikx} dx$$

Etudier l'application $f \rightarrow F$ lorsque g est fixé.

(on trouvera $F(x) = \int_{-\pi}^{+\pi} f(x-t)g(t)dt$ convolution de f et de g , opération étudiée au Chapitre 5).

EXERCICE N° 16.

Soit f une fonction de période 2π et appartenant à $L^2[-\pi, +\pi]$
Calculer la série de Fourier de la fonction

$$F_N(x) = \frac{1}{N} \sum_{k=0}^{N-1} f\left(x + k \frac{2\pi}{N}\right)$$

EXERCICE N° 17.

Caractériser l'ensemble des opérateurs P , définis sur $L^2[-\pi, +\pi]$
et à valeurs dans $L^2(-\pi, +\pi)$, qui possèdent les trois propriétés suivantes :

- (1) P est un opérateur linéaire
- (2) P est un opérateur continu
- (3) P est un opérateur stationnaire, c'est-à-dire que si $U_k(f)$ désigne la fonction translatée de f de pas k :

$$U_k(f)(x) = f(x-k)$$

alors

$$P(U_k(f)) = U_k(P(f)) \quad \text{pour tout } k \in \mathbb{R} .$$

EXERCICE N° 18.

Calculer les coefficients de Fourier généralisés de la fonction $f(x) = e^x$ relativement aux polynômes de Legendre normalisés de degré 0, 1 et 2.

Retrouver la meilleure approximation de la fonction $f(x) = e^x$
au sens de la norme $\|f\|^2 = \int_{-1}^{+1} |f(x)|^2 dx$

(On trouve $c_0 = \sqrt{\frac{1}{2}} \left(e - \frac{1}{e}\right)$; $c_1 = \sqrt{\frac{3}{2}} \left(\frac{2}{e}\right)$ et $c_2 = \frac{3\sqrt{10}}{4} \left(\frac{2e - \frac{14}{e}}{3}\right)$)

et donc

$$\begin{aligned} P_f(x) &= \sqrt{\frac{1}{2}} \left(e - \frac{1}{e}\right) \sqrt{\frac{1}{2}} + \sqrt{\frac{3}{2}} \frac{2}{e} \sqrt{\frac{3}{2}} x + \frac{3}{4} \sqrt{10} \left(\frac{2e - \frac{14}{e}}{3}\right) \frac{3}{4} \sqrt{10} \left(x^2 - \frac{1}{3}\right) \\ &= \frac{15}{4} \left(e - \frac{7}{e}\right) + \frac{3}{e} x + \frac{90}{16} \left(\frac{2e - \frac{14}{e}}{3}\right) (x^2 - 1) \end{aligned}$$

EXERCICE N° 19.

Soit e_1, e_2, \dots, e_n , n éléments d'un espace préhilbertien \mathcal{E} .
On appelle déterminant de Gram de ces éléments, le déterminant

$$G(e_1, \dots, e_n) = \begin{vmatrix} \langle e_1, e_1 \rangle & \dots & \langle e_1, e_n \rangle \\ \vdots & & \vdots \\ \langle e_n, e_1 \rangle & \dots & \langle e_n, e_n \rangle \end{vmatrix}$$

Montrer que

$$0 \leq G(e_1, \dots, e_n) \leq \|e_1\|^2 \dots \|e_n\|^2$$

Discuter les cas d'égalité.

(Il y a nullité si et seulement si les éléments e_1, \dots, e_n sont linéairement indépendants, l'autre égalité si et seulement si les éléments e_1, \dots, e_n sont orthogonaux).

EXERCICE N° 20.

Soit ABC un triangle en géométrie plane. Montrer qu'il existe un point M du plan tel que la somme des carrés des distances de M aux côtés du triangle soit minimale et donner une propriété géométrique de ce point.

EXERCICE N° 21.

Soit \mathcal{E} un espace préhilbertien et x_1, x_2, \dots, x_n , n vecteurs linéairement indépendants. Soit x un vecteur quelconque de \mathcal{E} . Montrer que la distance Δ de x au sous-espace vectoriel engendré par x_1, x_2, \dots, x_n est donnée par

$$\Delta = \sqrt{\frac{G(x_1, x_2, \dots, x_n, x)}{G(x_1, x_2, \dots, x_n)}}$$

où l'on a posé (déterminant de Gram : Cf Exercice N°19)

$$G(x_1, \dots, x_k) = \begin{vmatrix} \langle x_1, x_1 \rangle & \dots & \langle x_1, x_k \rangle \\ \vdots & & \vdots \\ \langle x_k, x_1 \rangle & \dots & \langle x_k, x_k \rangle \end{vmatrix}$$

(on écrira la relation de Pythagore

$$\Delta^2 + \langle y, y \rangle = \langle x, x \rangle$$

où y désigne la projection orthogonale de x sur le sous-espace engendré par x_1, x_2, \dots, x_n . On a aussi

$$\Delta^2 + \langle x, y \rangle = \langle x, x \rangle$$

On ajoute cette équation aux n équations linéaires fournissant y dans la base x_1, x_2, \dots, x_n et ensuite on écrit que le système d'ordre $(n+1)$ est de Cramer).

EXERCICE N° 22.

Soit P_n le $n^{\text{ième}}$ polynome de Legendre orthonormalisé. Montrer que

$P_n(0) = 0$ pour n impair et

$$P_{2n}(0) = (-1)^n \frac{1 \cdot 3 \cdot 5 \dots 2n-1}{2 \cdot 4 \dots 2n}$$

- CHAPITRE 3 -

LES FONCTIONS SPLINES

Dans ce Chapitre nous allons envisager une technique d'approximation dont l'origine est très intuitive. Une fonction f , représentant un certain phénomène physique, est en général connue par les valeurs prises par f en un certain nombre de points, mais on ne connaît pas exactement la valeur de f en tout point. En général, sur un dessin donnant les points connus de f , on interpole la fonction f en cherchant une courbe passant au mieux par les points donnés. L'interpolation peut se faire à la règle : on joint par un segment deux points d'abscisses consécutives, et il s'agit de l'interpolation linéaire. L'interpolation peut encore se faire avec un instrument appelé "pistolet" en français (spline en anglais) où la courbure joue son rôle. Si l'on peut donner une expression mathématique à l'opération mécanique ainsi faite, et ce n'est qu'affaire de technique, il paraît normal d'obtenir une approximation de la fonction f . Approximation en quel sens ? Il se trouve, qu'une fois encore, nous allons trouver une approximation minimisant une certaine expression quadratique.

1. INTERPOLATION LINEAIRE D'UNE FONCTION.

Pour fixer les idées, soit $[a, b]$ un segment de l'axe réel et $x_1 = a, x_2, \dots, x_n = b$, n points donnés distincts appartenant à ce segment. On qualifiera cette donnée de division Δ_n (d'ordre n) de $[a, b]$.

Définition : On dit qu'une fonction $g: [a, b] \rightarrow \mathbb{R}$ est linéaire par morceaux pour la division Δ_n si f est continue et linéaire sur tout segment $[x_k, x_{k+1}]$ pour $k = 1, 2, \dots, (n-1)$.

Appelons E_n l'espace des fonctions linéaires par morceaux pour la division Δ_n . Il s'agit d'un espace vectoriel dont la dimension vaut n précisément.

Appelons $g(x_k)$ les valeurs de g aux points de la division Δ_n . On obtient une expression analytique pour la valeur en x d'une fonction g de E_n .

$$g(x) = \sum_{k=1}^{n-1} \chi_k(x) \left[\frac{x_{k+1} - x}{x_{k+1} - x_k} f(x_k) + \frac{x - x_k}{x_{k+1} - x_k} f(x_{k+1}) \right]$$

où $\chi_k(x)$ est la fonction caractéristique du segment $[x_k, x_{k+1}]$.

On écrit encore

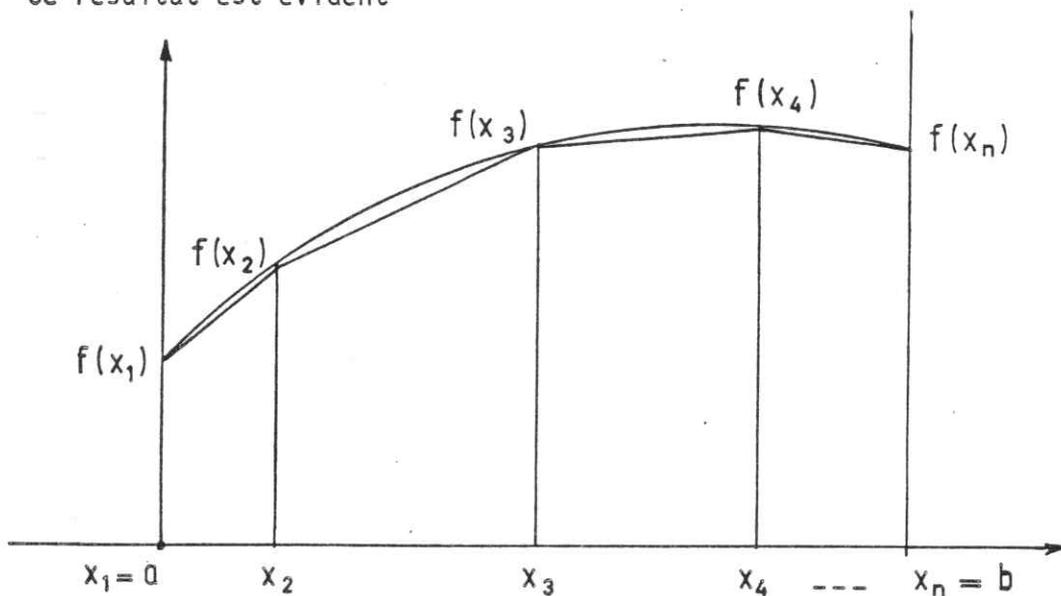
$$g(x) = \sum_{k=1}^n \left[\frac{x_{k+1}-x}{x_{k+1}-x_k} \chi_k(x) + \frac{x-x_{k-1}}{x_k-x_{k-1}} \chi_{k-1}(x) \right] f(x_k)$$

en convenant de noter $\chi_0(x) \equiv 0$ et $\chi_n(x) \equiv 0$.

Théorème : Soit f une fonction définie sur $[a,b]$ et à valeurs réelles et Δ_n une division de $[a,b]$. Il existe une unique fonction g linéaire par morceaux pour la division Δ_n et telle que

$$f(x_k) = g(x_k) \quad \forall k = 1, 2, \dots, (n-1)$$

Ce résultat est évident



Pour fixer les notations, appelons $P_n(f)$ la fonction g linéaire par morceaux pour la division Δ_n associée à f par le théorème précédent. Quelques propriétés de l'application $f \longrightarrow P_n(f)$ sont évidentes. On supposera que f appartient à un espace fonctionnel, disons $\mathcal{C}[a,b]$ l'ensemble des fonctions continues sur $[a,b]$, muni de la norme uniforme. Dès lors, $P_n(f)$ est un élément de E_n .

(1) P_n est linéaire

(2) $\|P_n f\| \leq \|f\|$ et même $\|P_n\| = 1$

$$(3) \quad P_n(1) = 1$$

(4) $P_n > 0$ c'est-à-dire si $f(x) > 0$ pour tout $x \in [a, b]$, on a également $P_n f(x) > 0$ pour tout $x \in [a, b]$.

$$(5) \quad P_n(f \cdot P_n g) = P_n(f \cdot g) \quad \forall f, g \in \mathcal{C}[a, b]$$

Cherchons maintenant l'expression analytique de $P_n(f)$:

$$P_n(f)(x) = \sum_{k=1}^{n-1} \chi_k(x) \left[\frac{x_{k+1} - x}{x_{k+1} - x_k} f(x_k) + \frac{x - x_k}{x_{k+1} - x_k} f(x_{k+1}) \right]$$

ou encore

$$P_n(f)(x) - f(x) = \sum_{k=1}^{n-1} \chi_k(x) \left[\frac{x_{k+1} - x}{x_{k+1} - x_k} (f(x_k) - f(x)) + \frac{x - x_k}{x_{k+1} - x_k} (f(x_{k+1}) - f(x)) \right]$$

Une majoration du reste est donc possible lorsque l'on suppose que f est une fonction dérivable et dont la dérivée est bornée : nous notons $\|f'\|_\infty = \sup_{x \in [a, b]} |f'(x)|$

Dès lors, en notant selon l'usage

$$\|P_n(f) - f\|_\infty = \sup_{x \in [a, b]} |P_n(f)(x) - f(x)|$$

et en appliquant deux fois le théorème des accroissements finis

$$|P_n(f)(x) - f(x)| \leq \sum_{k=1}^{n-1} \chi_k(x) \|f'\|_\infty \left(\frac{x_{k+1} - x}{x_{k+1} - x_k} (x - x_k) + \frac{(x - x_k)(x_{k+1} - x)}{x_{k+1} - x_k} \right)$$

Puis en majorant $x - x_k$ par $x_{k+1} - x_k$ et $x_{k+1} - x$ par $x_{k+1} - x_k$:

$$\leq \sum_{k=1}^{n-1} \chi_k(x) \|f'\|_\infty \left((x_{k+1} - x) + (x - x_k) \right)$$

Notons également $|\Delta_n|$, pas de la division Δ_n , le maximum de l'écart entre deux points consécutifs de la division

$$|\Delta_n| = \sup_{k=1, \dots, n-1} (x_{k+1} - x_k)$$

Il vient (pour une meilleure inégalité, Cf Exercice N° 31)

$$\|P_n(f) - f\|_\infty \leq \|f'\|_\infty |\Delta_n|$$

Mais $P_n(f)$ ne réalise pas en général une meilleure approximation de f par les

éléments de $P_n(f)$ au sens de la convergence uniforme (Cf Exercice N° 23).
Nous verrons d'ailleurs au Chapitre 4 ce que l'on peut dire lorsque $|\Delta_n|$ tend vers 0 pour n tendant vers l'infini.

2. INTERPOLATION POLYNOMIALE D'UNE FONCTION.

Au lieu d'interpoler par des fonctions linéaires par morceaux, nous allons maintenant interpoler par des polynômes. Le théorème suivant est bien connu.

Théorème de Lagrange : Soit Δ_n une division d'un segment $[a,b](x_1 = a, \dots, x_n = b)$ et soit f une fonction à valeurs réelles définie sur le segment $[a,b]$. Il existe un unique polynôme, noté $P_n(f)$ de degré au plus $(n-1)$, tel que

$$P_n(f)(x_k) = f(x_k) \quad \text{pour } k = 1, 2, \dots, n$$

Au lieu d'un calcul élémentaire bien connu du lecteur, donnons une démonstration susceptible de généralisation. Supposons que f appartienne à un espace vectoriel E réel. Pour chaque $x_k, k = 1, 2, \dots, n$, l'application

$$\delta_{x_k} : f \longrightarrow f(x_k)$$

est une forme linéaire sur E et en particulier une forme linéaire sur E_n , l'espace de dimension n des polynômes réels de degré $(n-1)$ au plus. Ces n formes linéaires sont linéairement indépendantes sur le dual E_n^* dont la dimension est également n (Cf E.A.F. Chapitre 0, Exercice N° 6). En outre E_n^{**} s'identifie canoniquement à E_n et il existe donc un unique polynôme $P \in E_n$ tel que $\delta_{x_k}(P)$ ait une valeur fixée pour $k = 1, 2, \dots, n$. On note $P_n(f)$ ce polynôme tel que $\delta_{x_k}(P(f)) = \delta_{x_k}(f)$ pour $k = 1, 2, \dots, n$.

L'expression analytique explicite de $P_n(f)$ est obtenue grâce à la construction sur E_n d'une base orthogonale à la base $\delta_{x_1}, \delta_{x_2}, \dots, \delta_{x_n}$ de E_n^* . Une telle base de polynômes L_1, L_2, \dots, L_n satisfait

$$\delta_{x_j}(L_i) = 0 \quad \text{si } i \neq j \quad \text{et} \quad \delta_{x_i}(L_i) = 1$$

Naturellement on a aussitôt :

$$L_i(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_1)(x_i-x_2)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)}$$

et

$$P_n(f) = \sum_{i=1}^n f(x_i) L_i$$

Si l'on suppose $E = \mathcal{C}[a,b]$ muni de la norme uniforme pour fixer les idées, on a quelques propriétés immédiates de l'application $P_n : \mathcal{C}[a,b] \longrightarrow E_n$

(1) P_n est un opérateur linéaire

(2) $\|P_n(f)\| < A\|f\|$ où la plus petite constante possible A vaut

$\sup_{x \in [a,b]} \left(\sum_{i=1}^n |L_i(x)| \right)$ mais il est techniquement difficile d'en donner une valeur plus commode

(3) $P_n(1) = 1$

(4) P_n n'est pas un opérateur positif (Cf Exercice N° 24)

(5) $P_n(f \cdot P_n g) = P_n(fg) \quad \forall f, g \in \mathcal{C}[a,b]$

Nous verrons au Chapitre 4 ce que l'on peut dire lorsque n tend vers l'infini, c'est-à-dire comment étudier $\|P_n f - f\|$.

Il peut être commode d'écrire autrement le polynôme $P_n(f)$, que l'on appelle quelquefois polynôme d'interpolation de Lagrange.

On remarque, grâce à la règle de dérivation d'un produit, que si l'on pose $\pi(x) = (x-x_1)(x-x_2)\dots(x-x_n)$, alors

$$L_i(x) = \frac{\pi(x)}{(x-x_i)\pi'(x_i)}$$

Donc

$$P_n(f)(x) = \sum_{i=1}^n \frac{\pi(x)}{(x-x_i)\pi'(x_i)} f(x_i)$$

Quelques relations supplémentaires, dites relations de Cauchy, jouent un certain rôle (Cf Exercice N° 25). Il est peut-être intéressant de noter que l'interpolation de Lagrange est mal commode en ce sens que l'on doit refaire tous les calculs pour passer de l'ordre n à l'ordre $(n+1)$.

Il existe de nombreux autres types d'interpolation : interpolation d'Hermite, de Newton, d'Abel-Gontcharoff, etc. (Cf. Exercice N° 28). Nous allons étudier une interpolation généralisant les deux paragraphes précédents et qui est très à la mode.

3. INTERPOLATION PAR LES FONCTION SPLINES CUBIQUES.

On se donne un segment $[a, b]$ et une division Δ de ce segment

$$\Delta_n = \{x_1 = a, x_2, \dots, x_n = b\}$$

Soit une fonction définie sur $[a, b]$ et à valeurs réelles pour fixer les idées. Dans chaque segment $[x_k, x_{k+1}]$, nous allons interpoler f par un polynôme de degré 3 au plus, prenant les mêmes valeurs que f en x_k et en x_{k+1} . En outre, nous allons imposer un bon raccordement des différents polynômes aux points de la division en supposant que ces portions de polynômes, une fois recollées, constituent une fonction deux fois continuellement dérivable.

Théorème des splines cubiques : Soit f une fonction réelle définie sur le segment $[a, b]$ et $\Delta_n = \{x_1 = a, \dots, x_n = b\}$ une division de ce segment. On suppose f dérivable (à gauche) en b et (à droite) en a . Il existe une unique fonction possédant les propriétés suivantes

- (1) g est deux fois continuellement dérivable sur $[a, b]$
- (2) la restriction de g à tout segment $[x_k, x_{k+1}]$ ($k = 1, \dots, (n-1)$) est un polynôme de degré trois au plus
- (3) $g(x_k) = f(x_k)$ pour $k = 1, 2, \dots, n$
- (4) $\frac{dg}{dx}(x_1) = \frac{df}{dx}(x_1)$; $\frac{dg}{dx}(x_n) = \frac{df}{dx}(x_n)$

Démonstration : Quelques étapes jalonnent cette démonstration technique mais facile.

Commençons par poser des notations. Pour un point x_k , on note par δ_{x_k} la forme linéaire

$$\delta_{x_k} : f \longrightarrow f(x_k) = \delta_{x_k}(f)$$

et par δ'_{x_k} la forme linéaire (définie pour les fonctions dérivables)

$$\delta'_{x_k} : f \longrightarrow \frac{df}{dx}(x_k) = \delta'_{x_k}(f)$$

1ère construction d'une base :

Considérons d'abord le segment $[0,1]$ et l'espace E de dimension 4 des p du troisième degré au plus sur ce segment. Les quatre formes linéaires $\delta_0, \delta_0', \delta_1$, et δ_1' sont linéairement indépendantes en tant qu'éléments duals de E , donc forment une base de cet espace dual E^* .
($\dim E^* = 4$).

En effet le déterminant de Gram associé n'est pas nul (Cf Exercice N° 27).
Cherchons une base de polynômes de E , orthogonale à la base $\delta_0, \delta_0', \delta_1$ et δ_1' du dual. Un calcul simple montre que les polynômes suivants conviennent

$$\begin{aligned} L_{0,0}(x) &= 2x^3 - 3x^2 + 1 \\ L_{1,0}(x) &= -2x^3 + 3x^2 \\ L_{0,1}(x) &= x^3 - 2x^2 + x \\ L_{1,1}(x) &= x^3 - x^2 \end{aligned}$$

puisqu'a

$$\begin{aligned} L_{0,0}(1) &= 1 & ; & \quad \delta_0(L_{0,1}) = \delta_0(L_{1,0}) = \delta_0(L_{1,1}) = 0 \\ L_{1,0}(0) &= 1 & ; & \quad \delta_0'(L_{0,0}) = \delta_0'(L_{1,0}) = \delta_0'(L_{1,1}) = 0 \\ L_{0,1}(0) &= 1 & ; & \quad \delta_1(L_{0,0}) = \delta_1(L_{0,1}) = \delta_1(L_{1,1}) = 0 \\ L_{1,1}(1) &= 1 & ; & \quad \delta_1'(L_{0,0}) = \delta_1'(L_{0,1}) = \delta_1'(L_{1,0}) = 0 \end{aligned}$$

Par là existe un unique polynôme L de degré au plus trois, tel que
 $L(0)L'(0) = \alpha_0'$; $L(1) = \alpha_1$ et $L'(1) = \alpha_1'$ où $\alpha_0, \alpha_0', \alpha_1$ et α_1'

sont des nombres réels donnés. Ce polynôme s'écrit :

$$L = \alpha_0 L_{0,0} + \alpha_0' L_{0,1} + \alpha_1 L_{1,0} + \alpha_1' L_{1,1}$$

2ème construction d'un polynôme :

Donnons-nous $(n-2)$ nombres réels notés $\beta_2, \dots, \beta_{n-1}$ et prenons $(\beta_1 = x_1)$, $\beta_n = \frac{df}{dx}(x_n)$. Considérons un intervalle $[x_k, x_{k+1}]$ de la n -ième Δ_n donnée pour $k = 1, 2, \dots, n-1$. Il existe un unique polynôme, noté tel que :

$$L_k(x_k) = f(x_k) \quad ; \quad L'_k(x_k) = \beta_k \quad ; \quad L_k(x_{k+1}) = f(x_{k+1}) \quad ; \quad L'_k(x_{k+1}) = \beta_{k+1}$$

Il s'agit du polynôme obtenu après une translation et une homothétie à partir de L et en tenant compte du comportement de la dérivation vis à vis d'une homothétie :

$$L_k(x) = f(x_k) L_{0,0}\left(\frac{x-x_k}{x_{k+1}-x_k}\right) + (x_{k+1}-x_k) \beta_k L_{0,1}\left(\frac{x-x_k}{x_{k+1}-x_k}\right) + f(x_{k+1}) L_{1,0}\left(\frac{x-x_k}{x_{k+1}-x_k}\right) + (x_{k+1}-x_k) \beta_{k+1} L_{1,1}\left(\frac{x-x_k}{x_{k+1}-x_k}\right)$$

3ème étape : Raccordement des polynômes :

Appelons g la fonction égale à $L_k(x)$ pour $x \in [x_k, x_{k+1}]$ où $k = 1, 2, \dots, (n-1)$. Grâce à la construction même, la fonction g satisfait la condition (2) du théorème, la condition (3) et la condition (4), et ceci quel que soit le choix de la suite $\beta_2 \dots \beta_{n-1}$. En outre, g possède une dérivée continue puisque seuls les points de la division sont douteux et l'on a précisément choisi

$$L'_k(x_k) = L'_{k-1}(x_k) = \beta_k \quad k = 2, \dots, (n-1)$$

Il reste à montrer qu'il existe un unique choix de ces $(n-2)$ nombres $\beta_2, \beta_3, \dots, \beta_{n-1}$ tel que g possède une dérivée seconde continue. Cette condition s'écrit

$$L''_k(x_k) = L''_{k-1}(x_k) \quad k = 2, \dots, (n-1)$$

Pour calculer les dérivées secondes, on remarque que

$$\left\{ \begin{array}{l} L''_{0,0}(x) = 12x - 6 \\ L''_{1,0}(x) = -12x + 6 \\ L''_{0,1}(x) = 6x - 4 \\ L''_{1,1}(x) = 6x - 2 \end{array} \right.$$

Donc, en posant $h_k = x_{k+1} - x_k$ pour alléger l'écriture

$$L''_k(x) = \frac{f(x_k)}{h_k^2} \left(12 \left(\frac{x-x_k}{h_k} \right) - 6 \right) + \frac{\beta_k}{h_k} \left(6 \left(\frac{x-x_k}{h_k} \right) - 4 \right) + \frac{f(x_{k+1})}{h_k^2} \left(-12 \left(\frac{x-x_k}{h_k} \right) + 6 \right) + \frac{\beta_{k+1}}{h_k} \left(6 \left(\frac{x-x_k}{h_k} \right) - 2 \right)$$

$$L''_k(x_k) = -6 \frac{f(x_k)}{h_k^2} - 4 \frac{\beta_k}{h_k} + 6 \frac{f(x_{k+1})}{h_k^2} - 2 \frac{\beta_{k+1}}{h_k}$$

tandis que

$$L''_{k-1}(x_k) = 6 \frac{f(x_{k-1})}{h_{k-1}^2} + 2 \frac{\beta_{k-1}}{h_{k-1}} - 6 \frac{f(x_k)}{h_{k-1}^2} + 4 \frac{\beta_k}{h_{k-1}}$$

Ce qui fournit un système linéaire avec $k = 2, \dots, (n-1)$.

$$\begin{aligned} + \frac{2}{h_{k-1}} \beta_{k-1} + 4 \left(\frac{1}{h_{k-1}} + \frac{1}{h_k} \right) \beta_k + \frac{2}{h_k} \beta_{k+1} &= -6 f(x_k) \left(\frac{1}{h_k^2} - \frac{1}{h_{k-1}^2} \right) \\ &+ 6 \frac{f(x_{k+1})}{h_k^2} - 6 \frac{f(x_{k-1})}{h_{k-1}^2} \end{aligned}$$

Il s'agit d'un système de $(n-2)$ équations linéaires à $(n-2)$ inconnues

$\beta_2, \dots, \beta_{n-1}$. Le déterminant de ce système s'écrit :

$$\begin{vmatrix} 4 \left(\frac{1}{h_1} + \frac{1}{h_2} \right) & \frac{2}{h_2} & 0 & \dots & 0 \\ \frac{2}{h_2} & 4 \left(\frac{1}{h_2} + \frac{1}{h_3} \right) & \frac{2}{h_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & 0 & \dots & \dots & \frac{2}{h_{n-2}} & 4 \left(\frac{1}{h_{n-2}} + \frac{1}{h_{n-1}} \right) \end{vmatrix}$$

On remarque alors que sur une même ligne le terme diagonal est strictement supérieur à la somme de tous les autres termes

$$4 \left(\frac{1}{h_{k-1}} + \frac{1}{h_k} \right) > \frac{2}{h_{k-1}} + \frac{2}{h_k}$$

C'est un résultat bien connu en analyse numérique (Cf aussi E.A.F., Exercice N° 46) que le déterminant n'est pas nul et donc que l'on a une unique famille

$\beta_2, \dots, \beta_{n-1}$. Cette assertion termine la démonstration du théorème et fournit d'ailleurs une formulation explicite de la fonction g à condition de résoudre le dernier système linéaire ; mais on dispose pour ce faire de techniques bien rodées sur ordinateurs.

L'espace E_n des fonctions qui vérifient les conditions (1) et (2) du théorème des splines cubiques est un espace vectoriel de dimension $(n+2)$. En effet, la démonstration précédente montre qu'une fonction de E_n dépend linéairement de la valeur aux n points X_1, X_2, \dots, X_n et de la valeur de sa dérivée en X_1 et en X_n . On pourrait même exhiber précisément une base de $(n+2)$ fonctions de E à l'aide des formules établies.

(d'ailleurs, d'un point de vue heuristique, l'ensemble des fonctions satisfaisant la condition (2) a une dimension égale à $4(n-1)$, la condition (1) fournit $3(n-2)$ conditions pour le raccordement, ce qui donne bien $4(n-1) - 3(n-2) = n+2$. La condition (3) donne n conditions et la condition (4) en donne 2 ce qui explique le théorème).

Appelons $P_n(f)$ l'unique fonction associée à f par le théorème

$$P_n : E \longrightarrow E_n$$

où E désigne l'ensemble des fonctions continuellement dérivables sur $[a, b]$.

On a quelques propriétés de P_n que l'on ordonne par analogie avec les paragraphes précédents.

(1) P_n est un opérateur linéaire

(2) Nous étudierons au Chapitre 4 les questions de normes et la précision de l'approximation de f par $P_n(f)$

$$(3) P(1) = 1$$

$$(4) P_n(f P_n g) = P_n(f g) \quad \forall f, g \in E$$

4. PROPRIETES DES FONCTIONS SPLINES CUBIQUES.

Lemme : Soit f une fonction à valeurs réelles, définie sur $[a, b]$ et deux fois continuellement dérivable. Soit $P_n(f)$ la fonction spline cubique relative à une division Δ_n . On a la relation

$$\|f''\|_2^2 = \|(P_n(f))''\|_2^2 + \|f'' - (P_n(f))''\|_2^2$$

où l'on a posé $\|f''\|_2^2 = \int_a^b (f''(x))^2 dx$

Démonstration : Il suffit bien sûr d'établir la relation d'orthogonalité

$$I = \int_a^b (P_n(f))''(x) (f''(x) - (P_n(f))''(x)) dx = 0$$

(rappelons à ce propos que les fonctions considérées sont à valeurs réelles)

Or l'intégrale I s'écrit

$$\begin{aligned}
 I &= \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} (P_n(f))''(x) (f''(x) - (P_n(f))''(x)) dx \\
 &= \sum_{k=1}^{n-1} \left[(P_n(f))''(x) (f'(x) - (P_n(f))'(x)) \right]_{x_k}^{x_{k+1}} \\
 &\quad - \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} (P_n(f))'''(x) (f'(x) - (P_n(f))'(x)) dx
 \end{aligned}$$

En utilisant une intégration par parties sur chaque segment $[x_k, x_{k+1}]$, c'est-à-dire en intégrant le terme $f''(x) - (P_n(f))''(x)$ et en dérivant le terme $(P_n(f))''$ afin de faire apparaître une constante.

Car $(P_n(f))''(x)$ est une constante α_k sur le segment $[x_k, x_{k+1}]$ grâce à la condition (2) du théorème des splines cubiques. D'où

$$I = \left[(P_n(f))''(x) (f'(x) - (P_n(f))'(x)) \right]_{x_1}^{x_n} - \sum_{k=1}^{n-1} \alpha_k \int_{x_k}^{x_{k+1}} (f'(x) - (P_n(f))'(x)) dx$$

Les conditions (3) et (4) de ce même théorème permettent de conclure.

Théorème : Soit f une fonction définie sur un segment $[a, b]$, à valeurs réelles et possédant une dérivée en a et en b . Soit Δ_n une division de $[a, b]$. On appelle \mathfrak{S} l'ensemble des fonctions g , définies sur $[a, b]$, à valeurs réelles, deux fois continuellement dérivables et telles que

$$g(x_k) = f(x_k) \quad k = 1, 2, \dots, n$$

$$g'(x_1) = f'(x_1) \quad ; \quad g'(x_n) = f'(x_n)$$

La fonction spline cubique $P_n(f)$ relative à Δ_n est l'unique fonction de \mathcal{S} qui rende minimale l'expression

$$\|g''\|_2^2 = \int_a^b [g''(x)]^2 dx$$

En effet, on peut appliquer le lemme précédent à la fonction g et à $P_n(f) = P_n(g)$.
Il vient

$$\|g''\|_2^2 = \|g'' - (P_n(f))''\|_2^2 + \|(P_n(f))''\|_2^2$$

donc si $g \neq P_n(f)$

$$\|g''\|_2^2 > \|(P_n(f))''\|_2^2$$

Nous venons donc de montrer que les fonctions splines cubiques sont parmi les fonctions régulières qui interpolent une fonction donnée suivant une division Δ_n , celles qui rendent minimale une certaine expression quadratique, à savoir l'intégrale du carré de la dérivée seconde.

5. GENERALISATION.

On peut naturellement définir des fonctions splines généralisant les fonctions splines cubiques. Nous ne donnons ici que les résultats et les définitions idoines.

Définition : soit x_1, x_2, \dots, x_n une suite strictement croissante de n nombres réels. Une fonction spline de degré m pour la division x_1, \dots, x_n est une fonction $g : \mathbb{R} \rightarrow \mathbb{C}$ telle que

(1) sur $]x_i, x_{i+1}[$, pour $i = 0, \dots, n$ en posant $x_0 = -\infty$ et $x_{n+1} = +\infty$ g est un polynôme de degré m au plus

(2) g est une fonction possédant $(m-1)$ dérivées continues.

Dans cette définition, nous avons voulu étudier le cas de l'axe réel entier et non celui d'un segment $[a, b]$ comme dans le cas des splines cubiques.

Définition : Une fonction spline naturelle est une fonction spline de degré impair $2k-1$, pour une division x_1, x_2, \dots, x_n , mais qui se réduit à des polynômes de degré $(k-1)$ au plus sur $]-\infty, x_1[$ ou $]x_n, +\infty[$

Théorème : Soit f une fonction définie sur \mathbb{R} et x_1, x_2, \dots, x_n une division donnée. Soit k un entier $1 \leq k \leq n$. Il existe une unique fonction spline naturelle g de degré $2k-1$, telle que

$$g(x_i) = f(x_i) \quad i = 1, 2, \dots, n$$

Théorème : Soit $f: [a, b] \rightarrow \mathbb{R}$ et soit $\Delta_{n+2} = (x_1 = a, x_2, \dots, x_{n+2} = b)$ une division de $[a, b]$. On appelle \mathcal{S} l'ensemble des fonctions $g: \mathbb{R} \rightarrow \mathbb{R}$, qui possèdent une dérivée d'ordre k ($0 \leq k \leq n$) continue et telles que

$$g(x_i) = f(x_i) \quad i = 2, \dots, n+1$$

La fonction spline naturelle d'ordre $2k-1$ donnée par le théorème précédent est l'unique fonction de \mathcal{S} qui minimise l'expression

$$\int_a^b [g^{(k)}(x)]^2 dx$$

Avec des conditions aux limites différentes, on a également

Théorème : Soit $f: [a, b] \rightarrow \mathbb{R}$ et soit $\Delta_{n+2} = (x_1 = a, x_2, \dots, x_{n+2} = b)$ une division de $[a, b]$; Soit k un entier, $1 \leq k \leq n$. On suppose que f possède des dérivées jusqu'à l'ordre $(k-1)$ en a et b . Il existe une unique fonction spline g d'ordre $(2k-1)$ telle que

$$g(x_i) = f(x_i) \quad i = 2, 3, \dots, n+1$$

$$\frac{d^r g}{dx^r}(a) = \frac{d^r f}{dx^r}(a) \quad r = 0, 1, \dots, (k-1)$$

$$\frac{d^r g}{dx^r}(b) = \frac{d^r f}{dx^r}(b) \quad r = 0, 1, \dots, (k-1)$$

On a également un théorème montrant qu'une telle fonction spline minimise une

intégrale $\int_a^b [g^{(k)}(x)]^2 dx$ parmi toutes les fonctions g satisfaisant des condi-

tions aux limites convenables et possédant une dérivée continue jusqu'à l'ordre k .

Nous renvoyons le lecteur à la bibliographie pour l'utilisation des fonctions splines, notamment en ce qui concerne les problèmes de valeurs initiales pour les équations différentielles.

Terminons par quelques indications.

On se donne une division $a < x_1 < \dots < x_n < b$ et on appelle $S_{n,k}$ l'ensemble des fonctions splines d'ordre k . Soit f une fonction continue sur $[a,b]$. On peut chercher l'unique fonction spline $s^* \in S_{n,k}$ qui minimise

$$\int_a^b |f(t) - s(t)|^2 dt$$

lorsque s parcourt $S_{n,k}$. Le problème est théoriquement très simple : il faut orthonormaliser une base quelconque de $S_{n,k}$. Un algorithme explicite et d'ailleurs un programme peuvent se trouver dans : C. de Boor, J.R. Rice, Least square cubic spline approximation, I fixed knots, Rpt CSD TR 20, Computer Science Department, Purdue University (1968).

On peut se poser le même problème en laissant libre la division Δ_n de $[a,b]$ et de même une référence utile est C. de Boor, J.R. Rice, Least square cubic spline approximations, II Variable knots, Rpt CSD TR 21, Computer Science Department, Purdue University (1968).

- EXERCICES DU CHAPITRE 3 -

EXERCICE N° 23.

Avec les notations du paragraphe 1, donner un exemple montrant que $P_n(f)$ n'est pas une meilleure approximation de f par les fonctions de E_n au sens de la convergence uniforme sur $\mathcal{C}[a, b]$ ou d'une norme $L^p[a, b]$.

EXERCICE N° 24.

Avec les notations du paragraphe 2, fournir un exemple montrant que P_n n'est pas un opérateur positif.

EXERCICE N° 25.

Soient $\{L_i\}_{i=1, \dots, n}$ les n polynômes qui interviennent dans l'interpolation de Lagrange (§ 2) d'ordre n .

$$L_i(x) = \frac{\pi(x)}{(x-x_i)\pi'(x_i)} \quad \text{avec} \quad \pi(x) = (x-x_1)(x-x_2)\dots(x-x_n)$$

Etablir les $(n+1)$ relations de Cauchy

$$\sum_{i=1}^n (x_i - x)^k L_i(x) \equiv 0 \quad k = 1, 2, \dots, n$$

et
$$\sum_{i=1}^n L_i(x) \equiv 1$$

(on pourra considérer l'interpolation du polynôme $(x-y)^k$ où y est un paramètre fixe, puis faire $y = x$).

EXERCICE N° 26. (interpolation trigonométrique)

Soient x_0, x_1, \dots, x_{2n} , $(2n+1)$ points du segment $[-\pi, +\pi[$ et une fonction réelle et 2π -périodique. Montrer qu'il existe un unique polynôme trigonométrique $P_n(f)$, c'est-à-dire une combinaison linéaire de $\cos kx$ et $\sin kx$ pour $k = 0, 1, \dots, n$ tel que

$$P(f)(x) = f(x) \quad i = 0, 1, \dots, 2n$$

(on construira une base orthogonale à la base de l'espace dual notée

$\delta_{x_0}, \delta_{x_1}, \dots, \delta_{x_{2n}}$:

$$T_i(x) = \frac{\prod_{\substack{k=0 \\ k \neq i}}^{2n} \sin\left(\frac{x-x_k}{2}\right)}{\prod_{\substack{k=0 \\ k \neq i}}^{2n} \sin\left(\frac{x_i-x_k}{2}\right)}$$

EXERCICE N° 27.

Soit X un espace vectoriel réel de dimension n . Soient L_1, L_2, \dots, L_n , n formes linéaires sur X et soit $\{x_1, x_2, \dots, x_n\}$ une base de X . On appelle déterminant de Gram $g(L_1, L_2, \dots, L_n)$ associé le déterminant

$$g(L_1, L_2, \dots, L_n) = \begin{vmatrix} L_1(x_1) & L_1(x_2) & \dots & L_1(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ L_n(x_1) & L_n(x_2) & \dots & L_n(x_n) \end{vmatrix}$$

Montrer qu'une condition nécessaire et suffisante de linéaire indépendance des n formes linéaires L_1, L_2, \dots, L_n (en tant qu'éléments du dual de X) est la non nullité du déterminant de Gram

EXERCICE N° 28.

Soit X un espace vectoriel réel de dimension n . Soient L_1, L_2, \dots, L_n , n formes linéaires sur X supposées linéairement indépendantes en tant qu'éléments du dual de X . Soit (y_1, y_2, \dots, y_n) une famille de n nombres réels. Montrer qu'il existe un unique vecteur x de X tel que

$$L_i(x) = y_i \quad i = 1, 2, \dots, n$$

(on pourra utiliser l'exercice N° 27).

Exemple 1 : $X =$ espace des polynômes réels de degré au plus $(n-1)$
 $L_1 = \delta_{x_1}, L_2 = \delta_{x_2}, \dots, L_n = \delta_{x_n}$ (notations du § 2)

(on obtient l'interpolation de Lagrange, les $x_i, i = 1, 2, \dots, n$ sont distincts)

Exemple 2 : $X =$ espace des polynômes réels de degré au plus $(n-1)$
 $L_1 = \delta_{x_1}$, $L_2 = \delta'_{x_2}$, ---- , $L_n = \delta_{x_n}^{(n)}$

où $\delta_{x_k}^{(k)}$ représente la forme linéaire qui à $f \in X$ associe le nombre réel $\frac{d^k f}{dx^k}(x_k)$. Ici les x_i peuvent être distincts ou non.
 (on obtient l'interpolation d'Abel-Gontcharoff).

EXERCICE N° 29.

Soit $\mathcal{C}[a,b]$ l'espace des fonctions continues et à valeurs réelles muni de la norme uniforme. Soit E_n le sous-espace des fonctions linéaires par morceaux pour une division donnée $\Delta_n = \{x_1 = a, x_2, \dots, x_n = b\}$

Montrer qu'il existe un unique opérateur linéaire P_n défini sur $\mathcal{C}[a,b]$ et à valeurs dans E_n tel que P_n soit de norme 1 avec $P_n(1) = 1$ et satisfasse $P_n(f P_n g) = P_n(fg)$ pour tous f et g de $\mathcal{C}[a,b]$.

(il s'agit de l'opérateur d'interpolation linéaire étudié au § 1).

EXERCICE N° 30.

Fournir un exemple montrant que dans le théorème des splines cubiques, et avec les notations du paragraphe 3, on n'a pas l'égalité

$$\frac{dg}{dx}(x_i) = \frac{df}{dx}(x_i) \quad i = 2, \dots, n-1$$

EXERCICE N° 31.

Soit f une fonction deux fois continuellement dérivable sur un intervalle $[a,b]$. Posons $\|f''\|_\infty = \sup_{x \in [a,b]} |f''(x)|$. Soit alors $P_n(f)$ l'interpolée linéaire de f par une fonction linéaire par morceaux pour une division Δ_n . On pose $|\Delta_n| = \sup_{k=0, \dots, (n-1)} (x_{k+1} - x_k)$

Montrer que

$$\|f - P_n(f)\|_\infty \leq \frac{1}{8} |\Delta_n|^2 \|f''\|_\infty^2$$

(il suffit d'obtenir une majoration sur un segment $[x_k, x_{k+1}]$. On fixe x sur ce segment et considère la fonction F de y définie sur ce segment par :

$$F(y) = f(y) - P_n(f)(y) - \frac{f(x) - P_n(f)(x)}{(x-x_k)(x-x_{k+1})} (y-x_k)(y-x_{k+1})$$

Comme $F(x_k) = F(x) = F(x_{k+1}) = 0$, il existe un point ξ où $F''(\xi) = 0$

Soit

$$F''(\xi) = f''(\xi) - 2 \frac{f(x) - P_n(f)(x)}{(x-x_k)(x-x_{k+1})} = 0$$

et donc pour tout x de $[x_k, x_{k+1}]$

$$\left| \frac{f(x) - P_n(f)(x)}{(x-x_k)(x-x_{k+1})} \right| \leq \frac{1}{2} \|f''\|_\infty$$

On conclut facilement).

EXERCICE N° 32.

Soit $\Delta_n = (x_1 = a \dots x_n = b)$ une division de $[a, b]$ et F_n l'espace des fonctions réelles deux fois continuellement dérivables sur $[a, b]$, dont la restriction à tout segment $[x_k, x_{k+1}]$ soit un polynôme de degré 3 au plus. Soit f une fonction réelle deux fois continuellement dérivable sur $[a, b]$. Cherchons une fonction de F_n qui rende minimale l'expression

$$\int_a^b (f''(x) - g''(x))^2 dx \quad \text{lorsque } g \text{ parcourt } F_n$$

(il s'agit naturellement de $P_n(f)$: la fonction spline cubique pour Δ_n).

EXERCICE N° 33.

Déterminer la meilleure approximation au sens de $L^2[0, 1]$ de la fonction $x \longrightarrow \sqrt{x}$ par des fonctions splines naturelles d'ordre trois pour la division $x_1 = \frac{1}{4}$ et $x_2 = \frac{1}{2}$ et $x_3 = \frac{3}{4}$.

- CHAPITRE 4 -

CONVERGENCE ET ESTIMATION DES METHODES D'APPROXIMATION

1. INTRODUCTION.

Jusqu'à présent, nous avons étudié l'approximation d'un vecteur d'un espace vectoriel E au moyen des éléments d'un sous-espace vectoriel de dimension finie. Il est important de savoir ce qui se passe lorsque la dimension de ce dernier espace augmente.

Deux problèmes se posent alors :

1) d'une part la dimension de l'espace E est en général infinie et de cardinal supérieur au dénombrable. Par suite, comme la réunion dénombrable d'espaces de dimension finie est au plus d'une dimension ayant la puissance du dénombrable, les résultats intéressants seront de nature topologique (théorèmes de densité) ;

2) d'autre part, il peut être intéressant de connaître le degré de l'approximation lorsque l'on s'arrête à un espace de dimension n , c'est-à-dire obtenir un majorant du reste R_n .

Il faut toutefois remarquer tout de suite que nous ne cherchons plus la meilleure approximation dans l'étude du problème (1). Donnons un exemple : soit $E = \mathcal{C} [0,1]$ l'espace vectoriel, normé par la norme uniforme, des fonctions définies sur $[0,1]$, à valeurs complexes et continues. Envisageons la famille $[E_n]$ des sous-espaces vectoriels de $E [0,1]$, chaque étant constitué par les polynômes de degré inférieur à n . L'espace E_n est alors de dimension n .

Le premier problème est de savoir si l'on peut représenter tout élément f de $[0,1]$ au moyen d'un élément de $\bigcup_{n=1}^{\infty} E_n$. La réponse directe est non : toute fonction continue ne s'écrit pas comme un polynôme d'un degré suffisamment élevé. Toutefois, lorsque f est analytique sur tout le plan complexe, la formule du développement de Taylor à l'origine.

$$f(x) = f(0) + xf'(0) + \dots + \frac{x^n}{n!} f^{(n)}(0) + R_n(x)$$

montre que la différence entre f et un certain polynôme de degré n est un

reste $R_n(x)$, lequel peut s'exprimer sous une forme intégrale de Lagrange

$$R_n(x) = \int_0^x \frac{(x-t)^n}{n!} f^{(n+1)}(t) dt$$

et on dispose de la majoration de ce reste à partir d'une majoration uniforme de la dérivée d'ordre $(n+1)$

$$|R_n(x)| \leq \frac{x^{n+1}}{(n+1)!} \|f^{(n+1)}\|_\infty$$

Soit, puisque l'on est placé sur $[0,1]$

$$\|R_n\|_\infty \leq \frac{\|f^{(n+1)}\|_\infty}{(n+1)}$$

Pour les fonctions entières, nous avons estimé le problème (2). Quant au problème (1), il se trouve dès lors résolu dès que l'on sait

$$\lim_{n \rightarrow +\infty} \|R_n\|_\infty = 0$$

c'est-à-dire si pour tout ε positif, il existe un entier n et un polynôme de degré n , tel que la différence entre f et ce polynôme soit en norme uniforme inférieure à ε . Un tel résultat est assuré lorsque f est la restriction à $[0,1]$ d'une fonction entière

Il est un cas où les problèmes (1) et (2) sont résolus: c'est le cas de l'approximation dans un espace de Hilbert (séparable). En effet soit \mathcal{H} un espace de Hilbert (séparable), $e_1, e_2, \dots, e_n, \dots$ une base hilbertienne de \mathcal{H} . Appelons E_n l'espace vectoriel (de dimension n) engendré par les n vecteurs e_1, e_2, \dots, e_n . Soit x un vecteur de \mathcal{H} . La meilleure approximation de x par les éléments de E_n est donnée par l'expression

$$P_{E_n}(x) = \sum_{k=1}^n \langle x, e_k \rangle e_k$$

d'après les résultats du Chapitre 2. En outre

$$\|x - P_{E_n}(x)\|^2 = \sum_{k>n} \langle x, e_k \rangle^2$$

ce qui fournit une valeur exacte du reste, lequel tend vers 0 lorsque n tend vers l'infini (Cf Analyse Harmonique).

Pour d'autres normes, les problèmes (1) et (2) s'avèrent nettement plus délicats. Deux techniques au moins sont possibles. On peut déduire un résultat de densité de celui obtenu pour une norme provenant d'un produit

Si mais en utilisant un procédé de régularisation, la convolution que l'on a vu au Chapitre 5. Une autre technique fructueuse, même pour la norme de la convergence uniforme, repose sur l'emploi du théorème de Korovkin. Ces deux thèmes se recoupent d'ailleurs (Cf Exercice N° 48). Enonçons d'abord une propriété, dite propriété K, pour la commodité de l'exposé.

Propriété K

Disons que n fonctions continues f_1, f_2, \dots, f_n , définies sur un interval fini $[a, b]$, ou plus généralement sur un espace topologique compact X admettent la propriété K, lorsqu'on peut trouver n fonctions continues g_1, \dots, g_n telles que

$$G(x, y) = \sum_{k=1}^{k=n} g_k(y) f_k(x)$$

soit une fonction positive, nulle si et seulement si $x = y$.

Exemple : soient les fonctions $f_1 = 1$, $f_2 = x$, $f_3 = x^2$ sur un intervalle $[a, b]$. Ces trois fonctions satisfont la condition K.

En effet, l'expression

$$(y-x)^2 = y^2 - 2xy + x^2 = y^2 f_1 - 2y f_2 + f_3$$

est positive nulle si et seulement si $x = y$. On choisit donc

$$g_1 = y^2$$

$$g_2 = -2y$$

et

$$g_3 = 1$$

La propriété K a pour intérêt de fournir une fonction strictement positive sur le carré $[a, b] \times [a, b]$, la diagonale étant exclue, et nulle sur cette diagonale. On peut en déduire une fonction non constante et strictement positive, formée à partir des fonctions f_n .

On pose à partir de deux points y_1 et y_2 distincts :

$$G(x) = G(x, y_1) + G(x, y_2)$$

La fonction $G(x)$ est strictement positive sur $[a, b]$, grâce aux hypothèses faites. Nous avons en plus besoin d'une définition générale :

Définition : on dit qu'un opérateur linéaire P , défini sur un espace fonctionnel, est positif lorsque la propriété $f(x) \geq 0$ pour tout x , implique

$$Pf(x) \geq 0$$

Exemple : sur $\mathcal{C}[0,1]$, l'opérateur P défini par

$$Pf(x) = \frac{1}{2\pi n} \int_{-\pi}^{+\pi} f(t) \left(\frac{\sin \frac{n(x-t)}{2}}{\sin \frac{x-t}{2}} \right)^2 dt$$

est un opérateur positif. C'est d'ailleurs l'opérateur de Féjer.

Théorème 1 : sur $\mathcal{C}[a,b]$, un opérateur linéaire positif est continu.

Pour une fonction réelle continue, on a les majorations

$$A \leq f(x) \leq B \quad \text{où} \quad A = \text{Min}_{x \in (a,b)} f(x)$$

$$\text{et} \quad B = \text{Max}_{x \in (a,b)} f(x)$$

donc $B - f(x) \geq 0$ et par suite en appliquant l'opérateur P :

$$BP(1)(x) - Pf(x) \geq 0$$

ce qui donne en faisant de même pour A

$$AP(1)(x) \leq Pf(x) \leq BP(1)(x)$$

donc

$$\|Pf\| \leq \|f\| P(1)$$

On obtient même la valeur de la norme de l'opérateur P : $\|P(1)\| = \|P\|$

Par contre, tout opérateur linéaire continu n'est pas un opérateur positif comme le montre par exemple l'opérateur

$$P: f \rightarrow Pf(x) = \int_{-\pi}^{+\pi} f(t) \sin(x-t) dt \text{ sur } \mathcal{C}[-\pi, +\pi]$$

2. THEOREME DE KOROVKIN.

Nous pouvons maintenant énoncer le théorème de Korovkin : dans ce qui suit, on pourra remplacer $\mathcal{C}[a,b]$ par $\mathcal{C}(X)$ où X est un espace topologique compact possédant au moins deux points. Rappelons que $\mathcal{C}[a,b]$ est muni de la norme uniforme.

Convention : dans ce chapitre, un opérateur linéaire L opérant sur une fonction de deux variables $G(x,y)$, sera considéré comme opérant sur la première variable, ce qui revient à dire :

$L(G(x,y))(z)$ est une fonction de z et de y .

Théorème 2 : soient n fonctions f_1, f_2, \dots, f_n satisfaisant la propriété K sur $\mathcal{E}[a,b]$.

Soit L_k une suite d'opérateurs linéaires positifs sur $\mathcal{E}[a,b]$ telle que

$$\lim_{k \rightarrow +\infty} \|L_k(f_i) - f_i\| = 0 \text{ (pour } i = 1, 2, \dots, n)$$

Alors on a pour tout f appartenant à $\mathcal{E}[a,b]$

$$\lim_{k \rightarrow +\infty} \|L_k(f) - f\| = 0$$

Calculons d'abord

$$\begin{aligned} L_k(G(x,y))(y) &= L_k\left(\sum_{i=1}^{i=n} g_i(y) f_i(x)\right)(y) \\ &= \sum_{i=1}^{i=n} g_i(y) L_k(f_i)(y) \end{aligned}$$

Or $L_k(f_i)$ converge uniformément vers f_i selon l'hypothèse du théorème, par suite $L_k(G(x,y))(y)$ converge uniformément, lorsque k tend vers l'infini, vers $\sum_{i=1}^{i=n} g_i(y) f_i(y)$, c'est-à-dire vers 0 grâce à la propriété K .

Grâce à ce résultat

$$\lim_{k \rightarrow +\infty} L_k(G(x,y))(y) = 0$$

et ceci uniformément, on peut démontrer beaucoup plus en prenant une fonction continue quelconque $F(x,y)$ sur $[a,b] \times [a,b]$, telle que $F(y,y) = 0$. On va en effet prouver que la fonction $L_k(F(x,y))(y)$ converge uniformément sur $[a,b]$ lorsque k tend vers l'infini.

Soit ϵ un nombre positif donné, il existe un nombre η tel que

$$|F(x,y)| < \epsilon \text{ pour } |y-x| < \eta$$

grâce à $F(y,y) = 0$ et à la continuité uniforme de F sur le carré $[a,b] \times [a,b]$.

De plus, pour $|y-x| \geq \eta$, $G(x,y)$ est une fonction continue strictement positive, donc $G(x,y) \geq \alpha > 0$. Donc, en notant $\|F\|$ le maximum de $|F|$ sur le carré $[a,b] \times [a,b]$, on dispose de l'inégalité

$$|F(x,y)| \leq \varepsilon + \frac{\|F\|}{\alpha} G(x,y)$$

Faisons agir l'opérateur L_k sur cette inégalité

$$|L_k F(x,y)|(y) \leq \varepsilon |L_k(1)(y)| + \frac{\|F\|}{\alpha} |L_k(G(x,y))|(y)$$

Or on peut facilement donner un majorant de $L_k(1)$ en utilisant la fonction $G(x)$ qui est strictement positive (Cf p. 58). Il existe donc un nombre β tel que :

$$G(x) \geq \beta > 0 \quad \text{sur } [a,b]$$

d'où

$$\frac{G(x)}{\beta} \geq 1$$

et donc

$$\frac{1}{\beta} L_k(G)(x) \geq L_k(1)$$

Mais $G(x) = G(x,y_1) + G(x,y_2)$ donc comme $L_k(G(x,y_i))(y)$ converge uniformément vers $G(y, y_i)$ pour $L = 1$ ou 2 , on obtient une majoration uniforme de $L_k(1)$ que nous noterons avec une certaine constante $\gamma > 0$

$$\|L_k(1)\| \leq \gamma \quad \forall k > 1$$

En passant, on remarque que $\lim_{k \rightarrow +\infty} L_k(G(x))$ converge uniformément vers la fonction G . Enfin, puisque

$$\lim_{k \rightarrow +\infty} L_k(G(x,y))(y) = 0$$

il est permis de choisir au nombre entier k suffisamment grand de sorte que $|L_k(G(x,y))(y)| \leq \varepsilon$. Finalement, on peut estimer $L_k(F(x,y))$ au point y selon :

$$|L_k(F(x,y))(y)| \leq \varepsilon \gamma + \frac{\|F\|}{\alpha} \varepsilon$$

ce qui montre bien que $\lim_{k \rightarrow +\infty} L_k(F(x,y))(y) = 0$

uniformément (en y) sur $[a,b]$.

Terminons maintenant la démonstration du théorème de Korovkin en posant

$$H(x,y) = f(x) - \frac{f(y)}{G(y)} G(x)$$

La fonction H est continue sur $[a,b] \times [a,b]$ puisque $G(y)$ est une fonction strictement positive. De plus $H(y,y) = 0$. Par suite :

$$\begin{aligned} \lim_{k \rightarrow \infty} L_k(H(x,y))(y) &= 0 \\ &= \lim_{k \rightarrow \infty} L_k(f)(y) - \frac{f(y)}{G(y)} \lim_{k \rightarrow \infty} L_k(G)(y) \end{aligned}$$

Or, comme nous avons déjà remarqué que

$$\lim_{k \rightarrow +\infty} L_k(G)(y) = G(y)$$

On a :

$$\lim_{k \rightarrow \infty} L_k(f)(y) = f(y) \text{ uniformément sur } [a,b]$$

ce qu'il fallait démontrer.

3. APPLICATION AU THEOREME DE WEIERSTRASS.

Donnons une conséquence du théorème de Korovkin.

Théorème de Weierstrass 3 : Toute fonction continue sur un intervalle $[a,b]$ peut être approchée, d'aussi près qu'on le désire au sens de la norme uniforme, par un polynôme.

Nous allons construire des opérateurs L_k convenables pour pouvoir appliquer le théorème précédent de Korovkin. On utilise un procédé ancien, dû à Bernstein :

on pose pour f appartenant à $\mathcal{C}[0,1]$ et pour tout entier $k = 1$

$$L_k(f)(x) = \sum_{i=0}^{i=k} c_k^i x^i (1-x)^{k-i} f\left(\frac{i}{k}\right)$$

(le cas $\mathcal{C}[a, b]$ se traiterait d'une manière analogue après une translation et une homothétie).

La transformation $f \longrightarrow L_k(f)$ est visiblement linéaire. De plus L_k est un opérateur positif, donc continu sur $\mathcal{C}[0, 1]$

Sur $[0, 1]$, nous avons pu constater que les trois polynômes $f_1 = 1$, $f_2 = x$ et $f_3 = x^2$ satisfaisaient la condition K. Or

$$L_k(f_1)(x) = \sum_{i=0}^{i=k} C_k^i x^i (1-x)^{k-i} = 1 = f_1(x)$$

Puis

$$\begin{aligned} L_k(f_2)(x) &= \frac{1}{k} \sum_{i=0}^{i=k} i C_k^i x^i (1-x)^{k-i} \\ &= \frac{x}{k} \sum_{i=1}^{i=k} i C_k^i x^{i-1} (1-x)^{k-i} \end{aligned}$$

Mais $i C_k^i = k C_{k-1}^{i-1}$, donc on peut encore écrire :

$$= \frac{x}{k} k \sum_{i=0}^{i=k-1} C_{k-1}^i x^i (1-x)^{(k-1)-i}$$

Soit $L_k(f_2)(x) = x$

Enfin

$$L_k(f_3)(x) = \sum_{i=0}^{i=k} C_k^i x^i (1-x)^{k-i} \frac{i^2}{k^2}$$

Mais on peut calculer facilement

$$\begin{aligned} k^2 L_k(f_3)(x) - k L_k(f_2)(x) &= \sum_{i=0}^{i=k} i(i-1) C_k^i x^i (1-x)^{k-i} \\ &= x^2 \sum_{i=2}^{i=k} i(i-1) C_k^i x^{i-2} (1-x)^{k-i} \end{aligned}$$

Or

$$\begin{aligned} i(i-1)c_k^i &= k(k-1)c_{k-2}^{i-2} \\ &= k(k-1)x^2 \sum_{i=0}^{i=k-2} c_{k-2}^i x^i (i-x)^{k-i-2} \end{aligned}$$

c'est-à-dire

$$L_k(f_3)(x) - \frac{1}{k} L_k(f_2)(x) = \frac{k-1}{k} x^2$$

ce qui tend vers $x^2 = f_3$ lorsque k tend vers l'infini. En résumé, $L_k(f_i)$ pour $i=1, i=2, i=3$, tend vers f_i . Nous sommes donc dans les conditions du théorème de Korovkin permettant d'assurer que

$$\lim_{k \rightarrow +\infty} \sum_{i=0}^{i=k} c_k^i x^i (1-x)^{k-i} f\left(\frac{i}{k}\right) = f(x)$$

uniformément en x sur l'intervalle $[0,1]$ (ce qui est la formulation de Bernstein du théorème de Weierstrass).

Remarque : $L_k(f)(x)$ est un polynôme de degré k qui approxime la fonction continue f . Il faut remarquer que $L_k(f)$ n'est pas un polynôme de meilleure approximation pour la norme uniforme ou une norme L_p et ne coïncide pas avec l'interpolation polynomiale de Lagrange à l'ordre n . D'ailleurs la correspondance entre f et sa meilleure approximation dans L_p ($p \neq 2$) n'est pas linéaire !

On peut considérablement préciser le théorème de Weierstrass. En fait, on n'a pas besoin d'avoir tous les termes en $x^1, x^2, x^3, \dots, x^n, \dots$ pour réaliser la densité. Le théorème suivant, que nous donnons sans démonstration, illustre ce propos.

Théorème de Müntz : Soit $[\Gamma_n]$ une suite croissante de nombres réels positifs telle que $\Gamma_0 = 0$ et $\lim_{n \rightarrow \infty} \Gamma_n = +\infty$. Pour que les combinaisons linéaires finies des fonctions $f_n = x^{\Gamma_n}$ soient denses dans $\mathcal{C}[0,1]$ il faut et il suffit que la série $\sum_{n=1}^{\infty} \frac{1}{\Gamma_n}$ diverge.

Nous venons donc de résoudre le problème (1) pour l'espace des polynômes et en utilisant la norme uniforme, ou d'ailleurs une norme $L^p [a,b]$ quelconque ($p \geq 1$) (Cf Exercice N° 37).

Un résultat identique a lieu pour l'espace des polynômes trigonométriques, en utilisant la norme uniforme ou une norme $L^p [-\pi, +\pi]$ ($p \geq 1$) (Cf Exercice N° 35 : théorème de Fejér que nous retrouverons au Chapitre 5 sous une autre démonstration).

L'étude du problème (2) est peut-être plus importante pour l'utilisateur, une fois résolu le problème (1). Cependant, on conçoit aisément que des propriétés fines spécifiques de la fonction à approximer f vont alors intervenir. Toute une faune d'espaces vectoriels surgit pour procurer des théorèmes de plus en plus précis. Nous allons nous contenter de donner quelques énoncés typiques, reportant les démonstrations en exercices (Cf notamment les Exercices N° 39 et 40).

4. ESTIMATION DE LA MEILLEURE APPROXIMATION.

Définition : on dit qu'une fonction f est de la classe de Lipschitz (M, α) lorsque $f: \mathbb{R} \rightarrow \mathbb{R}$

$$|f(x) - f(y)| \leq M |x - y|^\alpha \quad \forall x, y \in \mathbb{R}$$

(les seuls cas intéressants sont pour $0 < \alpha \leq 1$).

Définition : on appelle module de continuité $\omega(f, h)$ d'une fonction f , l'expression

$$\omega(f, h) = \sup_{\substack{x, t \\ |t| \leq h}} |f(x+t) - f(x)|$$

Pour une fonction de la classe de Lipschitz, $\omega(f, h) \leq Mh^\alpha$

4.1. Cas des fonctions périodiques.

Commençons par C^* , espace des fonctions 2π -périodiques, pour la norme uniforme (Cf Exercice N° 47).

Théorème de Jackson: (1) Il existe une constante H telle que si f est un élément de C^* de module de continuité ω , on a pour tout x

$$|f(x) - P_n f(x)| \leq H \omega\left(f, \frac{1}{n}\right)$$

où $P_n(f)$ désigne une meilleure approximation au sens de la norme uniforme de f par un polynôme trigonométrique d'ordre n au plus.

(2) De même, il existe des constantes A_r telles que si f est un élément de C^* possédant des dérivées continues jusqu'à l'ordre r, cette dérivée d'ordre r étant de la classe de Lipschitz (M, α) alors avec les mêmes notations, on a pour tout x

$$|f(x) - P_n f(x)| \leq A_r M / n^{r+\alpha}$$

On pose $E_n(f) = \inf \|f - g\|_\infty$ où g parcourt les polynômes trigonométriques d'ordre n au plus. La partie (1) du théorème de Jackson se note

$$E_n(f) \leq A \omega\left(f, \frac{1}{n}\right)$$

On dispose d'un théorème réciproque.

Théorème de Bernstein : Soit α ; $0 < \alpha < 1$, et r est un entier. Supposons que $f \in C^*$ satisfasse

$$E_n(f) \leq A / n^{r+\alpha}$$

pour une certaine constante A. Alors f possède r dérivées et la dérivée d'ordre r est d'une classe de Lipschitz d'ordre α .

Pour $\alpha = 0$, on a aussi un théorème.

Théorème de Zygmund : Soit r un entier. Une fonction f de C^* satisfait

$$E_n(f) \leq A/n^r \quad \text{pour une certaine constante A}$$

si et seulement si la dérivée d'ordre $(r-1)$ de f satisfait la relation

$$\omega_2(f^{(r-1)}, h) \leq Ah$$

pour une certaine constante A. Ici, ω_2 signifie

$$\omega_2(f, h) = \sup_{\substack{x, t \\ |t| \leq h}} |f(x+t) + f(x-t) - 2f(x)|$$

Un résultat d'ailleurs très précis est dû à S.B. Stečkin qui lie $E_n(f)$ et sa régularité. Il importe de noter que le comportement de $E_n(f)$ caractérise bien souvent la régularité de f . Ainsi f est lipschitzienne d'ordre α ($0 < \alpha < 1$) si et seulement si $E_n(f) \leq A/n^\alpha$ pour une constante A convenable.

4.2. Cas des polynômes ordinaires.

On prend l'espace $\mathcal{C}[-1,+1]$ des fonctions continues définies sur $[-1,+1]$ et pour E_n l'espace des polynômes de degré n au plus. Comme dans le théorème de Jackson, en posant

$$E_n(f) = \inf \|f - g\|_\infty$$

où g parcourt les polynômes de degré n au plus, on a que $E_n(f)$ est au plus d'ordre $1/n$. Mais en fait, le théorème de Timan précise que l'approximation par des polynômes s'améliore beaucoup aux points frontières de l'intervalle. On peut donner des réciproques à ces théorèmes permettant de montrer qu'ils sont les meilleurs possibles. Nous ne donnerons qu'un seul théorème de ce genre.

1er théorème de Timan : Il existe une constante M telle que pour toute fonction f appartenant à $\mathcal{C}[-1,+1]$, il existe une suite de polynômes P_n pour lesquels P_n est de degré n au plus et

$$|f(x) - P_n(x)| \leq M \omega(f, \Delta_n(x))$$

pour $-1 \leq x \leq +1$ où l'on a posé

$$\omega(f, h) = \sup_{\substack{x, t \\ |t| \leq h}} |f(x+t) - f(x)| \quad \text{module de continuité de } f$$

et

$$\Delta_n(x) = \sup \left(\frac{\sqrt{1-x^2}}{n}, \frac{1}{n^2} \right) \quad n = 1, 2, \dots$$

2ème théorème de Timan : Si une fonction continue admet p dérivées continues sur $[-1,+1]$, il existe une suite de polynômes $P_n(x)$ de degré n au plus et

$$\|f(x) - P_n(x)\| \leq M_p (\Delta_n(x))^p \omega(f^{(p)}, \Delta_n(x))$$

la constante M_p ne dépendant que du nombre p .

Remarque : le premier théorème de Timan précise la convergence de P_n vers f . De fait, on obtient par des procédés dus à Jackson (Cf Exercices N° 39 et 40), la majoration moins forte

$$\|f - P_n\| \leq M \omega(f, \frac{1}{n})$$

où M est une constante et ω le module de continuité de f . En particulier, lorsque f est une fonction dont la dérivée première est bornée

$$\omega(f, \frac{1}{n}) = \sup_{\substack{x, t \\ |t| \leq \frac{1}{n}}} |f(x+t) - f(x)| \leq \frac{\|f'\|_\infty}{n}$$

On a de même une réciproque, choisie parmi bien d'autres cas possibles.

Théorème : Soit $0 < \alpha \leq 1$, un nombre entier r et une fonction f continue sur $[-1, +1]$. Une condition nécessaire et suffisante pour qu'il existe une suite de polynômes P_n (de degré n au plus) telle que

$$|f(x) - P_n(x)| \leq (\Delta_n(x))^{r+\alpha} \quad \forall x \in [-1, +1]$$

est d'une part l'existence des r premières dérivées de f , d'autre part que la dérivée d'ordre r appartienne à une classe de Lipschitz (M_r, α) où M_r ne dépend que de r .

4.3. autres cas.

Nous venons de préciser le comportement asymptotique du reste entre une fonction et une meilleure approximation polynômiale (en norme uniforme) à partir de la régularité de cette fonction. On a des résultats analogues en analyse harmonique, comme on l'a vu en 4.1. On sait que la série de Fourier d'une fonction f de $L^2[-\pi, +\pi]$ converge vers f en moyenne quadratique. Si f a des propriétés de régularité, sa série de Fourier peut donc converger uniformément (Cf Analyse harmonique et Exercice N° 47). Un tel phénomène est général. Donnons sans démonstration un seul exemple, parmi les très nombreux légués par les mathématiciens du siècle dernier. Rappelons que si f est une fonction définie sur $[-1, +1]$, à valeurs complexes et appartenant à $L^2[-1, +1]$, on appelle série de Legendre de f la série qui correspond au développement hilbertien de f suivant la famille orthogonale et totale $\{P_n\}_{n \geq 0}$ des polynômes de Legendre (Cf Chapitre 2 et E.A.F.).

Théorème : Soit f une fonction analytique sur un domaine contenant le segment $[-1, +1]$. Le développement de f en série de Legendre converge dans la plus grande ellipse dont les foyers sont $+1$ et -1 et à l'intérieur de laquelle f est holomorphe.

Si $f \sim \sum_{n=0}^{\infty} a_n P_n$, la somme des demi-axes de l'ellipse de convergence est donnée par

$$R = \lim_{n \rightarrow \infty} \inf |a_n|^{1/n}$$

5. ESTIMATION DE L'APPROXIMATION AVEC DES FONCTIONS SPLINES CUBIQUES.

En règle générale, entre tous les cas pour la norme uniforme, il n'est pas commode d'obtenir une meilleure approximation. Le théorème de Korovkin ou les méthodes d'interpolation par des fonctions splines fournissent des approximations qui dépendent linéairement de la fonction à approximer. A-t-on aussi une idée du reste de l'approximation ? La réponse est positive et nous allons reprendre les notations du Chapitre 3, § 3, sur les fonctions splines cubiques.

Théorème : Soit f une fonction réelle définie sur le segment $[a, b]$ et $\Delta_n = \{x_1 = a, x_2, \dots, x_n = b\}$ une division de ce segment. On suppose f quatre fois continuellement dérivable. Soit $P_n(f)$ l'unique fonction spline cubique interpolant f fournie par le théorème des splines cubiques. Alors il existe des constantes A , B et C telles que

$$\|f - P_n(f)\|_{\infty} \leq A |\Delta_n|^4$$

$$\|f' - (P_n(f))'\|_{\infty} \leq B |\Delta_n|^3$$

$$\|f'' - (P_n(f))''\|_{\infty} \leq C |\Delta_n|^2$$

où

$$|\Delta_n| = \sup_{k=1, \dots, (n-1)} |x_{k+1} - x_k|$$

Nous admettrons ce résultat qu'il n'est pas difficile de démontrer à titre d'exercice.

6. AUTRES THEOREMES LIMITES.

6.1. D'abord avec l'interpolation linéaire, on a

$$\lim_{n \rightarrow \infty} \|f - P_n(f)\|_{\infty} = 0$$

en utilisant les notations du Chapitre 3 § 1. La démonstration directe tient compte de l'uniforme continuité de f . Une autre démonstration plus fonctionnelle est possible (Cf Chapitre 5, § 4(e)).

6.2. Examinons maintenant le cas de l'interpolation par un polynôme (Cf Chapitre 3, § 2 dont nous conservons les notations). Les choses se passent ici tout à fait différemment car $\|P_n\|$ tend vers l'infini avec n (Cf Chapitre 5, § 4(e)).

L'exercice N° 45 donne un exemple où $\|P_n(f) - f\|_\infty$ tend vers l'infini avec n pour une fonction f très régulière puisqu'analytique. En fait, on dispose d'un théorème négatif :

Théorème de Faber : Soit $\{\Delta_n\}_{n \geq 1}$ une famille de division de $[a,b]$ où $\Delta_n \subset \Delta_{n+1}$. Il existe une fonction f continue sur $[a,b]$, telle que $P_n(f)$, polynôme de Lagrange associé à Δ_n et f , ne converge pas uniformément vers f .

On dispose de tout un arsenal de tels résultats négatifs. Toutefois si Δ_n est constitué par les zéros d'une famille $\{Q_n\}_{n \geq 1}$ de polynômes orthonormaux, et si $\lim_{h \rightarrow 0} \omega(f, h) \log h = 0$, alors $P_n(f)$ associé à Δ_n

et f converge uniformément vers f (théorème de Bernstein). La raison de ce résultat dépend d'une estimation précise de $\|P_n\|$ comme le Chapitre 5, § 4(e) l'explique.

6.3. Polynômes de Bernstein.

Nous avons noté que si $f \in \mathcal{C}[0,1]$, alors le polynôme de Bernstein associé

$$P_n f(x) = \sum_{i=0}^n C_n^i f\left(\frac{i}{n}\right) x^i (1-x)^{n-i}$$

converge uniformément vers f .

En fait, on dispose de renseignements nettement plus forts qui précisent l'emploi des polynômes de Bernstein

Théorème : Soit f une fonction possédant des dérivées continues jusqu'à l'ordre k et soit $P_n(f)$ le polynôme de Bernstein d'ordre n . Alors

$$\frac{d^k P_n(f)}{dx^k}(x) \quad \text{converge uniformément vers } f^{(k)}.$$

Le polynôme de Bernstein suit remarquablement certains mouvements de f ;
notamment

Théorème f une fonction convexe sur $[0, +1]$; alors le polynôme de
Bernstein $P_n(f)$ est aussi convexe sur $[0, 1]$.

Contre, en règle générale, $P_n(f)$ ne converge pas très vite
vers f . On précise :

Théorème (Vasky) : Soit f une fonction bornée sur $[0, 1]$ possédant une
dérivée seconde en chaque point. On a en désignant par $P_n(f)$ le polynôme de
Bernstein d'ordre n

$$\lim_{n \rightarrow \infty} n (P_n(f)(x) - f(x)) = \frac{x(1-x)}{2} f''(x) \quad \forall x \in [0, 1]$$

La convergence en un point où $f''(x) \neq 0$ est donc de l'ordre de $1/n$, ce qui
n'est guère remarquable.

(l'étude des polynômes de Bernstein est très poussée dans G. Lorentz - Bernstein
polynomialson to 1953).

- EXERCICES DU CHAPITRE 4 -

EXERCICE N° 34.

Que peut-on dire de $P(|f|)$ en fonction de $|Pf|$ lorsque P est un opérateur linéaire positif sur un espace fonctionnel (on montrera que $|Pf| \leq P(|f|)$ en utilisant la remarque que $P(fg)$ est un produit scalaire généralisé).

EXERCICE N° 35.

A partir de la remarque que $1 - \cos(x-y)$ est une fonction positive et nulle si et seulement si $x = y$, donc que 1 , $\cos x$ et $\sin y$ satisfont la condition K , en déduire le théorème de Weierstrass suivant :

Théorème : toute fonction continue de période 2π peut être approximée, d'ausssi près qu'on le désire, par un polynôme trigonométrique.

Peut-on retrouver ce théorème à partir du théorème de Weierstrass démontré dans le texte ?

EXERCICE N° 36.

Soit f une fonction continue sur $[0, +\infty[$ et dont la limite est nulle lorsque x tend vers $+\infty$. Montrer que l'opérateur L défini par

$$L_k(f)(x) = \sum_{k=0}^{k=+\infty} f\left(\frac{k}{n}\right) \frac{n^k}{k!} x^k e^{-nx}.$$

est un opérateur positif, bien défini.

Montrer que $\lim_{n \rightarrow +\infty} L_n(f) = f$

uniformément sur le demi-axe positif $[0, +\infty[$.

EXERCICE N° 37.

Montrer que l'espace des polynômes est dense dans tout espace $L^p [0,1]$ pour tout nombre entier $p \geq 1$ (on utilise la densité de $\mathcal{C} [0,1]$ dans $L^p [0,1]$ et la densité de l'espace des polynômes dans $\mathcal{C} [0,1]$ démontrée par le théorème de Weierstrass).

EXERCICE N° 38.

Soit f une fonction périodique et de période 2π et T_n un polynôme trigonométrique de degré n

$$T_n(x) = a_0 + a_1 \cos x + b_1 \sin x + \dots + a_n \cos nx + b_n \sin nx$$

Soit $M_n(f)$ l'expression

$$M_n(f) = \min_{(T_n)} \|f - T_n\| \quad \text{minimum pris pour l'ensemble des polynômes } T_n \text{ de degré inférieur ou égal à } n.$$

Le problème (1) consiste à montrer que $\lim_{n \rightarrow \infty} M_n(f) = 0$ et le problème (2) consiste à préciser cette tendance vers 0.

Montrer que si l'on pose

$$S_n(f) = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \frac{\sin(2n+1) \frac{t-x}{2}}{2 \sin \frac{t-x}{2}} dt$$

c'est-à-dire la somme partielle d'ordre n de la série de Fourier de F , ou encore

$$S_n(f) = \frac{a_0}{2} + \sum_{k=1}^{k=n} (a_k \cos kx + b_k \sin kx)$$

avec pour $k \geq 0$:

$$a_k = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \cos kt dt \quad \text{et} \quad b_k = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \sin kt dt$$

alors

$$\|f - S_n(f)\| \leq K M_n(f) \log n \quad (n > 1)$$

où K est une constante convenable.

Lorsque f est suffisamment régulière pour que les théorèmes usuels sur les séries de Fourier soient applicables, que peut-on déduire de ce résultat relativement au comportement à l'infini de $M_n(f)$.

EXERCICE N° 39. (suite du précédent)

Montrer le théorème suivant dû à Jackson : il existe une constante K telle que pour toute fonction continue périodique de période 2π

$$M_n(f) \leq K \omega\left(f, \frac{1}{n}\right)$$

(on pose $L_n(t) = \frac{1}{\lambda_n} \left(\frac{\sin \frac{n t}{2}}{\sin \frac{t}{2}}\right)^4$ où λ_n est tel que $\int_{-\pi}^{+\pi} L_n(t) dt = 1$)

puis on pose

$$J_n(f)(x) = \int_{-\pi}^{+\pi} f(x+t) K_n(t) dt$$

où $K_n(t) = L_{n'}(t)$ avec $n' = \left[\frac{n}{2} \right] + 1$. Le symbole $[\cdot]$ signifiant partie entière de n . L'opérateur J_n s'appelle l'opérateur de Jackson et son noyau est K_n .

On forme

$$f(x) - J_n(f)(x) = \int_{-\pi}^{+\pi} (2f(x) - f(x+t) - f(x-t)) K_n(t) dt$$

et une majoration conduit au résultat après avoir noté que J_n est un polynôme trigonométrique).

EXERCICE N° 40. (suite du précédent)

Montrer le théorème suivant dû à Jackson : pour tout entier p , il existe une constante $R(p)$ telle que pour toute fonction périodique de période 2π , ayant p dérivées continues, on ait

$$M_n(f) \leq \frac{R(p)}{n^p} \omega\left(f^{(p)}, \frac{1}{n}\right)$$

(on utilisera une généralisation du noyau de Jackson en posant :

$$L_{n,2}(t) = \frac{1}{\lambda_{n,2}} \left(\frac{\sin \frac{nt}{2}}{\sin \frac{t}{2}} \right)^{2n} \text{ et } \int_{-\pi}^{+\pi} L_{n,2}(t) dt = 1$$

Comment les deux théorèmes de Jackson précédents peuvent-ils être utilisés pour obtenir des renseignements sur l'approximation d'une fonction par des polynômes ? Comparer les résultats avec les théorèmes de Timan.

EXERCICE N° 41.

Soit X un espace normé et x_n une famille d'éléments indépendants. Soit y un élément de Y , on pose

$$E_n(y) = \min_{a_i} \left\| y - \sum_{i=1}^n a_i x_i \right\|$$

1) montrer que $E_n(y)$ a les propriétés d'une norme sauf la propriété de nullité

2) lorsque X est complet, étant donnée une suite décroissante de nombres positifs α_n avec $\lim_{n \rightarrow \infty} \alpha_n = 0$, on peut trouver un élément y de tel que

$$E_n(y) = \alpha_n \text{ pour } n \geq 1$$

(ce résultat est dû à Bernstein).

EXERCICE N° 42.

Soit f une fonction dans le cercle unité.

Posons :

$$E_n(f) = \sqrt{\text{Min}_{a_i} \iint_D |f(z) - \sum_{i=0}^n a_i z^i|^2 dx dy}$$

Montrer que si f est analytique dans $|z| < \rho$ avec $\rho > 1$ et non analytique dans $|z| < \rho'$ pour tout $\rho' > \rho$ alors

$$\lim_{n \rightarrow \infty} \text{Sup} (E_n(f))^{\frac{1}{n}} = \frac{1}{\rho}$$

EXERCICE N° 43.

Soit $L^2([a,b], \mu(x) dx)$ l'espace des classes de fonctions à valeurs complexes, définies sur un intervalle et telles que

$$\int_a^b |f(x)|^2 \mu(x) dx \quad \text{existe}$$

où μ est une fonction mesurable positive (strictement) sur $[a,b]$

Montrer qu'une condition nécessaire et suffisante pour qu'une famille orthonormale $\{P_n\}_{n \geq 0}$ soit une base hilbertienne est que

$$\sum_{n=0}^{\infty} \left| \int_a^x P_n(t) \mu(t) dt \right|^2 = \int_a^x \mu(t) dt \quad \forall x \in [a,b]$$

(la condition est nécessaire comme on le voit en appliquant la formule de Parseval à la fonction caractéristique du segment $[a,x]$. La suffisance est plus élaborée. Pour démontrer la totalité du système, on prend $g \in L^2([a,b], \mu(t) dt)$ telle que $\langle g, P_k \rangle = 0$ pour tout $k > 0$.

On montre alors que la relation donnée implique $\langle g, f \rangle = 0$ pour toute fonction f qui est une fonction caractéristique d'un intervalle $[a,x]$. Dès lors pour tout $b \geq x \geq a$

$$\int_a^x g(t) \mu(t) dt = 0$$

et on peut conclure $g \equiv 0$ au sens de l'espace considéré (Ce point exige à vrai dire un peu de théorie de la mesure mais est bien intuitif). Ce théorème est connu sous le nom de théorème de Vitali. Il fournit une condition originale de totalité d'un système orthonormal).

EXERCICE N° 44.

Soient x_1, x_2, \dots, x_n les zéros du polynôme de Tchebichev, d'ordre n ($x_k = \cos \frac{2k-1}{2n} \pi$) et f une fonction continue sur $[-1, +1]$.

Soit $P_{2n-1}(f)$ le polynôme de degré inférieur ou égal à $2n-1$ tel que

$$\begin{cases} P_{2n-1}(f)(x_k) & = & f(x_k) \\ P'_{2n-1}(f)(x_k) & = & 0 \end{cases}$$

Montrer que $\lim_{n \rightarrow \infty} P_{2n-1}(f)$ converge uniformément vers la fonction f

EXERCICE N° 45.

Soit α un nombre strictement positif. On pose

$$f(z) = \frac{1}{z^2 + \alpha^2} \quad \forall z \in \mathbb{C}$$

et

$$g(z) = \frac{f(z) \pi_n(x)}{(z-x) \pi_n(z)}$$

où $\pi_n(y) = (y-a_1)(y-a_2)\dots(y-a_n)$ et $a_k = \frac{2k+1}{2n}$

avec $-n \leq k \leq n-1$ ($n=2m$) est une division du segment $[-1, +1]$. On appelle

$P_n(f)$ le polynôme de Lagrange interpolant f aux $n=2m$ points $\{a_k\}$.

En calculant $\int_{\Gamma} g(z) dz$ sur un cercle Γ de centre 0 et de rayon R , supérieurs à $|\alpha|$ et 1, et en estimant certaines quantités, établir que pour α assez petit

$$\lim_{n \rightarrow \infty} |f(1) - P_n(f)(1)| = +\infty$$

(phénomène de Runge, sur la divergence de l'interpolation de Lagrange d'une fonction analytique).

- CHAPITRE 5 -

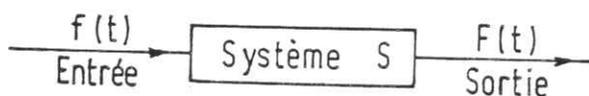
L'OPERATEUR DE CONVOLUTION

Dans le chapitre précédent, nous avons fourni une technique (théorème de Korovkin) permettant d'obtenir des résultats limites en théorie de l'approximation. Une autre technique malheureusement restreinte aux espaces de Hilbert est celle des bases hilbertiennes (Cf E.A.F.). Dans ce chapitre, nous allons envisager une troisième technique basée sur l'opérateur de convolution. Un tel opérateur est introduit à la fois pour des théorèmes limites et pour son rôle régularisant. Naturellement ces trois techniques permettent séparément de retrouver les théorèmes classiques de Weierstrass ou de Fejer. Pour définir l'opérateur de convolution, nous allons d'abord partir de considérations générales.

1. CONVOLUTION DE DEUX FONCTIONS PERIODIQUES DE CARRE SOMMABLE.

A. Introduction.

On rencontre fréquemment en physique des systèmes dont le fonctionnement global peut se schématiser de la façon suivante



Le système est une boîte qui réagit à une entrée en fournissant une sortie. Par exemple si S est une impédance, l'entrée est une intensité a priori variable avec le temps et la sortie est également une intensité. S peut être une télévision, un réacteur à piles ou un flipper. Toutefois des considérations particulières conduisent à supposer deux propriétés essentielles du comportement global :

(1) le système obéit à la règle de trois, plus précisément le système est linéaire. Si l'on multiplie l'entrée par λ , la sortie est multipliée par le même scalaire λ .

Si l'on ajoute deux entrées, la sortie est la somme des deux sorties correspondantes.

(2) le système est stable dans le temps : c'est-à-dire que si l'on décale l'entrée dans le temps, la sortie est décalée de la même manière. On dit aussi un système stationnaire.

Mathématiquement, on peut représenter le système S comme un opérateur qui fait passer de la fonction d'entrée f à la fonction de sortie F. On note

$$F = T(f)$$

Du point de vue théorique où nous nous plaçons, il importe de préciser la nature des fonctions f et F. Bien évidemment, l'espace fonctionnel des fonctions d'entrée et de sortie se doit de posséder deux propriétés :

- d'une part c'est un espace vectoriel, compte tenu de la propriété (1) ; existence d'une addition et d'une multiplication par un scalaire

- d'autre part, avec toute fonction f, il doit contenir la fonction f_h fonction translatée de la fonction f, définie par

$$f_h : t \longrightarrow f(t-h)$$

ceci compte tenu de la propriété (2) du système considéré.

Pour fixer les idées, nous allons nous placer dans l'espace $L^2(-\pi, +\pi)$, espace constitué par l'ensemble des fonctions périodiques et de période 2π , dont le carré est intégrable au sens de Lebesgue. L'expression

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(x) \overline{g(x)} dx$$

définit un produit scalaire sur cet espace et l'on sait qu'ainsi structuré $L^2[-\pi, +\pi]$ est un espace de Hilbert. La norme sur cet espace est donnée par l'expression

$$\|f\| = \langle f, f \rangle^{\frac{1}{2}}$$

On invoque fréquemment l'espace $L^2[-\pi, +\pi]$ lorsque l'on cherche des fonctions représentant l'énergie d'un système mécanique. En effet, cette énergie satisfait les équations

qu'explicite la mécanique analytique, mais si l'on ne veut pas, ou ne peut pas écrire la solution de ces équations, la première donnée sûre est l'appartenance à l'espace L^2 .

(la notion de périodicité est secondaire et utilisée pour éviter l'emploi de théorèmes plus délicats. En outre, il y a des avantages mathématiques à considérer des fonctions prenant des valeurs complexes).

Reprenons les données de notre étude,

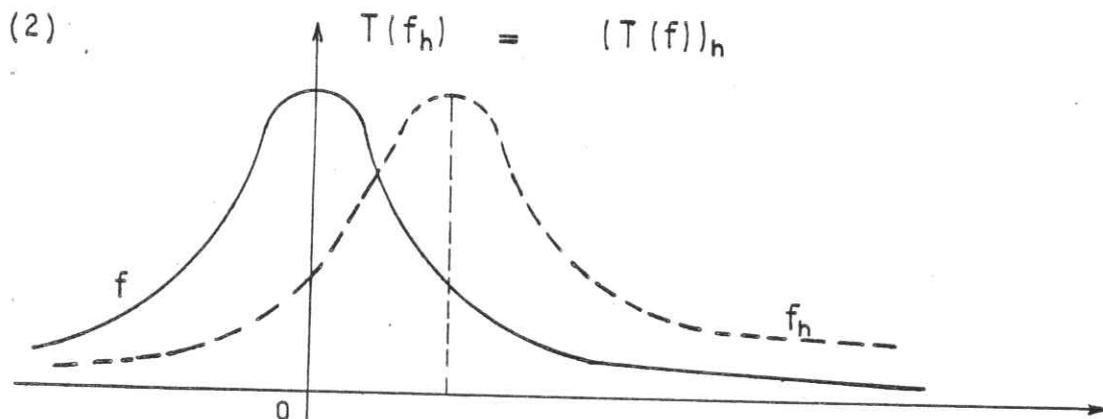
Nous avons un opérateur agissant sur L^2 et prenant ses valeurs dans L^2 . Traduisons maintenant les propriétés (1) et (2).

(1) T est un opérateur linéaire de L^2 dans L^2 .

(2) T commute avec les translations. Cette expression a la signification suivante :

soit f la fonction $t \mapsto f(t)$, la fonction f_h :

$f_h : t \mapsto f(t-h)$ est la fonction dite translatée de pas h de la fonction f et la propriété (2) de stabilité dans le temps signifie



Ces deux hypothèses étant posées, il paraît naturel de rechercher une représentation des opérateurs T , définis sur L^2 et satisfaisant (1) et (2). C'est ce que nous allons entreprendre.

B. Détermination des opérateurs T .

Pour étudier l'opérateur T , il convient d'examiner d'abord le comportement de T vis à vis de fonctions dont la translation s'exprime simplement à partir de la fonction de départ, par exemple les fonctions propres des opérateurs de translation.

Ainsi, prenons $f(t) = e^{i\omega t}$

$$f_h(t) = e^{i\omega(t-h)} = e^{-i\omega h} e^{i\omega t} = e^{-i\omega h} f(t)$$

$f_h(t)$ est un multiple de la fonction f .

Appliquons T à la fonction f_h :

$$T(f_h) = T(e^{-i\omega h} f) = e^{-i\omega h} T(f)$$

grâce à la linéarité de T . En outre, grâce à la propriété (2)

$$T(f_h) = (T(f))_h$$

Soit

$$(T(f))_h = e^{-i\omega h} T(f)$$

ou encore en posant $F = T(f)$

$$F(t-h) = e^{-i\omega h} F(t)$$

Changeant de variables (poser $h = -y$ et $t = 0$), il vient

$$F(y) = F(0)e^{i\omega y}$$

c'est-à-dire

$$(3) \quad T(f) = \lambda f \quad \text{où } \lambda \text{ est une fonction du paramètre } \omega (\lambda = F(0))$$

T transforme donc une exponentielle complexe en un multiple (peut-être nul) de cette même exponentielle complexe.

En optique, en électromagnétique ou en informatique, la fonction $\lambda(\omega)$ est connue sous le nom de réponse en fréquence du système S . Cette réponse en fréquence caractérise le système comme nous allons le prouver maintenant.

En effet, grâce aux séries de Fourier, nous pouvons calculer explicitement l'action de l'opérateur T sur toute fonction de L^2 , moyennant toutefois une condition supplémentaire de continuité.

Toute fonction f de l'espace L^2 s'écrit d'une façon unique sous la forme

$$f(t) = \sum_{n=-\infty}^{n=+\infty} C_n e^{int}$$

où la sommation s'entend au sens de la convergence en moyenne quadratique, c'est-à-dire

$$\lim_{N \rightarrow +\infty} \left\| f(t) - \sum_{n=-N}^{n=+N} C_n e^{int} \right\| = 0$$

où l'on a posé :

$$\left\| f(t) - \sum_{n=-N}^{n=+N} C_n e^{int} \right\| = \left(\frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| f(t) - \sum_{n=-N}^{n=+N} C_n e^{int} \right|^2 dt \right)^{\frac{1}{2}}$$

Par suite, si nous supposons que T est un opérateur continu (Cf E.A.F.), on peut écrire

$$F(t) = Tf(t) = T\left(\sum_{n=-\infty}^{n=+\infty} C_n e^{int}\right) = \sum_{n=-\infty}^{n=+\infty} C_n T(e^{int})$$

puisque la continuité permet de commuter T avec \sum . Soit

$$(a) f(t) = \sum_{n=-\infty}^{n=+\infty} C_n e^{int}$$

$$(b) Tf(t) = \sum_{n=-\infty}^{n=+\infty} C_n \lambda(n) e^{int}$$

Finalement, l'opérateur T consiste en la multiplication de chaque coefficient de Fourier C_n de la fonction d'entrée f par le nombre $\lambda(n)$. La réponse en fréquence détermine l'action du système. En fait, elle le caractérise.

Nous n'avons pas encore écrit que $T(f)$ appartient à l'espace L^2 et cette propriété va nous fournir un renseignement sur la fonction $\lambda(n)$.

Puisque T est un opérateur linéaire continu sur l'espace L^2 , on a l'inégalité fondamentale

$$(4) \quad \|T(f)\| \leq \|T\| \|f\| \quad (\text{Cf E.A.F.})$$

En outre la formule de Parseval nous permet de calculer la norme de f :

$$\|f\| = \left(\sum_{n=-\infty}^{n=+\infty} |c_n|^2 \right)^{\frac{1}{2}}$$

et

$$\|T(f)\| = \left(\sum_{n=-\infty}^{n=+\infty} |\lambda(n)|^2 |c_n|^2 \right)^{\frac{1}{2}}$$

L'inégalité (4) nous donne enfin

$$(c) \quad \left(\sum_{n=-\infty}^{n=+\infty} |\lambda(n)|^2 |c_n|^2 \right)^{\frac{1}{2}} \leq \|T\| \left(\sum_{n=-\infty}^{n=+\infty} |c_n|^2 \right)^{\frac{1}{2}}$$

cette inégalité étant valable pour toute fonction f , on peut en particulier prendre $f(t) = e^{int}$. Il vient $|\lambda(n)|^2 \leq \|T\|^2$

Soit (d) $|\lambda(n)| \leq \|T\|$

C'est-à-dire que les coefficients $\lambda(n)$ forment une suite bornée. Récapitulons les résultats acquis et énonçons une réciproque.

Théorème 1. Un opérateur linéaire défini sur l'espace L^2 et continu sur cet espace, commutant avec les translations, est déterminé par une suite bornée de nombres $\lambda(n)$ selon les formules (a) et (b).

Réciproquement, si l'on se donne une suite bornée de nombres réels ou complexes $\lambda(n)$, cette suite détermine, selon (a) et (b), un unique opérateur, linéaire, continu, commutant avec les translations sur $L^2[-\pi, +\pi]$

Pour la réciproque, il suffit de vérifier que (b) définit bien un élément de L^2 pour toute fonction f de L^2 . Ceci est une conséquence du théorème de Riesz-Fischer établissant un isomorphisme entre L^2 et ℓ^2 , et de la majoration

$$\sum_{n=-\infty}^{n=+\infty} |\lambda_n|^2 |c_n|^2 \leq \left(\sum_{n=-\infty}^{n=+\infty} |c_n|^2 \right) \sup_{n \in \mathbb{Z}} |\lambda_n|^2$$

La continuité et les propriétés (1) et (2) sont alors des conséquences faciles (Cf E.A.F.).

Essd'approfondir le théorème obtenu :

fonction f de $L^2[-\pi, +\pi]$ est déterminée par ses coefficients Fourier

$$\frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{-int} f(t) dt \quad (\text{théorème de Riesz-Fischer: Cf E.A.F.})$$

Supposons que les coefficients $\lambda(n)$, caractéristiques du système de l'opérateur T , soient eux-mêmes les coefficients de F d'une fonction g . Pour fixer les idées, nous supposons g appartient également à l'espace L^2 . Peut-on donner une expression de $T(f)$ selon une écriture simple en f et g ? Oui, à la formule généralisée de Parseval.

Il nous faut déterminer l'expression (b) et l'on pose

$$\lambda(n) \hat{g}(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} g(t) e^{-int} dt$$

Nous

$$(5) \quad \hat{h}(n) = \sum_{n=-\infty}^{+\infty} \hat{g}(n) \hat{f}(n) e^{inx} \quad \text{avec} \quad \hat{g}(n) = \lambda(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} g(t) e^{-int} dt$$

$$\text{et} \quad \hat{f}(n) = C_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) e^{int} dt$$

Posons $h(x) = \overline{g(x-t)}$, considérée comme une fonction de t , et calculons les coefficients de Fourier de h :

$$\hat{h}(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} h(t) e^{-int} dt = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \overline{g(x-t)} e^{-int} dt$$

$$= \frac{e^{-inx}}{2\pi} \int_{-\pi}^{+\pi} \overline{g(t)} e^{int} dt = \overline{\hat{g}(n)} e^{-inx}$$

$$\overline{\hat{h}(n)} = \hat{g}(n) e^{inx}$$

D'où

$$(6) \quad F(x) = \sum_{n=-\infty}^{n=+\infty} (\hat{f}(n) (\overline{\hat{h}(n)}))$$

Partons de la formule de Parseval (cf E.A.F.)

$$\|f + \mu h\|^2 = \sum_{n=-\infty}^{n=+\infty} |\hat{f}(n) + \mu \hat{h}(n)|^2$$

et développons en identifiant en μ , paramètre complexe arbitraire.

$$\text{Il vient : } \sum_{n=-\infty}^{n=+\infty} \hat{f}(n) \overline{\hat{h}(n)} = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) \overline{h(t)} dt$$

c'est ce que l'on appelle la formule de Parseval généralisée.

Donc en remplaçant h par sa valeur dans (6) et (7), on parvient à la formule

$$(8) \quad F(x) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) g(x-t) dt$$

Théorème 2 : La formule intégrale (8) montre que, sous les hypothèses faites, le système S est caractérisé par une fonction g et agit sur une fonction d'entrée f selon (8).

On convient de dire que F est la convolée des fonctions f et g . On note

$$F = f * g$$

Rémarque : On utilisait autrefois le mot produit de composition ou même "Faltung" pour désigner la convolution.

2. CONVOLUTION ET EXEMPLES FONCTIONNELS.

La définition précédente de la convolution est trop restrictive, en ce sens que l'intégrale (8) est définie quand bien même f et g ne sont pas des éléments de $L^2[-\pi, +\pi]$.

Définition 1 : On appelle convolution de deux fonctions périodiques de période 2π , f et g , la fonction définie par l'intégrale :

$$[f * g](x) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) g(x-t) dt$$

quand cette dernière existe pour toutes les valeurs de la variable x . Passons maintenant du groupe des nombres réels modulo 2π au groupe des nombres réels.

Définition 2 : On appelle convolution en x de deux fonctions f et g , définies sur l'axe réel, l'intégrale quand elle existe :

$$[f * g](x) = \int_{-\infty}^{+\infty} f(t) g(x-t) dt$$

Donnons quelques exemples.

(a) espace des fonctions de carré intégrable :

$L^2(\mathbb{R})$ est l'espace des fonctions définies sur l'axe réel et à valeurs réelles ou complexes telles que l'intégrale de Lebesgue,

$$\|f\|_2 = \left(\int_{-\infty}^{+\infty} |f(x)|^2 dx \right)^{\frac{1}{2}}, \text{ existe et soit finie (Cf E.A.F.)}$$

Pour deux éléments f et g de l'espace $L^2(\mathbb{R})$, la convolution est bien définie par

$$(9) \quad [f * g](x) = \int_{-\infty}^{+\infty} f(x-y) g(y) dy$$

En effet, grâce à l'inégalité de Schwarz, on a la majoration

$$\left| \int_{-\infty}^{+\infty} f(x-y) g(y) dy \right| < \left(\int_{-\infty}^{+\infty} |f(x-y)|^2 dy \right)^{\frac{1}{2}} \left(\int_{-\infty}^{+\infty} |g(y)|^2 dy \right)^{\frac{1}{2}}$$

$$|[f * g](x)| \leq \|f\|_2 \|g\|_2$$

Soit (10)

$$\boxed{\|f * g\|_{\infty} \leq \|f\|_2 \|g\|_2}$$

De plus, on montrerait d'une façon analogue à ce qui a été fait pour le cas périodique que :

$$(11) \quad [f * g](x) = \int_{-\infty}^{+\infty} e^{2i\pi xy} \hat{f}(y) \hat{g}(y) dy$$

où \hat{f} et \hat{g} sont les transformées de Fourier des fonctions f et g , dont l'existence est assurée par le théorème de Plancherel. Nous reviendrons plus loin, sur la transformée de Fourier et sur la formule (11) (cf 3b).

Comme \hat{f} et \hat{g} sont des éléments de l'espace L^2 , le produit $\hat{f}\hat{g}$ est absolument intégrable, ce qui prouve que $f * g$ est une fonction uniformément continue et tendant vers 0 à l'infini.

Cependant, en général $f * g$ n'est pas un élément de L^2 .

Théorème 3 : $f * g$ est une fonction uniformément continue, tendant vers 0 à l'infini lorsque f et g sont dans L^2 .

Quelques propriétés :

$$(12) \quad \|f * g\|_{\infty} \leq \|f\|_2 \|g\|_2$$

$$(13) \quad f * g = g * f \quad (\text{évident sur la formule (11)})$$

b) espace des fonctions définies sur le demi-axe positif :

Considérons la classe E des fonctions définies sur l'axe réel et nulles sur le demi-axe négatif. Cherchons à donner un sens à l'expression de la convolution. Pour éviter des difficultés, nous supposerons que les fonctions de la classe E sont continues.

Que vaut $[f * g](x) = \int_{-\infty}^{+\infty} f(x-y)g(y)dy$ lorsque f et g appartiennent à l'espace E ?

Cette intégrale peut s'écrire d'une autre façon puisque $g(y) = 0$ pour $y < 0$ et $f(x-y) = 0$ pour $y > x$. On trouve

$$[f * g](x) = \begin{cases} \int_0^x f(x-y)g(y)dy & \text{pour } x \geq 0 \\ 0 & \text{pour } x \leq 0 \end{cases}$$

Par suite $f * g$ est bien définie pour toutes les valeurs de la variable x et appartient à la classe E comme nous laissons le soin au lecteur de le vérifier.

On note dans la classe E

$$(14) \quad [f * g](x) = \int_0^x f(x-y)g(y)dy$$

Propriétés de la convolution :

$$(15) \quad f * g = g * f$$

$$(16) \quad (f * g) * h = f * (g * h)$$

On remarquera que mise à part la continuité, nous n'avons fait aucune hypothèse sur les éléments de la classe E (pas de borne ou d'appartenance à un espace du type L^p). L'essentiel est la nullité pour les valeurs négatives de la variable, voire seulement la nullité de f à gauche d'un certain nombre a, dépendant de f

c) classe des fonctions intégrables et des fonctions bornées.

Soit f une fonction de $L^\infty(\mathbb{R})$ et g une fonction de $L^1(\mathbb{R})$. Rappelons que l'espace $L^1(\mathbb{R})$ est constitué par l'ensemble des fonctions absolument intégrables au sens de Lebesgue (Cf. E.A.F.)

$$\|f\|_1 = \int_{-\infty}^{+\infty} |f(x)| dx$$

$L^\infty(\mathbb{R})$ est constitué par l'ensemble des fonctions bornées ; on note

$$\|g\|_\infty = \sup_{x \in \mathbb{R}} |g(x)|$$

En fait, les fonctions de L^∞ et L^1 doivent posséder en plus une certaine régularité (caractérisée par la mesurabilité) afin que l'intégrale $\int_a^b |g(x)| dx$ existe (au sens de Lebesgue), a et b étant des nombres réels finis. Cette remarque est de pure forme car toutes les fonctions bornées usuelles sont mesurables (Cf E.A.F.).

On démontre que les espaces $L^1(\mathbb{R})$ et $L^\infty(\mathbb{R})$ constituent des espaces de Banach (par construction, à partir des résultats de E.A.F.).

La fonction (17) $[f * g](x) = \int_{-\infty}^{+\infty} f(x-y) g(y) dy$ est bien définie puisque l'intégrale de la valeur absolue se majore selon

$$\begin{aligned} |[f * g](x)| &\leq \int_{-\infty}^{+\infty} |f(x-y)| |g(y)| dy \\ &\leq \sup_{y \in \mathbb{R}} |g(y)| \int_{-\infty}^{+\infty} |f(x-y)| dy \end{aligned}$$

Si f appartient à l'espace $L^1(\mathbb{R})$ et g à l'espace $L^\infty(\mathbb{R})$, il vient

$$(18) \quad \|f * g\|_\infty \leq \|f\|_1 \|g\|_\infty$$

La convolution d'une fonction de L^1 et d'une fonction de L^∞ est une fonction de L^∞ . De fait, on peut dire beaucoup plus sur $f * g$ grâce au théorème suivant dont l'importance est remarquable.

Soit f une fonction de l'espace $L^1(\mathbb{R})$ et f_h sa translatée de pas h . Lorsque h tend vers 0, $f_h(x)$ tend en chaque point x vers $f(x)$. Cependant, on peut obtenir un résultat plus précis selon le

Théorème 4 : Si f est un élément de L^1 , f_h converge vers f au sens de la norme L^1 lorsque h tend vers 0

$$\lim_{h \rightarrow 0} \|f - f_h\|_1 = 0$$

On sait que toute fonction absolument intégrable peut être approchée au sens de la norme L^1 , à $\frac{\epsilon}{3}$ près, par une fonction continue f_n nulle en dehors d'un intervalle borné $[a_n, b_n]$ (Cf E.A.F.)

$$\|f - f_n\|_1 = \int_{-\infty}^{+\infty} |f(x) - f_n(x)| dx \leq \frac{\epsilon}{3}$$

Dès lors,

$$\|f - f_h\|_1 \leq \|f - f_n\|_1 + \|f_n - f_{n,h}\|_1 + \|f_{n,h} - f_h\|_1$$

(dans cette expression $f_{n,h}$ désigne la translatée de pas h de la fonction f_n).

D'une part $\|f - f_n\|_1 = \|f_{n,h} - f_h\|_1$ grâce à l'invariance de la norme L^1 vis à vis de la translation.

En outre

$$\|f_n - f_{n,h}\|_1 = \int_{a_n - |h|}^{b_n + |h|} |f_n(x) - f_n(x-h)| dx$$

Comme f_n est continue sur $[a_n, b_n]$, cette fonction est uniformément continue, c'est-à-dire qu'il existe un nombre η choisi inférieur à $|b_n - a_n|$ pour lequel $|h| < \eta$ entraîne

$$\sup_{x \in \mathbb{R}} |f_n(x) - f_n(x-h)| \leq \frac{\epsilon}{4(b_n - a_n) + 4}$$

Finalement, pour $|h| \leq \eta$ et en prenant $2\eta < 1$

$$\|f - f_h\|_1 \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{4} \frac{b_n - a_n + 2n}{b_n - a_n + 1} + \frac{\varepsilon}{3} < \varepsilon$$

ce qui termine la démonstration.

On peut alors énoncer :

Théorème 5 : Si f est un élément de L^1 et g un élément de L^∞ , $f * g$ est une fonction bornée et uniformément continue.

La formule (18) établit que $f * g$ est bornée. Formons

$[f * g](x_1) - [f * g](x_2)$, il vient

$$\begin{aligned} |[f * g](x_1) - [f * g](x_2)| &\leq \int_{-\infty}^{+\infty} |f(x_1 - y) - f(x_2 - y)| |g(y)| dy \\ &\leq \|f - f_{x_1 - x_2}\|_1 \|g\|_\infty \end{aligned}$$

Le théorème 4 assure que $\lim_{h \rightarrow 0} \|f - f_h\|_1 = 0$, donc prouve la continuité uniforme de $f * g$.

Propriétés :

$$(18) \quad \|f * g\|_\infty \leq \|f\|_\infty \|g\|_\infty$$

$$(19) \quad f * g = g * f$$

(20) si f et h sont dans L^1 et g dans L^∞ , il y a associativité du produit de convolution

$$(f * g) * h = f * (g * h)$$

Exemple : calculons la convolution d'une fonction f de L^1 avec une exponentielle imaginaire $g(x) = e^{i\omega x}$

$$[f * g](x) = \int_{-\infty}^{+\infty} e^{i\omega(x-y)} f(y) dy = e^{i\omega x} \int_{-\infty}^{+\infty} e^{-i\omega y} f(y) dy$$

En notant

$$\hat{f}(y) = \int_{-\infty}^{+\infty} e^{-2i\pi xy} f(x) dx$$

la transformée de Fourier de la fonction f , on obtient la formule

$$f(x) * e^{i\omega x} = e^{i\omega x} \hat{f}\left(\frac{\omega}{2\pi}\right)$$

La convolution par une exponentielle imaginaire est un multiple de cette exponentielle. Cette dernière relation montre que si f et g sont différentes de 0, il se peut cependant que $f * g = 0$ (choisir f telle que $\hat{f}(\frac{\omega}{2\pi}) = 0$ et $g(x) = e^{i\omega x}$).

L^1 est un module sur L^∞ , mais n'est pas d'intégrité.

d) classe des fonctions absolument intégrables.

Soit f une fonction absolument intégrable et g une autre fonction ayant la même propriété. La fonction $x \longmapsto f(x-y)g(y)$ n'est pas toujours intégrable.

Par exemple si l'on prend

$$f(x) = g(x) = \frac{\exp(-|x|)}{\sqrt{|x|}} \text{ pour } x \neq 0 \text{ et } 0 \text{ pour } x = 0$$

le produit $f(x-y)g(y) = \frac{\exp|x-y|\exp|y|}{\sqrt{|x-y|} \sqrt{|y|}}$ considéré comme fonction de la variable y est intégrable de $-\infty$ à $+\infty$ pour toute valeur de x sauf pour $x = 0$. Toutefois

$$\iint_{\mathbb{R}^2} |f(x-y)g(y)| dx dy = \|f\|_1 \|g\|_1$$

Par suite, le théorème de Fubini (Cf E. A. F.), précise

que $x \longmapsto \int_{-\infty}^{+\infty} f(x-y)g(y) dy$ est une fonction absolument intégrable (bien que non définie partout!). En récapitulant :

Théorème 6 : Le produit de convolution de deux fonctions de L^1 est une fonction de L^1 et on a :

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1$$

Propriétés :

$$(21) \quad f * g = g * f$$

$$(22) \quad f * (g * h) = (f * g) * h$$

$$(23) \quad (f_1 + f_2) * g = f_1 * g + f_2 * g$$

$$(\lambda f) * g = \lambda (f * g)$$

L'opération de convolution est associative, commutative et distributive vis à vis de l'addition. Muni des lois + et * l'espace de Banach L^1 est structuré en une algèbre (Cf E.A.F. Chapitre 0), dite algèbre de Banach commutative.

Démontrons (22) pour donner un exemple de calcul sur la convolution

$$\begin{aligned} [f * (g * h)](x) &= \int_{\mathbb{R}} f(x-z)[g * h](z) dz \\ &= \int \int_{\mathbb{R}^2} f(x-z) g(z-y) h(y) dy dz \\ &= \int \int_{\mathbb{R}^2} f(x-z-y) g(z) h(y) dy dz \\ &= \int_{\mathbb{R}} [f * g](x-y) h(y) dy = [(f * g) * h](x) \end{aligned}$$

Bien d'autres cas sont envisageables pour assurer l'existence du produit de convolution. Il est inutile de donner des précisions supplémentaires. Les quatre exemples étudiés fournissent un bon aperçu des techniques. Il importe de bien noter que la convolution est une opération bien définie entre les espaces fonctionnels

$$L^2 \text{ et } L^2, \quad L^1 \text{ et } L^\infty \text{ ou } L^1 \text{ et } L^1$$

Remarque : dans ce paragraphe, nous nous sommes délibérément placés sur l'axe réel, toutefois le lecteur pourra aménager les théorèmes 3, 4, 5 et 6 dans le cas des fonctions périodiques. Nous donnerons plus loin un exemple. On pourra penser au cas de plusieurs variables également.

1. PROPRIETES DE LA CONVOLUTION.

Dans les quatre cas (a), (b), (c) ou (d) et lorsque ces expressions ont un sens, la convolution est commutative (12, 15, 16 et 21), associative (notamment 16, 20) et distributive vis à vis de l'addition (23).

Régularité.

D'autre part, le cas (c) montre que la convolée d'une fonction de L^1 et d'une fonction de L^∞ est plus "régulière" que les fonctions de départ. Nous allons voir comment cette propriété se réécrit en nous plaçant, par exemple, dans ce même cas

Théorème 7 : Soit f une fonction absolument intégrable et g une fonction bornée, n fois dérivable et dont la dérivée $n^{\text{ième}}$ est bornée. Alors la convolée $f * g$ est une fonction n fois dérivable, dont la dérivée $n^{\text{ième}}$ est fournie par la formule

$$(f * g)^{(n)}(x) = [f * g^{(n)}](x)$$

Constatons que la dérivabilité est une propriété transmise par hérédité par un seul des "parents" de la convolution !

Etablissons d'abord que si g et $g^{(n)}$ sont continues et bornées sur l'axe réel, alors il en est de même pour g' , g'' , ... et $g^{(n-1)}$.

Tout d'abord, on note par récurrence sur p que si I est un intervalle de longueur a et si $|g^{(p)}(x)| \geq \alpha > 0, \forall x \in I$, alors il existe un intervalle I_p inclus dans I , de longueur au moins $\frac{a}{4^p}$, tel que $|g^{(n-p)}(x)| \geq \alpha \frac{a^p}{2^{p(p+1)}}$ pour tout x dans I_p

(il suffit de découper I en quatre). Dès lors fixons un intervalle I de longueur a et notons

$$M_p = \sup_{x \in I} |g^{(p)}(x)|$$

Où bien $\frac{1}{2} M_{n-1} \leq \frac{a}{2} M_n$, ou bien $\frac{1}{2} M_{n-1} > \frac{a}{2} M_n$,

auquel cas $|f^{(k-1)}(x)|$ reste supérieure à $\frac{1}{2} M_{n-1}$ sur un intervalle de longueur au moins $\frac{a}{2}$ (théorème des accroissements finis) appliqué à $f^{(n-1)}$ et donc grâce à la remarque :

$$M_0 > \frac{M_{n-1}}{2} \cdot \frac{a^{n-1}}{2^{n(n-1)/2}}$$

En additionnant les majorations concernant les deux éventualités, on a :

$$\frac{1}{2} M_{n-1} \leq \frac{2^{(n-1)(n+1)}}{a^{n-1}} M_0 + \frac{a}{2} M_n$$

Le résultat s'en déduit puisque l'inégalité dépend seulement de la longueur de I

Nous pouvons nous contenter maintenant de montrer que $f * g$ est dérivable pour établir le théorème 7 puisque les n premières dérivées jouent un rôle comparable.

Or l'intégrale $\int_{-\infty}^{+\infty} f(y) g'(x-y) dy$, où f appartient à l'espace L^1 et g' à L^∞ , définit une fonction continue (Cf cas (c), du paragraphe 2) et de plus cette intégrale est uniformément

convergente sur tout intervalle borné parcouru par la variable x . Le théorème classique sur la dérivation sous le signe somme est applicable et procure

$$[f * g'](x) = \int_{-\infty}^{+\infty} f(y) g'(x-y) dy = (f * g)'(x)$$

(b) comportement vis à vis de la transformée de Fourier.

Dans notre introduction, nous avons directement relié le produit de convolution à la stabilité d'un opérateur face aux translations.

Cela nous incite à regarder les relations entre transformée de Fourier et produit de convolution. Nous le ferons seulement dans les deux cas (a) et (d) où la transformée de Fourier est classiquement définie pour les fonctions que l'on convole.

Cas des fonctions de carré intégrable :

Interprétons la formule (10) démontrée lors de l'étude du cas (a) des fonctions de carré intégrable

$$(10) \quad [f * g](x) = \int_{-\infty}^{+\infty} e^{2i\pi xy} \hat{f}(y) \hat{g}(y) dy$$

Cette formule indique, après utilisation du théorème de Plancherel qui est aussi un théorème d'inversion :

$$(25) \quad (f * g)^{\hat{}} = \hat{f} \cdot \hat{g}$$

On énonce :

Théorème 8 : la transformée de Fourier du produit de convolution de deux fonctions de L^2 est égale au produit des transformées de Fourier.

Cas des fonctions absolument intégrables :

Nous allons retrouver le théorème précédent au moyen d'un calcul direct. Pour un élément f de L^1 , on pose

$$(26) \quad \hat{f}(y) = \int_{-\infty}^{+\infty} e^{-2i\pi xy} f(x) dx$$

Calculons alors $(f * g)^{\hat{}}(y)$:

$$\begin{aligned}
 (f * g)^\wedge(y) &= \int_{-\infty}^{+\infty} e^{-2i\pi xy} [f * g](x) dx \\
 &= \int_{-\infty}^{+\infty} e^{-2i\pi xy} \left(\int_{-\infty}^{+\infty} f(x-t) g(t) dt \right) dx \\
 &= \iint_{\mathbb{R}^2} e^{-2i\pi(x-t)y} f(x-t) \cdot e^{-2i\pi ty} g(t) dt dx
 \end{aligned}$$

en utilisant la relation fonctionnelle de l'exponentielle complexe

$$(27) \quad e^{-2i\pi xy} = e^{-2i\pi(x-t)y} e^{-2i\pi ty}$$

Cependant l'intégrale double

$$\iint_{\mathbb{R}^2} e^{-2i\pi xy} f(x-t) g(t) dt dx$$

est absolument convergente, puisque f et g sont dans L^1 , et que

$$\iint_{\mathbb{R}^2} |f(x-t)| |g(t)| dx dt = \|f\|_1 \|g\|_1$$

On peut écrire

$$\begin{aligned}
 (f * g)^\wedge(y) &= \int_{-\infty}^{+\infty} e^{-2i\pi uy} f(u) du \int_{-\infty}^{+\infty} e^{-2i\pi ty} g(t) dt \\
 &= \hat{f}(y) \hat{g}(y)
 \end{aligned}$$

ce qui donne le théorème 8

$$(28) \quad \hat{f}(y) \hat{g}(y) = (f * g)^\wedge(y)$$

On devine combien cette propriété est fréquemment utilisée pour remplacer des équations où interviennent des convolutions par des équations où seuls figurent des produits.

4. QUELQUES UTILISATIONS DE LA CONVOLUTION.

(a) Régularisation des fonctions :

Soit f un élément de l'espace L^1 , nous savons théoriquement qu'il existe une fonction continue f_n approchant à moins de ϵ la fonction f , l'approximation ayant lieu au sens de la norme L^1 . Cependant, nous ne savons pas construire effectivement une fonction f_n . L'opérateur de convolution procure un procédé de manière pratique. Nous avons besoin d'un résultat préliminaire, assez long à établir, dont l'intérêt intrinsèque n'est toutefois pas négligeable.

noyularisants :

$K_h(x)$ une famille de fonctions de la variable x , continues sur \mathbb{R} et indexée par le paramètre h où $b > h > 0$.
Fais trois hypothèses suivantes :

1°) $K_h(x)$ est une famille absolument uniformément intégrable, ce qui signifie que pour tout $\varepsilon > 0$, il existe un $T_0 > 0$ tel que pour on ait

$$\int_{-\infty}^{-T_0} |K_h(x)| dx + \int_{T_0}^{+\infty} |K_h(x)| dx \leq \varepsilon$$

et d'autre part, il existe un M tel que pour tout h :

$$\int_{-\infty}^{+\infty} |K_h(x)| dx \leq M$$

$$2^\circ) \quad \lim_{h \rightarrow 0} \int_{-\infty}^{+\infty} K_h(x) dx = 1$$

3°) $K_h(x) = 0$ uniformément sur tout intervalle de la forme :
 $[\alpha, +\infty [$ ou $]-\infty, -\alpha]$

lorsque α est un nombre réel (strictement) positif.

Une telle famille K_h s'appelle une famille de noyaux régularisants. Son intérêt réside dans le théorème suivant :

Théorème : Lorsque f est une fonction absolument intégrable et continue à l'origine, et K_h une famille de noyaux régularisants, on a :

$$\lim_{h \rightarrow 0} \int_{-\infty}^{+\infty} K_h(x) f(x) dx = f(0)$$

Démonstration : Le comportement de la suite de fonctions K_h est bien clair sur le graphique. (cf fig. p.98). $K_h(x)$ tend à s'aplatir le long de l'axe des réels pour ne laisser subsister qu'un doigt de largeur en plus effilé le long de l'axe Oy .

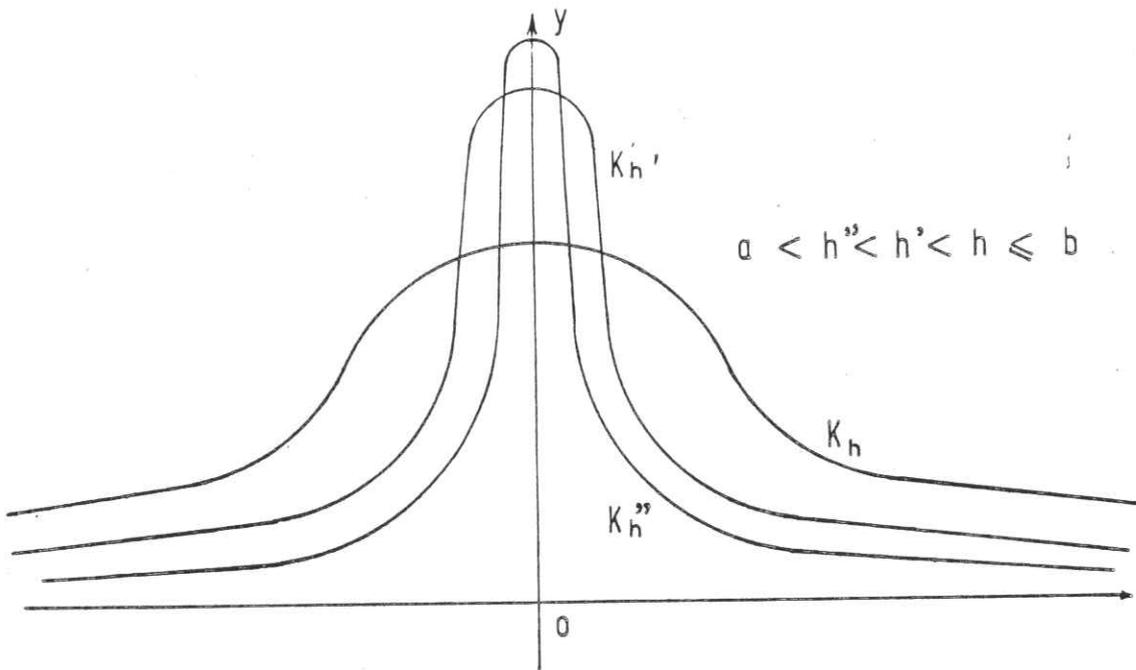
(La démonstration pourrait être un peu simplifiée en remplaçant (1) et (3) par

$$\int_{-\infty}^{+\infty} |K_h(x)| dx \leq M$$

et pour $\varepsilon > 0$ et tout $x_0 > 0$, il existe $H > 0$ tel que pour tout $h \geq H$

$$\int_{-x_0}^{x_0} |K_h(x)| dx + \int_{x_0}^{+\infty} |K_h(x)| dx < \varepsilon$$

En effet (1) et (3) impliquent ces conditions, mais pour conclure avec ces conditions il faudrait supposer f bornée en plus des autres propriétés).



Le graphique suggère une démonstration. Formons en effet :

$$(29) \quad J_h = \int_{-\infty}^{+\infty} K_h(x) (f(x) - f(0)) dx$$

Cette expression est bien définie pour toutes les valeurs positives de h comme nous allons l'établir :

Choisissons un nombre positif A . Grâce à l'hypothèse (3), pour un h assez petit, on a les majorations :

$$\sup_{x \in [\alpha, +\infty]} |K_h(x)| \leq A \quad \text{et} \quad \sup_{x \in]-\infty, -\alpha]} |K_h(x)| \leq A$$

D'une part :

$$\left| \int_{-\infty}^{+\infty} K_h(x) [f(x) - f(0)] dx \right| \leq \int_{-\infty}^{+\infty} |K_h(x)| |f(x)| dx + |f(0)| \int_{-\infty}^{+\infty} |K_h(x)| dx$$

D'autre part :

$$\int_{-\infty}^{+\infty} |K_h(x) f(x)| dx \leq \int_{-\infty}^{-\alpha} |K_h(x) f(x)| dx + \int_{-\alpha}^{+\alpha} |K_h(x) f(x)| dx + \int_{+\alpha}^{+\infty} |K_h(x) f(x)| dx$$

De plus :

$$\int_{-\infty}^{-\alpha} |K_h(x)| |f(x)| dx \leq A \int_{-\infty}^{-\alpha} |f(x)| dx$$

et

$$\int_{+\alpha}^{+\infty} |K_h(x) f(x)| dx \leq A \int_{+\alpha}^{+\infty} |f(x)| dx$$

$K_h(x)$ étant continue, il existe une constante M_h assurant la majoration

$$\sup_{x \in [-\alpha, +\alpha]} |K_h(x)| \leq M_h$$

Donc on obtient la majoration

$$\left| \int_{-\alpha}^{+\alpha} K_h(x) f(x) dx \right| \leq M_h \int_{-\alpha}^{+\alpha} |f(x)| dx$$

En définitive, on dispose d'une borne pour l'intégrale J_h , ce qui suffit à assurer sa convergence.

$$\int_{-\infty}^{+\infty} |(f(x) - f(0)) K_h(x)| dx \leq \left(\int_{-\infty}^{+\infty} |f(x)| dx \right) (A + M_h) + |f(0)| M$$

Passons à la démonstration du théorème. Puisque nous avons supposé f continue à l'origine, nous divisons l'intégrale J_h en trois parties :

$$(30) \quad \int_{-\infty}^{-x_0} K_h(x) (f(x) - f(0)) dx$$

$$(31) \quad \int_{-x_0}^{+x_0} K_h(x) (f(x) - f(0)) dx$$

et

$$(32) \quad \int_{x_0}^{+\infty} K_h(x) (f(x) - f(0)) dx$$

x_0 désignant un nombre (strictement positif) que nous allons bientôt déterminer. Soit ε un nombre positif donné. Traitons d'abord l'expression (31).

Grâce à la continuité de f à l'origine, nous pouvons fixer un nombre x_0 pour lequel

$$(33) \quad |x| < x_0$$

entraîne $|f(x) - f(0)| < \varepsilon$

(31) se majore selon

$$\left| \int_{-x_0}^{+x_0} K_h(x) (f(x) - f(0)) dx \right| < \varepsilon \int_{-x_0}^{+x_0} |K_h(x)| dx < M\varepsilon$$

Traitons maintenant l'expression (32). La fonction $K_h(x)$ ne converge pas uniformément vers la fonction 0 lorsque h tend vers 0, mais considérons l'intégrale :

$$J_h = \int_{x_0}^{+\infty} K_h(x) (f(x) - f(0)) dx$$

Cette dernière expression peut-elle être rendue inférieure à ε ? Malheureusement si l'on sort K_h du signe somme, il reste une expression non intégrable sur tout l'axe réel à cause de la présence du facteur constant $f(0)$. Cependant, d'après l'hypothèse (1)

$\int_{-\infty}^{+\infty} K_h(x) dx$ est une intégrale absolument uniformément convergente et l'on peut donc choisir un nombre T_0 tel que

$\int_{T_0}^{+\infty} |K_h(x)| dx < \varepsilon$ pour toute valeur de h . Ce nombre T_0 fixé, supérieur à 1 et au nombre x_0 , nous pouvons choisir, en utilisant l'hypothèse (3) de définition des noyaux régularisants, un nombre η tel que $0 < h < \eta$ implique :

$$(35) \quad \sup_{x \in [x_0, +\infty[} |K_h(x)| < \frac{\varepsilon}{T_0}$$

On obtient les majorations suivantes :

$$\int_{x_0}^{+\infty} |K_h(x) (f(x) - f(0))| dx \leq \int_{x_0}^{T_0} |K_h(x)| |f(x) - f(0)| dx + \int_{T_0}^{+\infty} |K_h(x)| |f(x) - f(0)| dx$$

$$\int_{T_0}^{+\infty} |K_h(x) (f(x) - f(0))| dx \leq \int_{T_0}^{+\infty} |K_h(x) f(x)| dx + |f(0)| \int_{T_0}^{+\infty} |K_h(x)| dx$$

$$\int_{x_0}^{T_0} |K_h(x) (f(x) - f(0))| dx \leq \int_{x_0}^{T_0} |K_h(x) f(x)| dx + |f(0)| \int_{x_0}^{T_0} |K_h(x)| dx$$

$$\int_{x_0}^{+\infty} |K_h(x) (f(x) - f(0))| dx \leq \sup_{x \in [x_0, +\infty[} |K_h(x)| \left[\int_{x_0}^{T_0} |f(x)| dx + |f(0)| (T_0 - x_0) \right]$$

$$+ \sup_{x \in [x_0, +\infty[} |K_h(x)| \int_{T_0}^{+\infty} |f(x)| dx + |f(0)| \int_{T_0}^{+\infty} |K_h(x)| dx$$

Finalement

$$(36) \quad \left| \int_{x_0}^{+\infty} K_h(x) (f(x) - f(0)) dx \right| \leq \frac{\varepsilon}{T_0} \int_{x_0}^{+\infty} |f(x)| dx + \varepsilon \frac{T_0 - x_0}{T_0} |f(0)| + \varepsilon |f(0)|$$

On majore de même l'expression (30) pour obtenir

$$(37) \quad \left| \int_{-\infty}^{-x_0} K_h(x) (f(x) - f(0)) dx \right| \leq \varepsilon \left(\int_{x_0}^{+\infty} |f(x)| dx + 2 |f(0)| \right)$$

en utilisant $T_0 > 1 > x_0$.

Additionnant les inégalités (34), (36) et (37) pour une estimation de J_h

$$(38) \quad |J_h| \leq \varepsilon \left(M + \int_{-\infty}^{+\infty} |f(x)| dx + 4 |f(0)| \right)$$

Ceci signifie que

$$\lim_{h \rightarrow 0} \int_{-\infty}^{+\infty} K_h(x) (f(x) - f(0)) dx = 0$$

Tenant compte de l'hypothèse (2), le résultat du théorème est acquis :

$$(39) \quad \lim_{h \rightarrow 0} \int_{-\infty}^{+\infty} K_h(x) f(x) dx = f(0)$$

Muni de ce théorème 9, sur les noyaux régularisants, nous allons envisager la convolution de la fonction f et d'un noyau régularisant pour ne plus particulariser l'origine. Les hypothèses faites assurent, lorsque f est une fonction absolument intégrable, que $f * K_h$ est une fonction continue et bornée (théorème 5). Quant au théorème 9, il ajoute que si f est une fonction continue, on a :

$$\lim_{h \rightarrow 0} [f * K_h](x) = f(x)$$

Pour s'en convaincre, il suffit d'appliquer le théorème 9 à la fonction $g_x(y) = f(x-y)$ qui est un élément de l'espace L^1 car :

$$\int_{-\infty}^{+\infty} |g_x(y)| dy = \int_{-\infty}^{+\infty} |f(y)| dy$$

De plus $g_x(y)$ est continue à l'origine puisque f l'est en tous points de l'axe réel. Il appert que

$$\int_{-\infty}^{+\infty} |g_x(y)| dy = \int_{-\infty}^{+\infty} |f(y)| dy$$

c'est-à-dire $f(x) = \lim_{h \rightarrow 0} \int_{-\infty}^{+\infty} K_h(y) g_x(y) dy$ pour tout nombre réel x .

Essayons d'approfondir ce résultat en tirant le maximum possible des inégalités établies dans la peine et l'effort lors du théorème 9. On pose $F_h = f * K_h$ et on cherche à préciser la convergence de la famille F_h lorsque h tend vers 0. On a d'abord :

$$|F_h(x) - g_x(0)| \leq \int_{-\infty}^{+\infty} |K_h(y)(g_x(y) - g_x(0))| dy + |g_x(0)| \left| \int_{-\infty}^{+\infty} K_h(y) dy - 1 \right|$$

Utilisons l'égalité (37)

$$|F_h(x) - g_x(0)| \leq \varepsilon \left(M + \int_{-\infty}^{+\infty} |g_x(y)| dy + 4|g_x(0)| \right) + \varepsilon |g_x(0)|$$

et quitte à diminuer le nombre η choisi en (35), on peut réaliser

$$\left| \int_{-\infty}^{+\infty} K_h(y) dy - 1 \right| < \varepsilon \text{ pour } 0 < h < \eta$$

Reprenant la notation $g_x(0) = f(x)$, il vient

$$|F_h(x) - f(x)| \leq \varepsilon \left(M + \int_{-\infty}^{+\infty} |f(x)| dx + 5|f(x)| \right)$$

Toutefois, nous avons utilisé l'inégalité (33) avec $f(x-y)$ au lieu de $f(y)$, c'est-à-dire que nous devons supposer la continuité uniforme de la fonction f . Ecrivons (41) sous une forme résumée :

$$(42) \quad \|f * K_h - f\|_{\infty} < \varepsilon (M + \|f\|_1 + 5\|f\|_{\infty})$$

On énonce le théorème :

Théorème 10 : Soit f une fonction définie sur l'axe réel, uniformément continue, bornée, et absolument intégrable. Si K_h est une famille de noyaux régularisants, $f * K_h$ converge uniformément vers f lorsque h tend vers 0.

La situation de ce théorème exige à la fois l'appartenance de f à L^∞ et la continuité. Que se passe-t-il en des points de discontinuité de f , cette dernière restant dans l'espace L^1 ?.

Nous savons que $F_h = f * K_h$ appartient à l'espace L^1 comme les fonctions composantes f et K_h . Calculons alors

$$\|f * K_h - f\|_1 = \int_{-\infty}^{+\infty} |f * K_h(y) - f(y)| dy$$

Il vient d'abord, d'après un groupement déjà effectué

$$(43) \int_{-\infty}^{+\infty} K_h(x) f(y-x) dx - f(y) = \int_{-\infty}^{+\infty} K_h(x) (f(y-x) - f(y)) dx + \left(\int_{-\infty}^{+\infty} K_h(x) dx - 1 \right) f(y)$$

Puis

$$\|f * K_h - f\|_1 \leq \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} |K_h(x) (f(y-x) - f(y))| dx \right] dy + \left| \int_{-\infty}^{+\infty} K_h(x) dx - 1 \right| \int_{-\infty}^{+\infty} |f(y)| dy$$

D'après l'hypothèse (2) sur les noyaux régularisants, la deuxième partie du second membre converge vers 0 lorsque h tend vers 0.

Quant à la première partie, posons :

$$p(x) = \int_{-\infty}^{+\infty} |f(y-x) - f(y)| dy$$

On encadre

$$0 < p(x) < 2\|f\|_1$$

et d'après le théorème 8, $p(x)$ est une fonction continue à l'origine.

Montrons que $\lim_{h \rightarrow 0} \int_{-\infty}^{+\infty} p(x) |K_h(x)| dx = 0$ selon la démonstration du théorème 8, bien que p n'appartienne pas à l'espace L^1 .

Toutefois la constatation $p(0) = 0$ va régler le problème :

$$\begin{aligned} \int_{-\infty}^{+\infty} p(x) |K_h(x)| dx &\leq \int_{-\infty}^{-T} p(x) |K_h(x)| dx + \int_{-T}^{-x_0} |K_h(x)| p(x) dx + \int_{-x_0}^{+x_0} |K_h(x)| p(x) dx \\ &\quad + \int_{-x_0}^{+T} |K_h(x)| p(x) dx + \int_T^{+\infty} p(x) |K_h(x)| dx \end{aligned}$$

Se donnant $\varepsilon > 0$, on choisit T tel que

$$\int_{-\infty}^{-T} |K_h(x)| dx + \int_T^{+\infty} |K_h(x)| dx < \varepsilon$$

(hypothèse (1) sur les noyaux) T fixé, on choisit $x_0 < T$ tel que $p(x) < \varepsilon$ pour $|x| < x_0$, grâce à la continuité à l'origine de p . Puis on choisit η pour que $0 < h < \eta$ entraîne

$$\sup_{x \in [x_0, T] \cup [-x_0, -T]} |K_h(x)| < \frac{\varepsilon}{T - x_0}$$

D'où en rappelant que

$$p(x) < 2\|f\|_1$$

Soit

$$\int_{-\infty}^{+\infty} p(x) |K_h(x)| dx \leq 4\varepsilon\|f\|_1 + 2\|f\|_1 \varepsilon + \varepsilon M$$

Ce qui termine la démonstration de

$$(44) \quad \lim_{h \rightarrow 0} \int_{-\infty}^{+\infty} p(x) |K_h(x)| dx = 0$$

Par suite

$$\|f * K_h - f\|_1 \leq \int_{-\infty}^{+\infty} p(x) |K_h(x)| dx + \left| \int_{-\infty}^{+\infty} K_h(x) - 1 \right| \int_{-\infty}^{+\infty} |f(y)| dy$$

En effet, comme $\iint K_h(x)(f(y-x) - f(y)) dx dy$ est absolument convergente, on peut intervertir les signes sommes dans le premier membre de (43) donnant finalement

$$(45) \quad \lim_{h \rightarrow 0} \|f * K_h - f\|_1 = 0$$

Énonçons le théorème :

Théorème 11 : Si K_h est une famille de noyaux régularisants et f une fonction absolument intégrable, $f * K_h$ tend vers f au sens de L^1 lorsque h tend vers 0.

Comme $f * K_h$ est une fonction continue, nous avons réalisé une approximation de f par une fonction continue. Il est fort possible de choisir K_h indéfiniment dérivable. Grâce au théorème 7, il en sera de même pour $f * K_h$ ce qui réalise une

approximation de f au sens de L^1 par des fonctions indéfiniment dérivables.

Exemple :

(α) on peut prendre pour K_h la fonction $K_h = \left(\frac{h}{h^2+x^2}\right) \frac{2}{\pi}$. On vérifiera facilement les propriétés requises pour les noyaux régularisants.

(β) On peut également prendre pour K_h la fonction $K_h = \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{x^2}{2h}\right)$ etc.

(b) Unité approchée dans L^1 .

On peut se poser le problème suivant, Existe-t-il une fonction f de L^1 telle que $f * g = g$ pour tout g dans L^1 ? Si tel est le cas, prenons les transformées de Fourier des deux membres, il vient :

$$\hat{f} \cdot \hat{g} = \hat{g} \text{ pour tout } g \text{ dans } L^1.$$

Seulement, en chaque point y , on peut trouver un élément g tel que $\hat{g}(y) \neq 0$. Par suite $\hat{f}(y) \equiv 1$. Or toute transformée de Fourier d'une fonction de L^1 tend vers 0 à l'infini, et le problème posé n'a pas de solution : il n'existe pas d'unité de convolution dans L^1 .

Toutefois, nous avons construit une famille K_h de fonctions telle que pour tout f de L^1 on a

$$\lim_{h \rightarrow 0} f * K_h = f$$

On convient de dire que K_h constitue une unité approchée dans L^1 .

Remarquons notamment que :

$$\lim_{h \rightarrow 0} (K_h)^\wedge(y) \hat{f}(y) = \hat{f}(y)$$

Mais pour tout y , on peut trouver f dans L^1 telle que $\hat{f}(y) \neq 0$ et par suite :

$$\lim_{h \rightarrow 0} (K_h)^\wedge(y) = 1$$

On construit quelquefois K_h à partir de cette propriété.

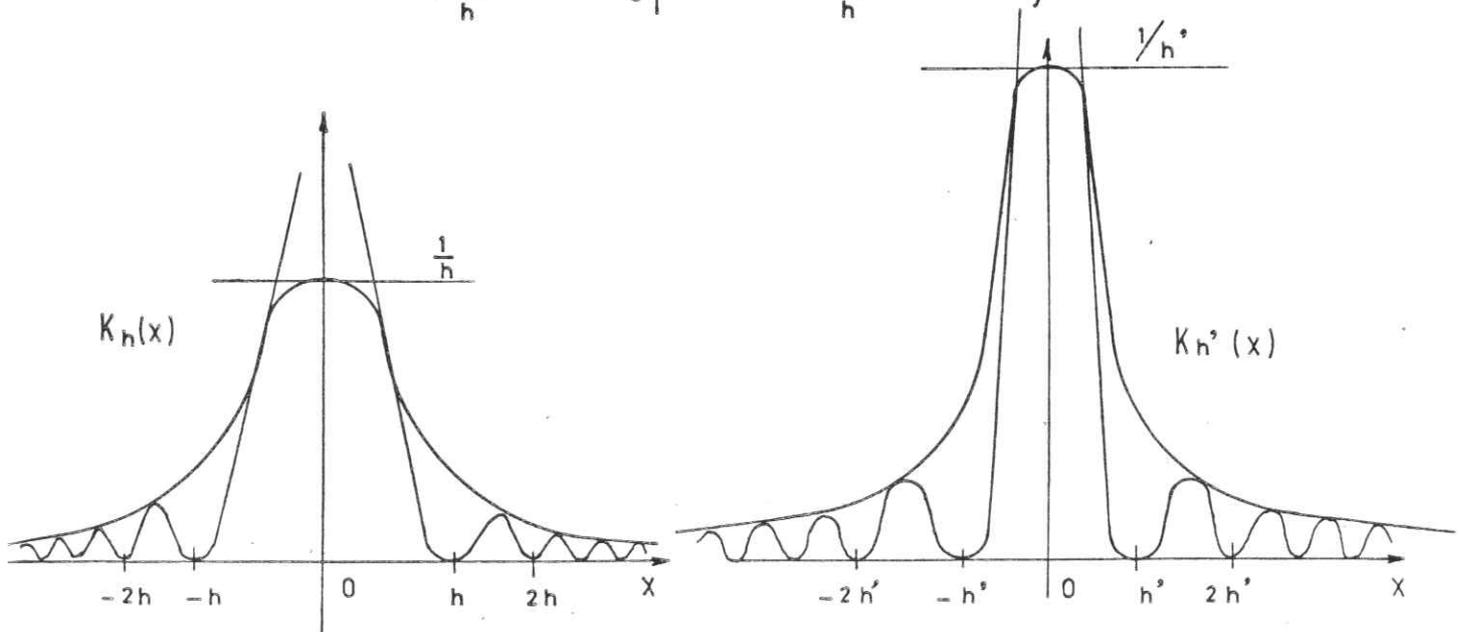
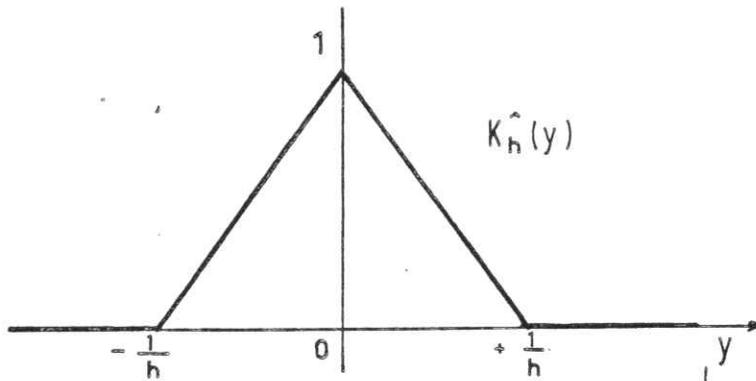
Exemple : on donne

$$(46) \quad (K_h)^\wedge(y) \begin{cases} 1 - h|y| & \text{si } |y| < \frac{1}{h} \\ 0 & \text{si } |y| \geq \frac{1}{h} \end{cases}$$

On calcule

$$(47) \quad K_h(x) = \int_{-\infty}^{+\infty} (K_h)^\wedge(y) e^{+2i\pi xy} dy$$
$$K_h(x) = \frac{1}{h} \left(\frac{\sin \frac{2\pi x}{2h}}{\frac{2\pi x}{2h}} \right)^2$$

c'est un élément de $L^1(\mathbb{R})$.



Figures Graphes de K_h et K_h^\wedge

On vérifie bien que $K_h(x) \geq 0$ et que $\int_{-\infty}^{+\infty} K_h(x) dx = (K_h)^\wedge(0) = 1$

Les autres propriétés sont faciles. La convolution de f par K_h supprime les fréquences de la transformée de Fourier supérieures à $1/h$. Le noyau K_h est appelé noyau de Fejer et permet de démontrer des théorèmes de convergence sur les séries et les intégrales de Fourier. (Cf Analyse Harmonique).

(c) Théorèmes de convergence.

Le noyau de Fejer va nous conduire à préciser le théorème d'inversion de la transformée de Fourier. Soit f une fonction de $L^1(\mathbb{R})$ et \hat{f} sa transformée de Fourier, laquelle peut ne pas appartenir à L^1 . Rappelons que :

$$(48) \quad \hat{f}(y) = \int_{-\infty}^{+\infty} e^{-2i\pi xy} f(x) dx$$

Sans hypothèse supplémentaire, $\int_{-\infty}^{+\infty} e^{-2i\pi xy} \hat{f}(y) dy$ ne converge pas en général vers $f(x)$. Toutefois, nous savons d'après le théorème 9 que $f * K_h$ converge uniformément vers f lorsque f est bornée et uniformément continue. Nous allons donc d'abord appliquer le théorème

d'inversion à la fonction $f * K_h$ puis nous passerons à la limite. On a constaté que $(f * K_h)^\wedge = \hat{f} \cdot (K_h)^\wedge$ d'après la formule (28).

$f * K_h$ par contre est une fonction indéfiniment dérivable à laquelle le théorème d'inversion usuel est applicable (cf Analyse Harmonique élémentaire) selon les formules

$$(49) \quad [f * K_h](x) = \int_{-\infty}^{+\infty} e^{2i\pi xy} \hat{f}(y) (K_h)^\wedge(y) dy$$

$$\text{Or} \quad [f * K_h](x) = \int_{-1/h}^{+1/h} e^{2i\pi xy} \hat{f}(y) (1-h|y|) dy$$

d'après la formule (47).

Posons $k = \frac{1}{h}$ pour simplifier l'écriture, il vient :

$$(50) \quad [f * K_h](x) = \int_{-k}^{+k} e^{2i\pi xy} \hat{f}(y) \left(1 - \frac{|y|}{k}\right) dy$$

Nous allons donner une autre forme à l'égalité (50). Pour cela, nous calculons d'abord :

$$\int_0^k e^{2i\pi xy} \hat{f}(y) \left(1 - \frac{y}{k}\right) dy = \int_0^k e^{2i\pi xy} \hat{f}(y) dy - \frac{1}{k} \int_0^k y \hat{f}(y) e^{2i\pi xy} dy$$

Intégrons par parties la deuxième intégrale du second membre :

$$-\frac{1}{k} \int_0^k y \hat{f}(y) e^{2i\pi xy} dy = -\frac{1}{k} \left[y \int_0^y \hat{f}(t) e^{2i\pi xt} dt \right]_0^k + \frac{1}{k} \int_0^k dy \int_0^y \hat{f}(t) e^{2i\pi xt} dt$$

Soit :

$$\int_0^k e^{2i\pi xy} \hat{f}(y) \left(1 - \frac{y}{k}\right) dy = \frac{1}{k} \int_0^k dy \int_0^y \hat{f}(t) e^{2i\pi xt} dt$$

De même

$$\int_{-k}^0 e^{2i\pi xy} \hat{f}(y) \left(1 + \frac{y}{k}\right) dy = \frac{1}{k} \int_0^k dy \int_{-y}^0 \hat{f}(t) e^{2i\pi xt} dt$$

D'où le résultat

$$(51) \quad \left[f * K_{\frac{1}{k}} \right](x) = \frac{1}{k} \int_0^k dy \int_{-y}^y \hat{f}(t) e^{2i\pi xt} dt$$

et

$$(52) \quad \lim_{k \rightarrow +\infty} \left[f * K_{\frac{1}{k}} \right](x) = f(x)$$

En remarquant que $\frac{1}{k} \int_0^k$ est une moyenne et en appelant transformée de Fourier tronquée la fonction valant $\hat{f}(t)$ pour $|t| \leq y$ et 0 ailleurs, on énonce le théorème suivant qui améliore le théorème d'inversion.

Théorème 12 : Soit f une fonction uniformément continue, bornée et absolument intégrable sur l'axe réel. La moyenne de l'inverse de Fourier de la transformée de Fourier tronquée converge uniformément sur l'axe réel vers la fonction f.

Ce théorème peut rendre des services lors de la manipulation des transformées de Fourier.

Par la même méthode de la convolution, on déduit facilement un théorème d'unicité.

Théorème 13 Deux fonctions absolument intégrables sur l'axe réel ayant la même transformée de Fourier sont égales presque partout

Traduisons maintenant le théorème 12 dans le cadre des fonctions périodiques et de période 2π .

La transformée de Fourier tronquée en N est la somme partielle :

$$S_N(x) = \sum_{n=-N}^{n=+N} C_n e^{inx}$$

où C_n est le $n^{\text{ième}}$ coefficient de Fourier de f , défini par

$$C_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) e^{-int} dt$$

La moyenne arithmétique des sommes partielles est l'expression

$$\frac{S_0(x) + \dots + S_{N-1}(x)}{N} . \text{ On obtient le théorème de Fejer.}$$

Théorème 14 : La moyenne arithmétique des sommes partielles de Fourier d'une fonction continue, périodique et de période 2π converge uniformément vers cette fonction.

(d) Applications aux équations différentielles.

Parce que la théorie en jeu fait intervenir les distributions tempérées, nous allons seulement opérer sur un exemple très simple. Soit l'équation différentielle

$$(53) \quad y''(x) + \omega^2 y(x) = f(x)$$

où ω désigne un nombre positif.

Nous cherchons une solution de (53) satisfaisant aux deux conditions aux limites.

$$(54) \quad \begin{cases} y(0) = a \\ y'(0) = b \end{cases}$$

Il s'agit de résoudre un problème de Cauchy. $\cos \omega x$ et $\sin \omega x$ sont deux solutions indépendantes de l'équation (53) sans second membre. Par la méthode de la variation des constantes, on obtient la solution générale de (52) selon :

$$(5.5) \quad y(x) = \left(\int_0^x \frac{\int_0^z f(t) \cos \omega t}{\cos^2 \omega z} dz \right) \cos \omega x + \alpha \left(\int_0^x \frac{dt}{\cos^2 \omega t} \right) \cos \omega x + \beta \cos \omega x$$

Compte tenu des conditions initiales, on trouve $\alpha = y'(0) = b$ et $\beta = y(0) = a$.

On peut mettre (55.) sous une forme différente en intégrant par parties :

$$y(x) = \left(-\frac{1}{\omega} \int_0^x f(t) \sin \omega t \, dt + \frac{1}{\omega} \left[\operatorname{tg.} \omega t \int_a^b f(z) \cos \omega z \, dz \right]_0^x \right) \cos \omega x + \frac{b}{\omega} \sin \omega x + a \cos \omega x$$

Il vient encore :

$$y(x) = \frac{1}{\omega} \int_0^x f(t) \left[\sin \omega t \cos \omega x - \sin \omega x \cos \omega t \right] dt + \frac{b}{\omega} \sin \omega x + a \cos \omega x$$

Soit :

$$y(x) = \frac{1}{\omega} \int_0^x f(t) \sin \omega(x-t) dt + \frac{b}{\omega} \sin \omega x + a \cos \omega x$$

Dans le cas particulier où $a = b = 0$ et où $f(t)$ est un élément de E (c'est-à-dire que f est nulle sur l'axe réel négatif : Cf § 3), on trouve la solution du problème sous la forme

$$(56) \quad y = \frac{1}{\omega} f(t) * \sin \omega t$$

On peut vérifier directement que (56) est solution (unique) de l'équation différentielle (53). La forme (56) prouve toutes les propriétés intéressantes de la solution de (53), mieux qu'une expression explicitement calculée. La situation présente est générale et l'on peut démontrer que la solution d'une équation différentielle linéaire à coefficients constants de la forme : $D(y) = 0$ où D est un polynôme de dérivation, pour un problème de Cauchy dans la classe E , peut s'écrire comme convolution de f . Ce résultat, dans le cas particulier où l'on recherche les solutions dans l'espace $L^2[-\pi, +\pi]$, peut directement se comprendre à partir de notre étude des opérateurs linéaires T commutant avec les translations (prendre ici $T = D$, mais D bien que commutant avec les translations, n'est pas défini sur tout l'espace L^2 puisque D est de la forme :

$$D(f) = \left(a_n \frac{d^n f}{dt^n} + a_{n-1} \frac{d^{n-1} f}{dt^{n-1}} + \dots + a_1 \frac{df}{dt} + a_0 \right)$$

On verra intervenir également la convolution pour la résolution du problème de Sturm-Liouville sur les équations différen-

tielles; g s'appelle alors la fonction de Green de l'équation différentielle (53), où g est la fonction qui, convolée avec f , donne la solution du problème de Sturm-Liouville. (cf E.A.F. chap 3)

(e) Application à la théorie de l'approximation.

La convolution par des noyaux régularisants fournit des opérateurs linéaires P_n qui peuvent servir à l'approximation d'une fonction. Mettons les choses en ordre. Soit $P_n : F \longrightarrow F_n$ une application linéaire, continue et idempotente, sur un espace normé F et à valeurs dans F_n , sous-espace vectoriel de dimension n de F . On pose :

$$E_n(f) = \inf_{g \in F_n} \|f - g\|$$

Alors

$$\|f - P_n(f)\| \leq E_n(f) (\|P_n\| + 1)$$

En effet, pour $g \in F_n$, on a puisque $P_n(g) = g$ pour tout g de F_n :

$$\begin{aligned} \|f - P_n(f)\| &\leq \|f - g + g - P_n(f)\| \\ &\leq \|f - g\| + \|P_n(g - f)\| \\ &\leq (1 + \|P_n\|) \|f - g\| \end{aligned}$$

Par suite, si $\lim_{n \rightarrow \infty} E_n(f) (\|P_n\| + 1) = 0$, on en déduit un théorème limite en théorie de l'approximation. Il y a donc grand intérêt à bien estimer $\|P_n\|$ tandis que $E_n(f)$ dépend de la régularité de f (Cf Chapitre 4).

Exemple : F est l'ensemble des fonctions continues réelles et 2π périodiques ; F_{2n+1} le sous-espace engendré par $1, \cos x, \sin x, \dots, \cos nx, \sin nx$. On prend la norme de la convergence uniforme

$$(1) \quad P_n(f) = f * D_n = \int_{-n}^{+n} f(x-t) D_n(t) dt$$

$$\text{où } D_n(t) = \frac{1}{2\pi} \frac{\sin(2n+1)t/2}{\sin t/2} \quad (\text{noyau de Dirichlet}).$$

On constate que P_n est un opérateur linéaire et continu de F sur F_{2n+1} , d'ailleurs idempotent (Puisque le coefficient de Fourier d'ordre h de D_n vaut 0 si $|h| > n$ et 1 si $h \leq n$).

On montre que (Cf Exercice N° 50) :

$$\|P_n\| = \int_{-n}^{+n} |D_n(t)| dt = \frac{4}{\pi^2} \text{Log} n + \epsilon_n \quad \text{où } \epsilon_n \text{ reste borné.}$$

Donc $\lim_{n \rightarrow \infty} \|f - P_n(f)\| = 0$ dès que $\lim_{n \rightarrow \infty} (E_n(f) \text{Log} n) = 0$

c'est-à-dire si f est suffisamment régulière (Cf Chapitre 4), par exemple si f possède une dérivée première bornée.

EXERCICES DU CHAPITRE 5

EXERCICE N° 45.

Soit E l'espace des fonctions continues, à valeurs réelles ou complexes, définies sur l'axe réel et nulles en dehors d'un certain intervalle (dépendant bien entendu de la fonction f choisie).

a) Montrer que E est un espace vectoriel normé par la norme de la convergence uniforme

$$\|f\| = \sup_{x \in \mathbb{R}} |f(x)|$$

b) On définit

$$[f * g](x) = \int_{\mathbb{R}} f(x-y) g(y) dy$$

Montrer que muni de la loi $*$, E constitue une algèbre normée. Est-ce une algèbre de Banach ?

c) Montrer que E est une algèbre sans diviseur de zéro, c'est-à-dire que $f * g = 0$ implique, soit $f = 0$, soit $g = 0$ (pour ce faire on pourra montrer que la transformée de Fourier de f , définie par

$$\hat{f}(z) = \int_{-\infty}^{+\infty} f(x) e^{-izx} dx,$$
 est une fonction prolongeable en une

fonction holomorphe de la variable z sur tout le plan complexe).

Le résultat obtenu en (c) présente-t-il un intérêt pour la résolution des équations de convolution ?

EXERCICE N° 46.

Soit f et g deux fonctions définies sur $[1, \infty[$. On pose lorsque l'intégrale à un sens

$$F \circledast g(x) = \int_1^{\infty} f\left(\frac{x}{t}\right) g(t) \frac{dt}{t}$$

Etudier, avec des hypothèses convenables sur f et g l'opération \circledast en comparant avec l'opération de convolution.

EXERCICE N° 47.

Soit k une fonction fixée dans $L^1[-\pi, +\pi]$. On définit un opérateur qui à tout f de $L^2[-\pi, +\pi]$ associe $Pf = k * f$ selon la formule (8) du chapitre 5.

(1) Montrer que $P : L^2[-\pi, +\pi] \longrightarrow L^2[-\pi, +\pi]$ est un opérateur linéaire et continu.

(2) On appelle spectre de P l'ensemble des valeurs complexes λ telles que $P - \lambda I$ ne possède pas d'inverse continu en tant qu'opérateur linéaire continu sur $L^2(\mathbb{R})$.

Montrer que le spectre de P s'identifie à l'ensemble composé de 0 et des valeurs $C_n(k)$ où n parcourt l'ensemble des entiers relatifs et où $C_n(k)$ est le coefficient de Fourier d'ordre n de la fonction k .

[On montrera que si λ n'appartient pas à ce dernier ensemble, $|\lambda - C_n(k)|$ est borné inférieurement lorsque $n \in \mathbb{Z}$].

(3) Que peut-on dire de l'ensemble des valeurs propres de l'opérateur P ?

(4) Interpréter les résultats de la théorie des opérateurs compacts auto-adjoints telle que développée en E.A.F. en supposant la fonction k réelle et paire.

EXERCICE N° 48.

Une famille $\{K_n\}_{n \geq 1}$ de fonctions 2π -périodiques forme une famille de noyaux ultra-réguliers si l'on a

$$(1) \quad \lim_{n \rightarrow \infty} \int_{-\pi}^{+\pi} K_n(t) dt = 1$$

$$(2) \quad |K_n(t)| \leq H_n(t) \quad \forall t \in [-\pi, +\pi] ; \quad \forall n > 1$$

où $H_n(t)$ est une fonction continuellement dérivable, 2π -périodique et qui satisfait les relations suivantes

$$\text{d'une part} \quad \int_{-\pi}^{+\pi} |t \frac{dH_n}{dt}(t)| dt \leq M \quad \forall n > 1$$

d'autre part
$$\lim_{n \rightarrow \infty} \sup_{\pi \geq |t| \geq \delta > 0} |t \frac{dH_n(t)}{dt}| = 0$$

pour tout nombre $\delta > 0$.

(3)
$$\lim_{n \rightarrow \infty} H_n(-\pi) = \lim_{n \rightarrow \infty} H_n(+\pi) = 0$$

1ère question: Montrer que la famille de noyaux $F_n(t) = \frac{1}{n+1} \left(\frac{\sin \frac{2n+1}{2} t}{\sin \frac{t}{2}} \right)^2$ constitue une famille de noyaux ultra-réguliers.

(on pourra vérifier que $H_n(t) = A \frac{n}{1+n^2 t^2}$ pour une certaine constante A, est convenable).

2ème question : Démontrer le théorème suivant, dû à Fatou

Soit f une fonction périodique et de période 2π , absolument intégrable sur $[-\pi, +\pi]$. Soit $\{K_n\}_{n \geq 1}$ une famille de noyaux ultra-réguliers. Pour presque tout x , on a

$$\lim_{n \rightarrow \infty} K_n * f(x) = f(x)$$

3ème question : Que donne le résultat de la deuxième question appliqué au noyau de Féjer ?

EXERCICE N° 49.

Retrouver une démonstration du théorème de Weierstrass à partir du théorème de Féjer.

EXERCICE N° 50.

Démontrer que

$$\alpha_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| \frac{\sin(2n+1)t/2}{\sin t/2} \right| dt = \frac{4}{\pi^2} \log n + \epsilon_n \quad \text{où } \epsilon_n \text{ reste borné}$$

(on écrit $\alpha_n = \frac{2}{\pi} \int_0^\pi \frac{|\sin nt|}{t} dt + \beta_n$ où $0 < \beta_n \leq M$)

puis

$$\int_0^\pi \frac{|\sin nt|}{t} dt = \sum_0^{n-1} \int_{k \frac{\pi}{n}}^{(k+1) \frac{\pi}{n}} \frac{|\sin nt|}{t} dt$$

et l'on compare enfin $\sum_{k=1}^n \frac{1}{k}$ à $\log n$).

- CHAPITRE 6 -

APPROXIMATION UNIFORME

1. INTRODUCTION

a) Position du problème

Nous particularisons le problème général de l'approximation en nous plaçant dans la situation suivante :

- l'espace E est l'espace de fonctions continues f définies sur le segment $[a, b]$ où $-\infty < a < b < +\infty$

- la norme utilisée sur E est la norme de la convergence uniforme, c'est-à-dire

$$\|f\| = \max_{x \in [a, b]} |f(x)|$$

- l'espace F des approximations est de dimension finie : cela pourra être l'espace des polynômes de degré inférieur ou égal à un entier donné n (espace de dimension n+1) ou plus généralement l'espace engendré par combinaisons linéaires à partir de n fonctions données $\varphi_1, \varphi_2, \dots, \varphi_n$, fonctions que nous supposerons fréquemment linéairement indépendantes.

b) Historique

Laplace, en 1799, posa le problème de trouver la meilleure approximation uniforme de la solution de systèmes linéaires avec des équations surabondantes. On s'aperçut très vite qu'il était beaucoup plus facile de résoudre un tel problème avec un ajustement aux "moindres carrés". En effet, on fait alors intervenir les techniques quadratiques des préhilbertiens (rôle de la linéarité). L'approximation uniforme en tant que telle ne fut pas développée.

C'est le chimiste Mendeleïev qui reposa au mathématicien russe Čebičev le problème de l'approximation uniforme d'une fonction continue par un polynôme de degré fixé. Celui-ci caractérisa cette approximation uniforme à la fin du XIXème siècle.

Pratiquement la commodité des calculs des approximations en moyenne quadratique fait que l'approximation uniforme ne fut guère utilisée au début du XXème siècle.

Les moyens modernes de calcul ont donné un regain d'activité aux problèmes d'approximation uniforme pour trois raisons :

1°) en rendant possibles les calculs effectifs des meilleures approximations uniformes qui n'étaient pas possibles manuellement

2°) en nécessitant le recours aux meilleures approximations uniformes pour générer les fonctions mathématiques (sinus, cosinus, exponentielles, etc.) figurant dans les "bibliothèques" des ordinateurs

3°) il y a de nombreuses analogies entre l'approximation uniforme et l'approximation au sens L^1 ,

La théorie est facile comme on pourra s'en apercevoir à la lecture de ce chapitre.

c) Exemple.

Pour montrer les idées essentielles de l'approximation uniforme nous allons examiner d'abord un exemple très simple.

Soit une fonction $f \in \mathcal{C}([0,1])$, pour fixer les idées $f(x) = e^x$ et cherchons une constante C telle que :

$$\|f - C\| = \max_{x \in [0,1]} |e^x - C|$$

soit minimal, c'est-à-dire une meilleure approximation de f par la famille des fonctions constantes ($\dim F = 1$)

Posons :

$$m = \min_{0 \leq x < 1} f(x) \quad \text{et} \quad M = \max_{0 \leq x < 1} f(x)$$

On trouve immédiatement que C doit être prise égale à :

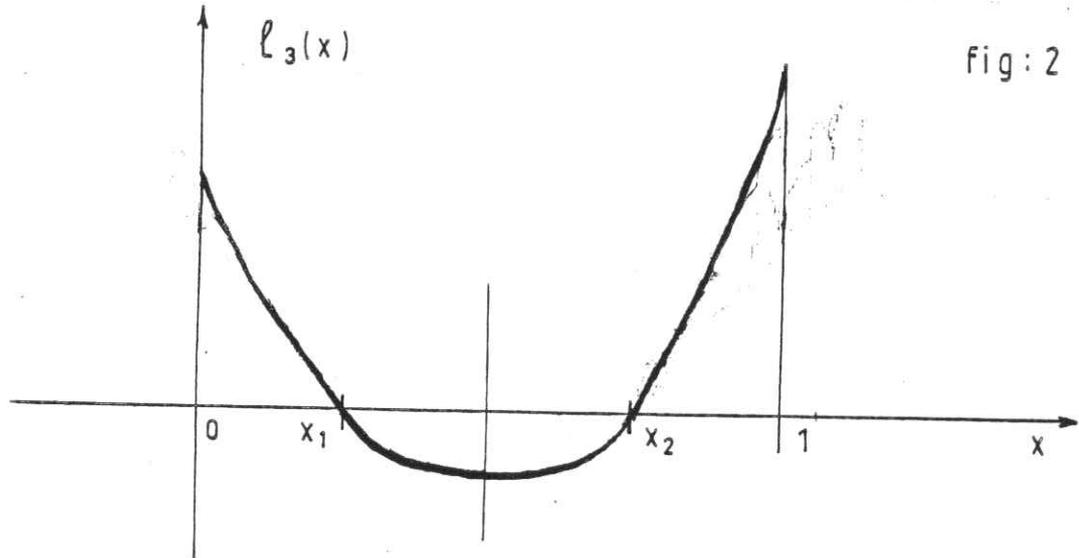
$$\hat{C} = \frac{1}{2} (m + M)$$

L'erreur étant mesurée par

$$\|f - \hat{C}\| = \frac{1}{2} (M - m)$$

trouver une droite A_3 encore plus proche de la courbe en se donnant deux points sur la courbe et en prenant A_3 passant par ces deux points : c'est un procédé d'approximation par interpolation.

La courbe d'erreur est alors schématiquement de la forme :



Si nous cherchons par tâtonnements, il est intuitif que l'on pourra réduire l'erreur maximale, lorsque le maximum est obtenu en un point, en déplaçant le segment pour diminuer ce maximum local sans "trop" augmenter l'erreur ailleurs. On le peut en faisant légèrement pivoter la droite autour d'un point d'erreur nulle. On peut même, s'il y a deux valeurs extrémales de f , avec des signes opposés, diminuer l'ordonnée de ces deux points en faisant encore tourner le segment de droite autour d'un point d'erreur nulle.

Mais il est bien clair alors que si nous avons trois maximums égaux en valeur absolue et alternant en signe on ne peut pas diminuer une des ordonnées sans augmenter l'autre. Il apparaît dans ce cas on ne puisse améliorer l'approximation et que cette propriété caractérise la meilleure approximation. Ceci nous amène au critère d'optimalité.

Critère de meilleure approximation : Une meilleure approximation d'une fonction continue par une fonction linéaire est telle que la courbe d'erreur a trois maximums égaux en valeur absolue et alternant en signe.

Nous remarquons alors que si nous traçons la courbe représentative de l'erreur, la valeur $\|f - \hat{C}\|$ est atteinte deux fois avec des signes différents. Réciproquement si une constante \hat{C} est telle que l'erreur présente cette propriété d'alternance, il est clair que \hat{C} est la meilleure approximation car tout léger déplacement de la droite $y = \hat{C}$ ne peut qu'accroître l'erreur maximale.

Nous avons donc obtenu une propriété caractéristique de la meilleure approximation... par une constante.

Examinons maintenant une meilleure approximation de e^x par une fonction linéaire d'équation :

$$y = C_0 + C_1 x$$

Soit par exemple l'approximation A_1 donnée par

$$A_1(x) = 1 + x$$

(inspirée du développement en série de Taylor en 0 de e^x)

Est-elle la meilleure ? Pour le voir examinons la courbe d'erreur d'équation :

$$\begin{aligned} \ell_1(x) &= f(x) - A_1(x) \\ &= e^x - 1 - x \end{aligned}$$

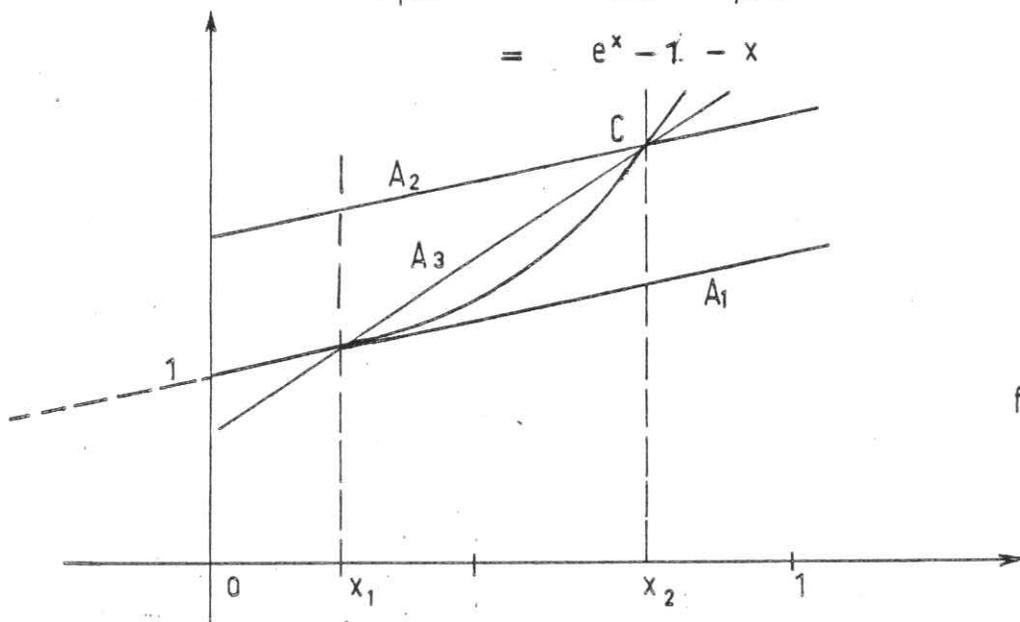


fig:1

Il est clair que l'on peut trouver une approximation meilleure que la tangente à l'origine car $\ell_1(x) \geq 0$.

Par addition d'une constante, une droite parallèle à A_1 permet de diviser l'erreur maximale par deux comme précédemment. On peut

La comparaison de ce que nous avons trouvé pour l'espace des constantes ($n = 1$) et pour l'espace des fonctions linéaires ($n = 2$) nous amène à chercher une généralisation.

Nous supposons que l'espace des approximations E_n est l'espace vectoriel engendré par les combinaisons linéaires de n fonctions continues sur $[a, b]$: $\varphi_1, \varphi_2, \dots, \varphi_n$.

Nous appellerons une telle combinaison linéaire un polynôme généralisé par analogie avec le cas où

$$\varphi_i(x) = x^{i-1} \quad i = 1, 2, \dots, n.$$

Pour que E_n soit de dimension n , il faut que les fonctions $\varphi_1, \varphi_2, \dots, \varphi_n$ soient linéairement indépendantes, c'est-à-dire qu'il n'existe pas de constantes $\lambda_1, \lambda_2, \dots, \lambda_n$, non toutes nulles, telles que pour tout x , on ait :

$$(1) \quad \lambda_1 \varphi_1(x) + \lambda_2 \varphi_2(x) + \dots + \lambda_n \varphi_n(x) = 0$$

D'après l'exemple introductif, il est utile d'introduire la fonction d'erreur

$$\ell(x) = f(x) - \sum_{i=1}^n a_i \varphi_i(x)$$

Et l'on cherche $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_n$ tels que l'expression

$$\|\ell\| = \max_{x \in [a, b]} |\ell(x)|$$

soit minimale.

Définition 1

On définit les points extrémaux de l'erreur, c'est-à-dire les abscisses x telles que

$$|\ell(x)| = \|\ell\|$$

2. THEOREME CARACTERISTIQUE.

Soit $\varphi(x) = \sum_{i=1}^n a_i \varphi_i(x)$. On dit que la courbe d'erreur

lorsque $\ell(x) = f(x) - \sum_{i=1}^n a_i \varphi_i(x)$ alterne en $n+1$ points x_1, x_2, \dots, x_{n+1}

$$|\ell(x_i)| = \sup_{x \in [a, b]} |\ell(x)| \quad i = 1, \dots, (n+1)$$

et lorsqu

$$l(x_i) = -l(x_{i+1}) \quad i = 1, 2, \dots, n$$

La mination d'une propriété caractéristique d'une meilleure approximation ne s'obtient agréablement que sous des conditions particulières sur les $[\varphi_i]$ puisque nous avons remarqué qu'il n'a pas unicité, en général, de la meilleure approximation. Pour fixer les idées, nous prendrons $a = 0$ et $b = 1$, travaillons sur $\mathcal{C}[0,1]$.

Définition: Un ensemble $[\varphi_i]_{i=1, \dots, n}$ est un ensemble de Čebičev de n éléments, lorsque $\sum_{i=1}^n a_i \varphi_i(x)$ a au plus $(n-1)$ zéros dans l'intervalle $[0,1]$ pour des a_i non simultanément nuls.

Exemple système $\varphi_i(x) = x^{i-1}$ pour i variant de 1 à n , c'est-à-dire l'espace de dimension n des polynômes (réels) de degré inférieur à n , constituée, d'après le théorème d'Alembert, un ensemble de Čebičev.

Théorème: Un ensemble de n éléments $[\varphi_i]$ forme un ensemble de Čebičev $[0,1]$ si et seulement si la matrice dont le coefficient général est $\varphi_i(x_j)$ a un déterminant non nul pour tout choix de n points distincts x_1, x_2, \dots, x_n dans l'intervalle $[0,1]$.

Le système d'équations linéaires à n inconnues a_1, a_2, \dots, a_n

$$(2) \quad \sum_{i=1}^n a_i \varphi_i(x_j) = 0$$

n'a de solution que $a_1 = a_2 = \dots = a_n = 0$ si, et seulement si, $\text{Det}[\varphi_i(x_j)] \neq 0$. Or le système (2) signifie que $\sum_{i=1}^n a_i \varphi_i(x)$ s'annule précisément en x_1, x_2, \dots, x_n . D'où le théorème.

La condition

$$\begin{vmatrix} \varphi_1(x_1) & \varphi_n(x_1) \\ \vdots & \vdots \\ \varphi_1(x_n) & \varphi_n(x_n) \end{vmatrix} \neq 0$$

s'appelle condition de Haar. On dit encore que le système des $[\varphi_i]$ est unisolvant lorsque cette condition est réalisée pour tout choix de points distincts x_1, x_2, \dots, x_n . Le corollaire 1 relie des problèmes d'approximation et de problèmes d'interpolation (Cf. Chapitre 3) et explique le mot unisolvant.

Coro.: Un système $[\varphi_i]$ de n fonctions est un système de Čebičev si, et si pour tout choix de $y_i \in \mathbb{R} (i = 1, \dots, n)$ et de points distincts $x_i \in I (i = 1, \dots, n)$ il existe un unique polynôme généralisé (élément de E_{n-1})

$$\varphi(x_i) = y_i \quad \text{pour} \quad i = 1, 2, \dots, n$$

La condition est évidente grâce aux règles de Cramer.

En venons au théorème essentiel de ce chapitre,

Théor. Soit f une fonction continue sur $[0, 1]$ et $[\varphi_i], i = 1, 2, \dots, n$, un système de Čebičev sur $[0, 1]$. Une condition nécessaire et suffisante pour $\varphi(x) = \sum_{i=1}^n a_i \varphi_i(x)$ soit une meilleure approximation de f au sens de l'uniforme sur $\mathcal{C}[0, 1]$ est que la fonction d'erreur $f(x) - \varphi(x)$ ait $n+1$ points au moins sur $[0, 1]$.

Rem. Il n'y a aucune difficulté à définir la meilleure approximation uniforme d'une fonction de plusieurs variables. On peut démontrer l'existence par le théorème général du Chap. Mais ainsi que nous allons le voir, l'unicité et les propriétés de la meilleure approximation uniforme sont liées à la condition de Haar qui n'est pas vérifiée pour les fonctions de plusieurs variables (Exercice N° 55).

La démonstration du théorème 2 se fera en deux étapes :

Première étape : nombre de points extrémaux:

Nous donnerons d'abord une première condition nécessaire et suffisante pour qu'une approximation soit optimale. On suppose pour éviter des ambiguïtés que f n'est pas une combinaison des fonctions φ_i .

Nous en déduirons ensuite que la courbe d'erreur

$$L(x) = f(x) - \varphi(x)$$

de l'approximation optimale a nécessairement $(n+1)$ points extrémaux au moins.

Deuxième étape : propriété de l'alternance: (condition suffisante)

Nous démontrerons (théorème de de la Vallée Poussin) que les approximations ayant $(n+1)$ points extrémaux avec alternance des signes sont nécessairement optimales.

Première étape : nombre de points extrémaux.

Théorème 3 : Une condition nécessaire et suffisante pour qu'un polynôme généralisé φ soit optimal est qu'il n'existe pas de polynôme généralisé de E_n qui soit strictement positif aux points extrémaux positifs de l'erreur et strictement négatif aux points extrémaux négatifs.

Démonstration : Supposons au contraire l'existence d'un tel polynôme généralisé P et montrons que φ n'est pas alors une meilleure approximation ce qui est contradictoire avec l'hypothèse. Cela démontrera la nécessité de la condition du théorème 3.

Nous désignons par $L(x)$ la fonction d'erreur $L(x) = f(x) - \varphi(x)$ et notons L le maximum de la valeur absolue de cette fonction lorsque x parcourt l'ensemble de définition $[0,1]$.

$$L = \sup_{x \in [0,1]} |f(x) - \varphi(x)| = \sup_{x \in [0,1]} |L(x)|$$

Soit P un polynôme généralisé strictement positif aux points extrémaux positifs de l'erreur ($P(x) > 0$ lorsque $L(x) = L$) et strictement négatif aux points extrémaux négatifs de l'erreur ($P(x) < 0$ lorsque $L(x) = -L$).

Soit x un point extrémal de l'erreur. Il existe un intervalle ouvert Δ_x , entourant le point x , et sur lequel $P(x)$ reste non nul et de signe constant. Appelons M le maximum de la valeur absolue de $L(x)$ sur le complémentaire de la réunion des intervalles Δ_x lorsque x parcourt l'ensemble des points extrémaux de $L(x)$. Ce complémentaire peut être vide, auquel cas on prendra $M = 0$. Soit M' le maximum de la valeur absolue du polynôme généralisé P intervenant dans l'énoncé de la propriété 1. Choisissons enfin un nombre positif ϵ de sorte que :

$$0 < \epsilon M' < L - M$$

ce qui est toujours possible puisque P n'est pas identiquement nul et $L > M$ par construction.

Pour x n'appartenant pas à $\bigcup_{x \in Y} \Delta_x$ où Y est l'ensemble des points extrémaux de l'erreur, on dispose de la majoration :

$$|f(x) - (\varphi(x) + \epsilon P(x))| \leq M + \epsilon M' < L$$

Pour x appartenant à $\bigcup_{x \in Y} \Delta_x$, puisque $P(x) \neq 0$ est alors du même signe que $f(x) - \varphi(x)$, on a également

$$|f(x) - (\varphi(x) + \epsilon P(x))| < L$$

donc, avec une inégalité stricte

$$\sup_{x \in [0,1]} |f(x) - (\varphi(x) + \epsilon P(x))| < L$$

Ainsi $\varphi + \epsilon P$ qui appartient à E_n réalise une meilleure approximation de f que la fonction φ . Donc φ n'est pas une meilleure approximation de f .

Déduisons-en que si φ est une meilleure approximation de f , il existe au moins $n+1$ valeurs distinctes de la variable x pour lesquelles l'erreur est extrémale en valeur absolue.

En effet, supposons qu'il existe seulement p ($p \leq n$) points extrémaux distincts x_1, \dots, x_p . Complétons au besoin par des points x_{p+1}, \dots, x_n de sorte que les points x_1, \dots, x_n soient distincts. Le système des $\varphi_1, \dots, \varphi_n$ étant un système de Čebičev, on sait trouver une combinaison linéaire des φ_i , où $i = 1, 2, \dots, n$, de sorte que

$$P(x_i) = \varepsilon(x_i) \sup_{x \in [0,1]} |L(x)|$$

avec

$\varepsilon(x) = +1$ lorsque l'erreur $L(x)$ est extrémale positive
 et $\varepsilon(x) = -1$ lorsque l'erreur $L(x)$ est extrémale négative.

La proposition 1 montre alors que φ ne saurait être une meilleure approximation de f par E_n .

(Nous venons d'établir qu'il existe au moins $n+1$ valeurs extrémales pour l'erreur, mais bien entendu, il peut en exister plus, par exemple toutes les valeurs x dans $[0, 1]$ lorsque f est de la forme

$$f(x) = \sum_{i=1}^n a_i \varphi_i(x)$$

Démontrons maintenant l'unicité de la meilleure approximation d'une fonction f par les éléments de E_n lorsque E_n est engendré par un système de Čebičev.

Théorème 4 : Soit $(\varphi_1, \dots, \varphi_n)$ un système de Čebičev sur $\mathcal{C}[0, 1]$.
Pour toute fonction numérique continue sur $[0, 1]$, il existe une unique meilleure approximation en norme uniforme par les éléments de l'espace E_n engendré par $\varphi_1, \dots, \varphi_n$

Soient φ' et φ'' deux meilleures approximations de f par des polynômes généralisés construits sur le système $\{\varphi_i\}$. Il est bien clair que

$$\varphi = \frac{\varphi' + \varphi''}{2}$$

est également une meilleure approximation grâce à l'inégalité triangulaire. Par suite, d'après la remarque précédente, il existe au moins $n+1$ points x_1, x_2, \dots, x_{n+1} dans $[0, 1]$, distincts tels que :

$$\varphi(x_i) - f(x_i) = \varepsilon(x_i)L \quad \text{pour } i = 1, 2, \dots, n+1$$

où $\varepsilon(x_i) = \pm 1$ et $L = \inf_{g \in E_n} \|f-g\|$

Ce qui fo

$$\frac{x_i - f(x_i)}{2} + \frac{\varphi''(x_i) - f(x_i)}{2} = \varepsilon(x_i) L$$

Mais $|\varphi'(x_i) - f'(x_i)| \leq L$ et de même $|\varphi''(x_i) - f''(x_i)| \leq L$. On en déduit don

$$\varphi'(x_i) - f'(x_i) = \varphi''(x_i) - f''(x_i)$$

ou encore ps n+1 points x_1, x_2, \dots, x_{n+1}

$$\varphi = \varphi''(x_i)$$

Ces égalitéiquent $\varphi' = \varphi''$ puisque $\varphi_1, \varphi_2, \dots, \varphi_n$ est un système de iv. D'où l'unicité annoncée. L'existence est déjà établie au chapitre 1 pe E_n est de dimension finie.

Deuxième: une propriété caractéristique d'alternance.

Enonid'abord un théorème dû au mathématicien belge de la Vallée Poussin :

Théorème 5 So un polynôme généralisé construit sur un système de
Čebičev de foons $\varphi_1, \varphi_2, \dots, \varphi_n$ tel que $f - \varphi$ prenne des valeurs
de signes alts non nulles (dont la valeur absolue est notée L_i) en
(n+1) points qincts x_i où i varie de 1 à (n+1). L'écart entre la fonc-
tion continue t sa meilleure approximation uniforme est supérieur ou
égal à Min_i
 $i = 1, \dots, n+1$

Supposons, en effet, que l'on puisse avoir la relation pour tout $x \in [0,1]$
 $|f(x) - \Psi(x)| \leq \|f - \Psi\| < \min_{i=1, \dots, n+1} L_i$ pour une combinaison linéaire Ψ des n
 fonctions φ_i . Le polynôme généralisé $\Psi - \varphi = (\Psi - f) - (f - \varphi)$ est du signe
 de $f - \varphi$ en chacun des $(n+1)$ points x_i , puisque

$$L_i = |\varphi(x_i) - f(x_i)|$$

Par suite, ces $(n+1)$ valeurs étant de signes alternés et $\Psi - \varphi$ étant
 une fonction continue, $\Psi - \varphi$ s'annule en n points distincts au moins.
 Le système $\{\varphi_i\}$ étant de Tchebichev (d'ordre n), on en déduit la nullité
 de $\Psi - \varphi$ donc $\Psi = \varphi$. Ce qui montre qu'en fait, Ψ est la meilleure
 approximation de f donc que l'inégalité stricte ne peut avoir lieu.

La suffisance de la caractérisation de la meilleure approximation (Th2)
 est une conséquence immédiate. Soit, en effet, φ une combinaison li-
 néaire des φ_i , telle que $L(x) = f(x) - \varphi(x)$ possède $(n+1)$ points ex-
 trêmes, égaux en valeur absolue, mais de signes alternés. Soit L' , la norme
 de l'erreur entre f et sa meilleure approximation par un élément de E_n . Bien
 évidemment, on a $L = \sup |L(x)| \geq L'$. Mais le théorème de la Vallée Poussin
 montre que $L' \geq L$ d'où $L = L'$. On déduit que φ est l'unique meilleure
 approximation de f par E_n , prouvant la suffisance de la condition du théorème 2.

Corollaire n° 1 : Soit f une fonction numérique continue sur $[a, b]$ et
 E_n l'espace des polynômes de degré au plus égal à n . Soit P un poly-
 nôme quelconque de E_n tel que

$$\sup_{x \in [a, b]} |f(x) - P(x)| = \delta$$

et tel qu'il existe $(n+2)$ points $a \leq x_1 < x_2 < \dots < x_{n+2} \leq b$ pour lesquels

$$f(x_i) - P(x_i) = \pm \delta \quad i = 1, 2, \dots, n+2$$

de façon alternée. Alors P est le polynôme de meilleure approximation de
 f dans E_n .

Naturellement les théorèmes du Chapitre 4 (notamment 4.5) fournis-
 sent des estimations de l'approximation en norme uniforme.

Exemple : Soit $f(x) = \exp x$. Le polynôme de degré inférieur ou égal
 à 5 qui en réalise la meilleure approximation en norme uniforme sur l'in-
 tervalle $[0,1]$ est donné par :

$$P(x) = 0,013 + 0,035x + 0,170x^2 + 0,500x^3 + 1,000x^4 + 1,000x^5$$

où l'on n'a donné que les trois premières décimales des coefficients.

On pourra comparer aux meilleures approximations au sens d'autres normes !

Corollaire N° 2 : Soit f une fonction numérique continue sur un intervalle borné [a, b]. Il existe une unique meilleure approximation par un polynôme de degré inférieur ou égal à n. Cette approximation est dite approximation de Čebičev de la fonction f.

3. PROCÉDE PRATIQUE DE CALCUL.

On dispose en fait d'une procédure approximative pour le calcul effectif de l'approximation de Čebičev d'une fonction f, due à Lanczos. Soit f une fonction continue sur [-1, 1], pour fixer les idées. D'après la théorie du Chapitre 3, il existe une famille {a_n} de coefficients telle que

$$\lim_{n \rightarrow \infty} \| f(x) - \left(\frac{a_0}{2} + \sum_{k=1}^n a_k T_k(x) \right) \|_2 = 0$$

où par T_k on désigne le k^{ième} polynôme de Čebičev et

$$\| f \|_2 = \left(\int_{-1}^{+1} |f(x)|^2 \frac{1}{\sqrt{1-x^2}} dx \right)^{\frac{1}{2}}$$

désigne la norme de l'espace L² $\left([-1, +1], \frac{1}{\sqrt{1-x^2}} \right)$ (Cf Exercice N° 36 et 37 -E.A.F.).

Ecrivons $f(x) = \frac{a_0}{2} + a_1 T_1(x) + \dots + a_n T_n(x) + a_{n+1} T_{n+1}(x) + R_{n+1}(x)$. La norme

uniforme de la différence $f(x) - \left(\frac{a_0}{2} + \dots + a_n T_n(x) \right) = a_{n+1} T_{n+1}(x) + R_{n+1}(x)$

est voisine de la norme uniforme de la différence entre f et sa meilleure approximation uniforme par des polynômes de degré n au plus. En effet, supposons que nous puissions négliger R_{n+1}(x) devant le terme précédent

a_{n+1} T_{n+1}(x). Alors a_{n+1} T_{n+1}(x) admet n+2 maximums ayant des

signes alternés d'après la remarque déjà effectuée sur les polynômes de Čebičev (Cf E.A.F, exercice N° 23). Par suite, d'après le théorème 3

$\frac{a_0}{2} + \dots + a_n T_n(x)$ est la meilleure approximation uniforme de f(x)

(à condition de négliger R_n(x)). Cette procédure peut être particulièrément utile lorsque l'on veut approximer des fonctions polynômiales

car l'on dispose de tableaux numériques donnant les coefficients a_n du développement en séries de Čebičev. Ainsi (exemple donné par P.J. Davis), calculons la meilleure approximation de $f(x) = 1 + \frac{x}{2} + \frac{x^2}{3} + \dots + \frac{x^5}{6}$

sur l'intervalle [-1, +1]. On calcule

$$f(x) = T_0 + \frac{T_1}{2} + \frac{1}{3} \cdot \frac{1}{2} (T_0 + T_1) + \frac{1}{4} \cdot \frac{1}{4} (3T_1 + T_3) + \frac{1}{5} \cdot \frac{1}{8} (3T_0 + 4T_2 + T_4) + \frac{1}{6} \cdot \frac{1}{16} (10T_1 + 5T_3 + T_5)$$

en utilisant l'expression de x^p en fonction des $p+1$ premiers polynômes de Čebičev, il vient

$$\frac{149}{120} T_0 + \frac{76}{96} T_1 + \frac{32}{120} T_2 + \frac{11}{96} T_3 + \frac{3}{120} T_4 + \frac{1}{96} T_5$$

Supposons que nous souhaitions une approximation à 5/100 près. Comme $|T_4|$ et $|T_5|$ sont majorés par 1 et que $\frac{3}{120} + \frac{1}{96} < \frac{5}{100}$, on peut prendre comme approximation à 0,05 près de f sur $[-1,+1]$.

$$\frac{149}{120} T_0 + \frac{76}{96} T_1 + \frac{32}{120} T_2 + \frac{11}{96} T_3$$

Autre exemple : La meilleure approximation de $\text{arc tg } x$ sur $[-1,+1]$ par un polynôme de degré n est (au sens de la norme uniforme)

$$\text{arc tg } x \sim 0,995\,3580 x - 0,288\,6902 x^3 + 0,007\,93380 x^5$$

et l'erreur maximale commise est de l'ordre de :

$$0,0006086$$

L'approximation obtenue par les coefficients du développement en séries de Čebičev est :

$$\text{arc tg } x \sim 0,994\,9494 x - 0,2870\,605 x^3 + 0,00780\,372 x^5$$

avec une erreur maximale de l'ordre de :

$$0,0006893$$

EXERCICES DU CHAPITRE 6

EXERCICE N° 51.

- 1) Montrer que le système $1, \cos x, \sin x, \dots, \cos nx, \sin nx$ est un système de Čebičev sur $[-\pi, +\pi]$
- 2) Montrer directement que $1, x, x^2, \dots, x^{n-1}$ satisfait la condition de Haar (relative à $(n+1)$ fonctions). On utilisera le déterminant de Van der Monde

$$\begin{vmatrix} x_1 & \dots & x_1^{n-1} \\ \vdots & & \vdots \\ x_n & \dots & x_n^{n-1} \end{vmatrix} = \prod_{1 \leq j < i \leq n} (x_i - x_j)$$

EXERCICE N° 52.

Soit le système d'équations linéaires à n inconnues et m équations ($m > n$)

$$\begin{array}{r} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n = f_1 \\ \vdots \\ a_{m1} x_1 + \dots + a_{mn} x_n = f_m \end{array}$$

Ce système est généralement impossible.

1°) Donner une définition de la "meilleure solution possible" de ce système pour une norme donnée. Appliquer avec la norme L^2 et retrouver les moindres carrés.

2°) Montrer que l'on se ramène à un problème de meilleure approximation d'une fonction $f(x)$ définie sur un nombre fini de points.

EXERCICE N° 53.

Soit f une fonction numérique continue sur l'intervalle $[0, 1]$ et composée de deux portions rectilignes. Déterminer la meilleure approximation

de f par une fonction linéaire. Interpréter géométriquement le résultat.

EXERCICE N° 54.

Soient n valeurs réelles a_1, a_2, \dots, a_n n'appartenant pas à $[-1, +1]$. Montrer que le système $\phi_l(x) = \frac{1}{x + a_l}$ pour $l = 1, 2, \dots, n$ est de Čebičev $[-1, +1]$

(on montrera que

$$\det \left[\frac{1}{x_l + a_j} \right] = \frac{\prod_{l>j} (x_l - x_j) (a_l - a_j)}{\prod_{l,j=1}^n (x_l + a_j)}$$

EXERCICE N° 55 (théorème de Haar)

Soit S un ensemble dans un espace euclidien de dimension $n \geq 2$. Si S contient un point intérieur, aucun ensemble de p fonctions ($p > 1$) ne peut être unisolvant.

(utiliser le principe des aiguillages sur le dessin).



C H A P I T R E 7

APPROXIMATION D'OPERATIONS LINEAIRES ET APPLICATIONS DIVERSES

Souvent la fonction que l'on cherche à approximer est solution d'une équation fonctionnelle (intégrale ou différentielle) ou bien intervient par les valeurs que prend un opérateur donné sur cette fonction. Des méthodes particulières d'approximation sont connues. Commençons par l'opérateur d'intégration usuel.

1. UNE APPLICATION DES POLYNOMES ORTHOGONAUX : LES METHODES DE QUADRATURE.

Soit E_{n+1} l'espace vectoriel de dimension finie $(n+1)$ constitué par les polynômes à coefficients réels dont le degré est inférieur ou égal à n . P est dans E_{n+1} si :

$$P(x) = \sum_{h=0}^n a_h x^h$$

Soit x_0^* l'application qui à tout élément P de E_{n+1} fait correspondre le nombre réel $(P(x_0))$ où x_0 désigne un nombre réel. Il est facile de constater que x_0^* est une forme linéaire sur E_{n+1} . En outre, si x_0, x_1, \dots, x_n sont nombres réels distincts, $x_0^*, x_1^*, \dots, x_n^*$ sont $(n+1)$ formes linéaires indépendantes sur l'espace dual E_{n+1}^* de E_{n+1} . (Cf Ex N° 29 E.A.F.). Or, l'espace E_{n+1} est de dimension $n+1$, donc toute forme linéaire L sur E_{n+1} s'écrit comme une combinaison linéaire des $n+1$ formes x_0^*, \dots, x_n^* . En particulier, si l'on plonge E_{n+1} dans $\mathcal{C}[-1, +1]$ pour fixer les idées, l'intégrale

$\int_{-1}^{+1} P(x) dx$ est une forme linéaire sur E_{n+1} . Il existe donc $n+1$ constantes $\alpha_0, \alpha_1, \dots, \alpha_n$, indépendantes de P , telles que pour tout polynôme P de E_{n+1} , on ait

$$\begin{aligned} (1) \quad \int_{-1}^{+1} P(x) dx &= \alpha_0 x_0^*(P) + \dots + \alpha_n x_n^*(P) \\ &= \alpha_0 P(x_0) + \dots + \alpha_n P(x_n) \end{aligned}$$

les $(n+1)$ valeurs réelles distinctes x_0, x_1, \dots, x_n ont été choisies dans l'intervalle $[-1, +1]$ précisément.

entendu, l'égalité (1) cesse d'être exacte si l'on remplace $P(x)$ par une fonction continue quelconque. Mais on peut imaginer que la formule (1) devient une formule de quadrature approchée. De fait, on peut améliorer la relation en précisant les $(n+1)$ abscisses qui interviennent. Les résultats obtenus sont essentiellement dus à Gauss et Jacobi et, pour fixer les idées resterons sur l'intervalle $[-1, +1]$ en ne cherchant que l'approximation de l'intégrale $\int_{-1}^{+1} f(x) dx$. On pourrait se placer sur $[a, b]$ et étudier l'intégrale définie pondérée $\int_a^{+b} f(x)h(x) dx$. Par suite, nous nous limitons uniquement aux polynômes orthogonaux sur $L^2[-1, +1]$, à savoir les polynômes de Legendre. On se reportera à E.A.F. (exercice N° 38) pour les propriétés remarquables des zéros de ces polynômes.

Théorème Soient x_1, x_2, \dots, x_n les n zéros distincts du polynôme de Legendre de degré n . Il existe n constantes $\alpha_1, \alpha_2, \dots, \alpha_n$ réelles et positives que

$$\int_{-1}^{+1} f(x) dx = \sum_{h=1}^n \alpha_h P(x_h) \quad \text{pour tout polynôme } P \text{ dont}$$

le degré est inférieur ou égal à $2n-1$.

Les nombres α_h sont appelés les nombres de Christoffel. Nous rappellerons les notations du Chapitre 3 sur l'interpolation de Lagrange :

$$l(x) = \prod_{h=1}^n (x - x_h) \quad \text{et} \quad L_h(x) = \frac{l(x)}{(x - x_h) l'(x_h)}$$

Soit $P(x)$ un polynôme quelconque de degré au plus égal à $2n-1$ et $Q(x)$ le polynôme de Lagrange de degré $(n-1)$ interpolant $P(x)$ aux n points x_1, \dots, x_n . Nous avons vu au Chapitre 3 § 1, que

$$Q(x) = \sum_{h=1}^n P(x_h) L_h(x)$$

On a $P(x) - Q(x) = R(x)$. Ce polynôme admet les n zéros x_1, \dots, x_n , donc, peut s'écrire grâce à une mise en facteur

$$R(x) = P_n(x) R_{n-1}(x) \quad \text{où } P_n(x) \text{ est le } n^{\text{ième}}$$

polynôme de Legendre dont les zéros sont x_1, x_2, \dots, x_n et $R_{n-1}(x)$ un polynôme de degré $(n-1)$ au plus. Calculons

$$\int_{-1}^{+1} P(x) dx$$

$$\int_{-1}^{+1} P(x) Q(x) dx + \int_{-1}^{+1} Q(x) dx = \int_{-1}^{+1} P_n(x) R_{n-1}(x) dx + \int_{-1}^{+1} Q(x) dx$$

Mais P_n est orthogonal à tout polynôme de degré inférieur à n , donc

$$\begin{aligned} \int_{-1}^{+1} P(x) dx &= \int_{-1}^{+1} Q(x) dx \\ &= \sum_{h=1}^n P(x_h) \left(\int_{-1}^{+1} L_h(x) dx \right) \end{aligned}$$

En posant $\alpha_h = \int_{-1}^{+1} L_h(x) dx$, nombre indépendant de P , on réalise bien le théorème

Donnons propriétés des nombres de Christoffel relatifs au cas des polynômes Legendre

$$(a) \quad \alpha_h = \int_{-1}^{+1} L_h(x) dx = \int_{-1}^{+1} (L_h(x))^2 dx$$

En effet $(L_h(x))^2$ est un polynôme de degré $2h-2$ dont les racines doubles sont $x_1, \dots, x_{h-1}, x_{h+1}, \dots, x_n$. En outre $L_h(x_h) = 1$.

Donc

$$\int_{-1}^{+1} (L_h(x))^2 dx = \sum_{h=1}^n \alpha_h L_h^2(x_h) = \alpha_h$$

ce qui assure la (stricte) positivité des nombres de Christoffel. On a aussi

$$(b) \quad \alpha_h = \frac{a_n}{a_{n-1}} \frac{1}{P_{n-1}(x_h) P_n'(x_h)}$$

où a_n est l'efficiant du terme de plus haut degré de $P_n(x)$.

La formule de Darboux-Christoffel fournit en effet

$$P_0(x) P_0'(y) P_n(x) P_n(y) = \frac{a_n}{a_{n+1}} \frac{P_{n+1}(x) P_n(y) - P_{n+1}(y) P_n(x)}{x-y}$$

et pour on peut intégrer cette formule entre -1 et $+1$. Compte tenu des relations d'orthogonalité, il vient

$$\frac{a_n}{a_{n+1}} \int_{-1}^{+1} \frac{P_{n+1}(x) P_n(x)}{x-x_h} dx = 1$$

Mais on remarque que

$$L_h(x) = \frac{P_n(x)}{P'_n(x_h)(x-x_h)}$$

ce qui permet d'écrire, en tenant compte de la relation ci-dessus :

$$-\frac{a_n}{a_{n+1}} P_{n+1}(x_h) P'_n(x_h) \int_{-1}^{+1} L_h(x) dx = 1$$

$$\alpha_h = -\frac{a_{n+1}}{a_n} \frac{1}{P_{n+1}(x_h) P'_n(x_h)}$$

que l'on transforme en tenant compte de la relation de récurrence entre trois polynômes successifs (Cf E.A.F.)

$$\boxed{\alpha_h = \frac{a_n}{a_{n-1}} \frac{1}{P_{n-1}(x_h) P'_n(x_h)} \quad h = 1, 2, \dots, n}$$

La méthode de Gauss-Jacobi trouve un regain de faveur avec les ordinateurs actuels. On dispose d'ailleurs des ordres de grandeur de l'approximation. Remarquons que la méthode de Gauss-Jacobi revient à l'approximation d'un opérateur (ici une forme $\int_{-1}^{+1} P(x) dx$) par des masses de Dirac: les $x_0^*, x_1^*, \dots, x_n^*$. On pose plus généralement le problème pour un opérateur linéaire quelconque et l'on dispose de résultats lorsque l'on se place dans un espace de Hilbert continuellement injecté dans $\mathcal{E}[-1, +1]$. De nombreux résultats de ce type sont connus pour l'approximation de distributions particulières (Cf Bibliographie).

Remarque : Le théorème 7.1 a été considérablement généralisé par Schoenberg au cas des fonctions splines naturelles (Cf Bibliographie).

2. UNE METHODE D'APPROXIMATION : LA METHODE DE GALERKINE.

On a vu, dans les éléments d'analyse fonctionnelle, que chercher la solution d'une équation différentielle avec conditions aux limites du type Sturm-Liouville équivaut à chercher la solution d'une certaine équation intégrale.

On retrouve des situations analogues pour résoudre des équations aux dérivées partielles. L'étude des équations intégrales a été conduite de façon abstraite par la décomposition d'un opérateur autoadjoint compact suivant ses sous-espaces propres sur un espace de Hilbert. Nous allons donner maintenant un procédé numérique utilisable pour les équations de type elliptique notamment (Cf Exercice N° 61).

La situation abstraite que nous considérons est la suivante. Soit H un espace de Hilbert (sur \mathbb{C}) muni du produit scalaire $\langle \cdot, \cdot \rangle$ et A un opérateur linéaire continu et auto-adjoint sur H . On suppose en outre que A est coercitif, d'est-à-dire satisfait la condition

$$\langle Ax, x \rangle \geq a \|x\|^2 \quad \forall x \in H$$

où a est une constante positive donnée. On veut résoudre l'équation $Ax = x'$ où x' est donné dans H . On a d'abord le résultat théorique suivant.

Proposition : Un opérateur linéaire, continu, auto-adjoint et coercitif sur un espace de Hilbert possède un inverse continu.

Démonstration : Puisque A est coercitif, il s'agit d'un opérateur injectif sur l'espace H . Pour démontrer que A est surjectif, on va d'abord construire un nouveau produit scalaire. Pour tout couple x, y d'élément de H , on pose

$$\langle\langle x, y \rangle\rangle = \langle Ax, y \rangle$$

Il s'agit bien d'un nouveau produit scalaire sur H :

$$(1) \quad \langle\langle \lambda x + \mu x', y \rangle\rangle = \lambda \langle\langle x, y \rangle\rangle + \mu \langle\langle x', y \rangle\rangle \quad (\text{linéarité de } A)$$

$$(2) \quad \langle\langle x, y \rangle\rangle = \overline{\langle\langle y, x \rangle\rangle} \quad (A \text{ est auto-adjoint})$$

$$(3) \quad \langle\langle x, x \rangle\rangle \geq 0 \quad \text{et} \quad \langle\langle x, x \rangle\rangle = 0 \quad \text{si et seulement si} \quad x = 0$$

(puisque A est coercitif).

Le nouveau produit scalaire induit une nouvelle norme sur H , à savoir

$\sqrt{\langle\langle x, x \rangle\rangle} = \|x\|$, mais cette norme est équivalente à la norme de départ grâce aux inégalités

$$(4) \quad \|x\| \geq \sqrt{a} \|x\| \quad (A \text{ est coercitif})$$

et (5) $\|Ax\| \leq \sqrt{|A|} \|x\|$ (A est un opérateur continu)

donc borné dont la norme sur H est notée $\|A\|$: (Cf. E.A.F.).

Notons \mathcal{H} le nouvel espace de Hilbert obtenu en munissant l'espace vectoriel H du produit scalaire $\langle\langle \dots \rangle\rangle$.

Fixons $x' \in H$ et considérons sur H l'application qui à tout y fait correspondre $\langle y, x' \rangle$. C'est une forme linéaire et continue sur l'espace \mathcal{H} , donc d'après le théorème de représentation de Riesz (Cf E.A.F.), il existe un unique z de H tel que

$$\langle y, x' \rangle = \langle\langle y, z \rangle\rangle$$

Soit pour tout z de H

$$\langle y, x' \rangle = \langle Ay, z \rangle = \langle y, Az \rangle$$

ce qui fournit un élément z de H tel que

$$Az = x'$$

et A est donc un opérateur surjectif.

Finalement A est un opérateur bijectif. Notons A^{-1} l'opérateur inverse. Il s'agit évidemment d'un opérateur linéaire. En outre A^{-1} est continu grâce à l'inégalité de coercitivité

$$\langle A(A^{-1}x), A^{-1}x \rangle \geq \|x\| \|A^{-1}x\|^2$$

Or

$$\langle x, A^{-1}x \rangle \leq \|x\| \|A^{-1}x\|$$

ce qui donne pour tout x de H la majoration de continuité

$$\|A^{-1}x\| \leq \frac{1}{a} \|x\|$$

Naturellement A^{-1} est auto-adjoint.

Remarque : Pour résoudre $Ax = x'$, nous avons écrit l'égalité équivalente

$$\langle Ax, y \rangle = \langle x', y \rangle \quad \forall y \in H$$

(Cf. Exercice N° 63).

Il restera à fournir un algorithme de calcul pour l'opérateur A^{-1} .
 Tel est de la méthode de Galerkin. On se donne alors une suite H_n
 de sous-espaces vectoriels de H tels que

- (1) H_n sous-espace vectoriel de dimension n
- (2) H_n inclus dans H_{n+1}
- (3) la réunion des H_n est dense dans H (cette dernière hypothèse suppose H séparable).

La façon de construire les H_n pour un H séparable donné est de prendre une base hilbertienne $[e_n]_{n \geq 1}$ et d'appeler H_n le sous-espace vectoriel engendré par e_1, e_2, \dots, e_n .

Exemple $\mathcal{B} = \mathcal{L}^2$ (et nous savons que tout espace de Hilbert séparable est isomorphe à \mathcal{L}^2 (Cf E.A.F.)), nous pouvons prendre pour H_n le sous-espace des suites (x_1, x_2, \dots) telles que

$$0 = x_{n+1} = x_{n+2} = \dots$$

H_n étant de dimension finie, est un sous-espace complet de \mathcal{B} , donc fermé. Nous pouvons appliquer le théorème de la projection (Cf. E.A.F.) et définir pour $x \in \mathcal{B}$ la projection orthogonale $P_n(x) \in H_n$ de x sur H_n . Ensuite en faisant varier n nous définissons une suite de points $[P_n(x)]_{n \geq 1}$ de \mathcal{B} . On constate $P_n(x) \rightarrow x$ quand n tend vers l'infini pour tout x dans \mathcal{B} .

Ceci est général. Appelons $P_n(x)$ la projection orthogonale d'un point x de H sur H_n . On pose alors le résultat suivant :

$$\text{si } n \longrightarrow \|P_n(x) - x\| \longrightarrow 0$$

En effet, si $n > m$, $H_m \subset H_n$, donc

$$\begin{aligned} \|x(x)\| &= \inf_{y \in H_n} \|x - y\| \leq \inf_{z \in H_m} \|x - z\| \\ &\leq \|x - P_m(x)\|. \end{aligned}$$

La suite de normes positives $\|P_n(x) - x\|$ est donc décroissante. Cette suite tend vers une limite λ et il suffit de démontrer que $\lambda = 0$.

S'il n'en était pas ainsi, en utilisant l'hypothèse de densité de $\bigcup_{n \geq 1} H_n$ dans \mathcal{H} , il existerait un entier p et un point $x_p \in H_p$ tel que :

$$0 < \|x_p - x\| < \lambda \leq \inf_{y \in H_p} \|y - x\|$$

ce qui est impossible, donc $\lambda = 0$, c.q.f.d..

Notre but est de résoudre $Ax = x'$ pour un x' donné de H .

L'idée de la méthode de Galerkin est de résoudre cette équation sur l'espace de dimension finie H_n avec pour second membre la projection orthogonale de x' sur H_n . Puis ensuite on passe à la limite.

1ère étape : Résolution de l'équation sur H_n .

P_n désigne l'opérateur de projection orthogonale de H sur H_n . L'équation en x :

$$P_n A(x) = P_n(x')$$

possède dans H_n une unique solution s , notée x_n . En effet, sur H_n , l'opérateur $P_n A$ est linéaire, continu et vérifie

$$\langle P_n A(x), x \rangle \geq a \|x\|^2 \quad \forall x \in H_n$$

Cette relation provient de $\langle P_n A(x), x \rangle = \langle A(x), P_n(x) \rangle$ puisque P_n est auto-adjoint. Mais $P_n(x) = x$ pour tout x de H_n , donc

$$\langle P_n A(x), x \rangle = \langle AP_n(x), P_n(x) \rangle \geq a \|P_n(x)\|^2 = a \|x\|^2$$

Par suite, on applique la proposition précédente qui assure l'existence et l'unicité de x_n . Notons $B_n(x') = x_n$. L'opérateur B_n est linéaire sur H et à valeurs dans H_n . Il satisfait par définition l'égalité suivante sur H : $P_n A B_n = P_n$, En outre B_n est continu puisque

$$\|B_n(x')\| \leq \frac{1}{a} \|P_n(x')\| \leq \frac{1}{a} \|x'\|$$

Soit

$$\|B_n\| \leq \frac{1}{a}$$

Pratiquement, la résolution sur H_n se fait en inversant une matrice. En effet soit (e_1, e_2, \dots, e_n) une base de H_n , Sur H_n , l'équation $P_n A(x) = P_n(x')$ est équivalente aux n équations :

$$\langle P_n A(x_n), e_h \rangle = \langle P_n(x'), e_h \rangle \quad h = 1, 2, \dots, n$$

soit en posant :

$$x_n = \alpha_{1,n} e_1 + \dots + \alpha_{n,n} e_n$$

et en notant que

$$\begin{aligned} \langle P_n A(x_n), e_h \rangle &= \langle A(x_n), e_h \rangle \\ \sum_{i=1}^{i=n} \alpha_{i,n} \langle A(e_i), e_h \rangle &= \langle P_n(x'), e_h \rangle \quad h = 1, 2, \dots, n \end{aligned}$$

C'est un système (de déterminant non nul) de n équations à n inconnues $\alpha_{1,n}, \dots, \alpha_{n,n}$, donc qui peut être résolu.

Naturellement, un choix judicieux de la base favorise la résolution du système (Cf. Exercice N° 61: possibilité d'une base orthonormale).

2ème étape : Résultat final.

Il nous faut démontrer que pour tout $x' \in H$, $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} B_n P_n(x') = A^{-1}(x')$ au sens de l'espace H bien entendu.

Pour ce faire, on remarque d'abord que pour tout $y \in H$, alors $y - AB_n y$ est orthogonal à tout vecteur z de H_n . En effet $Z = P_n(z)$, donc

$$\begin{aligned} \langle y - AB_n y, z \rangle &= \langle y - AB_n y, P_n(z) \rangle \\ &= \langle P_n(y) - P_n AB_n(y), z \rangle \\ &= 0 \end{aligned}$$

car $P_n(y) = P_n AB_n(y)$ pour tout y de H par définition de B_n . On écrit aussi pour tout z de H_n puisque A^{-1} est auto-adjoint

$$0 = \langle y - AB_n y, A^{-1} A z \rangle = \langle A^{-1} y - B_n y, A z \rangle$$

Prenons maintenant un vecteur z appartenant à un espace $A(H_n)$. On a aussi pour tout $n \geq h$ puisque $H_n \supset H_h$:

$$\langle A^{-1} y - B_n y, A z \rangle = 0$$

Donc si $z \in \bigcup_{h=1}^{\infty} A(H_h)$, on a encore :

$$\lim_{n \rightarrow \infty} \langle A^{-1}y - B_n y, Az \rangle = 0$$

Soit maintenant z un point quelconque de H et ε un nombre positif donné. Il existe un entier N tel que

$$\|z - P_h(z)\| < \varepsilon$$

Puisque pour tout $n \geq N$, on a également $|\langle A^{-1}y - B_n y, AP_h(z) \rangle| \leq \varepsilon$

Soit $|\langle A^{-1}y - B_n y, Az \rangle| \leq |\langle A^{-1}y - B_n y, AP_h(z) \rangle|$

$$+ |\langle A^{-1}y - B_n y, A(z - P_h(z)) \rangle|$$

$$\leq \varepsilon + (\|A^{-1}y - B_n y\|) \|A\| \varepsilon$$

$$\leq \varepsilon (1 + (\|A^{-1}\| + \|B_n\|) \|y\| \|A\|)$$

$$\leq \varepsilon (1 + \frac{2}{a} \|y\| \|A\|)$$

Ce qui prouve bien que pour tout $z \in H$, on a :

$$\lim_{n \rightarrow \infty} \langle A^{-1}y - B_n y, Az \rangle = 0 \quad \forall y \in H$$

On en déduit que $\lim_{n \rightarrow \infty} \langle A^{-1}y - B_n y, z \rangle = 0$ pour tout $z \in H$ puisque A est bijectif mais cela ne suffit pas à prouver le résultat limite (convergence faible, Cf E.A.F). Une astuce permet de lever le doute. On a grâce à la coercivité de A :

$$\langle A(A^{-1}y - B_n y), A^{-1}y - B_n y \rangle \geq a \|A^{-1}y - B_n y\|^2$$

Soit :

$$\langle y - AB_n y, A^{-1}y \rangle - \langle y - AB_n y, B_n y \rangle \geq a \|A^{-1}y - B_n y\|^2$$

Mais $\langle y - AB_n y, B_n y \rangle = 0$ puisque $B_n y \in H_n$

$$\text{Donc } \langle A^{-1} y - B_n y, y \rangle \geq \alpha \|A^{-1} y - B_n y\|^2$$

Or $\lim_{n \rightarrow \infty} \langle A^{-1} y - B_n y, y \rangle = 0$ pour chaque y fixé en faisant $z = y$

dans la relation d'orthogonalité. Finalement, pour chaque y fixé de H , on a bien :

$$\lim_{n \rightarrow \infty} \|A^{-1} y - B_n y\| = 0$$

Théorème 7.2 : Soit A un opérateur linéaire, continu, coercitif et auto-adjoint sur un espace de Hilbert H . L'unique solution de l'équation

$$Ax = x'$$

où x' est fixé dans H s'obtient comme limite dans H de l'unique solution dans H_n de l'équation

$$P_n A(x) = P_n(x')$$

3. CALCUL NUMERIQUE D'UNE APPROXIMATION PAR DISCRETISATION;

On peut quelquefois obtenir une valeur approchée d'une meilleure approximation en discrétisant le problème. Pour fixer les idées prenons une fonction f continue et à valeurs complexes sur un segment $[a, b]$. On se donne également une famille de n fonctions $\phi_1, \phi_2, \dots, \phi_n$, formant un système de Čebičev en tant que fonctions continues sur $[a, b]$. Nous avons déjà donné des formules exactes pour obtenir la meilleure approximation de f par des combinaisons linéaires des n fonctions ϕ_1, \dots, ϕ_n au sens d'une norme quadratique, c'est-à-dire au sens de $L^2([a, b])$. Pour un calcul approché, on peut discrétiser le problème en se donnant m points distincts $a < x_1 < x_2 < \dots < x_m \leq b$ formant une division Δ_m de $[a, b]$. On pose

$$|\Delta_m| = \sup_{h=1, \dots, (m-1)} |x_{h+1} - x_h|$$

et $\delta(x_h) = x_h - x_{h-1}$ pour $h = 1, 2, \dots, m$ où l'on pose $x_0 = a$

par convention. On rappelle que l'on a présupposé $x_1 > a$.

On pose ensuite $\rho_m(f) = \sqrt{\sum_{h=1}^{n-m} |f(x_h)|^2 \delta(x_h)}$

L'application $f \rightarrow \rho_m(f)$ a toutes les propriétés d'une norme sur $\mathcal{E}[a, b]$ sauf la propriété de nullité :

$\rho_m(f) = 0$ n'implique pas $f = 0$

On dit que ρ_m est une pseudo-norme. Il existe toutefois une unique combinaison linéaire f_m des $\phi_1, \phi_2, \dots, \phi_h$ qui minimise l'expression

$\rho_m(f - f_m)$

L'existence de cette meilleure approximation au sens de ρ_m s'obtient par la même méthode que dans le cas d'une norme (Cf. Chapitre 1). Pour l'unicité il faut pousser un peu l'analyse. Supposons donc que f_m et f'_m soient deux meilleures approximations au sens de ρ_m . On a a priori :

$\rho_m = \rho_m(f - f_m) = \rho_m(f - f'_m) \leq \rho_m(f - \frac{1}{2}(f_m + f'_m))$

Or

$\rho_m(f - \frac{1}{2}(f_m + f'_m)) = \rho_m(\frac{1}{2}(f - f_m) + \frac{1}{2}(f - f'_m))$

et

$\rho_m^2(\frac{1}{2}(f - f_m) + \frac{1}{2}(f - f'_m)) + \rho_m^2(\frac{1}{2}(f - f_m) - \frac{1}{2}(f - f'_m)) = \frac{1}{2}(\rho_m^2(f - f_m) + \rho_m^2(f - f'_m)) = \rho_m^2$

Ce qui implique

$\rho_m^2(\frac{1}{2}(f'_m - f_m)) = 0$

Mais f_m et f'_m sont des combinaisons linéaires des n fonctions ϕ_1, \dots, ϕ_n et sur l'espace vectoriel engendré par ces fonctions, ρ_m est en fait une norme dès que $m \geq n$ (Cf. Exercice N° 65) donc $f_m = f'_m$ ce qui termine la démonstration.

On note d'ailleurs la propriété d'orthogonalité caractéristique de f_m

$\sum_{h=1}^m (f(x_h) - f_m(x_h)) \phi_p(x_h) \delta_m(x_h) = 0$ pour $p = 1, 2, \dots, n$

L_n est explicitement calculable et s'appelle la meilleure approximation discrétisée de f . On dispose d'un théorème de convergence

Théorème [Δ_m] $m \geq 1$ une famille de divisions du segment $[a, b]$,
 avec et $\lim_{m \rightarrow \infty} |\Delta_m| = 0$. On suppose en outre que $\bigcup_{m=1}^{\infty} \Delta_m$ est
dérivable $] .$ Soit également $\Phi_1, \Phi_2, \dots, \Phi_n$ un système de Chebichev
sur $].$ Soit f un élément de $[a, b]$ et f_m la meilleure approximation
discrétisée à l'ordre m de la fonction f . Soit P_f la meilleure
approximation de f au sens de $L^2[a, b]$, par les
éléments du sous-espace vectoriel engendré par $\Phi_1, \Phi_2, \dots, \Phi_n$. Alors f_m conver-
ge vers f .

(grâce au N° 65, on peut remplacer l'hypothèse sur $\Phi_1, \Phi_2, \dots, \Phi_n$ par une de linéarité indépendante).

On appelle E_n le sous-espace vectoriel engendré par $\Phi_1, \Phi_2, \dots, \Phi_n$. D'une manière plus générale, dans toute meilleure approximation en norme, on dispose de

$$\rho_m(f) \leq 2 \rho_m(f) \quad (\text{Cf } \S 2.1)$$

Or

$$\rho_m = \left(\int_a^b |f(x)|^2 dx \right)^{1/2} = \rho(f) \quad (\text{grâce à la théorie de l'intégrale de Riemann puisque } \Delta_m \subset \Delta_{m+1} \text{ et } \lim_{m \rightarrow \infty} |\Delta_m| = 0 . \text{ Donc}$$

il existe une constante A dépendant pas de m telle que :

$$\rho_m(f_m) \leq A$$

Sur E_n les normes sont équivalentes et en particulier il existe une constante B telle que pour tout g de E_n on ait

$$\rho(g) \leq B \rho_m(g) \quad \text{En fait } \rho_m(g) = \left(\int_a^b |g(x)|^2 dx \right)^{1/2} \quad \text{on peut déduire que la suite des } B_m$$

est bornée par une constante B' . Donc :

$$\rho(f_m) \leq B' \rho_m(f_m) \leq AB'$$

La suite f_m est bornée au sens de la norme ρ sur E_n . Il existe alors une sous-suite f_{m_h} qui converge vers $g \in E_n$ (Cf Théorème de Bolzano-Weierstrass).

Montrons que

$$\lim_{h \rightarrow \infty} \rho_{m_h}(f - f_{m_h}) = \rho(f - g)$$

On a :

$$(1) |\rho_{m_h}(f - f_{m_h}) - \rho(f - g)| \leq |\rho_{m_h}(f - f_{m_h}) - \rho(f - f_{m_h})| + |\rho(f - f_{m_h}) - \rho(f - g)|$$

La deuxième expression du terme de droite dans (1) tend vers 0 lorsque h tend vers l'infini. Or en supposant $x_m = b$, pour simplifier, on a :

$$|\rho_{m_h}(F) - \rho(F)|^2 \leq \sum_{h=1}^{m_h} \int_{x_{h-1}}^{x_h} |F(x_h)|^2 - |F(x)|^2 dx$$

Si donc F est continue, cette fonction est uniformément continue et

$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \forall x, y \quad |x - y| \leq \eta \Rightarrow |F(x) - F(y)| < \varepsilon$$

d'où

$$|\rho_{m_h}(F) - \rho(F)|^2 \leq \varepsilon^2 (b - a)$$

On peut montrer que les fonctions $f - f_{m_h}$ sont également uniformément équi continues, (même η pour tout ε et ce pour tout $h \geq 1$: ceci provient d'un théorème d'équivalence des normes sur E_n qui montre que les coefficients de f_{m_h} décomposés dans la base $\phi_1, \phi_2, \dots, \phi_n$ sont uniformément bornés).

Par suite le premier terme de l'expression de droite dans (1) tend également vers 0 lorsque h tend vers l'infini.

Par définition de la meilleure approximation Pf de f au sens de ρ , on a

$$\rho(f - g) \geq \rho(f - Pf)$$

Mais

$$\rho(f - Pf) = \lim_{h \rightarrow \infty} \rho_{m_h}(f - Pf)$$

et

$$\rho_{m_h}(f - Pf) \geq \rho_{m_h}(f - f_{m_h})$$

D'où à la limite

$$\rho(f - g) = \rho(f - Pf) \quad Pf \text{ et } g \in E_n$$

Par le théorème d'unicité, on en déduit $Pf = g$ et comme cette égalité est exacte pour toute sous-suite convergente de f_n (il en existe au moins une d'après d'après ce qui a été vu) on a bien $\lim_{m \rightarrow \infty} f_m = g$.

La convergence a lieu au sens de \mathcal{Q} ou au sens de la convergence uniforme puisque ces normes sont équivalentes sur E_n . Ceci termine la démonstration.

4. APPROXIMATION NON LINEAIRE.

Il peut y avoir intérêt, notamment du point de vue de la rapidité des calculs, à effectuer des approximations qui ne soient pas des combinaisons linéaires de n fonctions données. L'exemple le plus usuel est l'approximation d'une fonction f au moyen de fractions rationnelles :

$$R(x) = \frac{a_0 + a_1 x + \dots + a_n x^n}{b_0 + b_1 x + \dots + b_m x^m}$$

Il s'agit donc de déterminer les $(n+m+2)$ constantes $a_0, a_1, \dots, a_n, b_0, \dots, b_m$, de sorte que $f-R$ soit, en norme, optimalement petit. De fait, la pratique montre que du point de vue de la capacité d'une fraction $R(x)$ à représenter une fonction $f(x)$, les possibilités sont les mêmes que celles d'un polynôme de degré $n+m$. Cependant, le calcul de la valeur d'une fraction rationnelle est plus court que le calcul de la valeur d'un polynôme équivalent. En effet :

Le calcul d'un polynôme de degré $(n+m)$ requiert $n+m$ additions et $n+m-1$ multiplications, si l'on utilise l'algorithme de Hörner.

Le calcul d'une fraction rationnelle, après avoir mis cette dernière sous forme de fraction continue, exige au plus $\text{Sup}(n,m)$ divisions et multiplications.

Exemple :

$$\begin{aligned} R(x) &= \frac{2x^4 - 4x^3 - 2x^2 + 12x - 4}{x^3 - 2x^2 - x + 5} \\ &= 2x + \frac{2x - 4}{x^3 - 2x^2 - x + 5} \\ &= 2x + \frac{2}{(x^3 - 2x^2 - x + 5) / (x - 2)} \\ &= 2x + \frac{2}{x^2 - 1 + 3 / (x - 2)} \end{aligned}$$

Par division, successives, on a obtenu une expression qui nécessite 4 opérations "longues" (multiplication et division) et 4 opérations "courtes" (addition et soustraction) au lieu de 6 et 7 pour le polynôme.

Les théorèmes démontrés dans ce texte sur l'existence des approximations et leurs caractérisations dans le cas linéaire ne sont pas applicables directement mais on peut montrer que :

1°) il existe une meilleure approximation uniforme par les fractions rationnelles d'ordre donné (n,m)

2°) la meilleure approximation uniforme par les fractions rationnelles d'ordre (n,m) est telle que la courbe de l'erreur $R(x)-f(x)$ atteint son maximum de façon alternée en au moins $(m+n-2)$ points si les degrés du dénominateur et du numérateur sont exactement n et m . De nombreux algorithmes d'approximation ont été mis au point et nous ne pouvons insister, renvoyant le lecteur à la bibliographie.

EXERCICES SUR LE CHAPITRE 7

EXERCICE N° 57 :

A partir de la relation de Darboux-Christoffel, montrer que

$$\alpha_k^{-1} = \sum_{i=0}^n (P_i(x_k))^2 = K_n(x_k, x_k)$$

avec les notations du paragraphe consacré à la quadrature mécanique et aux polynômes orthogonaux.

EXERCICE N° 58 :

Soient (x_1, x_2, \dots, x_n) les n zéros du $n^{\text{ième}}$ polynôme orthogonal $P_n(x)$ associé à un poids $h(x)$ sur un intervalle $[a, b]$. Soient $\alpha_1, \dots, \alpha_n$ les nombres de Christoffel associés. On pose

$$\langle f, g \rangle = \sum_{k=1}^n \lambda_k f(x_k) g(x_k)$$

Montrer qu'il s'agit d'un produit scalaire et que la base orthogonale obtenue à partir des n fonctions $1, x, x^2, \dots, x^{n-1}$ est précisément la base formée par les polynômes orthogonaux de départ P_n .

EXERCICE N° 59 : (A. MARKOV)

Montrer que pour une fonction $2n$ fois continuellement dérivable

$$\int_{-1}^{+1} f(x) dx - \sum_{h=1}^n \alpha_h f(x_h) = \frac{f^{(2n)}(\xi)}{(2n)! \alpha_n^2} \text{ où } -1 < \xi < +1$$

avec les notations de ce chapitre.

EXERCICE N° 60 :

Vérifier les formules suivantes d'intégration mécanique :

$$\left\{ \begin{array}{l} n=0 \quad \int_{-1}^{+1} f(x) dx \sim 2 f(0) \\ n=1 \quad \int_{-1}^{+1} f(x) dx \sim f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right) \\ n=2 \quad \int_{-1}^{+1} f(x) dx \sim \frac{1}{9} [5f(-\sqrt{\frac{3}{5}}) + 8f(0) + 5f(\sqrt{\frac{3}{5}})] \end{array} \right.$$

et plus généralement :

$$\alpha_n = \frac{2}{1-x_h^2} \frac{1}{(P'_n(x_h))^2}$$

EXERCICE N° 61 :

Soit $x : t \rightarrow x(t)$, une fonction continue sur $[0,1]$ et à valeurs réelles.

(1) Montrer que la solution y du problème de Sturm-Liouville

$$y''(t) = -x(t) \quad ; \quad y(0) = y(1) = 0$$

vérifie

$$y(t) = \int_0^1 K(s,t) x(s) ds$$

$$\begin{array}{ll} \text{ou } K(s,t) = s(1-t) & \text{pour } 0 \leq s \leq t \leq 1 \\ = t(1-s) & \text{pour } 0 \leq t \leq s \leq 1 \end{array}$$

(2) On appelle également K l'opérateur qui à x associé la fonction y . Montrer que K s'étend en un opérateur linéaire, continu et auto-adjoint sur $L^2[0,1]$.

Montrer que $\langle Kx, x \rangle \geq 0$ et ce pour toute fonction x de $L^2[0,1]$.

(3) On définit $A = I + K$ où I désigne l'opérateur identité sur $L^2[0,1]$ et K l'opérateur précédent. Montrer que A est un opérateur linéaire, continu, auto-adjoint et coercitif sur $L^2[0,1]$.

(4) Appliquer la méthode de Galerkin en prenant comme base hilbertienne de $L^2[0,1]$, la base

$$e_0 = 1; \quad e_{2h}(t) = \sqrt{2} \cos(2\pi ht) \quad e_{2h+1}(t) = \sqrt{2} \sin(2\pi ht)$$

EXERCICE N° 62 :

Soit A un opérateur linéaire sur un espace de Hilbert H . On définit $\langle\langle x, y \rangle\rangle = \langle Ax, y \rangle$. Donner un ensemble nécessaire et suffisant de conditions sur A de sorte que l'on définisse ainsi un produit scalaire sur H et une norme équivalente à la norme de départ.

EXERCICE N° 63 :

Soit H un espace de Hilbert réel et $a(\dots)$ une forme bilinéaire sur \mathcal{H} telle qu'il existe deux constantes positives α et β et

$$(1) \quad |a(x, y)| \leq \beta \|x\| \|y\| \quad \forall x, y \in H, \beta > 0$$

$$(2) \quad a(x, x) \geq \alpha \|x\|^2 \quad \forall x \in H, \alpha > 0$$

Montrer que pour toute forme linéaire continue L sur H , il existe un unique x de H tel que $a(x, y) = L(y) \quad \forall y \in H$

EXERCICE N° 64 :

Démontrer le théorème suivant qui généralise la méthode de Galerkin :
Soit $a(\dots)$ une forme bilinéaire définie sur un espace de Hilbert réel \mathcal{H} et vérifiant pour tout u et tout v les inégalités :

$$|a(u, v)| \leq K_1 \|u\| \|v\|$$

$$|a(u, u)| \geq \alpha \|u\|^2 \quad (\alpha > 0)$$

Soit $L(\cdot)$ une forme linéaire définie sur \mathcal{H} et vérifiant pour tout v l'inégalité :

$$|L(v)| \leq K_2 \|v\|$$

Il existe un unique $u \in \mathcal{F}$ vérifiant pour tout $v \in \mathcal{F}$ l'égalité :

$$a(u, v) = L(v)$$

Construire un algorithme fournissant la solution (on peut procéder comme suit ; qui est une version différente de celle donnée pour la méthode de Galerkin).

1ère étape : démonstration directe de l'unicité de la solution si elle existe.

2ème étape : démonstration de l'existence et de l'unicité de la suite $[u_{m_n}]_{n \geq 1}$ des solutions approchées obtenues en se plaçant dans un sous-espace vectoriel H_m de dimension finie comme dans la méthode de Galerkin

3ème étape : démonstration de la convergence au sens faible d'une sous-suite $[u_{m_n}]_{n \geq 1}$ vers une limite \hat{u} .

4ème étape : démonstration du fait que \hat{u} est une solution u (ceci établit l'existence de la solution u).

5ème étape : démonstration du fait que la suite $[u_m]$, entière converge vers u au sens faible.

6ème étape : démonstration de la convergence dans \mathcal{F} de u_m vers la solution u du problème P.

EXERCICE N° 65:

(1) Avec les notations du § 3 du Chapitre 7, montrer que \mathcal{C}_m est une norme sur l'espace vectoriel engendré par les n fonctions $(\Phi_1, \Phi_2, \dots, \Phi_n)$ qui forment un système de Cebicev :

(si $f = \sum_{i=1}^n a_i \Phi_i$ est telle que $\mathcal{C}_m(f) = 0$, alors on a

$$f(x_j) = 0 \quad j = 1, \dots, m$$

On rappelle que $x_i \neq 0$. Ceci donne le système :

$$\sum_{i=1}^n a_i \Phi_i(x_j) = 0 \quad j = 1, \dots, m$$

lorsque $m \geq n$, le déterminant de terme général $\Phi_i(x_j)$ pour $i = 1, 2, \dots, n$ et $j = 1, 2, \dots, n$ est différent de zéro puisque le système est de Čebičev. Donc $a_1 = a_2 = \dots = a_n = 0$ est l'unique solution possible.

(2) On suppose maintenant que les n fonctions $\Phi_1, \Phi_2, \dots, \Phi_n$, sont linéairement indépendantes en tant qu'éléments de $\mathcal{C}[a,b]$. On se donne une famille Δ_m de divisions telle que

(a) $\Delta_m \subset \Delta_{m+1}$

(b) $\Delta = \bigcup_{m=1}^{\infty} \Delta_m$ est un sous-ensemble dense de $[a,b]$

En déduire que pour m assez grand, ρ_m est une norme sur l'espace vectoriel engendré par les n fonctions $\Phi_1, \Phi_2, \dots, \Phi_n$.

(la relation $\sum_{i=1}^n a_i \Phi_i(x) = 0$ pour tout x de Δ implique $\sum_{i=1}^n a_i \Phi_i = 0$ parce que les fonctions Φ_i sont continues donc $a_1 = a_2 = \dots = a_n = 0$)

EXERCICE N° 66 :

Soit $R(x) = \frac{a+bx}{1+cx}$ où a, b, c sont trois constantes.

Existe-t-il une fonction R telle que

$$R(0) = \alpha ; R'(0) = \beta ; R''(0) = \gamma$$

où α, β et γ sont fixées ? Peut-on généraliser (approximants de Padé d'une fonction f) ?

CHAPITRE 8

MISE EN OEUVRE PRATIQUE DE CERTAINS CALCULS D'APPROXIMATION

8-1 BUT DU TRAVAIL

Le but du travail que le groupe s'est proposé est d'obtenir par la méthode dite des moindres carrés, c'est-à-dire par les méthodes hilbertiennes :

- soit le lissage de courbes définies discrètement,
- soit le remplacement d'une fonction définie analytiquement mais "complexe" par une fonction polynôme facilement exploitable.

Le principe théorique est celui exposé au chapitre 2 : approximation en moyenne quadratique.

Nous posons que E est un espace vectoriel de fonctions (réelles) définies sur un intervalle I de \mathbb{R} , muni d'un produit scalaire \langle , \rangle .

On prend E_n , le sous-espace vectoriel des polynômes réels de degré inférieur ou égal à n . Cet E_n est de dimension $(n+1)$, c'est un espace euclidien (cf E.A.F. IA 4.4.5).

Soit f une fonction de E . La meilleure approximation de f , par E_n au sens de la norme déduite sur E du produit scalaire, existe, est unique et s'obtient comme la projection orthogonale f_0 de f sur E_n (cf chapitre II, § 3)

Supposons le produit scalaire sur E défini au moyen de p points x_1, x_2, \dots, x_p de I selon la formule

$$\langle u, v \rangle = \sum_{i=1}^p u(x_i) v(x_i) w(x_i)$$

où w est un poids donné (ici p valeurs réelles). La projection de f sur E_n est dite approximation de f au sens des moindres carrés relativement au poids w .

Pour déterminer pratiquement cette projection, différentes méthodes sont possibles (cf chapitre 2). Nous étudierons les calculs relatifs aux deux méthodes suivantes :

1ère méthode

On peut repérer E_n au moyen d'une base quelconque : $\mathcal{B} = (e_0, e_1, \dots, e_n)$ par exemple le système e_k où $e_k : x \rightarrow x^k$, fonction définie de I dans \mathbb{R} et exprimer le produit scalaire d'un élément quelconque g de E_n avec f en fonction des composantes a_0, a_1, \dots, a_n de g dans \mathcal{B} . Puis écrire que, pour rendre $\|f-g\|$ minimum, il est nécessaire que les dérivées partielles de

la fonction $\begin{matrix} \mathbb{R}^{n+1} & \longrightarrow & \mathbb{R} \\ (a_0, \dots, a_n) & \longmapsto & \|f-g\| \end{matrix}$ soient nulles. On obtient

ainsi un système (S) d'équations, dites équations normales de Gram. La théorie nous enseigne que ces équations sont linéaires, et c'est là, bien sûr, l'avantage considérable de la méthode hilbertienne.

Si l'on cherche l'approximation au sens des moindres carrés, le système (S) est, en outre, symétrique. On peut utiliser, pour résoudre ce système (S), des méthodes numériques appropriées (par exemple Cholesky).

2ème méthode

a) (Polynômes orthogonaux). On peut déterminer une base orthogonale $\mathcal{B}_1 = (P_0, P_1, \dots, P_n)$ de E_n . Si l'on appelle s_0, s_1, \dots, s_n les composantes du polynôme f_0 cherché dans \mathcal{B}_1 , on a, en posant $f' = f - f_0$ ($f' \perp E_n$ par définition de la projection orthogonale) :

$$\begin{aligned} \langle f, P_i \rangle &= \langle f_0 + f', P_i \rangle \\ &= \langle f_0, P_i \rangle \quad (\text{car } f' \perp P_i) \\ &= \langle s_0 P_0 + s_1 P_1 + \dots + s_n P_n, P_i \rangle \\ &= s_i \langle P_i, P_i \rangle \quad (\text{car } \langle P_k, P_i \rangle = 0 \text{ si } k \neq i). \end{aligned}$$

D'où :

$$s_i = \frac{\langle f, P_i \rangle}{\langle P_i, P_i \rangle}$$

onc calculer les composantes s_i de f_0 et en déduire :

$$f_0 = \alpha_0 P_0 + \alpha_1 P_1 + \dots + \alpha_n P_n$$

(re II, § 4)

Si la détermination d'une base orthogonale n'est pas déjà faite, on prend le système linéaire correspondant (cf chapitre II, formule (5))

$$f_0 = y_0 P_0 + \dots + s_n P_n$$

$$\begin{aligned} \langle f_0, P_i \rangle &= \langle f, P_i \rangle \\ &= s_0 \langle P_0, P_i \rangle + \dots + s_n \langle P_n, P_i \rangle \end{aligned}$$

Soit

$$\begin{bmatrix} \langle P_0, P_0 \rangle & \dots & \langle P_0, P_n \rangle \\ \langle P_1, P_0 \rangle & \dots & \langle P_1, P_n \rangle \\ \vdots & & \vdots \\ \langle P_n, P_0 \rangle & \dots & \langle P_n, P_n \rangle \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} \langle f, P_0 \rangle \\ \langle f, P_1 \rangle \\ \vdots \\ \langle f, P_n \rangle \end{bmatrix}$$

Nous examiner ces diverses méthodes sur quelques cas concrets :

s d'une fonction indéfiniment dérivable : $\sin x$

s d'une fonction continue, dérivable sauf en un point

: $|x|$

s d'une fonction non continue, à discontinuités de première

°

: fonction en escalier.

Une remarque s'impose d'abord, quant au choix d'un outil particulier de programmation :

8-2 POURQUOI PL/1 ?

Après avoir été initiés à l'Informatique, il était normal que des stagiaires particulièrement intéressés désirassent poursuivre et approfondir leur expérience de la programmation en l'appliquant à un plus vaste domaine, tel que celui des problèmes d'approximation. Encore fallait-il disposer d'un outil de programmation adéquat, plus souple et plus riche que le simple BASIC, malgré tout un peu sommaire pour des programmes de cette envergure.

PL/I fut donc choisi parce que, seul, il offrait les qualités d'"universalité" et de souplesse nécessaires.

Pour comprendre cela, il faut savoir que, dès sa conception, PL/I a repris les possibilités des autres langages. C'est un ensemble du meilleur de FORTRAN, BASIC, COBOL, ALGOL ou l'Assembleur. On y retrouve les mêmes facilités de contrôle des boucles FOR-NEXT du BASIC, mais considérablement étendues et améliorées. Ce que l'on ne sait pas toujours, c'est que PL/I, initialement conçu afin de tirer profit des qualités des ordinateurs de 3ème génération, permet, entre autres avantages, de gérer les interrupteurs-machines aussi bien que l'Assembleur. Il contient aussi ses propres outils de mise au point qui lui donnent la faculté de préciser et de localiser les erreurs de programmation.

De plus, une grande souplesse est permise dans la gestion de la mémoire des variables. En distinguant 4 classes de mémoire : STATIC, AUTOMATIC, CONTROLLED et BASED, le langage donne au programmeur les moyens d'ajuster au plus près la taille des tableaux, structures ou tableaux de structures, en même temps que la possibilité de contrôler le moment de leur implantation en mémoire et celui de leur libération.

Ainsi, des bornes d'un tableau peuvent être spécifiées par des variables. Convenablement valorisées au cours de l'exécution du programme, celles-ci serviront à réserver au tableau l'emplacement mémoire nécessaire et suffisant. Par exemple :

```
DCL A (*,*) CONTROLLED ;
```

```
GET LIST (N) ;
```

```
ALLOCATE A(N,2*N) ;
```

Lors de la réalisation de l'ALLOCATE, $2n^2$ positions sont réservées pour A .

De même, la classe de mémoire CONTROLLED permet l'organisation et le traitement des "piles".

Mieux encore, PL/I permet facilement l'emploi de routines écrites dans d'autres langages. Une procédure PL/I peut donc utiliser directement un sous-programme scientifique de la S.S.P.* ou bien les sous-programmes de la table traçante CALCOMP. A partir des 2 tableaux des coordonnées des points de la courbe à tracer, l'appel des sous-programmes SCALE et AXIS permet la mise en place des échelles et le tracé des axes alors que l'appel de la subroutine LINE trace le graphe point par point. D'autres subroutines servent à améliorer la présentation du dessin, par exemple en y adjoignant divers commentaires directement tracés par la plume de la table traçante.

En fin de compte, ces quelques facilités rapidement mentionnées, ajoutées aux autres possibilités habituelles déjà connues de FORTRAN, BASIC ou COBOL, montrent que PL/I est le seul vrai langage à vocation universelle. Si nous ajoutons que le compilateur, dans sa version "Optimiseur", ne nécessite que 44 K de mémoire centrale (ce qui est peu par rapport aux compilateurs d'autres langages comparables), que le code résultant est particulièrement performant et que de nouvelles instructions adaptées aux techniques de programmation structurée viennent d'être incorporées au langage, on comprendra sans peine que cette richesse et cette souplesse puissent séduire un esprit ouvert, pour qui la programmation ne se réduit pas à un banal codage, plus ou moins automatique.

* Scientific Subroutine Package : ensemble de sous-programmes scientifiques résolvant les problèmes classiques d'analyse numérique ou de statistique.

8-3 PREMIERE METHODE

On suppose f définie de façon discrète par $N+1$ couples (x_i, y_i) .

La méthode des moindres carrés revient à minimiser

$$\epsilon = \sum_{i=1}^{N+1} (y_i - P_n(x_i))^2$$

où

$$P_n(x) = a_0 + a_1 x + \dots + a_n x^n$$

Il s'agit de calculer a_0, a_1, \dots, a_n , coefficients du polynôme. Le calcul se fait en écrivant $(n+1)$ équations linéaires en $(n+1)$ inconnues.

$$\frac{\partial \epsilon}{\partial a_k} = 0 \quad k = 0, 1, \dots, n$$

Le système symétrique a été résolu par la méthode de Cholesky (décomposition de A sous la forme $R.R^T$ où R est une matrice triangulaire supérieure) :

$$A = R.R^T \implies R.R^T.X = B$$

$$\begin{cases} R^T.X = Y \\ RY = B \end{cases}$$

les, à partir du 4ème ou 5ème degré, la courbe initiale et la courbe à deux ou trois décimales près).

, l'approximation est évidemment moins ait nécessaire d'augmenter le nombre (figure 1, figure 2).

escalier, le phénomène de GIBBS aux points (pour les séries de Fourier) Figure 3.

Pour les degrés importants (par exemple > 15 avec 50 à 100 points d'appui), la méthode de résolution par Cholesky devient instable et les courbes d'approximation se mettent à osciller. Ceci est inhérent en partie au calcul par ordinateur (cumul d'erreur) et à la méthode de Cholesky elle-même. Des essais ultérieurs ont permis de conserver une stabilité jusqu'à des degrés plus élevés avec des méthodes utilisant des transformations orthogonales du type Householder (QU).

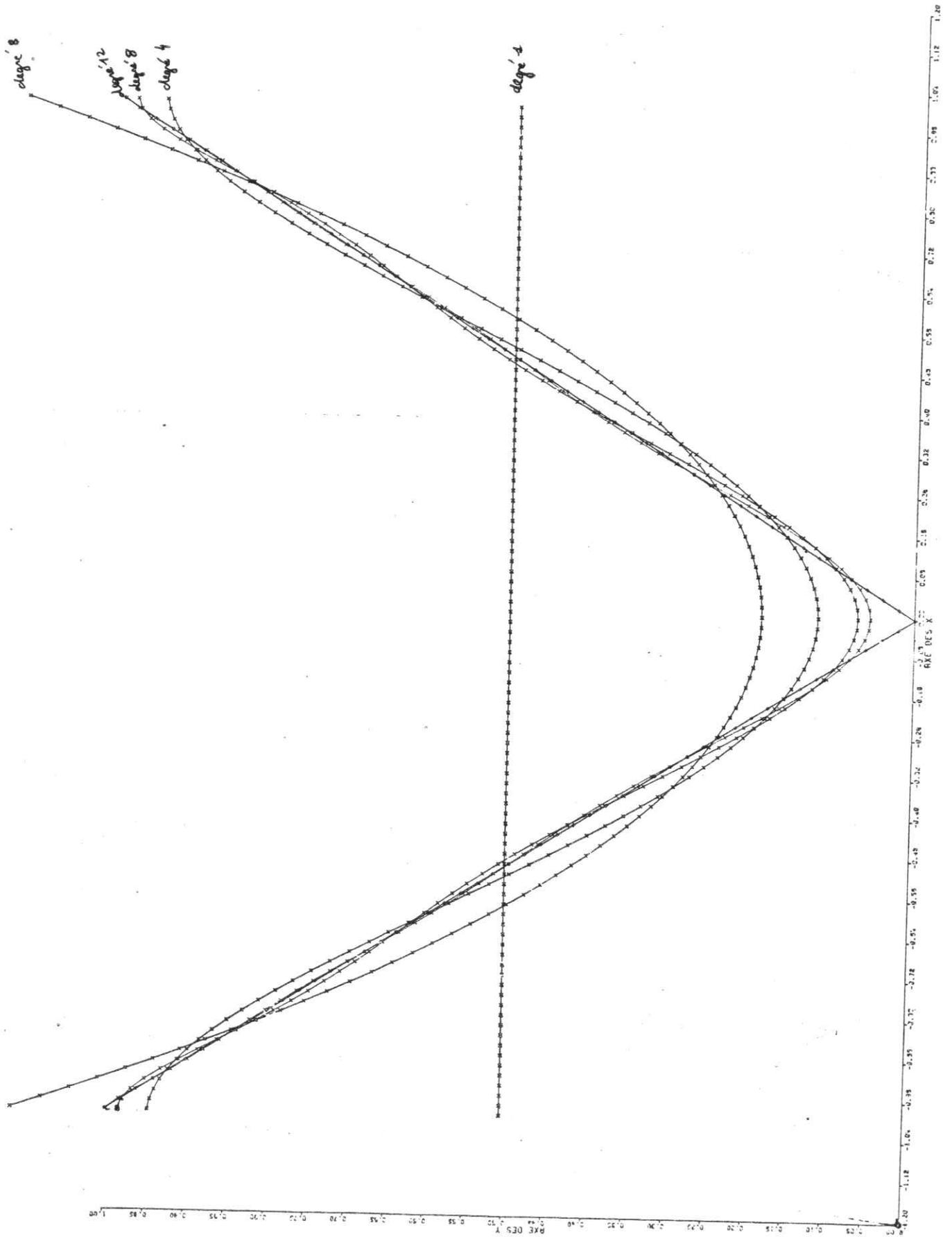


Figure 1

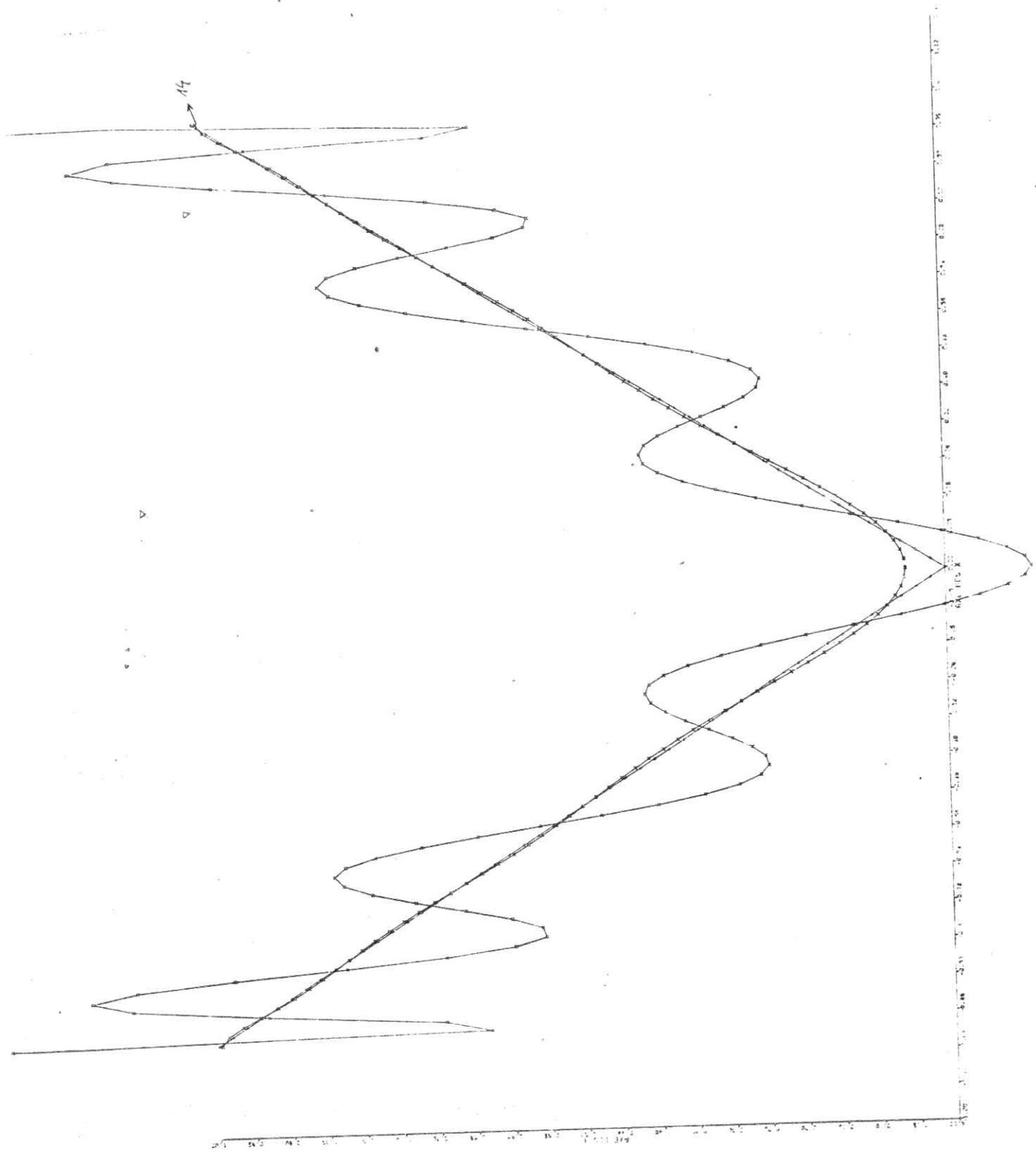


Figure 2

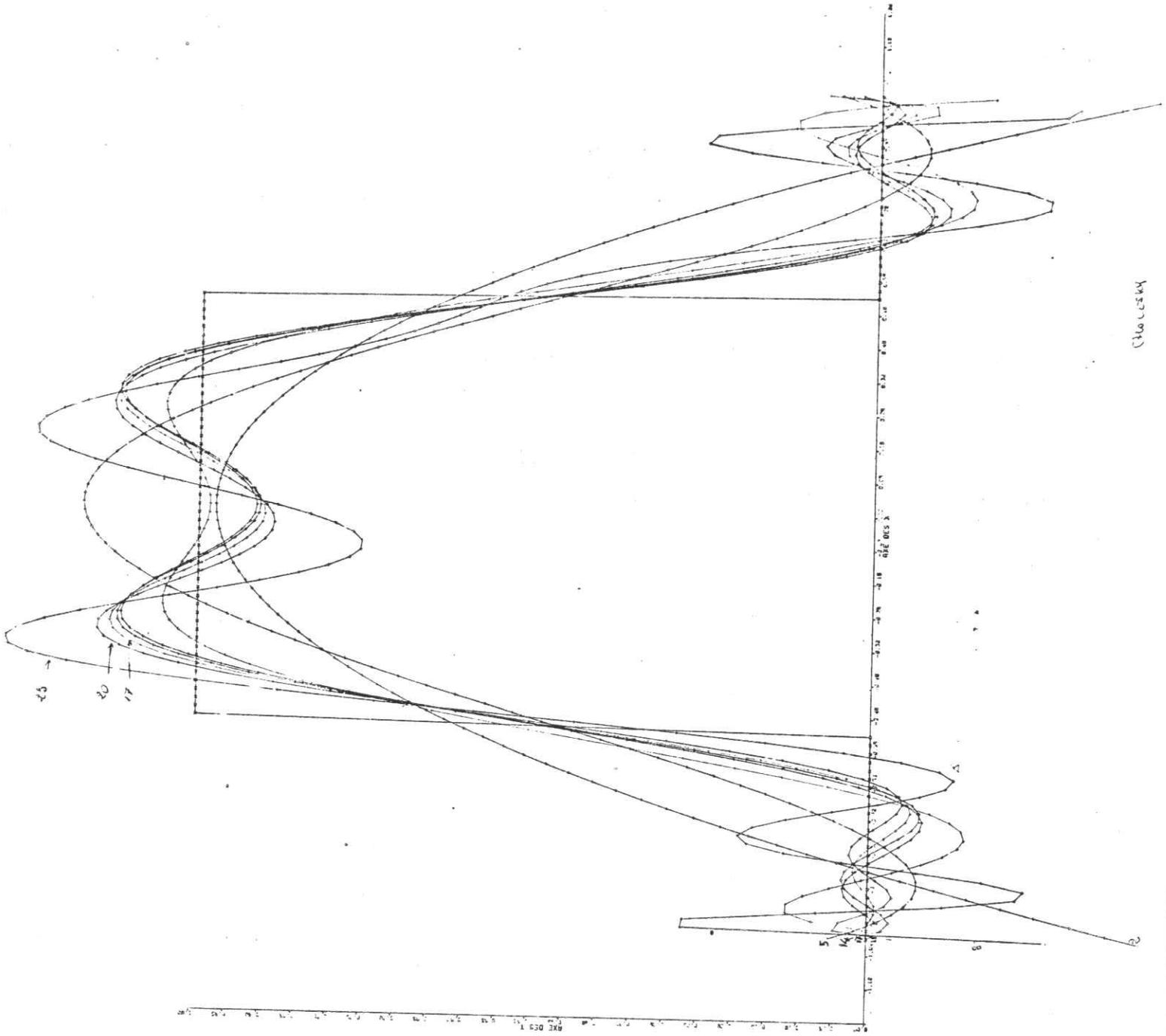


Figure 3

Résultats obtenus

Dans le cas de fonctions continues, on a très rapidement une bonne approximation.

Fonction indéfiniment dérivable

Exemple

$f(x) = \frac{1-x^2}{(1+x^2)^2}$, approximée par des polynômes de degrés 1, 2, 4, 8, 12 connaissant 101 points.

Pour le degré 12, l'approximation est presque confondue avec la courbe.

Fonctions

On a fourni le tracé pour 101 points et les degrés 1, 12, 40 (figure 4). Pour le degré 40, les résultats sont très proches de la courbe et le tracé de l'approximation est à peine distinct de celui de la fonction.

Fonction en escalier $y = \text{ent}(x)$

On a fourni le tracé pour les degrés 1, 12, 40 et 101 points (figure 5). Excepté au point de discontinuité, l'approximation oscille en restant proche de la courbe, voire confondue aux extrémités.

Conclusions

Les résultats sont "meilleurs" que ceux de la méthode précédente. Les polynômes étant orthogonaux, il n'y a plus de système à résoudre, d'où évidemment une très bonne stabilité des résultats.

Il est à prévoir cependant des problèmes numériques dans le cas de bornes d'intervalles $\gg 1$ et pour des degrés élevés.

PL/I OPTIMIZING COMPILER

APPROX:PROC OPTIONS(MAIN):

SIMT LEV NT

113	1	1	CALL PLOT(0,0,0,0-3);
114	1	1	END;
115	1	0	END APPROX;

PL/I OPTIMIZING COMPILER

APPROX:PROC OPTIONS(MAIN):

ATTRIBUTE AND CROSS-REFERENCE TABLE (FULL)

ATTRIBUTES AND REFERENCES

DCL NO.	IDENTIFIER	ATTRIBUTE AND REFERENCES
14	A	(*) CONTROLLED ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 67,67,67,73,75,78,81,107
1	APPROX	EXTERNAL ENTRY RETURNS(DECIMAL /* SINGLE */ FLOAT (6))
7	AXIS	EXTERNAL ENTRY(ALLOCATED DECIMAL /* SINGLE */ FLOAT (6)), ALIGNED DECIMAL /* SINGLE */ FLOAT (5), (4) ALIGNED BINARY FIXED (31,0), ALIGNED BINARY FIXED (31,0), ALIGNED DECIMAL /* SINGLE */ FLOAT (6), RETURNS(DECIMAL /* SINGLE */ FLOAT (6))
14	B	(*) CONTROLLED ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 68,68,68,79,81,108
17	E	AUTOMATIC ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 20,29,31
17	F	AUTOMATIC ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 21,30,31
9	FACTOR	EXTERNAL ENTRY(ALLOCATED DECIMAL /* SINGLE */ FLOAT (6)) RETURNS(DECIMAL /* SINGLE */ FLOAT (6))
17	H	AUTOMATIC ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 31,33

PL/I OPTIMIZING COMPILER

APPROX:PROC OPTIONS(MAIN):

ATTRIBUTES AND REFERENCES

DCL NO.	IDENTIFIER	ATTRIBUTES AND REFERENCES
18	I	AUTOMATIC ALIGNED DECIMAL FIXED (3,0) 35,36,38,39,44,44,48,48,48,48,74,74,74,75,75,98,98,99, 99,104,104,104,104
3	IBUF	(1000) AUTOMATIC ALIGNED BINARY FIXED (31,0) 53
18	J	AUTOMATIC ALIGNED DECIMAL FIXED (3,0) 77,77,77,78,78,78,79,79,81,81,81,81,81,81,81,90,90,90,91,91,91,93,93,93,94,94, 94
18	K	AUTOMATIC ALIGNED DECIMAL FIXED (3,0) 32,32,32,33,33,80,80,81,81,81,81,81
13	L	AUTOMATIC ALIGNED DECIMAL FIXED (3,0) 7
6	LINE	EXTERNAL ENTRY(*) ALIGNED DECIMAL /* SINGLE */ FLOAT (6), (*) ALIGNED DECIMAL /* SINGLE */ FLOAT (6); ALIGNED BINARY FIXED (31,0); ALIGNED BINARY RETURNS(BINARY FIXED (15,0))
11	N	AUTOMATIC ALIGNED BINARY FIXED (31,0) 63,63,63,63,65,67,67,68,68,69,69,77,85,87,90,93 51,57
12	NK	(4) AUTOMATIC ALIGNED BINARY FIXED (31,0) 52,58
12	NY	(4) AUTOMATIC ALIGNED BINARY FIXED (31,0) 23,25,26,27,30,31,32,35,38,38,44,44,48,48,55,56,57,57,58,58,59,61,61,62,62, 69,69,71,72,73,73,74,80,96,98,101,101,102,102,104,104,112
11	P	AUTOMATIC ALIGNED BINARY FIXED (31,0) 23,25,26,27,30,31,32,35,38,38,44,44,48,48,55,56,57,57,58,58,59,61,61,62,62, 69,69,71,72,73,73,74,80,96,98,101,101,102,102,104,104,112
PL/I OPTIMIZING COMPILER		
APPROX:PROC OPTIONS(MAIN):		
ATTRIBUTES AND REFERENCES		
5	PLOT	EXTERNAL ENTRY(ALIGNED DECIMAL /* SINGLE */ FLOAT (6)) ALIGNED DECIMAL /* SINGLE */ FLOAT (6); ALIGNED BINARY FIXED (31,0); RETURNS(DECIMAL 60,113
4	PLOTS	EXTERNAL ENTRY(*) ALIGNED BINARY FIXED (31,0); ALIGNED BINARY FIXED (31,0) ; ALIGNED BINARY FIXED (31,0); RETURNS(DECIMAL /* SINGLE */ FLOAT (6)) 53
22	Q	(*,*) CONTROLLED ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 69,69,70,71,72,75,78,78,79,79,81,81,81,81,91,94,99
96	R	(*) CONTROLLED ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 87,87,88,91,94
84	S	(*) CONTROLLED ALIGNED DECIMAL /* DOUBLE */ FLOAT (16) 85,85,89,94,99,109
8	SCALE	EXTERNAL ENTRY(*) ALIGNED DECIMAL /* SINGLE */ FLOAT (6); ALIGNED DECIMAL /* SINGLE */ FLOAT (6); ALIGNED BINARY FIXED (31,0); ALIGNED BINARY FIXED 55,56
***** SUM		
***** SYSIN		
***** SYSPRINT		
15	T	BULLETIN 73,78,78,79,79,91,94,99 EXTERNAL FILE 20,21,23,50 EXTERNAL FILE PRINT 23,24,42,43,44,45,45,47,48,49,65,50,103,104,105,106,107,108,109 (*) CONTROLLED ALIGNED DECIMAL /* SINGLE */ FLOAT (6)

PL/I OPTIMIZING COMPILER APPROX:PROC OPTIONS(MAIN):

UCL NO. IDENTIFIER ATTRIBUTES AND REFERENCES

- 12 TTRX AUTOMATIC UNALIGNED CHARACTER (16)
- 50
- 12 TTRY AUTOMATIC UNALIGNED CHARACTER (16)
- 50
- 12 TIX (4) DEFINED UNALIGNED CHARACTER (4)
- 51
- 12 TIY (4) DEFINED UNALIGNED CHARACTER (4)
- 52

***** UNSPEC BULLETIN 51,52

- 16 X (*) CONTROLLED ALIGNED DECIMAL /* SINGLE */ FLOAT (6)
- 25,25,29,30,33,33,44,55,57,57,59,73,73,73,75,78,81,112
- 16 Y (*) CONTROLLED ALIGNED DECIMAL /* SINGLE */ FLOAT (6)
- 26,26,28,36,39,41,48,50,58,58,59,61,62,64,91,101,102,110,111,112
- 19 YY (*) CONTROLLED ALIGNED DECIMAL /* DOUBLE */ FLOAT (16)
- 27,27,41,61,62,64,110
- 14 Z (*) CONTROLLED ALIGNED DECIMAL /* DOUBLE */ FLOAT (16)
- 96,96,97,99,101,102,104,111

PL/I OPTIMIZING COMPILER APPROX:PROC OPTIONS(MAIN):

COMPILER DIAGNOSTIC MESSAGES

ERROR ID L STMT MESSAGE DESCRIPTION

COMPILER INFORMATORY MESSAGES

```

IELU5331 I NO *DECLARE* STATEMENT(S) FOR *SYSIN*, *SYSPRINT*, *UNSPEC*, *SUM*.
IELG9721 I *BUFFERS(1)* DEFAULTED FOR FILE *SYSRINT*. EXECUTION TIME CAN BE IMPROVED BY DECLARING
*BUFFERS(2)*.
IELG9721 I *BUFFERS(1)* DEFAULTED FOR FILE *SYSIN*. EXECUTION TIME CAN BE IMPROVED BY DECLARING
*BUFFERS(2)*.

```

END OF COMPILER DIAGNOSTIC MESSAGES

COMPILE TIME 2.05 MINS SPILL FILE: 44 RECORDS, SIZE 3491

PLIOPT CPU=C0143+IC-4H NSID=001910 #0013 F #

LNKEDT CPU=C0042*10-4H NSID=004905 #0018 F #

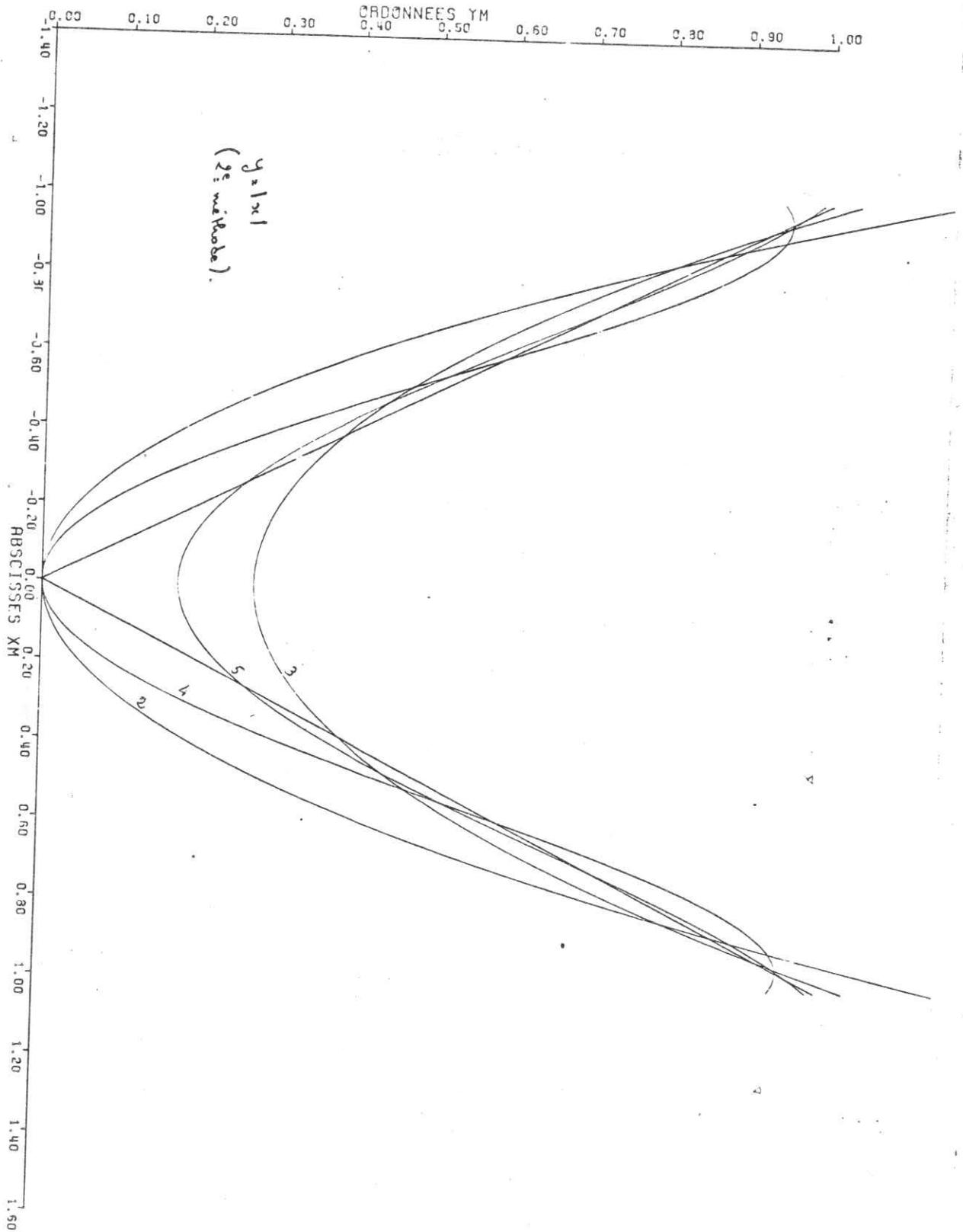


Figure 4 bis



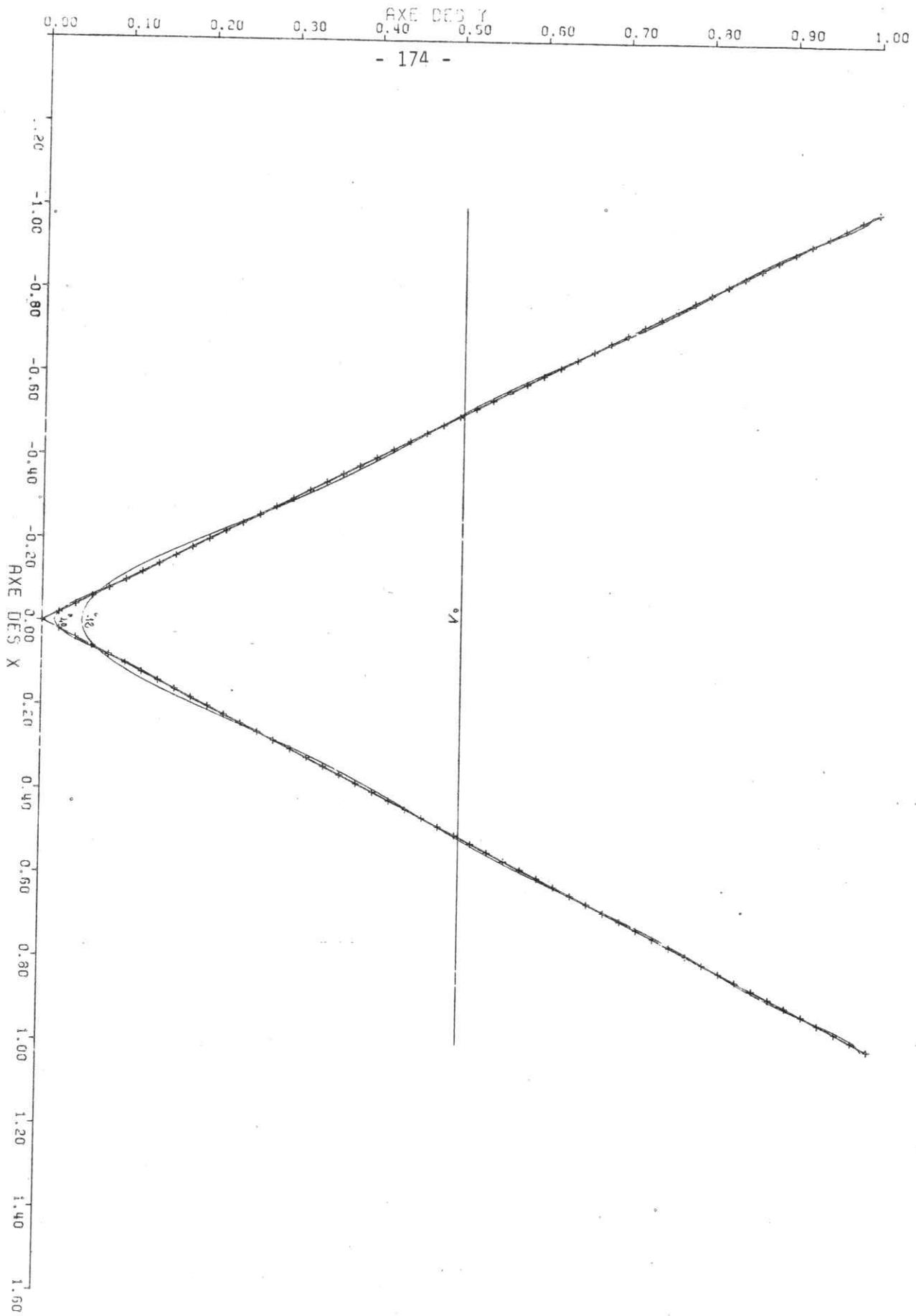


Figure 4

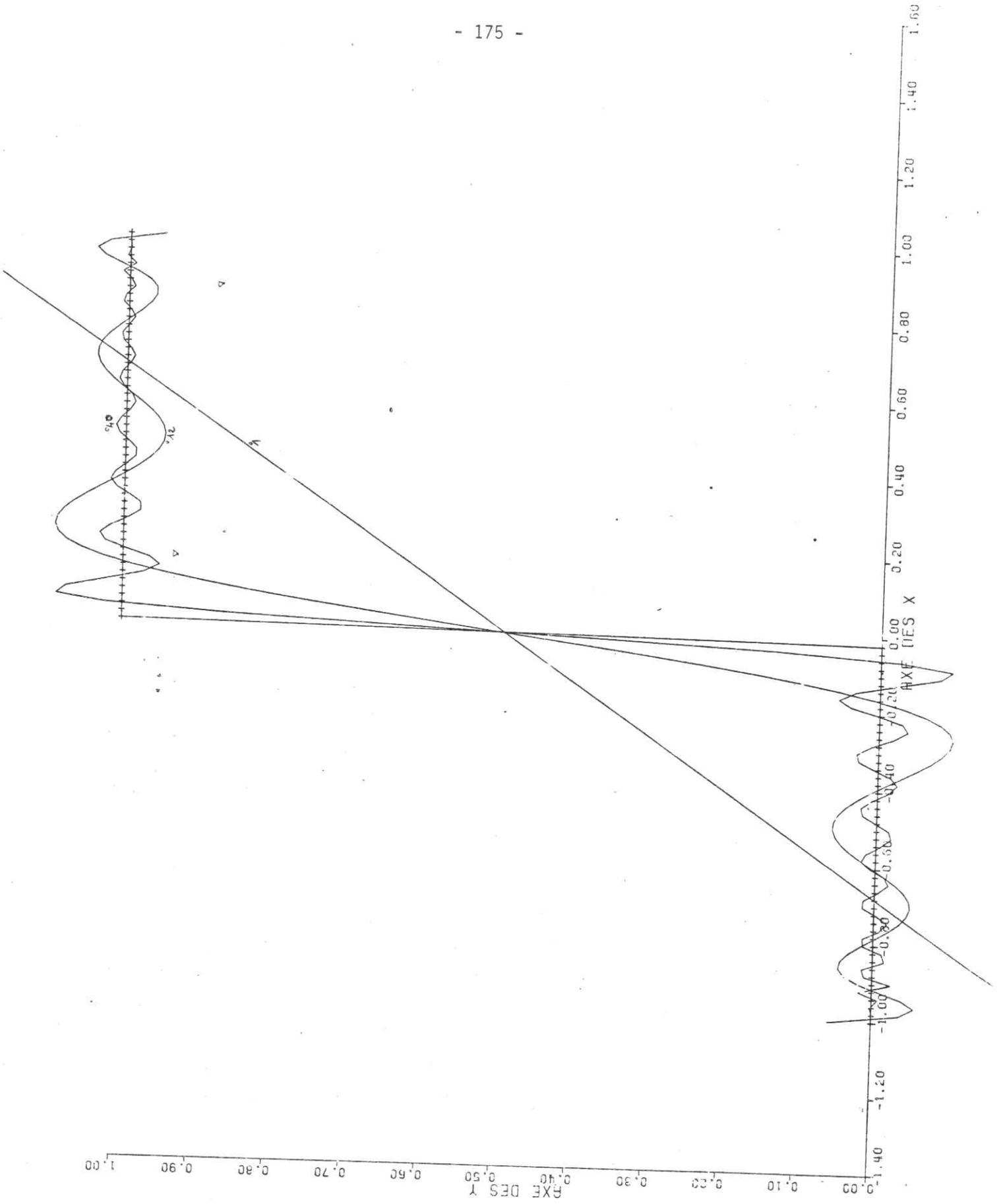


Figure 5

8-5 APPROXIMATION PAR LES POLYNOMES DE TCHEBYCHEFF

8-5-1 POLYNOMES DE TCHEBYCHEFF

Soit la fonction f connue par les points $(x_i; y_i = f(x_i))_{i=0 \text{ à } N}$, $(N+1)$ points donnés d'abscisses également réparties.

On cherche un polynôme de degré p : $\sum_{\alpha=0}^p C_\alpha T_\alpha(x)$ qui minimise $E = \sum_{i=0}^N \left[y_i - \sum_{\alpha=0}^p C_\alpha T_\alpha(x_i) \right]^2$ où les $(T_\alpha(x))_{\alpha=0 \text{ à } p}$ sont les $(p+1)$ premiers polynômes de Tchébycheff. Ces polynômes forment un système de polynômes orthogonaux sur $[-1, +1]$ pour la fonction de poids $u(x) = \sqrt{\frac{1}{1-x^2}}$ et sont donnés par $T(x) = \cos [\alpha \text{ Arc cos } (x)]$. Ils vérifient la relation de récurrence ($\alpha \geq 1$) :

$$T_{\alpha+1}(x) - 2x T_\alpha(x) + T_{\alpha-1}(x) = 0$$

avec $T_0(x) = 1$, $T_1(x) = x$, $T_2(x) = 2x^2 - 1$, et la propriété d'orthogonalité suivante (cas discret) :

$$\sum_{i=1}^{\alpha+1} T_k(\bar{x}_i) T_l(\bar{x}_i) = \begin{cases} 0 & \text{si } k \neq l \\ \frac{\alpha+1}{2} & \text{si } k = l \neq 0 \\ \alpha+1 & \text{si } k = l = 0 \end{cases}$$

avec $0 \leq k \leq \alpha$, $0 \leq l \leq \alpha$ et $\bar{x}_i = \cos \left[\frac{(2i-1)\pi}{2(\alpha+1)} \right]$, $i = 1, 2, \dots, (\alpha+1)$.

(Les \bar{x}_i sont les zéros du polynôme $T_{\alpha+1}(x)$ et sont encore appelés abscisses de Tchébycheff).

Comme dans le cas général, les $(p+1)$ inconnues C_j sont données par un système symétrique d'équations linéaires normales :

$$\begin{bmatrix} \sum_{i=1}^N T_0^2(x_i) & \dots & \sum_{i=1}^N T_0(x_i) T_p(x_i) \\ \sum_{i=1}^N T_1(x_i) T_0(x_i) & \dots & \vdots \\ \vdots & & \vdots \\ \sum_{i=1}^N T_p(x_i) T_0(x_i) & \dots & \sum_{i=1}^N T_p^2(x_i) \end{bmatrix} \times \begin{bmatrix} C_0 \\ \vdots \\ C_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i T_0(x_i) \\ \vdots \\ \sum_{i=1}^N y_i T_p(x_i) \end{bmatrix}$$

Deux procédés ont ensuite été utilisés.

8-5-2 PREMIERE METHODE : ABSCISSES QUELCONQUES

- 1) On ramène les abscisses (x_i) dans l'intervalle $[-1,+1]$ par une transformation linéaire.

Le système est ensuite résolu dans un sous-programme par la méthode de Cholesky et on obtient le polynôme :

$$\sum_{\alpha=0}^p C_{\alpha} T_{\alpha}(x) = \sum_{\alpha=0}^p d_{\alpha} x^{\alpha} .$$

On calcule les (d_{α}) à partir des (C_{α}) et des coefficients du polynôme $T_{\alpha}(x)$.

A l'aide du polynôme obtenu, on calcule les ordonnées d'environ 100 points, d'abscisses également réparties, ce qui donne la courbe d'approximation. Les écarts entre les ordonnées exactes et les ordonnées d'approximation donnent la courbe des écarts.

- 2) La méthode a été appliquée à différents types de fonctions : pour la fonction continue $f(x) = \frac{1-x^2}{(1+x^2)^2}$, une bonne approximation a été obtenue pour un nombre de points restreint avec un degré d'approximation faible (cf. courbe 15 points, degré 12). (Figure 6)

Fonction valeur absolue. (Figure 7)

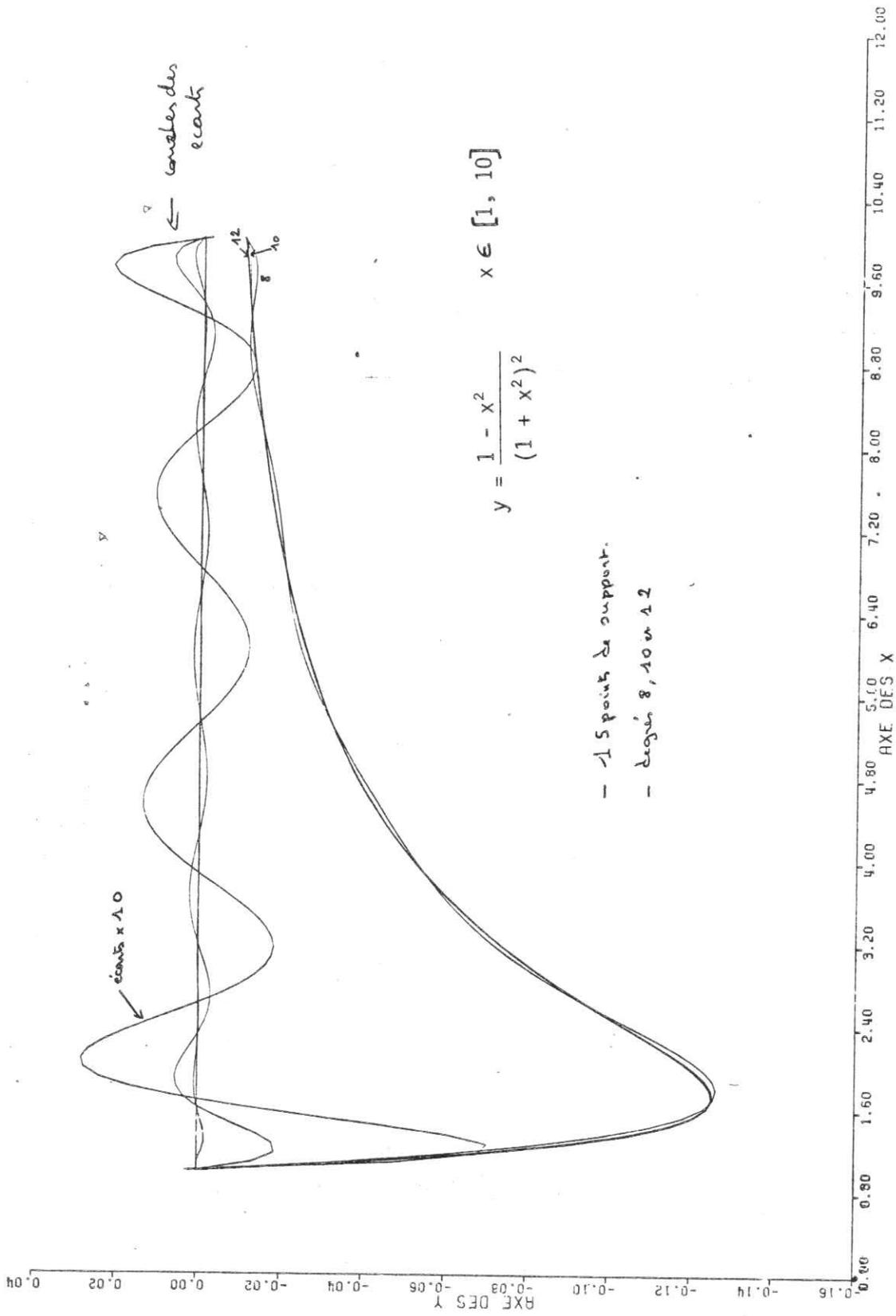


Figure 6

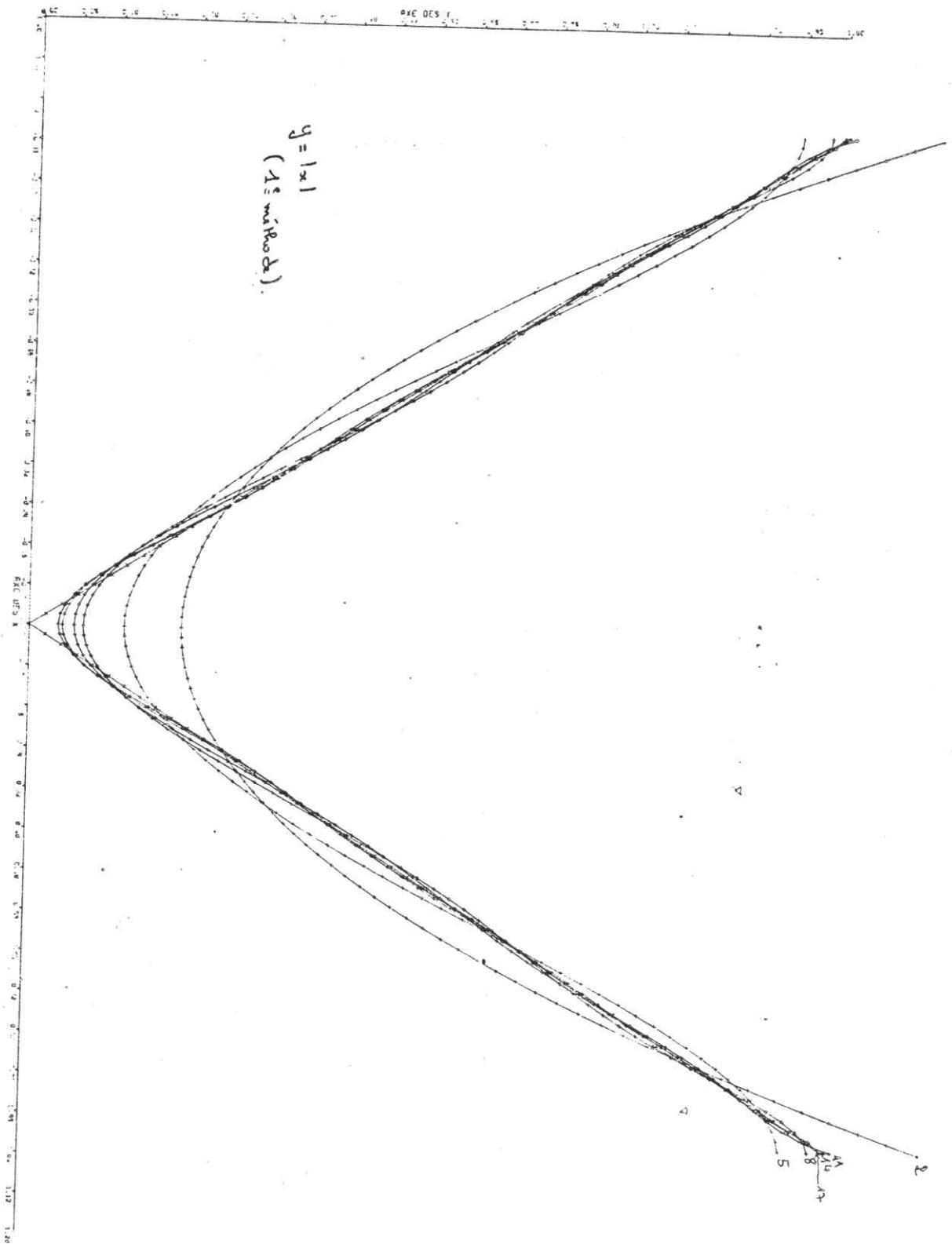


Figure 7

8-5-3 DEUXIEME METHODE : UTILISATION DES ABSCISSES DE TCHEBYCHEFF

1) Le système est transformé en un système diagonal :

$$\begin{bmatrix} \sum_{i=1}^N T_0^2(\bar{x}_i) & & \\ & \dots & \\ & & \sum_{i=1}^N T_p^2(\bar{x}_i) \end{bmatrix} \times \begin{bmatrix} C_0 \\ \vdots \\ C_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \bar{y}_i T_0(\bar{x}_i) \\ \vdots \\ \sum_{i=1}^N \bar{y}_i T_p(\bar{x}_i) \end{bmatrix}$$

les (\bar{x}_i) étant calculés par $\bar{x}_i = \cos \frac{(2i-1)\pi}{2p}$, pour $i = 1$ à $p+1$, soit $(p+1)$ points.

On obtient les résultats suivants :

$$C_0 = \frac{1}{p+1} \sum_{i=1}^{p+1} \bar{y}_i T_0(\bar{x}_i)$$

et

$$C_j = \frac{2}{p+1} \sum_{i=1}^{p+1} \bar{y}_i T_j(\bar{x}_i), \quad j = 1 \text{ à } p.$$

Les ordonnées d'approximation sont ensuite données par les relations :

$$y_i = b_0 - \bar{x}_i b_1, \quad i = 1 \text{ à } p+1$$

avec

$$b_j = C_j + 2 \bar{x}_i b_{j+1} - b_{j+2}, \quad j = p, p-1, \dots, 0$$

et

$$b_{p+1} = b_{p+2} = 0.$$

2) Cette méthode a été elle aussi appliquée à plusieurs types de fonctions :

- dans le cas $f(x) = |x|$, les courbes obtenues ont la particularité de passer par le point anguleux pour les degrés pairs ;
- dans le cas $f(x) = |x^2 - 0,25|$, on remarque une très bonne approximation pour le degré 60. (Figure 8)
- dans le cas d'une fonction en escalier figure 9.

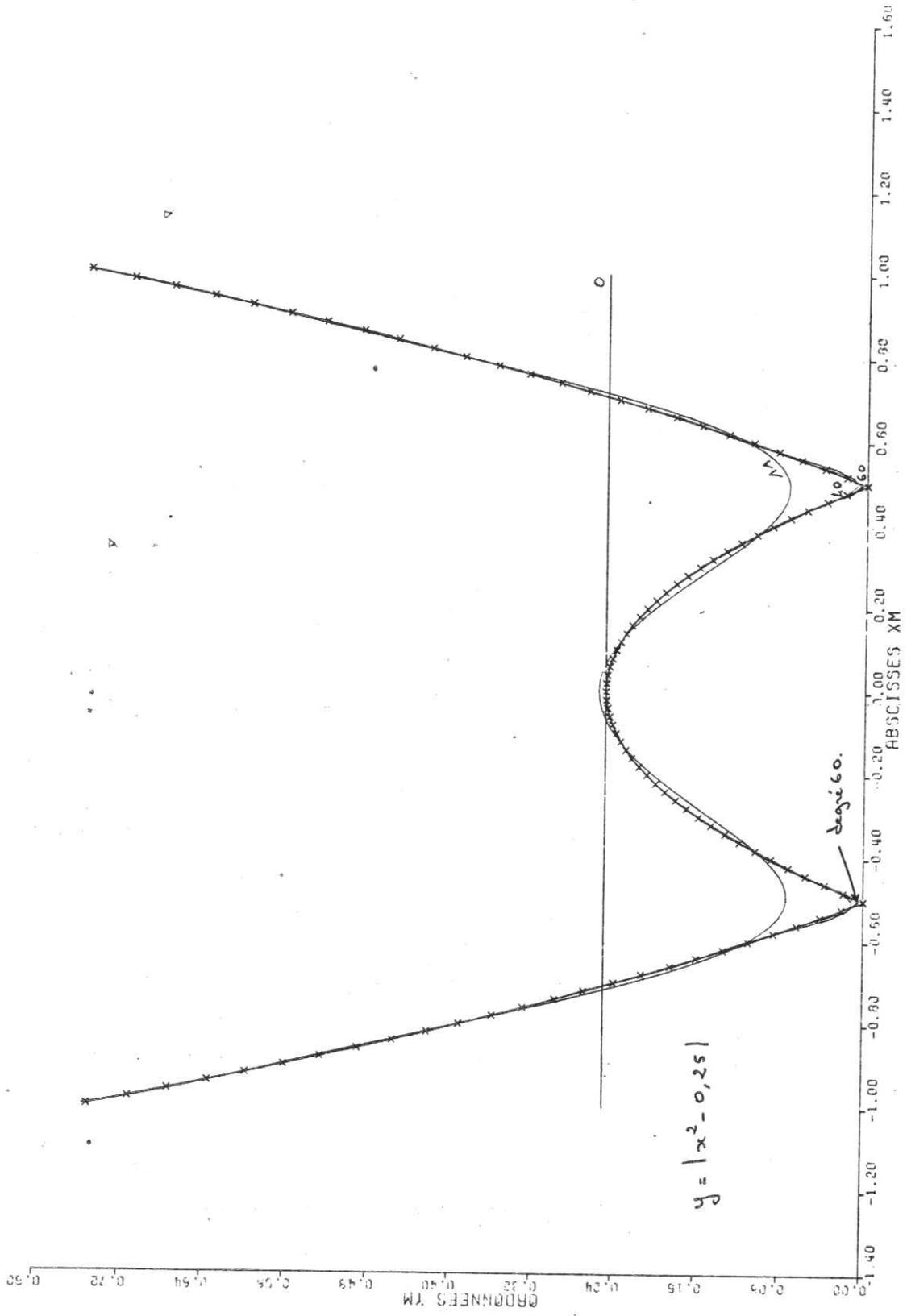


Figure 8

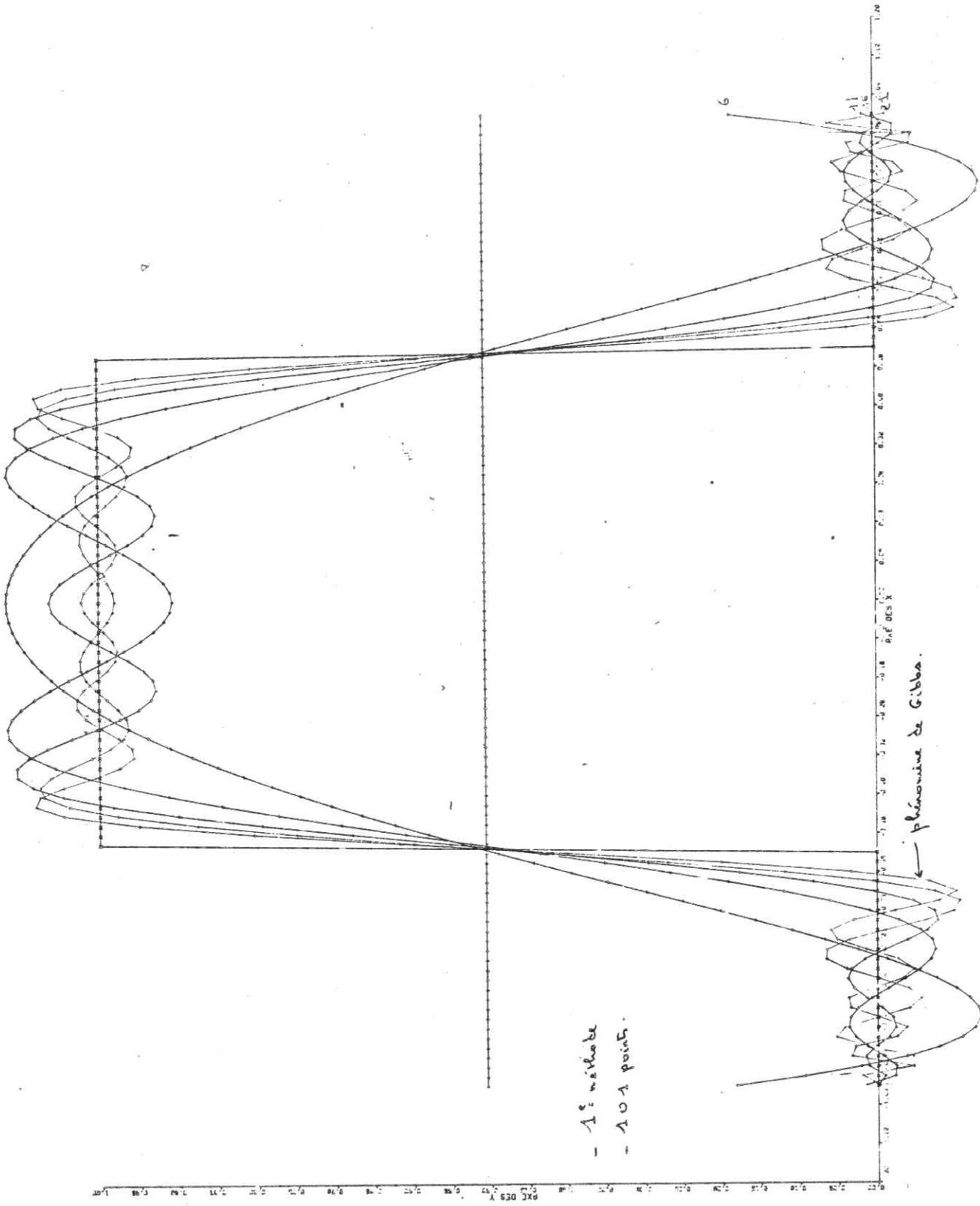


Figure 9

8-5-4 CONCLUSIONS

Pour les fonctions en escalier, on retrouve le phénomène de Gibbs dans les deux méthodes.

On a constaté que la 1ère méthode n'est plus stable pour des degrés supérieurs à 17 environ (cela est dû à la méthode de Cholesky) contrairement à la seconde.

Les deux méthodes sont nettement différenciées par le fait que la 1ère permet de faire varier indépendamment le nombre de points de support et le degré d'approximation, alors que dans le second cas on ne peut choisir qu'un de ces 2 paramètres.

APPENDICE
ET PAN DANS LE MILLE*

A titre d'illustration des deux premiers chapitres, nous reproduisons ici un exemple géométrique.

* Reproduction d'un texte paru dans le Bulletin de l'Association des Professeurs de Mathématiques (Avril 1976).

A propos de l'un de ces stands de tir où l'on gagne de grosses poupées roses entre des effluves de frites ou de gaufres, je ne parlerai pas de mon score, pour éviter de susciter crainte ou espoir de me provoquer, mais de la remarque du bonimenteur. "Le tir était remarquablement réparti".

Qu'est-ce à dire ?

Partons donc de n points, M_1, M_2, \dots, M_n , du plan euclidien, repérés par leurs affixes complexes, disons z_1, z_2, \dots, z_n . Une première idée est de considérer l'isobarycentre G , à savoir le point d'affixe z_G défini par

$$(1) \quad z_G = \frac{z_1 + \dots + z_n}{n}$$

Une deuxième idée consiste à envisager s'il existe un point M dont l'affixe z_M réalise un certain minimum, à savoir par exemple

$$(2) \quad \sum_{i=1}^n |z_M - z_i| = \inf_{z \in \mathbb{C}} \left(\sum_{i=1}^n |z - z_i| \right)$$

Une troisième idée consiste à chercher s'il existe un point C , centre d'un disque de rayon minimal et contenant les n points M_1, M_2, \dots, M_n . Analytiquement, nous pouvons encore traduire ceci par un certain minimum. Pour tout z de \mathbb{C} on définit :

$$d(z) = \sup_{i=1, \dots, n} |z - z_i|$$

et $\delta = \inf_{z \in \mathbb{C}} d(z)$ où la borne inférieure est prise sur tout le plan complexe.

Le point C d'affixe z_C devrait alors satisfaire

$$(3) \quad d(z_C) = \delta$$

Par ces quelques notes, nous allons vagabonder autour de ces trois idées, les rattacher à des notions connues, imaginer des généralisations, bref

baguenauder. Le fond des démonstrations restera très simple et si le jargon mathématique non redéfini paraît pédant à certains, ils pourront vérifier que ce jargon n'intervient qu'en vue d'énoncer des généralisations pour faire le pont avec les théories peut-être déjà étudiées par le lecteur au cours de sa carrière universitaire.

Cas du barycentre

L'existence et l'unicité de z_G sont bien évidentes. Une généralisation naturelle conduit à associer des masses α_i ($i = 1, \dots, n$) aux points z_1, z_2, \dots, z_n de sorte que

$$\alpha_i \geq 0 \quad \text{et} \quad \sum_{i=1}^n \alpha_i = 1$$

On considère alors le barycentre défini par l'affixe :

$$\sum_{i=1}^n \alpha_i z_i$$

et chacun sait que l'ensemble des barycentres ainsi obtenu coïncide avec le plus petit ensemble convexe contenant les n points z_1, z_2, \dots, z_n . Un pas de plus dans la généralisation consiste à envisager non plus une famille finie de points, mais un ensemble E plus riche, un continu par exemple. La théorie de l'intégration par rapport à une mesure permet d'envisager des expressions du type

$$(1)\text{bis} \quad z_G = \int_E z \, d\alpha(z)$$

où α est une mesure de Radon positive telle que $\int_E d\alpha(z) = 1$. A ce niveau,

E peut ne plus être un sous-ensemble du plan complexe mais un sous-ensemble universellement mesurable d'un espace \mathbb{C}^n ou \mathbb{R}^n . On démontre encore que z_G appartient au plus petit convexe fermé contenant tous les points de E et que réciproquement tout point de ce dernier ensemble peut être obtenu comme un z_G . La cinématique, avec les centres de gravité, ou les probabilités avec les moments, nous ont familiarisés avec des expressions du type (1)bis.

Cas du point M

L'existence d'un point M n'est guère difficile à établir puisque la fonction définie sur \mathbb{C} par

$$z \longmapsto \sum_{i=1}^n |z - z_i| = f(z)$$

est continue : on a même l'inégalité déduite de l'inégalité triangulaire

$$|f(z) - f(z')| \leq n |z - z'|$$

En outre

$$\lim_{|z| \rightarrow +\infty} f(z) = +\infty$$

Donc la borne inférieure $\inf_{z \in \mathbb{C}} f(z)$ est égale à la borne inférieure

$\inf_{|z| \leq R} f(z)$ pour un R assez grand. Mais comme le disque fermé de centre 0

et de rayon R est compact, la fonction continue f atteint son minimum en au moins un point z_0 de ce disque

$$\inf_{z \in \mathbb{C}} f(z) = f(z_0)$$

L'unicité du point M exige une analyse un peu plus étoffée. Appelons K le sous-ensemble non vide des points de \mathbb{C} où f atteint son minimum. C'est évidemment un sous-ensemble fermé et borné de \mathbb{C} donc compact. Plus intéressante est la propriété de convexité de K :

Si Z et Z' sont dans K, $\lambda Z + (1-\lambda)Z'$ est encore dans K et ce pour tout λ satisfaisant $0 \leq \lambda \leq 1$. Géométriquement, ceci signifie que le segment joignant Z à Z' appartient tout entier à K pour Z et Z' dans K.

En effet :

$$f(\lambda Z + (1-\lambda)Z') = \sum_{i=1}^n |\lambda(Z - z_i) + (1-\lambda)(Z' - z_i)|$$

Soit :

$$\begin{aligned}
 f(\lambda Z + (1-\lambda)Z') &\leq \lambda \sum_{i=1}^n |Z-z_i| + (1-\lambda) \sum_{i=1}^n |Z'-z_i| \\
 &\leq \lambda f(Z) + (1-\lambda) f(Z') \\
 &\leq (\lambda+1-\lambda) f(z_0) = f(z_0)
 \end{aligned}$$

L'inégalité stricte ne pouvant avoir lieu, toutes les inégalités précédentes deviennent des égalités. En particulier, pour tout i variant entre 1 et n , on a

$$(4) \quad |\lambda(Z-z_i) + (1-\lambda)(Z'-z_i)| = |\lambda(Z-z_i)| + |(1-\lambda)(Z'-z_i)|$$

Par suite $(Z-z_i)$ et $(Z'-z_i)$ sont sur une même demi-droite, c'est-à-dire qu'il existe des nombres α_i, β_i positifs ou nuls et $\alpha_i(Z-z_i) = \beta_i(Z'-z_i)$ avec $\alpha_i + \beta_i > 0$

Soit $\alpha_i Z - \beta_i Z' = (\alpha_i - \beta_i) z_i$

S'il existe un indice i pour lequel $\alpha_i = \beta_i$ on déduit aussitôt $Z = Z'$. On peut alors supposer que $\alpha_i \neq \beta_i$ pour tout $i = 1, 2, \dots, n$, auquel cas tous les points z_i doivent être situés sur la droite passant par Z et Z' ; on est ramené à un problème rectiligne analogue et on pourra supposer $Z-z_i$ et $Z'-z_i$ réels. On note alors que dans ce cas $Z-z_i$ et $Z'-z_i$ ont même signe, c'est-à-dire qu'il existe des $\varepsilon_i (\varepsilon_i = \pm 1)$ pour lesquels

$$|Z-z_i| = \varepsilon_i (Z-z_i) \quad ; \quad |Z'-z_i| = \varepsilon_i (Z'-z_i)$$

d'où

$$\sum_{i=1}^n \varepsilon_i (Z-z_i) = \sum_{i=1}^n \varepsilon_i (Z'-z_i)$$

soit $(Z-Z') \left(\sum_{i=1}^n \varepsilon_i \right) = 0$

Si n est impair, $\sum_{i=1}^n \varepsilon_i$ est toujours différent de 0 et par suite

$Z = Z'$. Par contre, si n est pair, on ne peut rien conclure. D'ailleurs en prenant $z_1 = 0$ et $z_2 = 1$, on constate que tous les points situés entre 0 et 1 réalisent le minimum de $|x| + |1-x|$. On peut alors énoncer l'amusant théorème : Soient z_1, z_2, \dots, z_n , n points du plan complexe. Supposons que ces points ne sont pas alignés ou que n soit impair. Il existe alors un unique point M d'affixe z_M tel que

$$\sum_{i=1}^n |z_M - z_i| = \inf_{z \in \mathbb{C}} \left(\sum_{i=1}^n |z - z_i| \right)$$

On peut généraliser ce dernier cas en supposant que z_1, z_2, \dots, z_n sont n points d'un espace normé X sur le corps des réels ou des complexes et en cherchant z_M tel que

(2)bis

$$\sum_{i=1}^n \|z_M - z_i\| = \inf_{z \in X} \left(\sum_{i=1}^n \|z - z_i\| \right)$$

Bien que des résultats nettement plus généraux soient accessibles, nous nous contenterons de supposer X préhilbertien, c'est-à-dire de supposer qu'il existe un produit scalaire \langle, \rangle sur X de sorte que pour tout z de X

$$\|z\|^2 = \langle z, z \rangle \quad (\text{cf E.A.F.})$$

L'unicité de z_M , sous les mêmes conditions, suit la même démonstration que précédemment grâce à la remarque.

(5) $\|x+y\| = \|x\| + \|y\|$ si et seulement s'il existe des coefficients α, β positifs ou nuls tels que $(\alpha+\beta > 0)$ et $\alpha x = \beta y$

En effet, si l'égalité (5) est vraie, on doit avoir en élevant au carré⁽¹⁾

$$\operatorname{Re} \langle x, y \rangle = \|x\| \|y\|$$

Dont on déduit, puisqu'en général $\operatorname{Re} \langle x, y \rangle \leq |\langle x, y \rangle| \leq \|x\| \|y\|$, que $\langle x, y \rangle = \|x\| \|y\|$

(1) "Re" signifiant "Partie réelle de".

et donc facilement le résultat en étudiant la forme quadratique

$$\lambda \longrightarrow \lambda^2 \|y\|^2 + 2\lambda \langle x, y \rangle + \|x\|^2 = \|x + \lambda y\|^2 \geq 0$$

(dans le cas où E est pris sur le corps des réels pour fixer les idées).

L'existence de z_M requiert une astuce puisque la boule unité de X n'a plus de propriété de compacité. Cependant soit z un point n'appartenant pas au sous-espace vectoriel E engendré par les n points z_1, z_2, \dots, z_n de X . On sait qu'il existe un point z_p , projection orthogonale de z sur E et tel que pour tout z' de E on ait grâce au théorème de Pythagore

$$\|z - z_p\|^2 + \|z_p - z'\|^2 = \|z - z'\|^2$$

Donc l'inégalité

$$\|z_p - z'\| \leq \|z - z'\| \quad \text{pour tout } z' \text{ de } E.$$

D'où en particulier

$$\sum_{i=1}^n \|z_p - z_i\| \leq \sum_{i=1}^n \|z - z_i\|$$

Par suite

$$\inf_{z \in X} \left(\sum_{i=1}^n \|z - z_i\| \right) = \inf_{z \in E} \left(\sum_{i=1}^n \|z - z_i\| \right)$$

Comme E est de dimension finie, l'argument de compacité réapparaît.

Une autre généralisation consiste à modifier la nature de la somme envisagée. On peut prendre par exemple

$$\left(\sum_{i=1}^n \|z - z_i\|^2 \right)^{1/2}$$

voire plus généralement toute expression où 2 est remplacé par un nombre réel p ($p \geq 1$)

$$d_p(z) = \left(\sum_{i=1}^n \|z - z_i\|^p \right)^{1/p}$$

Restons dans le plan complexe $X = \mathbb{C}$ pour simplifier. Le lecteur constatera que les démonstrations précédentes restent valables, mutatis mutandis, et fournissent un point unique $M^{(p)}$. On utilisera une inégalité dite de Minkowski.

$$(6) \quad \left(\sum_{i=1}^n |z_i + z'_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^n |z_i|^p \right)^{1/p} + \left(\sum_{i=1}^n |z'_i|^p \right)^{1/p}$$

L'égalité ayant lieu, si, et seulement si $\alpha z_i = \beta z'_i$ pour $\alpha \geq 0, \beta \geq 0$ et $\alpha + \beta > 0$ lorsque $p > 1$. Pour $p = 2$, le point $M(2)$ est l'isobarycentre des n points M_1, M_2, \dots, M_n .

En faisant tendre p vers l'infini, on montre sans peine que l'expression $d_p(z)$ tend vers

$$d(z) = \sup_{i=1, \dots, n} |z - z_i| \quad \text{ce qui nous amène au dernier cas.}$$

Cas du point \mathbb{C}

Appelons A l'ensemble constitué par les points z_1, z_2, \dots, z_n et notons

$$d(z) = \sup_{z' \in A} |z - z'|$$

Il est bien clair, puisque A est un ensemble borné, que $d(z)$ est un nombre fini pour tout z de \mathbb{C} . L'existence d'un point z_c tel que

$$d(z_c) = \inf_{z \in \mathbb{C}} d(z) = \delta$$

peut se démontrer comme pour l'existence du point M par un argument de continuité sur un compact. Toutefois, je préfère une analyse plus percutante qui va fournir simultanément l'existence et l'unicité de \mathbb{C} .

D'abord, partons d'une remarque de calcul bien facile

(7) Si z et z' sont deux nombres complexes de module inférieur ou égal à 1, on a l'inégalité

$$\left| \frac{z + z'}{2} \right|^2 \leq 1 - \left| \frac{z - z'}{2} \right|^2$$

Pour chaque entier $n \geq 1$ puisque, δ est une borne inférieure, il existe un nombre complexe noté $z^{(n)}$ tel que

$$d(z^{(n)}) \leq \delta + \frac{1}{n}$$

Montrons que la suite $\{z^{(n)}\}_{n \geq 1}$ est une suite de Cauchy de nombres complexes : c'est-à-dire que nous avons l'implication suivante :

Pour tout $\varepsilon > 0$, il existe un entier N et pour tous entiers n, m supérieurs ou égaux à N , on a :

$$|z^{(n)} - z^{(m)}| < \varepsilon$$

Raisonnons par l'absurde en supposant que la suite ne soit pas de Cauchy. Par suite, il existe un nombre $\varepsilon > 0$ et pour toute valeur entière de N , on peut trouver deux entiers n et m qui dépendent de N , ($n \geq N$; $m \geq N$), de sorte que

$$|z^{(n)} - z^{(m)}| \geq \varepsilon.$$

Nous choisirons un N convenable un peu plus tard.

Dès lors, on remarque que pour tout z' (de A) on a la minoration

$$|(z^{(n)} - z') - (z^{(m)} - z')| \geq \varepsilon$$

Donc en posant $\delta_{n,m}(z') = \max(|z^{(n)} - z'|, |z^{(m)} - z'|)$, $\leq \delta + 1/n + 1/m$

on minore selon

$$\left| \frac{z^{(n)} - z'}{\delta_{n,m}(z')} - \frac{z^{(m)} - z'}{\delta_{n,m}(z')} \right| \geq \frac{\varepsilon}{\delta+2}$$

Par suite en utilisant (7)

$$\left| \frac{z^{(n)} + z^{(m)}}{2} - z' \right|^2 \leq 1 - \left(\frac{\varepsilon^2}{4(\delta+2)^2} \right) \delta_{n,m}^2(z')$$

Ce que l'on majore encore par

$$\left| \frac{z^{(n)} + z^{(m)}}{2} - z' \right|^2 \leq \left(1 - \frac{\varepsilon^2}{4(\delta+2)^2} \right) \left(\delta + \frac{1}{n} + \frac{1}{m} \right)^2 \leq \left(\delta + \frac{2}{N} \right)^2 \left(1 - \frac{\varepsilon^2}{4(\delta+2)^2} \right)$$

En prenant N assez grand, l'expression du second membre est strictement inférieure à δ^2 , ce qui fournit une contradiction en passant à la borne supérieure au premier membre par rapport à z' .

$$d\left(\frac{z^{(n)} + z^{(m)}}{2}\right) < \delta$$

La suite $\{z^{(n)}\}_{n \geq 1}$ est donc de Cauchy dans \mathbb{C} . Elle converge vers un z_c puisque \mathbb{C} est un espace complet. On doit alors voir

$$d(z_c) = \lim_{n \rightarrow \infty} d(z^{(n)}) = \delta$$

Ce qui termine la démonstration de l'existence. L'unicité est évidente car si z_c et z'_c conviennent, la suite alternativement égale à z_c et z'_c doit être de Cauchy donc converge. Donc $z_c = z'_c$. Cédons une fois encore à la tentation de généraliser. Il est clair que le même raisonnement convient pour tout sous-ensemble borné A et pour un espace normé X qui possède une propriété du type (7) et soit complet. On posera

$$d(z) = \sup_{z' \in A} \|z - z'\|$$

$$3(\text{bis}) \quad \text{et} \quad \delta = \inf_{z \in X} d(z)$$

La bonne généralisation de la propriété (7) s'appelle uniforme convexité de X elle s'énonce.

(7)bis Il existe une fonction croissante $\delta :]0, 2] \rightarrow]0, 1]$ telle que pour tout couple (z, z') de points distincts de X avec

$$\|z\| \leq 1 \quad \text{et} \quad \|z'\| \leq 1 \quad \text{on ait}$$

$$\left\| \frac{z + z'}{2} \right\| \leq 1 - \delta(\|z - z'\|)$$

Les espaces préhilbertiens sont uniformément convexes comme on le constate sans peine en utilisant l'égalité du parallélogramme (d'ailleurs caractéristique de ces espaces) :

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

C'est bien autour des propriétés de convexité que nous tournons depuis le début de cet exposé. Ainsi, la propriété (4) cruciale pour la démonstration de l'unicité de M et d'ailleurs mise sous forme plus générale en (5), peut s'énoncer sous une forme équivalente :

(5)bis Si $\|x\| \leq 1$ et $\|y\| \leq 1$ alors $\left\| \frac{x+y}{2} \right\| = 1$ implique $x = y$

Géométriquement cette propriété signifie que le milieu de deux points distincts d'une boule d'un espace normé X ne peut être sur la sphère, frontière de cette boule. Un espace normé possédant une telle propriété est dit strictement convexe. C'est cette propriété qu'il faut évoquer pour le cas d'égalité dans (6) en établissant que pour $p > 1$ l'espace \mathbb{F}_n^p est strictement convexe. (Rappelons que \mathbb{F}_n^p est l'espace \mathbb{C}^n muni de la norme

$\|z\|_p$ définie par

$$\|z\|_p = (|z_1|^p + \dots + |z_n|^p)^{1/p}.$$

La propriété (7)bis améliore la propriété de stricte convexité, par suite un espace uniformément convexe est strictement convexe.

Il est assez instructif de spécialiser plus ou moins les propriétés de l'espace géométrique euclidien pour effectuer de la géométrie en dimension infinie comme nous venons de le faire. On ne saurait toutefois terminer cet exposé sans des remarques de géométrie du triangle.

Un exemple

Sur l'exemple de trois points M_1, M_2 et M_3 formant un triangle non dégénéré du plan euclidien, nous allons interpréter géométriquement les points $G, M^{(p)}$ et C . Notons z_1, z_2 et z_3 les affixes de M_1, M_2 et M_3 .

G est le centre de gravité du triangle, point de rencontre des médianes du triangle.

$M^{(1)}$ est le point de Fermat. C'est l'unique point M , d'affixe z , réalisant le minimum de

$$|z-z_1| + |z-z_2| + |z-z_3|$$

Il n'est pas difficile d'établir, lorsque tous les angles du triangle sont aigus, que le point $M^{(1)}$ se trouve à l'intersection des 3 droites issues de chaque sommet du triangle et joignant le sommet opposé du triangle équilatéral construit sur le côté opposé (et extérieur).

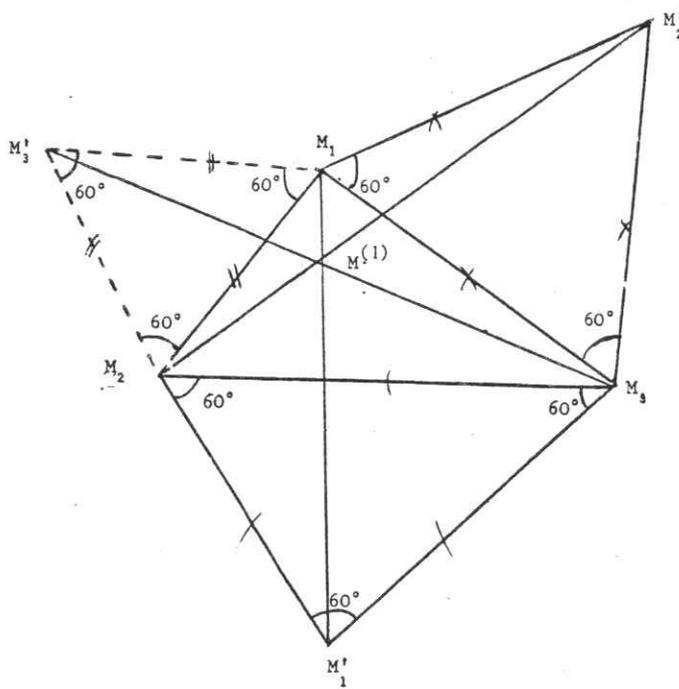


Figure 1 Construction du point de Fermat

$M^{(2)}$ C'est l'unique point M , d'affixe z , réalisant le minimum $\sqrt{|z-z_1|^2 + |z-z_2|^2 + |z-z_3|^2}$. Il est classique de vérifier qu'il s'agit du centre de gravité G . On a en effet pour tout M du plan

$$\|\vec{MM}_1\|^2 + \|\vec{MM}_2\|^2 + \|\vec{MM}_3\|^2 = 3 \|\vec{MG}\|^2 + \|\vec{GM}_1\|^2 + \|\vec{GM}_2\|^2 + \|\vec{GM}_3\|^2$$

Point C (ou encore $M^{(\infty)}$)

C'est le centre du disque de plus petit rayon contenant M_1 , M_2 et M_3 .

On remarque que le diamètre d'un tel disque est au moins égal au plus grand côté du triangle et que la circonférence d'un tel disque passe par au moins deux des trois points M_1, M_2, M_3 . On doit distinguer deux cas.

Si le sommet opposé au plus grand côté est obtus (ou droit), le cercle dont un diamètre est ce plus grand côté est le cercle cherché. Le point C est au milieu du plus grand côté. On note que dans le cas d'un angle droit, C est le centre du cercle circonscrit (cf. figure 2).

Si le sommet opposé au plus grand côté est aigu, et donc si tous les angles sont aigus, on constate sans peine que C est sur la médiatrice de chaque côté, donc coïncide avec le centre du cercle circonscrit.

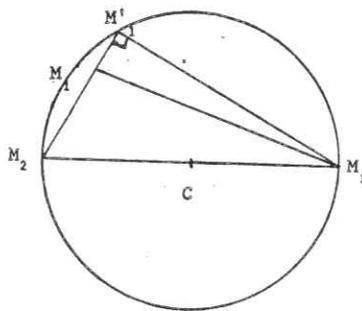


Figure 2 Construction de C : Cas où l'angle en M_1 est obtus

Point de Lemoine

Ce point réalise une sorte de dual du point $M^{(2)}$

On pose $\delta_{z_i, z_j}(z) = \inf_{\lambda \in \mathbb{R}} |z - (\lambda z_i + (1-\lambda)z_j)|^2$ et

$$\delta(z) = \delta_{z_1, z_2}(z) + \delta_{z_2, z_3}(z) + \delta_{z_3, z_1}(z)$$

Il existe un point qui minimise la somme $\delta(z)$ des carrés des distances aux trois côtés d'un triangle. Ce point, dit point de Lemoine, est à l'inter-

section des trois symédianes. Nous ne donnerons pas ici de démonstration. Rappelons que la symédiane issue d'un sommet d'un triangle est la symétrique de la médiane par rapport à la bissectrice. Je ne céderai pas à la tentation de généraliser la notion de point de Lemoine en diverses directions.

Conclusions

Revenons à la foire et au bonimenteur. Tout dépend de la notion de répartition choisie.

Celle de l'isobarycentre tient simultanément compte de la répartition en direction et en distance. On peut être piètre tireur et avoir le mille comme point central !

Celle du point $M^{(1)}$ ne tient compte que de la répartition en distance. Cela fait gagner des points.

Celle du point C me paraît celle de mon bonimenteur, car il utilisa une pièce de cent sous pour s'expliquer.

BIBLIOGRAPHIE

Nous ne donnons ici que les livres portant essentiellement sur la Théorie de l'Approximation, renvoyant à la bibliographie des "Eléments d'Analyse Fonctionnelle" pour ce qui concerne les généralités d'analyse. Bien que de développement récent, la théorie de l'approximation compte à son actif de très nombreux et excellents livres.

N.I. ACHIESER - Theory of approximation - New-York 1956, 307 pages
Traduit du russe (il existe une édition russe augmentée, 1965)

Ce livre permet une approche moderne, mais non trop abstraite, à de très nombreux problèmes et théorèmes d'approximation sur des espaces fonctionnels variés. L'approximation en moyenne quadratique est très détaillée.

E.W. CHENEY - Introduction to approximation theory , McGraw Hill 1966, 259 pages

Livre concis et clair, fournissant les théorèmes principaux de ce cours

Y. CHERRUAULT - Approximation d'opérateurs linéaires et applications - Dunod (Paris) 1968

P.P. KOROVKIN - Linear operators and approximation theory - Hindustan Publishing Corp. India Dehli (1960)

La traduction anglaise est défectueuse.

RALSTON & WIL F. Mathematical methods for digital computers, Vol. 1 & 2 - John Wiley & Sons (New-York)

J. RICE - The approximation of functions, Vol. I : linear theory - Vol. II : general theory - Addison-Wesley Reading (Ma) USA 1964

Traduction française : approximation linéaire des fonctions, Vol. I, Dunod 1969

Ouvrage introductif très clair.

Il existe un volume II, également traduit en français et qui traite de l'approximation non linéaire.

Il existe beaucoup d'autres ouvrages mathématiques. On peut citer notamment les ouvrages suivants, dont les bibliographies sont très fournies :

- LORE - The approximation of continuous functions by positive linear operators - Lecture Notes, Springer-Verlag N° 293 (1970)
- LORENTZ - Approximation of functions, Holt, Rinehart - New-York 1966, 188 pages
- Livre remarquable par son élégance et sa concision, mais de lecture un peu ardue pour un non mathématicien
- SINGER - Best approximation in normed linear spaces - Springer-Verlag 1970
- Traduit du roumain "Cea mai bună aproximare..." et possède des références intéressantes en russe
- P.L. BLATZER - 1°. Fourier analysis and approximation, Vol. I - R.J. NESSEL - Birkhäuser Verlag (533 pages)
- Concerne essentiellement l'approximation des fonctions périodiques et fournit des renseignements très riches sinon complets en la matière
- 2°. Semi-groups of operators and approximation - Springer-Verlag 1967 (318 pages)
- H.S. SHAPIRO - Smoothing and approximation of functions - Van Nostrand 1969 (134 pages)
- G. MEINARDUS - Approximation of functions : theory and numerical methods (traduit de l'allemand) - Springer Tracts Vol 13 (1967)
- A. SARD - Linear approximation Math. Surveys N° 9 Providence American Math. Soc. 1963 (544 pages)
- A.F. TIMAN - Theory of approximation of functions of a real variable - New-York, McMillan 1963 (631 pages)

Les références du symposium suivant peuvent être utiles :

Approximation of function : proceedings of the Symposium, General Motors Research Laboratories 1964 - Amsterdam Elsevier 1975 (220 pages)

Sur la théorie des polynômes orthogonaux , on peut consulter :

- G. SZEGÖ - Orthogonal polynomials - American Mathematical Soc. - Colloquium Publications Vol. 23, 3e Edition 1967

Sur les opérateurs d'interpolation :

J.G. DHOMBRES - Some convergence theorems in averaging theory -
Asian Institute of Technology (Bangkok)
Technical Note N° 34 (1971)

En ce qui concerne l'opérateur de convolution en général, et non pas en tant qu'objet de la théorie de l'approximation, on peut consulter principalement :

J. WIDDER - The convolution transform (traduit en français chez Dunod)

En ce qui concerne la théorie des fonctions splines, il existe une littérature déjà énorme. Citons parmi les plus accessibles, outre les ouvrages précédemment cités :

J.H. AHLBERG, - The theory of splines and their applications -
E.N. NILSON Math. in Science and Engineering, Vol. 13
J.L. WALSH Academic Press 1967 (284 pages)

T.N.E. GREVILLE - Theory and applications of spline functions -
(editor) Academic Press 1969

Il s'agit d'un recueil des contributions de divers auteurs à un séminaire de recherche tenu à Madison en 1968

I.J. SCHOENBERG - Cardinal interpolation and spline functions -
Journal of Approximation Theory 2 (1969)
pp 167-206

La théorie des fonctions splines est placée dans un cadre hilbertien, à la suite des travaux d'Attéia, dans le livre suivant qui combine les méthodes de l'analyse fonctionnelle et les méthodes d'approximation.

P.J. LAURENT - Approximation et Optimisation
Hermann 1972

Un autre ouvrage, prenant les choses à un niveau plus élémentaire, mais pénétrant, même à la fois analyse fonctionnelle et théorie de l'approximation.

I.P. NATANSON - Constructive theory of functions (En russe),
Gostekhizdat 1949
(Il existe une traduction anglaise.. et même une traduction chinoise fort bon marché).

INDEX DES EXPRESSIONS

Le premier numéro donne le numéro du chapitre, etc. Le sigle EAF signifie que la référence est à chercher dans le texte : Eléments d'Analyse Fonctionnelle. (Nanta Iremica, volume N° 7).

Abélien (groupe)	0.2.1	EAF
Adjoint (opérateur)	2.6	EAF
Affine (espace)	0.2.2	EAF
Algèbre	0.3.5	EAF
Algorithme de Polya	1.4	
Analyse harmonique	3.3	EAF et 2.5
Anneau	0.3.5	EAF
Application	0.1	EAF
Application inverse	0.1	EAF
Appolonius (égalité d'..)	1.A 4.4.9	EAF
Approximation en moyenne quadratique	2 (chapitre entier)	et EAF
Approximation uniforme	6 (chapitre entier)	
Associative (loi)	0.1	EAF
Auto-adjoint		EAF
Automorphisme	0.3.8	EAF
Banach (espace de..)	1.B 6.3	EAF
Base algébrique	0.2.4	EAF
Base hilbertienne	3.8	EAF
Beppo-Lévi (théorème de..)	1.B 6.7	EAF
Bernstein (théorème de..)	4.4 et exercice N° 41	
Bessel (inégalité de..)	3.8.3	EAF
Bidual algébrique	0.4	EAF
Bijective	0.1	EAF
Bilinéaire (forme)	1.4.4.2	EAF
Biunivoque	0	EAF
Borne d'une fonction	1.A 2.3	EAF
Borné (opérateur)	1.B 3	EAF
Borné (ensemble)	1.B 3	EAF
Bosse glissante (exemple de la..)	A.B 6.8	EAF
Boule unité d'espaces	2.6	EAF et 1.5

Carré intégrable (fonctions de..) 1.B 6.6.2 EAF
 Cauchy (critère de..) 1.B 6 EAF
 Cauchy (relations de..) 3.2
 Čebičev (polynômes de..) 1ère et 2e espèces 3.6 EAF
 Čebičev (système de..) 6.2
 Changement de base 0.3.9 EAF
 Christoffel (nombres de..) 7.1
 Classe d'équivalence 1.B 6.6.2 EAF
 Côtes (nombre de..) 7.1
 Codimension de l'opération linéaire P 0.3.4 EAF
 Coïncidences (théorème des..) 3.4.2 EAF
 Commutatif 0.2.1 EAF
 Complémentaire 0.1. EAF
 Complément orthogonal 3 EAF
 Complet (espace) 1.B 6 EAF
 Complétion d'un espace vectoriel normé 1.B 6.7 EAF
 Complétion d'un espace métrique 1.B. 6.6 EAF
 Condition de Haar 6.2
 Continue (fonction) 1.B 2 EAF
 Convexe (fonction) 1.B 6.10 et exercice N° 48 EAF
 Convexe (ensemble) EAF
 Convexité (inégalités) 1.6.10 EAF
 Convolution (opérateur de..) chap. 5 et exercice N° 47
 Coordonnées hilbertiennes 2.A 2 EAF
 Coercitif (opérateur..) 7.1

 Darboux-Christoffel 3.4.2 et exercices N° 35 & 38 EAF
 Décomposition d'un vecteur 0.2.4 EAF
 de la Vallée Poussin (théorème de...) 6.2
 Déterminant persymétrique 2.2
 Division d'un intervalle 3.1
 Dimension d'un espace vectoriel 0.2.5 EAF
 Distance 1.A 1 EAF
 Distributivité 0.2.1 EAF
 Dirichlet (noyau de..) 5.4
 Discrétisation 7.3
 Droite réelle 1.A 2 EAF
 Dual (espace) 0.4 EAF
 Dualité topologique 1.B 4 EAF
 Dualité (théorème de la..) 4 EAF

Ecart 1.A 1 EAF
 Élément neutre 0.1 EAF
 Endomorphisme 0.3.1 homomorphisme EAF
 Ensemble de mesure nulle 1.B. 6.7 EAF
 Ensemble dense 1.B 5.4 EAF
 Ensemble de Čebičev 6.2
 Equivalence des normes 1.B 5.3.3 EAF
 Espace affine 0.2.2 EAF
 Espace hermitien 1.A 4.4.5 EAF
 Espace dense 3.8 EAF
 Espace de Hilbert 1.B 6.5 EAF
 Espace préhilbertien 1.A 4.4.5 EAF
 Espace métrique 1.A 2 EAF
 Espace vectoriel 0.2 réel 0.2.1 complexe 0.2.1 EAF
 Espace normé 1.A 4 EAF
 Espace euclidien 1.A 4.4.5 EAF
 Espace hermitien 1.A 4.4.5 EAF
 Espace l^2 A.1 4.5.3 EAF
 Espace vectoriel normé complet 1.B 6 EAF
 Espace de Banach 1.B 6.5 EAF
 Espace L^2 avec poids 1.B 6.7.5 EAF
 Espace strictement convexe 1.3
 Espace de Hilbert séparable 3.8.2 EAF
 Espace des fonctions splines chap. 3
 Euclidien (espace) 1.A 4.4.5 EAF
 Exponentielle d'une matrice Exercice N° 15 EAF

 Faber (théorème de) 4.6
 Famille orthonormale 3.2.3 EAF
 Fermé (ensemble) 1.B 5.4 EAF
 Fermeture d'un ensemble 1.B.5.4 EAF
 Fejer (théorème de..) exercice N° 35
 Fonction génératrice 3.4.4 EAF
 Fonction en escalier 1.B 6.6.5 EAF
 Fonctions équivalentes 1.B 6.7.2 EAF
 Fonction intégrable au sens de Lebesgue 1.B 6.6.2 EAF
 Forme linéaire 0.3.6 EAF
 Forme bilinéaire 0.1.4.4.2 EAF
 Forme diagonale d'un opérateur 2 EAF

Galerkine (Méthode de..) 7.2
 Gauss-Jacobi (méthode de..) 7.1
 Gram (déterminant de..) exercices N° 19, 21 & 27
 Groupe 02.1 EAF

Haar (condition de..) 6.2
 Hahn-Banach (théorème de..) 1.6
 Hermite (polynôme d'..) 3.6 EAF
 Hermitien (espace) 1.A 4.4.5 EAF
 Hilbert (espace de..) 1.B 6.5 EAF
 Hirschfeld (théorème de..) 2.1 et exercice N° 11
 Homomorphisme 0.3 EAF
 Hyperplan 0.4 EAF

Idempotent 3.1 EAF
 Image d'un opérateur 0.3.2 EAF
 Image inverse 0.1 EAF
 Image directe 0.1 EAF
 Indépendance linéaire 0.2.4 EAF
 Inégalité de Hölder 1.B 6.10 EAF
 Inégalité de Minkowsky 1.B 4.4.2 EAF
 Inégalité de Schwarz 1.A 4.4.2 EAF
 Inégalité de Ptolémée Exercice N° 12 EAF
 Inégalité triangulaire 1.A.1 EAF
 Injectif 0.1. EAF
 Intégrale de Lebesgue 1.B 6.6.2 EAF
 Interpolation (polynôme d'..) 3.1
 Interpolation d'une fonction chap. 3 & 7
 Interpolation linéaire 3.1
 Interpolation de Lagrange 3.2
 Interpolation trigonométrique exercice N° 26
 Interpolation d'Abel-Gontcharoff exercice N° 28
 Intersection 0.1 EAF
 Inverse (opérateur) 0.3.8 EAF
 Isométrie 1.B.5 EAF
 Isomorphisme 1.B.5 EAF

Jackson (théorèmes de..) et (noyau de..) 4.4 & exercices 38, 39, 40
 Jacobi (polynôme de..) 3.6 EAF

Korovkin (théorème de..) 4.2
 Lagrange (polynômes d'interpolation de..) 3.1
 Laguerre (polynômes de..) 3.6 EAF
 Lanczos (méthode de..) 6.3
 Linéarité de la meilleure approximation 2.1
 Lebesgue (intégrale de..) 1.B 6.6.2 EAF
 Legendre (polynôme de..) 3.2 EAF & 5.3
 Legendre (série de..) 4.4.3
 Lemoine (point de..) exercice N° 20
 Limite 1.B.1. EAF
 Linéaire (forme) 0.3.6 EAF
 Linéaire indépendance 0.2.4 EAF
 Lipschitz (condition de..) 4.3

 Mal conditionné (système) 2.4
 Matrices transposée, conjuguée, adjointe 0.3.9 EAF
 Mesurable 1.B 6.6.2 EAF
 Méthode de quadrature 7.2
 Métrique (espace) 1.A.1 EAF
 Meilleure approximation 1.1
 Minkowsky (inégalité de..) 1.A 4.4.3 EAF
 Module de continuité 4.4
 Moyenne quadratique (approximation en..) chap. 2 & 3
 Müntz (théorème de..) 4.4

 Neutre (élément) 0.2.1 EAF
 Nombres de Christoffel 7.2
 Normes 1.A.4 EAF
 Norme d'un opérateur 1.B.3 EAF
 Normes équivalentes 1.B.5 EAF
 Noyau reproduisant (théorème du..) 3.4.2 EAF
 Noyaux régularisants 5.4
 Noyaux ultrarégularisants exercice N° 46
 Noyau de Féjer 5.4

Opérateur continu 1.B.3 EAF
 Opérateur linéaire 0.3 EAF
 Opérateur positif 4.1
 Opérateur hermitien 3.2 EAF
 Opérateur symétrique (opérateur hermitien sur un espace préhilbertien réel) EAF
 Opérateur d'interpolation chap. 3
 Orthogonalité 1.A.4 EAF et exercice NO 10
 Orthogonal (sous-espace) 0.4.1 EAF
 Orthogonaux (polynômes) 3.3 EAF
 Orthonormée (base) 3.2 EAF et 2.5
 Orthonormalisation de Gram-Schmidt 5.3 et EAF

 Parseval (formule de..) 3.8.3 EAF
 Persymétrique (déterminants) 2.4
 Préhilbertien (espace) 1.A 4.4.5 EAF
 Presque-partout 1.B 6.6.2 EAF
 Propriété K 4.1
 Propriété de l'alternance 6.2
 Projection orthogonale 2.2
 Prolongement (théorème de..) 1.B 6.6 EAF
 Projection (théorème de la..) 2.2
 Polya (algorithme) 1.4
 Polynômes de Bernstein 4.6.3
 Polynômes orthogonaux 3.3 EAF
 Pseudo-norme 7.3
 Pythagore (théorème de..) 1.A 4.4.7 EAF

 Rang d'une application linéaire 0 EAF
 Régularisation d'une fonction 5.4
 Récurrence (formules de..) 3.4.1 EAF
 Réponse en fréquence du système 5.1
 Restriction d'un opérateur 0.3.4 EAF
 Réunion 0.1 EAF
 Rodriguès (formule de..) 3.4.2 EAF
 Runge (phénomène de..) exercice N° 44

Scalaire (produit) 1.A 4.4.1 EAF
 Schwarz (inégalité de..) 1.A 4.4.2 EAF
 Semblable (matrice) 0.3.9 EAF
 Série de Fourier 2.5 et EAF
 Somme directe 0.2.5 EAF
 Sous-espace métrique 1.A.3 EAF
 Sous-espace vectoriel 0.2.3 EAF
 Sous-espace vectoriel engendré par A 0.2.4 EAF
 Sous-espace supplémentaire 0.2.3 EAF
 Spectre d'un opérateur exercice N° 49 EAF
 Splines (fonctions..) chap. 3 Splines naturelles 3.4
 Stricte convexité 1.3
 Supplémentaire (sous-espace) 0.2.3 EAF
 Surjectif 0.1 EAF
 Symétrique (élément) 0.1 EAF
 Système de générateurs 0.2.4 EAF
 Système libre 0.2.4 EAF
 Système total 3.8 EAF

 Théorème des bases hilbertiennes 3.8.3 EAF
 Théorème de Haar exercice N° 57
 Théorème de Markov exercice N° 60
 Timan (théorèmes de..) 4.3
 Transformations (produits de..) 0.1 EAF
 Transposé (opérateur) 0.4.3 EAF
 Totalité des polynômes orthogonaux 3.8.6 EAF
 Topologie 1.B

 Unisolvent (système) 6.2
 Unifère (anneau) 0.3.5 EAF
 Unité (élément) 1.2.1 EAF
 Unité approchée 5.4

 Vectoriel (espace) 0.2 EAF
 Vitali (théorème de..) exercice N° 42
 Voronovsky (théorème de) 4.6.3

 Weierstrass-Bolzano (théorème de) 2.A.5 EAF
 Weierstrass (théorème de..) 4.3

 Zygmund (théorème de..) 4.3

